Journal of
Animal Breeding and Genetics

Genome wide survey on three Sicilian horse populations with a focus on Runs of homozygosity pattern

Journal:	*Journal of Animal Breeding and Genetics*
Manuscript ID	JABG-21-0131
Manuscript Type:	Original Article
Date Submitted by the Author:	04-Aug-2021
Complete List of Authors:	Criscione, Andrea; University of Catania, Dipartimento of Agricoltura, Alimentazione e Ambiente Mastrangelo, Salvatore; University of Palermo, Dipartimento Scienze Agrarie, Alimentari e Forestali D'ALESSANDRO, ENRICO; Universita degli Studi di Messina, Dipartimento di Scienze Veterinarie Tumino, Serena; University of Catania, Dipartimento of Agricoltura, Alimentazione e Ambiente Di Gerlando, Rosalia; University of Palermo, Dipartimento Scienze Agrarie, Alimentari e Forestali Zumbo, Alessandro; Università degli Studi di Messina Dipartimento di Scienze Veterinarie Marletta, Donata; University of Catania, Dipartimento of Agricoltura, Alimentazione e Ambiente Bordonaro, Salvatore; University of Catania, Dipartimento of Agricoltura, Alimentazione e Ambiente
Subject Area:	diversity, horse, SNP

Genome wide survey on three Sicilian horse populations with a focus on Runs of homozygosity pattern

Andrea Criscione[1†], Salvatore Mastrangelo[2†], Enrico D'Alessandro[3], Serena Tumino[1*], Rosalia Di Gerlando [2], Alessandro Zumbo[3], Donata Marletta[1], Salvatore Bordonaro[1]

[1]Dipartimento of Agricoltura, Alimentazione e Ambiente, Università di Catania, Via Valdisavoia 5, 95123 Catania, Italy

[2]Dipartimento Scienze Agrarie, Alimentari e Forestali, Università di Palermo, 90128 Palermo, Italy

[3]Dipartimento di Scienze Veterinarie, Università di Messina, 98168 Messina, Italy

[*]corresponding author: serena.tumino@unict.it

[†]These authors contributed equally

Running title: Biodiversity and autozigosity of horse populations

Abstract

Purosangue Orientale Siciliano, Sanfratellano and Siciliano represent the Sicilian equine genetic resource. This work aimed at investigating the genetic diversity, population structure and pattern of autozygosity of Sicilian horse populations using genome-wide single nucleotide polymorphism data (SNP) generated with Illumina Equine SNP70. SNP data of Arab, Maremmano and Norwegian Fjord breeds were also included in the study. Patterns of genetic differentiation, model-based clustering, and Neighbor-Net showed the close connections between the Purosangue Orientale Siciliano and the Arab, as well as between Sanfratellano, Siciliano and Maremmano. The highest He and Ne were reported in Siciliano (He = 0.323 Ne = 400), the lowest in Purosangue Orientale Siciliano (He = 0.277 Ne = 10). The analysis of the Runs of Homozygosity and the relative derived F_{ROH} highlighted the high internal homogeneity of Purosangue Orientale Siciliano and Arab horses, intermediate values in Maremmano and Sanfratellano and the high heterogeneity of the Siciliano population. The gene level analysis showed the selective pressure to which the Purosangue Orientale Siciliano seems to be subjected towards the traits related to endurance performance and the genetic proximity of this with the Arab. Our results underline the importance of planning adequate conservation and exploitation programs with the means to reduce the level of inbreeding and therefore the loss of diversity.

Keywords: autochthonous horses, genetic diversity, Runs of homozygosity, SNPs

1. Introduction

Throughout history, horses have played an important role in human civilization for their influence on agriculture, warfare, trade and transportation (Al Abri et al., 2021). For the past 400 years, the establishment of formal breed registries has focused on the conservation of local populations and improvement of traits related to riding, draft, aesthetics, and performance (Zhang et al., 2018). Today in Sicily there are about 15,000 animals belonging to the *equidae* family, of which less of 10% are native horses. Three populations (Sanfratellano, Siciliano and Purosangue Orientale Siciliano) that boast an ancient history and an origin that can be traced back to Greek domination (600BC), represent the Sicilian equine heritage (Guastella, Zuccaro, Criscione, Marletta, & Bordonaro, 2011). The total consistencies of the three populations poorly explain the relative importance of the different genetic types in the Sicilian equine framework; Purosangue Orientale Siciliano and Siciliano are about 200 individuals each, while Sanfratellano counts 1300 horses (PSR Regione Sicilia 2014-2020, ARACSI). The origins of the Sanfratellano horse date back to the Middle Ages, when the Sicilian native horses were crossed with North African, Oriental and subsequently Iberian populations (Fogliata, 1910). Limited introgression of Thoroughbred and Oriental stallions was practiced in 1925 to improve the morphological structure of Sanfratellano (Hendricks, 1995). More recently, from the 1930s and occasionally until the end of the century, Maremmano stallions were used in the planned mating to improve withers height and size (Chiofalo, Portolano, Liotta, Rundo Sotera, & Finocchiaro, 2003; Zuccaro et al., 2008). Sanfratellano is a meso-doligomorphic horse suitable for saddle and draft. Today the breed is successfully engaged in trekking, sports and hippotherapy activities. Purosangue Orientale Siciliano, is a genetic type of Arab-Oriental matrix belonging to the Italian Stud Book since 1875; it represents a Sicilian nucleus of Oriental horses imported from Syria and Mesopotamia since 1864 (Balbo, 1995). It is a mesomorphic and meso-doligomorphic type horse. The morphological characteristics of the Purosangue Orientale Siciliano make it suitable as a saddle horse and for light draft, with particular predisposition for running and endurance performance over long distances. These horse populations possess valuable traits such as disease resistance, longevity and adaptation

to harsh conditions and poor-quality feed. The Siciliano horse, which took origin from the crossbreed between the Asiatic and the North African horses that were reared in Sicily until the 16th century (Guastella et al., 2011), is a heterogeneous population, reared in an extensive and semi-extensive system and not yet officially recognized as breed. This population includes mesomorphic type horses, more widespread in the central areas of Sicily, and meso-dolicomorphic horses, reared mainly in the eastern part of the island. Overall, it has a conformation that adapts to the saddle and to draft, of a docile and submissive character.

With the development of molecular technology in recent years and in particular the use of microarray platforms, investigation techniques to define the genomic structure and evolutionary history of populations have become increasingly widespread, also for the horse breeds (Pereira et al., 2017; Petersen et al., 2013). However, compared to the livestock species, only a limited number of genetic diversity studies were conducted in horses, leaving the population structure of local breeds undisclosed, as for the three Sicilian horse populations. Genetic diversity is a key measure for the monitoring of genetic parameters that are important for the prevention of genetic erosion, inbreeding and other deleterious processes that may lead to population extinction. A valuable method, called Runs of Homozygosity (ROH), has been used in livestock for the identification of homozygous genomic regions and as predictors of whole-genome inbreeding levels (Marras et al., 2015; Mastrangelo et al., 2018). ROH are consecutive homozygous genotypes of variable length distributed across the genome with prevalence in those regions affected by low recombination rate. ROH arise from identical-by-descendent haplotypes transmitted by common ancestors whose length appears to be proportional to the level of inbreeding and directly linked to the generation of parental transmission of the homozygous genotypes. (Ceballos, Joshi, Clark, Ramsay, & Wilson, 2018; Curik, Ferencakovic, & Solkner, 2014; Kim et al., 2013). The characterization of the distribution and lengths of ROH within a population can help to reveal its evolutionary history, reveal incorrect mating schemes that end in an increased level of inbreeding, as well as identify close genomic associations with phenotypic characters. In recent years, studies focused on detection of positive selection, using

ROH signals have been also carried out in horse species (Druml et al., 2018; Grilz-Seger, Druml, Neuditschko, Dobretsberger, et al., 2019; Metzger et al., 2015). In this study, a medium density SNP genotyping panel was used to characterize the three Sicilian horse populations, with the aim of investigating the genetic diversity, the population structure and the patterns of ROH. For comparative purposes in relation to their origins and their evolutionary history, the SNP genotyping data of three additional horse breeds (Maremmano, Arab and Norwegian Fjord), were also included in the analyses.

2. Materials and Methods

2.1. DNA sampling, genotyping and quality control

Blood samples were collected from 46 horses belonging to Sanfratellano (SAN = 17), Purosangue Orientale Siciliano (SOP = 12) and Siciliano (SIC = 17). Whole blood samples (10 mL) were obtained from the jugular vein in tubes containing ethylenediamine tetra-acetic acid (EDTA) as anticoagulant. Sampling procedure was carried out, according to Directive 2010/63/EU, by authorized personnel during the periodic veterinary control, therefore, no pain, suffering, distress and lasting harm was caused to the animals involved in the present study. DNA was extracted from leukocytes using the Illustrablood genomic Prep Mini Spin kit (GE Healthcare, Little Chalfont,UK). Individual samples were genotyped with the Illumina Equine SNP70K Beadchip (Illumina Inc., San Diego, CA, USA), which consist of 65,157 SNPs. Chromosome assignment and position for each marker are referred to the equine *Equ Cab 2.0* genome assembly. The raw data of Sicilian horses have been merged with the genotyping data of three horse breeds, Arab (ARR = 24), Maremmano (MARM = 24) and Norwegian Fjord (NORF = 21), retrieved from a previous study (Petersen *et al*. 2013). Two data sets were generated, one that includes the Sicilian horses, the Maremmano and the Arab breeds (5POP), the other which also includes the NORF as an outgroup breed (6POP). The program PLINK ver.1.9 (Purcell et al., 2007) was used to perform quality control. SNPs were filtered to exclude loci assigned to unmapped contigs, and only those SNPs located on autosomes were considered. Quality control

included call frequency ⩾0.98 and minor allele frequency (MAF) ⩾0.01. Animals with more than 2% missing SNPs were also removed from the analysis. After quality control, 40,715 (6POP) and 40,601 (5POP) SNPs were retained, respectively.

2.2. Genetic diversity and population structure

PLINK ver.1.9 (Purcell et al., 2007) was used to estimate within-population diversity (Ho and He). The software Arlequin ver. 3.5.2.2 (Excoffier & Lischer, 2010) was implemented to infer genetic relationships between populations by pairwise Reynolds' genetic distances. Neighbor-net was constructed from the estimated genetic distances using SplitsTree4 software ver. 4.14.8 (Huson & Bryant, 2006). According to the random mating option within the LD method (Waples & Do, 2010), the contemporary effective population size (Ne) was estimated using NeEstimator V2.1 (Do et al., 2014). PLINK software was also used to calculate pairwise identical by-state (IBS) distances between populations, graphically represented by multidimensional scaling (MDS) analysis. The population structure of SOP, SAN, SIC, ARR and MARM populations was investigated by applying the model-based clustering algorithm run in ADMIXTURE (Alexander, Novembre, & Lange, 2009) from K = 2 to 10; cross-validation procedure was applied (cv=10). Circle plot of Admixture results was obtained through the 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).

2.3. Runs of homozygosity detection

Runs of homozygosity (ROH) were detected by means of the R package detectRUNS ver. 0.9.6 (Biscarini, Cozzi, Gaspa, & Marras, 2018). The ROH statistics were inferred using the method of consecutive runs (Marras et al. (2015). In detail, ROH were obtained by setting the minimum number of SNPs to 15, not allowing neither missing nor heterozygous SNPs, setting the minimum length of run to 1 Mbps and the maximum gap between consecutive SNPs in a run to 1 Mb. The mean number (N_{ROH}) and average length (L_{ROH}) of ROH per individual per population, as well as the sum of all

ROH segments (S_{ROH}) per animal were estimated. Each ROH was categorized based on its physical length as follows: <2 Mb, 2 to <4 Mb, 4 to <8 Mb, 8 to <16 Mb, and ≥16 Mb. For each of the ROH length categories, the mean sum of ROH per population was calculated by summing all ROH per animal in that category and averaging this per population. The total length of the genome covered by ROH was divided by the total horse autosomal genome length covered by the SNP array to evaluate the individual genomic inbreeding coefficient using the ROH data (F_{ROH}). The most common ROH (ROH islands), which showed a within breed occurrence ⩾50%, were further investigated. The genomic coordinates of these regions were examined through the Ensemble browser for horse genome, according to the assembly EquCab2 (https://oct2018.archive.ensembl.org/index.html) to retrieve annotated gene lists. Horse Quantitative Trait Locus Database (Horse QTLdb) (https://www.animalgenome.org/cgi-bin/QTLdb/EC/index) was then interrogated to search for possible associations between aforementioned markers and reported QTL in horse species, as well as to clarify the gene's identity and functions. Gene Ontology (GO) and enrichment analysis of annotated genes was conducted using the open source Database for Annotation, Visualization, and Integrated Discovery v.6.8 package (https://david.ncifcrf.gov) (Huang da, Sherman, & Lempicki, 2009). For the Gene Ontology (GO) terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways analysis, the *Equus caballus* annotation file as background was used.

3. Results

3.1. Genetic diversity and population structure

The genetic diversity indices are shown in Table 1. The highest expected heterozygosity value (He) was reported in SIC, the lowest in SOP; the observed heterozygosity (Ho) was highest in MARM and lowest in ARR. The effective population size (Ne) was 10 and 31 in SOP and SAN respectively, while notably higher values were recorded in ARR (195), MARM (294) and SIC (400). The reduction of SNP matrix's variability by the first two component (which accounted for 30.8% of the total variation) of MDS analysis is represented in Fig. 1. As expected, SOP and ARR populations were

spatially close, SIC and SAN formed a cluster together with MARM breed, and NORF (which is the outgroup of the data set), migrated towards an isolated part of the figure. In particular, the first component (19.2%), clearly separated the Oriental type horses cluster (ARR and SOP), the group consisting of meso-doligomorphic horses (SIC, SAN and MARM breeds) and the NORF horse. The second component, which accounted for 11.6% of the variation, did not discriminate NORF from the oriental mesomorphic type breeds. The Neighbor-Net based on Reynolds' pairwise genetic distance (Fig. 2) recalled the output of the first dimension in the MDS analysis and reported ARR and SOP connected to the same split node, SIC, SAN and MARM close to a common reticulation, with the NORF outgroup breed connected to the same split node. The analysis of population structure, performed on the Sicilian horses together with ARR and MARM breeds, gave results comparable to that of MDS survey (Fig. 3). The results indicated that the most probable number of inferred populations was K = 4 (Fig. S1). At K = 2 the admixture analysis underlined shared ancestral components between ARR and SOP as well as between SAN, SIC and MARM; at K = 3, the MARM breed forms a separate group, the horses of the Oriental type (ARR and SOP) maintain the common clustering, while SIC and SAN shared a similar genetic background. Finally, at K = 4, almost all populations have their own identity, with moderate level of admixture for SIC (Fig. 3).

3.2. Runs of homozygosity detection

Table 2 summarize the sum of ROH length expressed in Mb (S_{ROH}), the number of ROH (N_{ROH}), the length of homozygosity runs expressed in Mb (L_{ROH}) and the inbreeding coefficient estimated from ROHs (F_{ROH}). The parameters were highly variable, especially if we consider the ARR and SIC samples, which showed the highest and lowest values, respectively. In particular, the mean length of ROH distributed over the 31 chromosomes (S_{ROH}) was highest in ARR (419.57±134.97) and SOP (299.45±90.15) breeds, followed by the values of MARM (223.06±48.27), SAN (205.63±38.53) and SIC (159.05±45.80) horses. In the whole sample, 3 ARR and 1 SOP horses showed an S_{ROH} higher than 500 Mb, while 12 individuals (9 SIC, 2 MARM, 1 SAN) reported values lower than 150 Mb.

The N_{ROH} and L_{ROH} mean values were highest in ARR, followed by MARM, SOP, SAN and SIC. The mean F_{ROH} varied between 19% (ARR) and 7% (SIC) and retraced the same breed ranking as the sum of ROH length; the highest within-breed F_{ROH} value per individual were in ARR (40%) and SOP (25%), while the lowest value was in SIC (5%). Average breeds' and individual inbreeding coefficients are plotted in Fig. 4: ARR showed the highest values and the highest internal variability, followed by SOP horse, while MARM, SAN and particularly SIC showed lower values and a higher within sample homogeneity. The large majority of the ROH detected in the five populations showed a length not exceeding 8 Mb (Table 3), from 94.8% in ARR to 98.2% in SIC: the Arab horse highlighted the lowest percentage for ROHs included in the bottom class of length (0-2 Mb), while SIC and SAN showed the highest value. The medium length class (4-8 Mb) reported ARR and MARM the breeds with the percentage above 7% while the Sicilian horses showed lower values (5.5% in SOP - 4.5% in SIC). The highest percentage of ROHs with length above 8 Mb was registered in ARR (5.2%), followed by MARM and SAN (3.7%), then SOP (2.2%) and SIC (1.8%). In the same table, the F_{ROH} values per class of ROH's length are reported: the inferred inbreeding coefficients decreased with the increasing length of ROHs with the exception of SOP and SAN, which showed a slight increase corresponding to the > 8Mb class. The ARR sample reported the highest F_{ROH} values, considering both the most recent and the oldest inbreeding, whilst SIC showed the lowest values, particularly for the longest classes where F_{ROH} tended to zero. The F_{ROH} percentage incidence (F_{ROH}%) of the two lowest length classes (<4 Mb) was always above 55% of the total F_{ROH} per breed (lowest % in ARR) and reached the highest value in SIC (75%). In the SIC horse the remaining portion of F_{ROH} is equally distributed between the middle (4-8) and long (>8 Mb) length classes, SOP reported a slight percentage increase from the intermediate class to the two major ones, in ARR, MARM and SAN the incidence of F_{ROH}> 8Mb is always higher than 24%. The markers involved in ROHs showed a percentage of recurrence within breed ranging from 4% to 100% (Fig. S2-S6). We examined and further investigated the case of those markers within ROH islands that showed an incidence per breed ⩾ 50%. Table S1 reports the genomic coordinates of the ROH islands, the number

of SNPs per ROH, as well as the annotated genes and QTL traits. A total of 115 ROH islands harbouring 1770 markers, were identified. The highest number of ROH islands was identified in SOP, in which by 50% up to 100% of individuals shared 60 ROHs harbouring 1029 SNPs detected on 25 chromosomes. In particular, 339 markers are located within intronic regions and 4 markers are detected within exon sequences of 157 known genes (data not shown). In ARR, 50% up to 100% of horses shared 47 ROH islands, identified in 20 chromosomes; within above mentioned ROHs 628 markers were identified: 204 SNPs are located in intronic portions and 9 within exon sequences of 111 known genes. SIC and SAN samples showed 3 ROHs per breed, respectively. In SIC sample, ROH islands were located in chromosomes ECA9 and ECA7, in which 54 markers were identified at a population's frequency ranging between 52.9% and 58.8%: 51 markers are inter-genic variants, 1 intronic variant and 2 exonic variants. Whilst in SAN, ROH islands were identified on ECA11, ECA15 and ECA17 at a population's frequency ranging between 52.9% and 58.8%; in this case, 39 markers were detected: 10 markers are located on intronic regions, the rest are inter-genic variants. Within the MARM breed, 2 ROHs in ECA4 and ECA17 were identified with a frequency between 50% and 54.2% of the individuals; 20 markers in a sequence of ROHs were detected: 1 marker is an intronic variant, the rest are located on inter-genic regions. The search on Horse Quantitative Trait Locus Database (Horse QTLdb) revealed 67 different markers within ROH islands in association with 76 QTL belonging to 11 different traits (Table S2). The highest number of QTL-associated markers was detected in SOP and ARR breeds. In particular, in the SOP breed 44 markers were identified in association with 9 traits (osteochondrosis dissecans, withers height, insect bite hypersensitivity, alternate gaits, white markings, guttural pouch tympany, male fertility, altitude adaptation, temperament). The ARR breed showed 34 markers associated with 8 traits (number of progressively motile sperm, guttural pouch tympany, altitude adaptation, sperm progressive motility, insect bite hypersensitivity, withers height, white markings, alternate gaits). The SIC population showed two markers associated with withers height and alternate gaits traits, whilst MARM and SAN breeds reported one marker each, associated with insect-bite-hypersensitivity and withers-height,

respectively. Twenty-one different markers of the abovementioned 67 SNPs fall within intronic regions of 17 known genes. In particular, the marker rs68871178 on ECA9 (45279882 bps), corresponding to the intronic region of the *VPS13B* gene (vacuolar protein sorting 13 homolog B), is a flanking marker of QTL #119813 associated with the temperamental expression in Tennessee Walking horse (Staiger, Albright, & Brooks, 2016). The variant located on ECA18 at position 49758616 bps (rs69171012), sequenced in the intronic region of the *MYO3B* (myosin IIIB) gene, is related to QTL #29459 associated with the capacity to adapt at high altitudes of the Andean horse (Hendrickson, 2013). The results of the GO and enrichment analysis on breeds' annotated genes, shown in Table S3, revealed 215 genes enriched in 93 biological processes, molecular and cellular component functions. In particular, in SOP 65 genes were enriched in 27 biological processes, 27 genes in 9 molecular function and 12 in 4 cellular components. ARR harboured 109 genes, 35 of them were found to be enriched in 44 biological processes, 2 in 1 molecular function and 8 in 4 cellular components while in SAN 4 biological process involved 2 genes. GO analysis revealed no enrichment for SIC and MARM due to the low number of annotated genes. GO terms evidences were also corrected for multiple testing (Bonferroni adjusted $p<0.05$) showing significant enrichment for 5 molecular function related to the nucleotide-binding process within the SOP sample. The KEGG analysis highlighted 8 biological pathways each in ARR and SOP.

4. Discussion

Sicily, in the centre of the Mediterranean area, has always been the crossroads of a continuous flow of animal germplasm that accompanied various dominations. Historically, in the Sicilian equine sector there has been an intense interchange of breeding animals, as well as the succession of different equestrian schools with different methods of training and breeding strategies. Since the distant past (600 BC) up to the 16th century, the equine genetic basis present in the island has been influenced and shaped by various horse breeds from North Africa and Middle East, from Northern Europe with the Norman invasion, from Iberian countries during the Spanish domination (Fogliata, 1910). In more

recent times, it was worth of note the contribution made by breeds such as the Thoroughbred and Maremmano to the evolution of Sanfratellano (Zuccaro et al., 2008). Arab stallions contributed to the origin of the Purosangue Orientale Siciliano breed and are still used as breeding animals, and also have partly influenced the evolution of the Siciliano horse. The advent of high-throughput genotyping arrays has greatly facilitated the study of genetic structure in livestock species, giving also the possibility to investigate the old and recent relationships among populations. Previous studies (Criscione, Moltisanti, Chies, Marletta, & Bordonaro, 2015; Guastella et al., 2011; Zuccaro et al., 2008), have already focused on the genetic characterization of Sicilian breeds by implementing nuclear and DNAmt markers. In this paper, we presented for the first time the results of the genomic characterization of Sicilian horse populations.

The expected heterozygosity has always been lower than observed, with the exception of Arab breed, in which the He and Ho values almost overlap. The observed heterozygosity in Arab is consistent to that reported by Cosgrove et al. (2020) who highlighted a range of 0.30-0.33 in different Arab strains and 0.26 in Straight Egyptian, and consistent to that reported by Schaefer et al. (2017). Cosgrove et al. (2020) also reported Ho values of other 18 breeds, ranging between 0.32 and 0.36, including the Maremmano horse (Ho=0.36), and consistent with our results. Lower values, both for Ho and He, were reported by Druml et al. (2018) in Haflinger, Noriker, Arab and Bosnian Mountain Horse, (0.256-0.326 Ho; 0.258-0.311 He). Effective population size (Ne) is one of the variables to be taken into account in breed conservation (Verrier et al., 2015) and is defined as the size of an idealized population that would produce the same genetic variation as the population under study (Wright, 1969). The maintenance of Ne at, or above, 50 to 100 is a principle of breed conservation (Meuwissen, 2009). The effective population size indicated a high risk of inbreeding and reduced genetic diversity in Sanfratellano and Purosangue Orientale Siciliano, thus suggesting an appropriate investigation on the breeds' actual census to confirm this evidence. Bayesian model-based clustering algorithm and MDS were used to visualize and explore the genetic relationships between Sicilian populations and the other horse breeds. The results have pointed out the relationship within two group of horses (ARR-

SOP, SAN-MARM-SIC), according to their genetic origin and breeding history. Reynolds' genetic distance represented by the NeighborNet algorithm gave highly coherent picture of the breeds' relationships confirming both the results obtained by MDS and the genomic admixture analysis and the historical records that explain many of the connections between these genetic types.

Arab and Purosangue Orientale Siciliano partially share a common ancestry: the Purosangue Orientale Siciliano represents the evolution guided by the selection of a nucleus of oriental horses imported from Syria and Mesopotamia in 1864 directly from the Bedouin tribes and belonging to the Hamdani, Saglawi, Kuhaylan and Abayan lines (Balbo, 1995). During the early years of the twentieth century, oriental stallions continued to be imported from the Middle East, Hungary, France and Poland (studbook source). Since the formation of the Purosangue Orientale Siciliano breed, Arabian stallions have been fundamental in mating plans and still represent an important source of genomic diversity for oriental horses reared in Sicily. The most recent use of Arab stallions as breeding animals date back to 2016 (studbook source). Guastella et al. (2011), in a study on Sicilian horses carried out by mtDNA characterization, identified in SOP a unique haplotype which corresponds to the Dafina matrilineal line founder of the Keilan el Krush Arab strain. The Purosangue Orientale Siciliano sums up the physical characteristics of Arab, with the exception of the pure Egyptian lines most voted for performance shows; the morphology developed over the course of its evolution makes it suitable as saddle and light draft horse, with particular predisposition for running and endurance over long distances. The Maremmano horse has significantly influenced the evolution of the Sanfratellano breed: starting from 1934 and for the next 10 years, seven Maremmano stallions were used in the Sanfratellano mating plans. This process of genetic introgression constituted the basic structure of the current Sanfratellano breed. The aim was to soften the shapes of the population, improving its character, increasing the height at the withers without however removing the innate frugality, the robustness of the skeletal structure, the resistance to fatigue, typical of this autochthonous breed and transmitted by the maternal lines. Selective hybridisation was practiced on the progeny of this group of stallions until 1958. At the end of the sixties, two other Maremmano stallions were used in the

selective mating of Sanfratellano (Chiofalo et al., 2003; Zuccaro et al., 2008). The genomic admixture between the Sanfratellano and Siciliano horses can be explained by the common origins of the two Sicilian autochthonous populations influenced by Oriental and North Africa horses, documented by historical data (Fogliata, 1910; Zuccaro et al., 2008), as well as by occasional gene flow between the two populations. Siciliano is a very heterogeneous and largely unmanaged population, probably derived from a primitive strain of Sicilian horses and largely influenced by the breed "Real Casa di Ficuzza" (Borbon domination XIX sec.) which was strictly related to the Napoletano, Persano and Arab horses (Balbo, 1995). The relationship between Siciliano and Maremmano can trace back to the introgression of Thoroughbred genetics into both populations (Balbo, 1995; Giontella et al., 2020; Hendricks, 1995). In recent years, the globalization of equine breeding has strongly oriented this species as a sporting animal (Waran, 2007). The preferential breeding of breeds with high sporting and economic potential and the use of sperm from selected stallions is a threat to the genetic diversity of local populations and therefore of the equine species (Bowling & Ruvinsky, 2000). Local populations, such as Sicilian horses, often have a small effective size, which implies difficulties related to the management of inbreeding and intra-breed genetic diversity. The risk of extinction is recognized in the Sanfratellano (extinction state) and Purosangue Orientale Siciliano (critical state) both by local (PSR Regione Sicilia 2014-2020) and international authorities (http://www.fao.org/dad-is/browse-by-country-and-species/en/). Population genetics studies performed by analyzing the distribution, prevalence and location of ROHs provide useful information about population structure, evolutionary history and breeding selection. The inbreeding index estimated on molecular autozygosity is one of the parameters obtainable from genetic characterization using SNPs panels, and particularly useful where genealogical records are lacking or absent. Our results showed that the Arab reached the highest values F_{ROH}, followed by Purosangue Orientale Siciliano. As reported by Cosgrove et al. (2020), the Arab breed have been dispersed widely across the globe but kept a unique genetic identity; the studbook, one of the oldest in the equestrian world, imposes a very restrictive standard that has made the Arab the horse it is today. The F_{ROH} was higher than that reported by

Druml et al. (2018) in Shagya Arabians (F_{ROH} = 0.16) and Purebred Arabians (F_{ROH} = 0.18), lower than the reported inbreeding coefficient (F_{PLINK}) in Straight Egyptian horses (0.30) by Cosgrove et al. (2020) who also reported a range of F_{PLINK} varying between 0.12 and 0.30 in 6 different lineages of Arabian horses. The Purosangue Orientale Siciliano breed is a genetic type whose Stud Book was established with Royal Decree No. 2690 in 09/19/1875. The breed has always maintained a high degree of morphological and genetic homogeneity during its evolution and despite the very low consistency (today about 200 horses), it has maintained not excessively high degree of inbreeding thanks to the periodic introduction of Arab blood. The F_{ROH} value in Purosangue Orientale Siciliano (0.13) was substantially lower than in Arab sample (ARR) and lower than the values reported by Druml et al. (2018) in Arab. Furthermore, F_{ROH} of Purosangue Orientale Siciliano was comparable to the F_{PLINK} values in the Arab lineages of Poland and Iran, as well as F_{PLINK} of multi-origin Arabs (Cosgrove et al., 2020), and comparable to the F_{PLINK} value reported by Schaefer et al. (2017). The Maremmano and Sanfratellano saddle horses showed intermediate values of the ROH parameters which, especially when compared with the Arab and Purosangue Orientale Siciliano breeds, corroborate the different history of formation of these breeds that have undergone the influence of genetic types such as the Thoroughbred and Iberian horses and report a more recent closure of the registers. The F_{ROH} values of Maremmano (0.10) and Sanfratellano (0.09) are comparable to those reported in Slovenian Haflinger (0.12) (Grilz-Seger et al., 2018), in Lipizzan (mean 0.13) which showed a variation between 0.07 and 0.15 in the 4 lineages analyzed (Grilz-Seger, Druml, Neuditschko, Dobretsberger, et al., 2019), in the Noriker breed with an average of 0.10 and a range of variation between the 6 coat colour lineages of 0.08-0.13 (Grilz-Seger, Druml, Neuditschko, Mesaric, et al., 2019). The Siciliano horse, an equine population that currently does not have breed recognition and for which there is no selective plan, showed the lowest F_{ROH} index (0.07). The census population, recorded by the Association of breeders (ARACSI), currently stands at around 200 horses, a number that would make us wait for higher inbreeding values. Probably the common genomic basis of this population has maintained a high degree of variability among the different family lines kept

by different breeders in Sicily, also by virtue of unsystematic crossbreeding involving a population of breeding animals larger than the recorded one. The F_{ROH} values in Siciliano are lower than those found in Bosnian Mountain Horse which has less than 200 heads (0.13) and comparable to those of Posavje Horse which has about 600 heads (0.09), horse breeds whose selective recovery programs have only been started in the last 30 years (Grilz-Seger et al., 2018). The inbreeding index derived from the analysis of ROH by length classes allows us to hypothesize the number of generations back in time to which the autozygosity segments refer. The expected length of an autozygous segment follows an exponential distribution with mean equal to 1/2 g Morgans, where g is the number of generations since the common ancestor (Howrigan, Simonson, & Keller, 2011). In particular, ROH segments of 16 MB in length are estimated to reflect inbreeding up to three generations in the past, while short ROH (1 MB) are related to ancient inbreeding, up to 50 generations in the past. Assuming an average generational interval of 10 years in the equine species, as reported by various authors (Valera, Molina, Gutiérrez, Gómez, & Goyache, 2005), the F_{ROH}, calculated for each length class, trace back the common inbreeding in a time interval from 30 to 500 years. In Siciliano, inbreeding is mainly attributable to distant ancestors and date back to the Spanish domination (XVI-XVII century), a period in which the equine genetic basis in Sicily was influenced by Iberian horses and the historical period in which the differentiation between genetic types that we know today (SOP, SAN and SIC) had its beginning. Guastella et al. (2011) reported in Siciliano one haplotype that traces back to a Bronze Age archaeological site (Inner Mongolia; DQ900929). The distribution of length-class inbreeding has shown that in Arab, Maremmano and Sanfratellano, a considerable percentage of the total F_{ROH} dates back to 70 years in the past (ROH length> 8Mb). The Sanfratellano horse, therefore, reports most of its autozygosity in correspondence with the hybridisation process (1950s) that followed the first introduction of Maremmano stallions (1934) and the last introduction of the Maremmano blood into the breed in 1969. Purosangue Orientale Siciliano, after Siciliano, showed the highest percentage of F_{ROH}% for the 0-4 Mb length class, showing also in this case a considerable share of inbreeding attributable to the distant past (500-120 years). Among the breeds analyzed in

relation to the level of autozygosity, only Arab and Purosangue Orientale Siciliano have showed ROH islands with intra-breed percentages ⩾ 75%, probably a result linked to a high intra-breed homogeneity. Interestingly, the ROH islands on ECA19 and ECA23 and shared by 50% of Purosangue Orientale Siciliano's individuals overlapped with QTL for alternating gaits. Specifically, the ROH island on ECA19 harbouring the gene *FBXO40*, already reported for gait type in Tennessee Walking Horses by Staiger, Abri, Silva, and Brooks (2016), suggest the potential association between this gene and gait phenotype. The *FBXO40* gene is also expressed in skeletal muscle and belongs to the F-box protein family that are key components of SCF (Skp1-Cullin1-F-box protein) E3 ubiquitin ligase complexes, in which they act as protein-ubiquitin ligases. In the ROH islands in Purosangue Orientale Siciliano, mapped the *MYO3B* (ECA18), a gene reported in association with the QTL altitude adaptation in a study on Andean horse (Hendrickson, 2013). High altitude exposes animals to intense pressure as permanent oxidative stress and extreme temperature exposure requiring the adaption of the blood, cardiovascular, pulmonary and muscle systems. Different performance disciplines, including prolonged or high-intensity exercise, may result in oxidative stress involving the skeletal muscle fibers. Performing breeds influenced by the Arabian gene pool were known for their heat tolerance and athletic endurance, trait that is well expressed in Purosangue Orientale Siciliano. The *MYO3B* gene was also reported in ROH islands in other breeds, such as French Trotter, Gidran, Selle Francais Shagya Arabian, Trakehner, Holsteiner, Hanoverian, and Oldenburger (Grilz-Seger, Neuditschko, et al., 2019; Nolte, Thaller, & Kuehn, 2019). In the ROH island on ECA9, mapped the *VPS13B* gene (vacuolar protein sorting 13 homolog B), related to a QTL for temperament (Staiger, Albright, et al., 2016). This gene encodes a potential transmembrane protein that may function in vesicle-mediated transport and sorting proteins within the cell. This protein may play a role in the development and function of the eye, haematological system, and central nervous system. Our results give reason to suppose that the traits of temperament and predisposition to endurance performance have been subjected to selective pressure in the Purosangue Orientale Siciliano breed, a consideration that is reflected in the morphological characteristics and behaviour of the breed as

reported by historical data and by the breeders themselves. ROH island located in ECA3 and shared by at least 50% of both Purosangue Orientale Siciliano and Arab breeds overlapped with a dense QTL region associated with four traits: white markings, guttural pouch tympany, withers heigh and insect bite hypersensitivity. In particular, the ROH island on ECA3 (35.6–36.9 Mbp) harboured the *NFKB1* gene, a member of the NF-kB transcription factor family, stimulates the expression of many genes involved in a wide variety of biological functions. Inappropriate activation of persistent inhibition of *NFKB* has been implicated in the pathogenesis of several inflammatory diseases, among which skin disease (Wullaert, Bonnet, & Pasparakis, 2011). The GO analysis in Arab breed confirmed the involvement of the *NFKB* gene in the negative regulation of inflammatory and defence response (GO:0050728, GO:0031348). The *NFKB1* gene was annotated in ROH of chestnut horses investigated by Grilz-Seger, Neuditschko, et al. (2019), highlighting its involvement in the reported higher susceptibility of chestnut phenotype for skin disorders (Bellone et al., 2017). *NFKB1* was also annotated in Straight Egyptian subgroup investigated by Cosgrove et al. (2020); the same authors reported the *SLC9B2* gene in Straight Egyptian subgroup also annotated in Purebred Arabian and Gidran breeds investigated by Grilz-Seger, Neuditschko, et al. (2019).

5. Conclusion

Based on genome-wide data, we investigated for the first time the genetic diversity, population structure and autozygosity pattern of three autochthonous equine populations, including the Maremmano and Arab breeds that have been and still are an important genomic source in the current structure of Sicilian horses. The present study confirmed the historical data that relate Sanfratellano and Maremmano and the close link that exists between Purosangue Orientale Siciliano and Arab horse. We also showed the close genetic relationship between the Sanfratellano and the Siciliano populations and between these and the Maremmano breed. The analysis of the autozygosity pattern of Sicilian equine populations indicated decreasing values, from Purosangue Orientale Siciliano to Sanfratellano and Siciliano, for that part of the genome covered by homozygous sequences and the

estimated inbreeding index. The ROH parameters, in total and calculated by classes of length, reflect the consequences linked to the actual size of the populations and their selective histories. Effective population size values are of concern in Sanfratellano and Purosangue Orientale Siciliano. Gene level investigation has placed the accent on the selective pressure to which the Purosangue Orientale Siciliano seems to be subjected, particularly with regard to performance traits, a result that is reflected by the morphology and by the description of the breed made both by breeders and breed experts. As in general for all livestock species, also in the equine species the widespread use of breeding animals of highly selected breeds represents a threat to the survival of local breeds and therefore to the maintenance of an adequate level of specific diversity. The presence on the Sicilian territory of these equine populations constitutes a precious reservoir of genetic variability that is particularly suited to supporting the increasing demand of the equestrian tourism sector. Therefore, the opportunity arises to identify the subjects currently reared to develop a qualitative conservation program, while contributing to the maintenance and exploitation of the territory. In this context, the genomic information and genealogical data have a crucial role in assisting the management of small populations with the prior target of planning correct matings and reducing the inbreeding rate.

.

Acknowledgements

This work has been possible thanks to a collaboration between "Dipartimento of Agricoltura, Alimentazione e Ambiente – Di3A (University of Catania)", "Istituto di Incremento Ippico per la Sicilia" and the horse breeders associations: Associazione Regionale Cavallo da Sella Siciliano Indigeno (A.R.A.C.S.I.), Associazione Nazionale Italiana Cavallo Orientale (ANICO), Associazione Allevatori Cavallo Sanfratellano (A.N.A.C.S.).

Conflict of interest statement

The authors declare no conflict of interest.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Al Abri, M. A., Brooks, S. A., Al-Saqri, N., Alkharousi, K., Johnson, E. H., Alqaisi, O., . . . Al Marzooqi, W. (2021). Investigating the population structure and genetic diversity of Arabian horses in Oman using SNP markers. *Anim Genet, 52*(3), 304-310. doi:10.1111/age.13056

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

Balbo, S. M. (1995). L'influenza dell'arabo-orientale sul cavallo siciliano. In R. R. Editrice (Ed.), *L'Asil Arabo il cavallo nobile d'Arabia* (pp. 169-175).

Bellone, R. R., Liu, J., Petersen, J. L., Mack, M., Singer-Berk, M., Drogemuller, C., . . . Lassaline, M. (2017). A missense mutation in damage-specific DNA binding protein 2 is a genetic risk factor for limbal squamous cell carcinoma in horses. *Int J Cancer, 141*(2), 342-353. doi:10.1002/ijc.30744

Biscarini, F., Cozzi, P., Gaspa, G., & Marras, G. (2018). detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. CRAN (The Comprehensive R Archive Network).

Bowling, A. T., & Ruvinsky, A. (2000). Genetic Aspects of Domestication, Breeds and Their Origins. In A. T. Bowling & A. Ruvinsky (Eds.), *The Genetics of the Horse*. New York, USA.: CABI Publishing.

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet, 19*(4), 220-234. doi:10.1038/nrg.2017.109

Chiofalo, L., Portolano, B., Liotta, L., Rundo Sotera, A., & Finocchiaro, R. (2003). Demographic characterization, inbreeding and genetic variability within Sanfratellano population horse from genealogical data. *Ital. J. Anim. Sci., 2*(sup1), 592-594. doi:10.4081/ijas.2003.11676086

Cosgrove, E. J., Sadeghi, R., Schlamp, F., Holl, H. M., Moradi-Shahrbabak, M., Miraei-Ashtiani, S. R., . . . Brooks, S. A. (2020). Genome Diversity and the Origin of the Arabian Horse. *Sci Rep, 10*(1), 9702. doi:10.1038/s41598-020-66232-1

Criscione, A., Moltisanti, V., Chies, L., Marletta, D., & Bordonaro, S. (2015). A genetic analysis of the Italian Salernitano horse. *Animal, 9*(10), 1610-1616. doi:10.1017/S1751731115001019

Curik, I., Ferencakovic, M., & Solkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livest Sci, 166*, 26-34. doi:10.1016/j.livsci.2014.05.034

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. *Mol Ecol Resour, 14*(1), 209-214. doi:10.1111/1755-0998.12157

Druml, T., Neuditschko, M., Grilz-Seger, G., Horna, M., Ricard, A., Mesaric, M., . . . Brem, G. (2018). Population Networks Associated with Runs of Homozygosity Reveal New Insights into the Breeding History of the Haflinger Horse. *J Hered, 109*(4), 384-392. doi:10.1093/jhered/esx114

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. *Mol Ecol Resour, 10*(3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x

Fogliata, G. (1910). *Tipi e razze equine in rapporto con la produzione equina in Italia, con l'aggiunta della produzione del mulo*. Pisa: Tipografia editrice cav. F. Mariotti.

Giontella, A., Cardinali, I., Lancioni, H., Giovannini, S., Pieramati, C., Silvestrelli, M., & Sarti, F. M. (2020). Mitochondrial DNA Survey Reveals the Lack of Accuracy in Maremmano Horse Studbook Records. *Animals, 10*(5). doi:10.3390/ani10050839

Grilz-Seger, G., Druml, T., Neuditschko, M., Dobretsberger, M., Horna, M., & Brem, G. (2019). High-resolution population structure and runs of homozygosity reveal the genetic architecture of complex traits in the Lipizzan horse. *BMC Genomics, 20*(1), 174. doi:10.1186/s12864-019-5564-x

Grilz-Seger, G., Druml, T., Neuditschko, M., Mesaric, M., Cotman, M., & Brem, G. (2019). Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. *Anim Genet, 50*(4), 334-346. doi:10.1111/age.12797

Grilz-Seger, G., Mesaric, M., Cotman, M., Neuditschko, M., Druml, T., & Brem, G. (2018). Runs of Homozygosity and Population History of Three Horse Breeds With Small Population Size. *J. Equine Vet. Sci., 71*, 27-34. doi:10.1016/j.jevs.2018.09.004

Grilz-Seger, G., Neuditschko, M., Ricard, A., Velie, B., Lindgren, G., Mesaric, M., . . . Druml, T. (2019). Genome-Wide Homozygosity Patterns and Evidence for Selection in a Set of European and Near Eastern Horse Breeds. *Genes (Basel), 10*(7). doi:10.3390/genes10070491

Guastella, A. M., Zuccaro, A., Criscione, A., Marletta, D., & Bordonaro, S. (2011). Genetic analysis of Sicilian autochthonous horse breeds using nuclear and mitochondrial DNA markers. *J Hered, 102*(6), 753-758. doi:10.1093/jhered/esr091

Hendricks, B. L. (1995). International encyclopedia of horse breeds. In N. University of Oklahoma Press, OK, USA (Ed.).

Hendrickson, S. L. (2013). A genome wide study of genetic adaptation to high altitude in feral Andean Horses of the paramo. *BMC Evol Biol, 13*, 273. doi:10.1186/1471-2148-13-273

Howrigan, D. P., Simonson, M. A., & Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics, 12*, 460. doi:10.1186/1471-2164-12-460

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc, 4*(1), 44-57. doi:10.1038/nprot.2008.211

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol, 23*(2), 254-267. doi:10.1093/molbev/msj030

Kim, E. S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassell, C. P., Crooker, B. A., . . . Sonstegard, T. S. (2013). Effect of artificial selection on runs of homozygosity in u.s. Holstein cattle. *PLoS One, 8*(11), e80813. doi:10.1371/journal.pone.0080813

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., . . . Macciotta, N. P. (2015). Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim Genet, 46*(2), 110-121. doi:10.1111/age.12259

Mastrangelo, S., Ciani, E., Sardina, M. T., Sottile, G., Pilla, F., Portolano, B., & Bi.Ov. Ita, C. (2018). Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. *Anim Genet, 49*(1), 71-81. doi:10.1111/age.12634

Metzger, J., Karwath, M., Tonda, R., Beltran, S., Agueda, L., Gut, M., . . . Distl, O. (2015). Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics, 16*, 764. doi:10.1186/s12864-015-1977-3

Meuwissen, T. (2009). Genetic management of small populations: A review. *Acta Agric. Scand. A Anim. Sci., 59*(2), 71-79. doi:10.1080/09064700903118148

Milanesi, M., Capomaccio, S., Vajana, E., Bomba, L., Garci, J. F., Ajmone-Marsan, P., & Colli, L. (2017). *BITE: an R package for biodiversity analyses*. bioRxiv 181610.

Nolte, W., Thaller, G., & Kuehn, C. (2019). Selection signatures in four German warmblood horse breeds: Tracing breeding history in the modern sport horse. *PLoS One, 14*(4), e0215913. doi:10.1371/journal.pone.0215913

Pereira, G. L., Chud, T. C. S., Bernardes, P. A., Venturini, G. C., Chardulo, L. A. L., & Curi, R. A. (2017). Genotype Imputation and Accuracy Evaluation in Racing Quarter Horses Genotyped Using Different Commercial SNP Panels. *J. Equine Vet. Sci., 58*, 89-96. doi:10.1016/j.jevs.2017.07.012

Petersen, J. L., Mickelson, J. R., Cothran, E. G., Andersson, L. S., Axelsson, J., Bailey, E., . . . McCue, M. E. (2013). Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One, 8*(1), e54997. doi:10.1371/journal.pone.0054997

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet, 81*(3), 559-575. doi:10.1086/519795

R Development Core Team. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL http://www.R-project.org. Retrieved from http://www.R-project.org

Schaefer, R. J., Schubert, M., Bailey, E., Bannasch, D. L., Barrey, E., Bar-Gal, G. K., . . . McCue, M. E. (2017). Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics, 18*(1), 565. doi:10.1186/s12864-017-3943-8

Staiger, E. A., Abri, M. A., Silva, C. A., & Brooks, S. A. (2016). Loci impacting polymorphic gait in the Tennessee Walking Horse. *Journal of Animal Science, 94*(4), 1377-1386. doi:10.2527/jas.2015-9936

Staiger, E. A., Albright, J. D., & Brooks, S. A. (2016). Genome-wide association mapping of heritable temperament variation in the Tennessee Walking Horse. *Genes Brain Behav, 15*(5), 514-526. doi:10.1111/gbb.12290

Valera, M., Molina, A., Gutiérrez, J. P., Gómez, J., & Goyache, F. (2005). Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. *Livest. Prod. Sci., 95*(1-2), 57-66. doi:10.1016/j.livprodsci.2004.12.004

Verrier, E., Audiot, A., Bertrand, C., Chapuis, H., Charvolin, E., Danchin-Burge, C., . . . Sabbagh, M. (2015). Assessing the risk status of livestock breeds: a multi-indicator method applied to 178 French local breeds belonging to ten species. *Anim. Genet. Resour., 57*, 105-118. doi:10.1017/s2078633615000260

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. *Evol Appl, 3*(3), 244-262. doi:10.1111/j.1752-4571.2009.00104.x

Waran, N. (2007). *The Welfare of Horses* (N. Waran Ed.): Springer, Dordrecht.

Wright, S. (1969). *Evolution and the genetics of populations: Vol.2 The Theory of Gene Frequencies.*: Univ. of Chicago Press, Chicago.

Wullaert, A., Bonnet, M. C., & Pasparakis, M. (2011). NF-kappaB in the regulation of epithelial homeostasis and inflammation. *Cell Res, 21*(1), 146-158. doi:10.1038/cr.2010.175

Zhang, C., Ni, P., Ahmad, H. I., Gemingguli, M., Baizilaitibei, A., Gulibaheti, D., . . . Zhao, S. (2018). Detecting the Population Structure and Scanning for Signatures of Selection in Horses (Equus caballus) From Whole-Genome Sequencing Data. *Evol Bioinform Online, 14*, 1176934318775106. doi:10.1177/1176934318775106

Zuccaro, A., Bordonaro, S., Criscione, A., Guastella, A. M., Perrotta, G., Blasi, M., . . . Marletta, D. (2008). Genetic diversity and admixture analysis of Sanfratellano and three other Italian horse breeds assessed by microsatellite markers. *Animal, 2*(7), 991-998. doi:10.1017/S1751731108002255

TABLES

Table 1. Breed's acronym, sample size (n.), expected heterozygosity (He), observed heterozygosity (Ho) with relative standard deviations (s.d), and effective population size (Ne) of the three Sicilian horses (Sanfretellano-SAN, Siciliano-SIC and Purosangue Orientale Siciliano-SOP), Arab (ARR) and Maremmano (MARM) breeds.

Breed	n.	He	s.d.	Ho	s.d.	Ne
ARR	24	0.297	0.170	0.292	0.182	195
MARM	24	0.317	0.154	0.382	0.175	294
SOP	12	0.277	0.184	0.315	0.231	10
SAN	17	0.300	0.163	0.314	0.191	31
SIC	17	0.323	0.147	0.333	0.175	400

Table 2. Breed's acronym and parameters' results of runs of homozygosity (ROH) analysis on Sanfratellano (SAN), Siciliano (SIC), Purosangue Orientale Siciliano (SOP), Arab (ARR) and Maremmano (MARM) samples. Parameters show mean values over individuals and chromosomes of the sum of ROH in Mb (S_{ROH}), of the number of detected ROHs (N_{ROH}), of the length of ROH in Mb (L_{ROH}), of inbreeding coefficient (F_{ROH}) with respective standard deviations (s.d.) and minimum and maximum values.

Breed	Parameters	Mean	s.d.	Min.	Max.
ARR	S_{ROH}	419.57	134.97	275.49	890.35
	N_{ROH}	120.58	56.63	45	281
	L_{ROH}	2.65	0.43	1.99	3.58
	F_{ROH}	0.19	0.06	0.12	0.40
MARM	S_{ROH}	223.06	48.27	140.98	314.92
	N_{ROH}	72.35	35.65	22	173
	L_{ROH}	2.43	0.41	1.71	3.04
	F_{ROH}	0.10	0.02	0.06	0.14
SOP	S_{ROH}	299.45	90.15	215.43	554.80
	N_{ROH}	53.06	26.18	10	118
	L_{ROH}	2.15	0.46	1.49	3.32
	F_{ROH}	0.13	0.04	0.10	0.25
SAN	S_{ROH}	205.63	38.53	135.53	281.53
	N_{ROH}	49.81	27.32	12	121
	L_{ROH}	2.26	0.50	1.38	3.27
	F_{ROH}	0.09	0.02	0.06	0.13
SIC	S_{ROH}	159.05	45.80	113.76	298.55
	N_{ROH}	45.77	24.01	9	118
	L_{ROH}	1.86	0.29	1.39	2.48
	F_{ROH}	0.07	0.02	0.05	0.13

Table 3. Breed's acronym and parameters' results of runs of homozygosity (ROH) analysis per class of ROH's length (in Mb) on Sanfratellano (SAN), Siciliano (SIC), Purosangue Orientale Siciliano (SOP), Arab (ARR) and Maremmano (MARM) samples. Parameters show the percentage distribution of ROHs (ROH%), inbreeding coefficient (F_{ROH}) and the F_{ROH} percentage incidence on total F_{ROH} (F_{ROH}%) per class of ROH's length (in Mb).

		Classes of ROH's length in Mb				
Breed	Parameters	0-2	2-4	4-8	8-16	>16
ARR	ROH %	64.10	22.80	7.90	3.80	1.40
	F_{ROH}	0.061	0.043	0.03	0.029	0.024
	F_{ROH}%	32.62	22.99	16.04	15.51	12.83
MARM	ROH%	71.90	17.00	7.30	2.70	1.00
	F_{ROH}	0.04	0.019	0.017	0.013	0.011
	F_{ROH}%	40.00	19.00	17.00	13.00	11.00
SOP	ROH%	72.00	20.20	5.50	1.50	0.80
	F_{ROH}	0.058	0.032	0.018	0.01	0.013
	F_{ROH}%	44.27	24.43	13.74	7.63	9.92
SAN	ROH %	77.80	13.50	5.00	2.30	1.40
	F_{ROH}	0.042	0.015	0.011	0.01	0.014
	F_{ROH}%	45.65	16.30	11.96	10.87	15.22
SIC	ROH%	77.90	15.80	4.50	1.40	0.40
	F_{ROH}	0.038	0.016	0.009	0.006	0.003
	F_{ROH}%	52.78	22.22	12.50	8.33	4.17

FIGURE LEGENDS

Figure 1. Genetic relationship defined with multidimensional scaling (MDS) analysis among Sicilian (Sanfretellano-SAN, Siciliano-SIC and Purosangue Orientale Siciliano-SOP) and other three horse breeds, Arab (ARR), Maremmano (MARM) and Norwegian Fjord (NORF). The individual spatial coordinates of 115 samples are plotted taking into account the first (x-axis) and the second component (y-axis) of the total variance. Only Sicilian horses are colour-plotted (SAN=red, SIC=blue, SOP=green).

Figure 2. NeighborNet phylogenetic network estimated from Reynolds' pairwise genetic distances calculated between the three Sicilian horses (Sanfretellano-SAN, Siciliano-SIC and Purosangue Orientale Siciliano-SOP), Maremmano (MARM), Arab (ARR) and Norwegian Fjord (NORF) breeds.

Figure 3. Circle plot representing K=2, K=3 and K=4 ancestral clusters inferred by Admixture analysis of Sicilian horses (Sanfratellano-SAN, Siciliano-SIC and Purosangue Orientale Siciliano-SOP), Arab (ARR) and Maremmano (MARM) breeds. The colours, which are consistent between the different K values, represent each of the genomic group to which the 94 individuals belong.

Figure 4. Scatter plot of individual (circles) inbreeding coefficient (F_{ROH}) and breeds' mean F_{ROH} (squares) estimated from runs of homozygosity (ROH) analysis of the three Sicilian horses (Sanfretellano-SAN, Siciliano-SIC and Purosangue Orientale Siciliano-SOP), Maremmano (MARM) and Arab (ARR) breeds. Y-axis represents F_{ROH} values' gradient, x-axis distributes the 94 individuals grouped per population (coloured alternatively black and white).

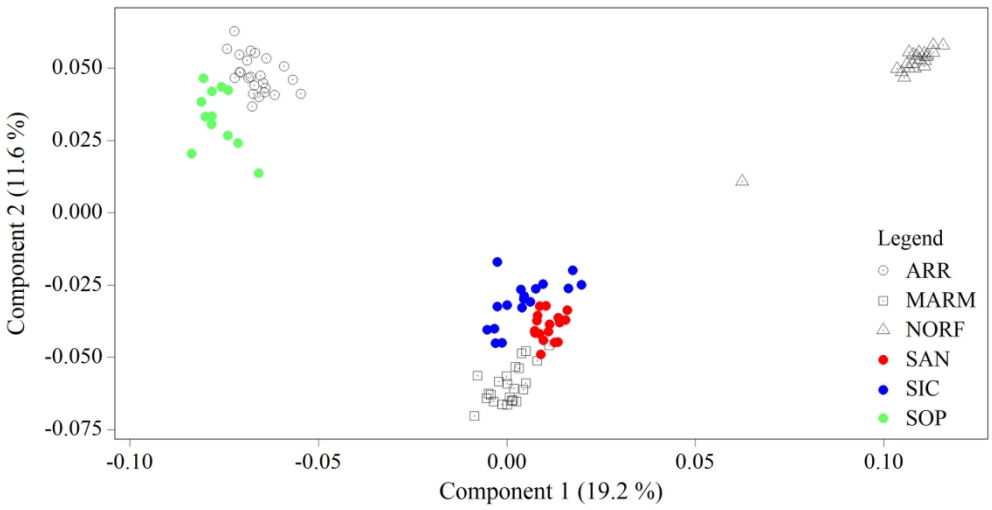

Fig.1

199x102mm (300 x 300 DPI)

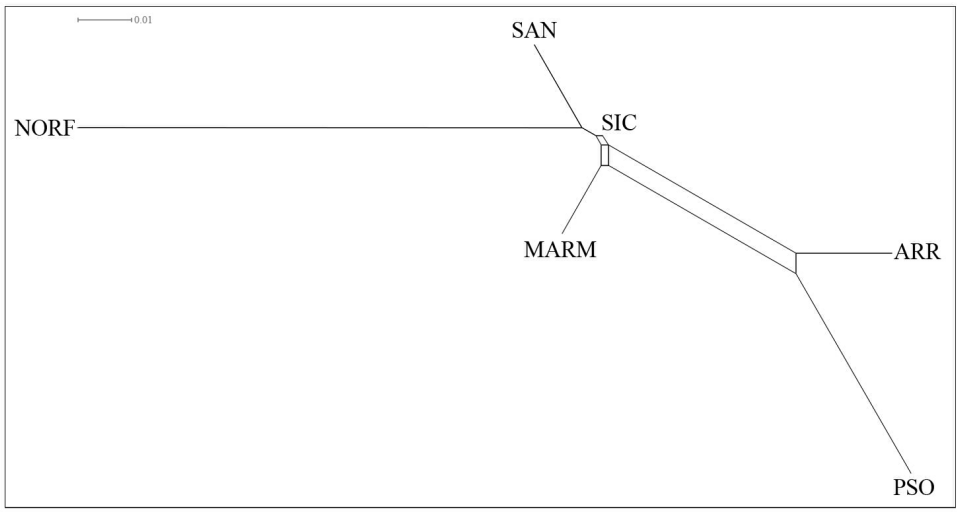

Fig.2

201x111mm (191 x 191 DPI)

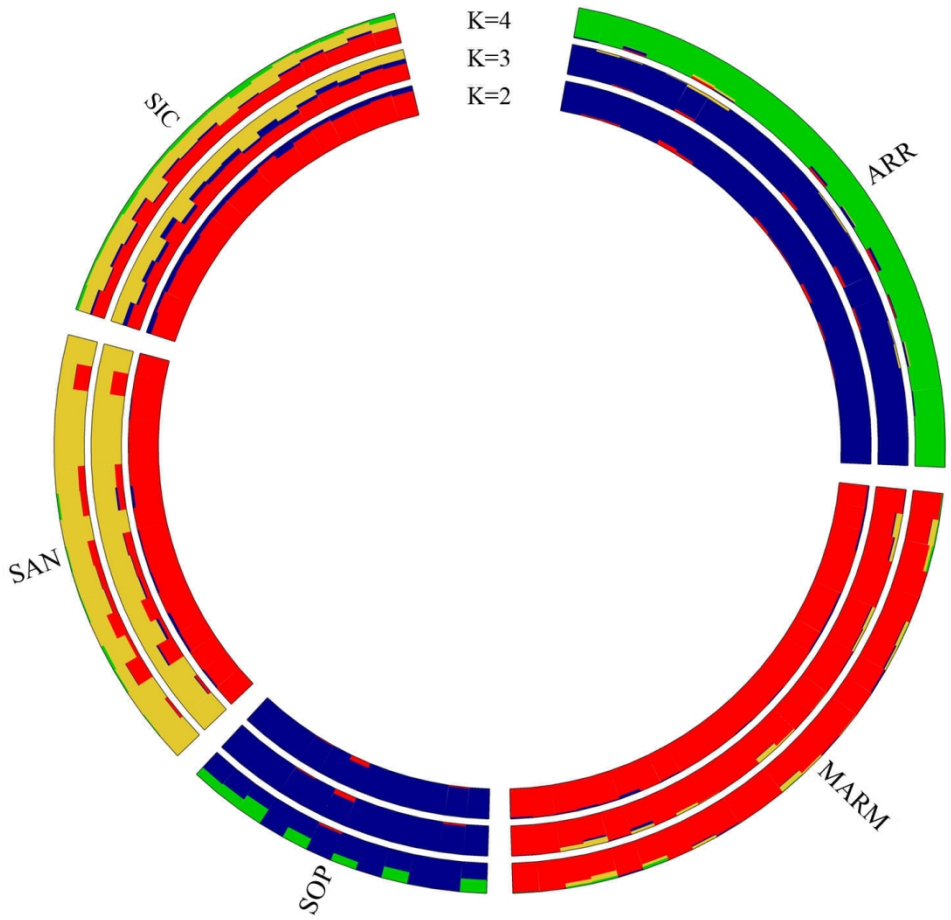

Fig.3

199x191mm (195 x 195 DPI)

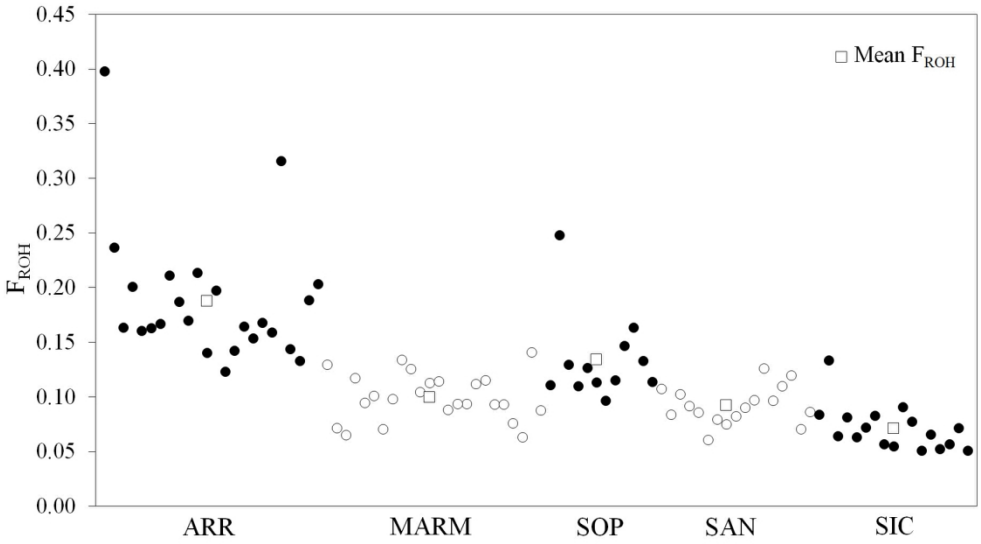

Fig.4

199x109mm (237 x 237 DPI)

Table S1. Runs of homozygosity (ROH) islands with a population frequency ⩾ 50% in Purosangue Orientale Siciliano (SOP), Arab (ARR), Sanfratellano (SAN), Siciliano (SIC) and Maremmano (MARM) samples. The table reports the chromosome (Chr), start and end (in bp), the number of SNPs, the annotated genes and the quantitative trait loci (QTL) associated to each ROH island, per breed.

Breed	Chr	Start (bp)	End (bp)	SNPs	Annotated Genes	QTL trait
SOP	1	127,474,273	128,704,337	18	*CSNK1G1, DAPK2, HERC1*	Osteochondrosis dissecans
	2	60,502,324	60,766,468	6		
	2	83,204,952	83,434,179	5	*DCLK2*	
	2	100,516,619	101,659,052	27	*HSPA4L, INTU*	
	2	105,334,216	105,441,404	2	*ENSECAG00000014829*	
	3	29,459,812	30,401,672	17	*PLCG2, ENSECAG00000012114, CDH13*	
	3	34,703,671	36,977,339	28	*GALNS, CDH15, SPG7, SLC9B2, ZC3H18*	Guttural pouch tympany, white markings, insect bite hypersensitivity
	3	41,779,288	42,076,097	10		Guttural pouch tympany
	3	51,318,319	51,566,821	3	*MAPK10, ARHGAP24*	
	3	59,371,429	59,766,260	9	*ENSECAG00000019008, CCDC158*	Withers height, white markings
	4	48,681,810	48,966,208	6		
	4	49,416,818	49,416,818	1		
	4	54,362,183	55,679,757	20	*IGF2BP3, TRA2A, CCDC126*	
	4	69,983,201	71,088,704	28	*ZNF277, BMT2*	
	4	74,481,553	75,753,615	32	*CFTR, CTTNBP2*	Male fertility
	5	53,875,750	54,030,212	4	*SYCP1, CSDE1*	
	6	1,200,345	2,122,149	20	*CPS1, ERBB5*	
	6	28,910,075	30,196,869	24	*ERC1*	
	6	34,928,006	36,213,188	22	*CLEC6A, ENSECAG00000024719, RIMKLB, PHC1, ENSECAG00000024571, A2M*	
	7	49,544,091	49,970,817	11	*ANGPTL6, COL5A3, OLFM2*	Alternate gaits
	7	49,974,860	53,032,339	49	*PIN1, OR7D2, ENSECAG00000009441*	
	7	53,101,032	54,218,962	26	*PANX1, HEPHL1, ENSECAG00000011283, ENSECAG00000018114, DEUP1, FAT3*	Alternate gaits
	7	96,262,423	96,314,375	2		
	8	21,896,311	24,528,856	51	*RNF34, KDM2B, ORAI1, RHOF, WDR66, MLXIP, KNTC1, PITPNM2, RILPL1, TCTN2, DNAH10*	Withers height, male fertility

8	33,243,738	34,901,335	25	*PTPRM,ENSECAG00000000531, RALBP1, RAB31, ENSECAG00000008508*	
8	35,725,364	38,656,972	64	*CEP192, PTPN2, PRELID3A, CIDEA, IMPA2, MPPE1, GNAL, NPC1, ENSECAG00000014999*	Withers height
8	38,832,157	39,474,433	14		Withers height
8	49,803,688	51,065,161	19		
8	53,749,645	54,124,770	6	*CCDC178*	
8	85,930,077	86,092,501	4		
8	86,117,039	86,364,039	8		
9	32,378,712	33,295,888	19	*SNTG1*	
9	43,736,188	45,743,060	35	*LAPTM4B, MATN2, POP1, STK3, VPS13B*	Temperament
9	45,982,954	46,122,801	4	*FBXO43*	
9	49,578,518	49,738,362	6	*LRP12*	
10	26,051,935	26,216,055	3	*DUXA*	
10	26,496,507	27,365,794	15	*ENSECAG00000013505*	Withers heigh
12	29,818,453	31,065,419	13	*NADSYN1, ENSECAG00000024486, KCNQ1*	Withers heigh
13	38,534,765	39,154,679	8	*DNAJA3, SRL, ADCY9, CREBBP, TRAP1*	
14	32,061,942	32,169,961	6		
14	52,318,224	52,700,641	6		
14	52,875,538	53,242,786	3		
14	61,204,489	61,973,905	12	*MAN2A1-201*	
15	37,421,951	40,570,157	51	*WDPCP, VPS54, MDH1, XPO1, PUS10, BCL11A, EHBP1, COMMD1, CCT4*	
16	72,540,233	73,393,693	19	*DZIP1L, DBR1, FAIM*	Withers height
17	19,151,572	20,399,417	21	*WDFY2, ENSECAG00000000195*	Insect bite hypersensitivity
18	47,965,811	50,793,288	54	*STK39, ERICH2, GORASP2, TLK1, METTL8, CYBRD1, CERS6, SPC25, LRP2, CCDC173, UBR3, MYO3B*	Altitude adaptation
19	2,251,539	4,000,799	25	*MLF1, PPM1L, OTOL1,*	
19	37,285,309	38,562,870	27	*CD86, ILDR1, SLC15A2, FBXO40, POLQ, STXBP5L, FSTL1*	Alternate gaits
20	40,213,287	41,520,518	16	*ENSECAG00000010368, BYSL, TRERF1*	Insect bite hypersensitivity
21	10,543,222	10,725,689	5	*IPO11*	
21	42,859,380	43,142,721	7		

		22	22,921,048	24,479,470	21	*TTLL9,ENSECAG00000013313, CCML2, HCK, POFUT1, ASXL11, NOLAL, SUN5, BPIFB2*	
		22	28,918,236	29,613,756	13	*ACTR5, DHX35*	
		23	5,210,643	5,473,131	5	*NTRK2*	
		23	25,631,961	25,707,521	2	*GLIS3*	Alternate gaits
		24	7,128,259	7,160,531	3	*ENSECAG00000024305*	
		24	20,554,036	22,028,066	36	*ZC2HC1C, TMED10, JDP2, TTLL5, TGFB3, ESRRB*	
		25	36,455,019	37,004,731	13	*RXRA, COL5A1*	
		27	9,978,838	10,960,672	20		Withers height
	ARR	1	5,153,748	5,153,748	1		
		2	99,963,374	101,909,442	44	*JADE1, INTU, HSPA4L*	
		2	102,209,678	102,209,678	1		
		2	102,695,553	102,695,553	1		
		2	103,589,120	103,595,092	2		
		3	22,462,831	23,837,815	29	*CALB2, HYDIN, VAC14, SF3B3, COG4, GLG1, RFWD3*	
		3	35,408,432	38,417,961	38	*BANK1, CDH15, SPG7, ENSECAG00000024558, ENSECAG00000007487, SLC9B2, MANBA, NFKB1, SLC39A8*	Guttural pouch tympany, white markings
		3	38,637,820	39,121,166	11	*PPP3CA*	
		4	50,011,701	50,011,701	1		
		4	54,362,183	55,679,757	20	*IGF2BP3, TRA2A, CCDC126*	
		4	94,066,308	94,066,308	1		
		4	94,198,392	94,198,392	1		
		5	22,315,621	23,121,913	14	*HMCN1, ENSECAG00000020648, TPR*	
		6	832,388	2,676,477	27	*KANSL1L, CPS1, ERBB4*	Insect bite hypersensitivity
		7	39,560,106	41,315,018	29	*NTM, OPCML*	
		7	46,673,819	53,032,339	88	*ENSECAG00000010608, ACP5, ANGPTL6, COL5A3, OLFM2, PIN1, OR7D2, ENSECAG00000009441*	Withers height, alternate gaits
		8	22,113,249	22,113,249	1		
		8	41,122,996	42,224,095	17	*CLUL1*	
		8	86,117,039	87,037,673	20		
		9	43,736,188	44,008,061	4	*LAPTM4B*	
		9	44,302,586	44,540,142	4	*POP1*	
		9	45,810,452	46,597,593	10	*RGS22, FBXO43, SNX31, ENSECAG00000020699*	
		10	23,498,268	23,696,143	8	*ENSECAG00000017969, ENSECAG00000018216, TTYH1*	

	10	26,496,507	27,584,754	17	*ENSECAG00000013505*	Withers height
	11	26,466,594	26,466,594	1		
	11	28,575,714	28,649,035	3		
	11	29,388,966	29,388,966	1		
	11	36,847,420	37,304,818	6	*HNF1B, HEATR6, ENSECAG00000024790*	
	13	38,534,765	39,934,602	25	*DNAJA3, SRL, ADCY9, CREBBP, TRAP1, SLX4, ENSECAG00000004481, ZNF213, ZNF205, PRSS33, SRRM2*	
	14	28,801,037	29,364,494	7	*SH3TC2, HTR4, FBXO38*	
	14	56,479,907	57,300,613	8	*COMMD10, CCDC112*	
	15	38,092,583	38,262,144	6	*WDPCP*	
	15	54,542,270	55,482,156	23	*LRPPRC, ABCG8, ABCG5, DYNC2LI1, THADA*	Number of progressively motile sperm, guttural pouch tympany
	15	61,359,879	61,576,938	2		
	15	61,951,648	62,377,510	10		
	15	64,311,775	65,374,455	17	*BIRC6, SPAST, MEMO1,*	
	15	70,724,205	71,990,181	23	*DNMT3A, ADCY3, NCOA1, ENSECAG00000014593, ATAD2B, KLHL29*	Guttural pouch tympany
	16	5,107,537	5,538,239	7	*SYN2, VGLL4*	
	18	10,340,383	11,395,285	15	*GLI2, PTPN4*	
	18	40,048,064	40,048,064	1		
	18	48,993,925	50,094,655	22	*CCDC173, UBR3, MYO3B,*	Altitude adaptation
	19	57,657,810	58,012,795	5	*EPHA6*	
	25	38,243,869	39,526,987	27	*RABL6, TRAF2, ENSECAG00000012141, NELFB, EXD3, PNPLA7, ZMYND19, EHMT1, CACNA1B*	Withers height
	26	40,613,690	40,702,991	2	*COL18A1*	
	30	28,307,647	28,307,647	1	*ENSECAG00000013767*	
	30	28,592,376	28,621,849	2		
	30	28,783,979	30,032,173	25	*NAV1, RNPEP, LGR6, PPP1R12B, KDM5B, KLHL12,*	
SIC	6	28,910,075	29,192,657	5		
	7	48,775,910	49,199,227	11	*ACP5*	Withers height
	7	49,972,297	51,565,018	38	*PIN1, OR7D2*	Alternate gaits
SAN	11	25,501,083	25,620,897	6	*FAM117A*	
	15	39,172,156	39,550,827	12	*COMMD1, CCT4*	
	17	29,819,882	30,634,556	21		Withers height
MARM	4	22,327,297	23,335,529	11		
	17	19,151,572	19,689,865	9	*WDFY2*	Insect bite hypersensitivity

Table S2. Markers within runs of homozygosity (ROH) islands with a population frequency ⩾ 50% in Sanfratellano (SAN), Siciliano (SIC), Purosangue Orientale Siciliano (SOP), Arab (ARR) and Maremmano (MARM) samples. Table reports reference SNP (rs) of markers, chromosome (Chr), chromosome location (bp), breed, reported associated traits and annotated genes.

rs	Chr	bp	Breed	Associated trait	Annotated gene
rs68607941	1	128,248,812	SOP	Osteochondrosis dissecans	*DAPK2*
rs68458700	3	34,786,570	SOP	Guttural pouch tympany	
rs68458763	3	35,140,330	SOP	Guttural pouch tympany	*ZC3H18*
rs68458777	3	35,279,221	SOP	Guttural pouch tympany	
rs68458808	3	35,408,432	SOP ARR	White markings; insect bite hypersensitivity; guttural pouch tympany	
rs68458828	3	35,502,786	SOP ARR	Guttural pouch tympany	
rs68458833	3	35,639,914	SOP ARR	White markings; guttural pouch tympany	
rs68458837	3	35,661,804	SOP ARR	White markings	*CDH15*
rs68596182	3	59,588,660	SOP	Withers height	
rs68596194	3	59,602,599	SOP	Withers height	
rs68598177	3	59,730,107	SOP	White markings	
rs68599438	3	41,842,702	SOP	Guttural pouch tympany	
rs68599460	3	41,928,376	SOP	Guttural pouch tympany	
rs68600833	3	42,002,976	SOP	Guttural pouch tympany	
rs68649649	3	36,840,265	SOP ARR	Guttural pouch tympany	*ENSECAG00000024558*
rs68649674	3	36,977,339	SOP ARR	White markings; guttural pouch tympany	*SLC9B2*
rs68649680	3	36,979,198	ARR	White markings	*SLC9B2*
rs68650924	3	37,148,447	ARR	Guttural pouch tympany; white markings	
rs68650979	3	37,398,507	ARR	Guttural pouch tympany; white markings	*NFKB1*
rs68651622	3	37,540,091	ARR	White markings; guttural pouch tympany	
rs68651668	3	37,635,869	ARR	Guttural pouch tympany	
rs68651688	3	37,725,713	ARR	White markings	
rs68651692	3	37,739,907	ARR	White markings	
rs68653018	3	37,806,266	ARR	White markings	*BANK1*
rs68653070	3	37,957,776	ARR	White markings	*BANK1*
rs68676998	3	38,157,802	ARR	Guttural pouch tympany	
rs68678317	3	38,344,351	ARR	White markings; guttural pouch tympany	
rs68678325	3	38,417,961	ARR	Guttural pouch tympany	
rs68678383	3	38,767,613	ARR	Guttural pouch tympany	
rs68679821	3	38,930,295	ARR	Guttural pouch tympany	

rs68679833	3	39,054,277	ARR	Sperm progressive motility; guttural pouch tympany	
rs69577536	4	75,575,124	SOP	Male fertility	
rs69578919	4	75,637,735	SOP	Male fertility	
rs68773137	6	1,212,183	ARR	Insect bite hypersensitivity	
rs68718309	7	53,950,659	SOP	Alternate gaits	
rs68838523	7	49,199,227	ARR SIC	Withers height	
rs68838528	7	49,414,532	ARR	Alternate gaits	
rs68838533	7	49,544,091	SOP ARR	Alternate gaits	
rs68838539	7	49,701,722	SOP ARR	Alternate gaits	
rs68910287	7	50,851,826	SOP ARR SIC	Alternate gaits	
rs68740994	8	22,640,886	SOP	Withers height	
rs68741974	8	24,411,688	SOP	Male fertility	
rs68796373	8	24,528,856	SOP	Male fertility	
rs68819080	8	38,851,571	SOP	Withers height	
rs68819093	8	38,853,229	SOP	Withers height	
rs68862716	8	36,632,827	SOP	Withers height	
rs68863891	8	37,153,931	SOP	Withers height	*CEP192*
rs68871178	9	45,279,882	SOP	Temperament	*VPS13B*
rs68997447	10	26,986,368	SOP ARR	Withers height	
rs68885394	12	31,065,419	SOP	Withers height	*KCNQ1*
rs68980399	15	71,588,357	ARR	Guttural pouch tympany	*ENSECAG00000014593*
rs69017496	15	54,632,135	ARR	Number of progressively motile sperm	
rs69018798	15	55,230,333	ARR	Guttural pouch tympany	*THADA*
rs69068932	16	73,393,693	SOP	Withers height	
rs69078496	16	73,235,184	SOP	Withers height	
rs69060192	17	30,209,609	SAN	Withers height	
rs69125108	17	19,689,865	SOP MARM	Insect bite hypersensitivity	*WDFY2*
rs69171012	18	49,758,616	SOP ARR	Altitude adaptation	*MYO3B*
rs69171035	18	50,094,655	SOP ARR	Altitude adaptation	
rs69250631	19	37,674,757	SOP	Alternate gaits	*FBXO40*
rs69250632	19	37,674,807	SOP	Alternate gaits	*FBXO40*
rs69250662	19	37,719,196	SOP	Alternate gaits	*POLQ*
rs69176260	20	41,031,989	SOP	Insect bite hypersensitivity	
rs69330873	23	25,631,961	SOP	Alternate gaits	*GLIS3*
rs69330876	23	25,707,521	SOP	Alternate gaits	*GLIS3*
rs69312200	25	38,546,385	ARR	Withers height	
rs69359229	27	10,676,851	SOP	Withers height	

Table S3. Gene Ontology (GO) and enrichment analysis based on annotated genes within ROH islands (frequency ⩾ 50%) in Purosangue Orientale Siciliano (SOP), Arab (ARR) and Sanfratellano (SAN) samples. The table reports the type of process involving genes (category), the GO analysis output (term), the significance level of the gene-term enrichment (p-value), genes involved in given term (genes), the measure of the enrichment's magnitude (Fold Enrichment) and the correction of significance levels for multiple observations (Bonferroni p-value).

Purosangue Orientale Siciliano - SOP horse annotated gene list

In the SOP breed from the 157 identified genes, GO analysis underlined that 65 genes were enriched in 27 biological processes, 27 genes in 9 molecular function and 12 genes in 4 cellular components.

Significant overrepresentation of biological processes related to eye morphogenesis and development, camera-type eye development, sensory organ morphogenesis, embryo development ending in birth or egg hatching, endocytosis and secretion, was found. Five GO molecular function related to the nucleotide-binding process showed significant corrected p-values (Bonferroni $p<0.05$). Further, two GO cellular component (GO:0098590; GO:0005588) and five KEGG pathways were significantly highlighted.

Category	Term	p-value	Genes	Fold Enrichment	Bonferroni p-value
Biological process	GO:0048592~eye morphogenesis	0.011	*NTRK2, COL5A1, KDM2B, MAN2A1, PTPRM*	5.72	1
	GO:0001654~eye development	0.015	*NTRK2, RXRA, COL5A1, KDM2B, MAN2A1, PTPRM, WDPCP*	3.48	1
	GO:0090596~sensory organ morphogenesis	0.022	*NTRK2, COL5A1, KDM2B, MAN2A1, PTPRM, WDPCP*	3.73	1
	GO:0043010~camera-type eye development	0.028	*NTRK2, RXRA, KDM2B, MAN2A1, PTPRM, WDPCP*	3.50	1
	GO:0046903~secretion	0.028	*PANX1, NTRK2, OLFM2, RAB31, ERBB4, TGFB3, ILDR1, STK39, CIDEA, CFTR, STXBP5L*	2.17	1
	GO:0006897~endocytosis	0.029	*HCK, RALBP1, RAB31, NPC1, PLCG2, LRP2, CSNK1G1*	2.97	1
	GO:0009792~embryo development ending in birth or egg hatching	0.034	*RNASEH2B, RXRA, INTU, KDM2B, POFUT1, MAN2A1, UBR3, BYSL, STK3*	2.37	1
	GO:0043067~regulation of programmed cell death	0.053	*NTRK2, RNF34, KDM2B, TGFB3, DAPK2, CIDEA, STK3, HCK, ERBB4, DNAJA3, PLCG2, PTPN2, FAIM*	1.79	1
	GO:1900118~negative regulation of execution phase of apoptosis	0.063	*RNF34, CIDEA*	30.30	1
	GO:0071404~cellular response to low-density lipoprotein particle stimulus	0.063	*NPC1, CDH13*	30.30	1
	GO:0032940~secretion by cell	0.069	*PANX1, NTRK2, OLFM2, RAB31, TGFB3, ILDR1, CIDEA, CFTR, STXBP5L*	2.05	1
	GO:0006464~cellular protein modification process	0.070	*RNF34, PPM1L, CCDC126, STK39, PTPRM, FBXO43, STK3, MAN2A1, ERBB4, ERC1,*	1.39	1

			RIMKLB, NTRK2, CREBBP, PHC1, KDM2B, TGFB3, DCLK2, HCK, TTLL5, NPC1, WDFY2, JDP2, TTLL9, PTPN2, CSNK1G1		
	GO:0036211~protein modification process	0.070	*RNF34, PPM1L, CCDC126, STK39, PTPRM, FBXO43, STK3, MAN2A1, ERBB4, ERC1, RIMKLB, NTRK2, CREBBP, PHC1, KDM2B, TGFB3, DCLK2, HCK, TTLL5, NPC1, WDFY2, JDP2, TTLL9, PTPN2, CSNK1G1*	1.39	1
	GO:0043069~negative regulation of programmed cell death	0.070	*NTRK2, HCK, RNF34, KDM2B, ERBB4, DNAJA3, PLCG2, CIDEA, FAIM*	2.04	1
	GO:0045909~positive regulation of vasodilation	0.071	*CPS1, PTPRM*	26.93	1
	GO:0071402~cellular response to lipoprotein particle stimulus	0.071	*NPC1, CDH13*	26.93	1
	GO:1903532~positive regulation of secretion by cell	0.072	*PANX1, TGFB3, ILDR1, CFTR, STXBP5L*	3.16	1
	GO:0009306~protein secretion	0.081	*PANX1, OLFM2, TGFB3, CIDEA, CFTR, STXBP5L*	2.58	1
	GO:0021953~central nervous system neuron differentiation	0.081	*NTRK2, HERC1, ERBB4, DCLK2*	3.91	1
	GO:0050714~positive regulation of protein secretion	0.082	*PANX1, TGFB3, CFTR, STXBP5L*	3.88	1
	GO:0051047~positive regulation of secretion	0.084	*PANX1, TGFB3, ILDR1, CFTR, STXBP5L*	2.98	1
	GO:0072358~cardiovascular system development	0.085	*NTRK2, RXRA, COL5A1, ERBB4, POFUT1, CDH13, PTPRM, WDPCP, LRP2, STK3*	1.85	1
	GO:0072359~circulatory system development	0.085	*NTRK2, RXRA, COL5A1, ERBB4, POFUT1, CDH13, PTPRM, WDPCP, LRP2, STK3*	1.85	1
	GO:0055098~response to low-density lipoprotein particle	0.086	*NPC1, CDH13*	22.03	1
	GO:0007423~sensory organ development	0.094	*NTRK2, RXRA, COL5A1, KDM2B, MAN2A1, PTPRM, WDPCP*	2.21	1
	GO:0042981~regulation of apoptotic process	0.096	*NTRK2, HCK, RNF34, KDM2B, ERBB4, TGFB3, DNAJA3, DAPK2, CIDEA, PTPN2, FAIM, STK3*	1.67	1
	GO:0060548~negative regulation of cell death	0.096	*NTRK2, HCK, RNF34, KDM2B, ERBB4, DNAJA3, PLCG2, CIDEA, FAIM*	1.90	1
Molecular function	GO:0032550 ~purine ribonucleoside binding	0.0002	*POLQ, TRAP1, NTRK2, DAPK2, STK39, DCLK2, SPG7, RHOF, SRL, NADSYN1, STK3, MAPK10, HCK, GNAL, RAB31, CPS1, ERBB4, DHX35, DNAJA3, TLK1,*	2.19	0.0155

			RIMKLB, CFTR, CSNK1G1, CCT4		
	GO:0032555~purine ribonucleotide binding	0.0003	*POLQ, TRAP1, NTRK2, DAPK2, STK39, DCLK2, SPG7, RHOF, SRL, NADSYN1, STK3, MAPK10, HCK, GNAL, RAB31, CPS1, ERBB4, DHX35, DNAJA3, TLK1, RIMKLB, CFTR, CSNK1G1, CCT4*	2.15	0.0195
	GO:0017076~purine nucleotide binding	0.0003	*POLQ, TRAP1, NTRK2, DAPK2, STK39, DCLK2, SPG7, RHOF, SRL, NADSYN1, STK3, MAPK10, HCK, GNAL, RAB31, CPS1, ERBB4, DHX35, DNAJA3, TLK1, RIMKLB, CFTR, CSNK1G1, CCT4*	2.15	0.0208
	GO:0032553~ribonucleotide binding	0.0003	*POLQ, TRAP1, NTRK2, DAPK2, STK39, DCLK2, SPG7, RHOF, SRL, NADSYN1, STK3, MAPK10, HCK, GNAL, RAB31, CPS1, ERBB4, DHX35, DNAJA3, TLK1, RIMKLB, CFTR, CSNK1G1, CCT4*	2.14	0.0217
	GO:0005524~ATP binding	0.0006	*POLQ, TRAP1, NTRK2, DAPK2, STK39, DCLK2, SPG7, NADSYN1, STK3, MAPK10, HCK, CPS1, ERBB4, DHX35, DNAJA3, TLK1, RIMKLB, CFTR, CSNK1G1, CCT4*	2.28	0.0437
	GO:0032559~adenyl ribonucleotide binding	0.0007	*POLQ, TRAP1, NTRK2, DAPK2, STK39, DCLK2, SPG7, NADSYN1, STK3, MAPK10, HCK, CPS1, ERBB4, DHX35, DNAJA3, TLK1, RIMKLB, CFTR, CSNK1G1, CCT4*	2.25	0.0521
	GO:0030554~adenyl nucleotide binding	0.0008	*POLQ, TRAP1, NTRK2, DAPK2, STK39, DCLK2, SPG7, NADSYN1, STK3, MAPK10, HCK, CPS1, ERBB4, DHX35, DNAJA3, TLK1, RIMKLB, CFTR, CSNK1G1, CCT4*	2.24	0.0544
	GO:0004672~protein kinase activity	0.0150	*MAPK10, NTRK2, HCK, ERBB4, DAPK2, STK39, DCLK2, TLK1, CSNK1G1, STK3*	2.53	0.6674
	GO:0004540~ribonuclease activity	0.0925	*DBR1, RNASEH2B, POP1*	5.78	0.9992
Cellular component	GO:0098590~plasma membrane region	0.012	*NTRK2, OLFM2, ERBB4, DNAJA3, STK39, CDH13, ORAI1, LRP2, CDH15, CFTR*	2.668	0.895
	GO:0005588~collagen type V trimer	0.026	*COL5A1, COL5A3*	74.535	0.994
	GO:0005583~fibrillar collagen trimer	0.085	*COL5A1, COL5A3*	22.360	1.000
	GO:0098643~banded collagen fibril	0.085	*COL5A1, COL5A3*	22.360	1.000

KEEG Pathway	ecb04611:Platelet activation	0.001	*ADCY9, COL5A1, LOC100054420, COL5A3, PLCG2, ORAI1*	7.049	0.192
	ecb04024:cAMP signaling pathway	0.010	*MAPK10, CREBBP, ADCY9, LOC100066125, ORAI1, CFTR*	4.401	0.799
	ecb05200:Pathways in cancer	0.013	*MAPK10, CREBBP, RALBP1, ADCY9, RXRA, TGFB3, DAPK2, PLCG2*	3.074	0.860
	ecb04020:Calcium signaling pathway	0.031	*ADCY9, GNAL, ERBB4, PLCG2, ORAI1*	4.105	0.992
	ecb05146:Amoebiasis	0.041	*GNAL, COL5A1, TGFB3, COL5A3*	5.115	0.998
	ecb05212:Pancreatic cancer	0.073	*MAPK10, RALBP1, TGFB3*	6.568	1.000
	ecb04976:Bile secretion	0.077	*ADCY9, RXRA, CFTR*	6.375	1.000
	ecb04971:Gastric acid secretion	0.085	*ADCY9, KCNQ1, CFTR*	6.021	1.000

Arab - ARR horse annotated gene list

The annotated gene list of the ARR breed harboured 109 genes, 35 of them were found to be enriched in 44 biological processes, 2 in 1 molecular function and 8 in 4 cellular components.

Gene ontology and enrichment analysis highlighted the overrepresentation of *ABCG8* and *ABCG5* genes, members of the superfamily of ATP-binding cassette (ABC) transporters, involved in the regulation of digestive system process (GO:0060457) and in several biological functions related to the regulation of intestinal lipid absorption.

Category	Term	p-value	Genes	Fold Enrichment	Bonferroni p-value
Biological process	GO:1904730~negative regulation of intestinal lipid absorption	0.011	*ABCG8, ABCG5*	184.75	1.00
	GO:0045796~negative regulation of intestinal cholesterol absorption	0.011	*ABCG8, ABCG5*	184.75	1.00
	GO:0010949~negative regulation of intestinal phytosterol absorption	0.011	*ABCG8, ABCG5*	184.75	1.00
	GO:0046903~secretion	0.016	*KDM5B, OLFM2, ERBB4, BANK1, ADORA1, CACNA1B, TRAF2, HNF1B, SYN2*	2.70	1.00
	GO:1904479~negative regulation of intestinal absorption	0.016	*ABCG8, ABCG5*	123.17	1.00
	GO:1905114~cell surface receptor signaling pathway involved in cell-cell signaling	0.017	*PPP3CA, ADORA1, PIN1, KLHL12, LGR6, NFKB1*	3.93	1.00
	GO:0032465~regulation of cytokinesis	0.026	*SPAST, PIN1, BIRC6*	11.79	1.00
	GO:0010629~negative regulation of gene expression	0.029	*KDM5B, BANK1, TPR, EHMT1, ZNF205, HNF1B, PABPC1, NFKB1, GLG1, ATAD2B*	2.24	1.00
	GO:0032372~negative regulation of sterol transport	0.032	*ABCG8, ABCG5*	61.58	1.00

	GO:0032375~negative regulation of cholesterol transport	0.032	*ABCG8, ABCG5*	61.58	1.00
	GO:0030299~intestinal cholesterol absorption	0.032	*ABCG8, ABCG5*	61.58	1.00
	GO:0010605~negative regulation of macromolecule metabolic process	0.033	*KDM5B, SF3B3, EHMT1, HNF1B, FBXO43, NFKB1, GLG1, ATAD2B, BANK1, TPR, PIN1, ZNF205, PABPC1*	1.90	1.00
	GO:0010564~regulation of cell cycle process	0.035	*SPAST, RFWD3, TPR, PIN1, BIRC6, FBXO43*	3.23	1.00
	GO:0032368~regulation of lipid transport	0.036	*KDM5B, ABCG8, ABCG5*	9.90	1.00
	GO:1904729~regulation of intestinal lipid absorption	0.037	*ABCG8, ABCG5*	52.79	1.00
	GO:0030300~regulation of intestinal cholesterol absorption	0.037	*ABCG8, ABCG5*	52.79	1.00
	GO:0060457~negative regulation of digestive system process	0.042	*ABCG8, ABCG5*	46.19	1.00
	GO:1904478~regulation of intestinal absorption	0.042	*ABCG8, ABCG5*	46.19	1.00
	GO:1900180~regulation of protein localization to nucleus	0.049	*PPP3CA, ERBB4, TPR, PIN1*	4.77	1.00
	GO:0051051~negative regulation of transport	0.050	*ABCG8, ABCG5, BANK1, TPR, ADORA1*	3.54	1.00
	GO:0042307~positive regulation of protein import into nucleus	0.051	*PPP3CA, ERBB4, TPR*	8.15	1.00
	GO:1904591~positive regulation of protein import	0.052	*PPP3CA, ERBB4, TPR*	8.03	1.00
	GO:0045947~negative regulation of translational initiation	0.057	*BANK1, TPR*	33.59	1.00
	GO:0098856~intestinal lipid absorption	0.057	*ABCG8, ABCG5*	33.59	1.00
	GO:0050728~negative regulation of inflammatory response	0.059	*ADORA1, ACP5, NFKB1*	7.49	1.00
	GO:0051253~negative regulation of RNA metabolic process	0.062	*KDM5B, TPR, EHMT1, ZNF205, HNF1B, PABPC1, NFKB1, ATAD2B*	2.22	1.00
	GO:0031327~negative regulation of cellular biosynthetic process	0.067	*KDM5B, BANK1, TPR, EHMT1, ACP5, ZNF205, HNF1B, NFKB1, ATAD2B*	2.03	1.00
	GO:0032695~negative regulation of interleukin-12 production	0.067	*ACP5, NFKB1*	28.42	1.00
	GO:1903047~mitotic cell cycle process	0.068	*PPP3CA, SPAST, TTYH1, RFWD3, TPR, FBXO43*	2.68	1.00
	GO:0051254~positive regulation of RNA metabolic process	0.070	*NCOA1, PPP3CA, CREBBP, ERBB4, PIN1, HNF1B, PABPC1, NFKB1, ATAD2B,*	2.02	1.00

	GO:0045935~positive regulation of nucleobase-containing compound metabolic process	0.070	*NCOA1, PPP3CA, CREBBP, SLX4, ERBB4, PIN1, HNF1B, PABPC1, NFKB1, ATAD2B*	1.90	1.00
	GO:0046824~positive regulation of nucleocytoplasmic transport	0.074	*PPP3CA, ERBB4, TPR*	6.60	1.00
	GO:0030177~positive regulation of Wnt signaling pathway	0.075	*PIN1, LGR6, NFKB1*	6.52	1.00
	GO:0000280~nuclear division	0.076	*SPAST, SLX4, TTYH1, TPR, FBXO43*	3.07	1.00
	GO:0032369~negative regulation of lipid transport	0.077	*ABCG8, ABCG5*	24.63	1.00
	GO:0030111~regulation of Wnt signaling pathway	0.078	*PIN1, HNF1B, LGR6, NFKB1*	3.93	1.00
	GO:0051248~negative regulation of protein metabolic process	0.088	*SF3B3, BANK1, TPR, PIN1, FBXO43, NFKB1, GLG1*	2.23	1.00
	GO:0043065~positive regulation of apoptotic process	0.090	*NCOA1, COL18A1, ERBB4, TRAF2, ZNF205*	2.89	1.00
	GO:0043068~positive regulation of programmed cell death	0.093	*NCOA1, COL18A1, ERBB4, TRAF2, ZNF205*	2.86	1.00
	GO:0008217~regulation of blood pressure	0.094	*RNPEP, ADORA1, CACNA1B*	5.71	1.00
	GO:0031348~negative regulation of defense response	0.097	*ADORA1, ACP5, NFKB1*	5.60	1.00
	GO:0099536~synaptic signaling	0.098	*PPP3CA, ADORA1, CACNA1B, HTR4, SYN2*	2.80	1.00
	GO:0099537~trans-synaptic signaling	0.098	*PPP3CA, ADORA1, CACNA1B, HTR4, SYN2*	2.80	1.00
	GO:0010942~positive regulation of cell death	0.099	*NCOA1, COL18A1, ERBB4, TRAF2, ZNF205*	2.79	1.00
Molecular function	GO:0017127~cholesterol transporter activity	0.065	*ABCG8, ABCG5*	29.08	1.00
Cellular component	GO:0043190~ATP-binding cassette (ABC) transporter complex	0.013	*ABCG8, ABCG5*	155.55	0.91
	GO:1902495~transmembrane transporter complex	0.037	*ABCG8, ABCG5, OLFM2, TTYH1, CACNA1B*	3.93	1.00
	GO:0098533~ATPase dependent transmembrane transport complex	0.073	*ABCG8, ABCG5*	25.93	1.00
	GO:0071013~catalytic step 2 spliceosome	0.085	*SRRM2, SF3B3, PABPC1*	6.06	1.00
KEEG Pathway	ecb04020:Calcium signaling pathway	0.017	*PPP3CA, ADCY9, ERBB4, CACNA1B, HTR4*	4.93	0.90
	ecb04024:cAMP signaling pathway	0.024	*CREBBP, ADCY9, ADORA1, HTR4, NFKB1*	4.40	0.97
	ecb04380:Osteoclast differentiation	0.033	*PPP3CA, ACP5, TRAF2, NFKB1*	5.50	0.99

	ecb04622:RIG-I-like receptor signaling pathway	0.050	*PIN1, TRAF2, NFKB1*	8.13	1.00
	ecb05166:HTLV-I infection	0.052	*VAC14, PPP3CA, CREBBP, ADCY9, NFKB1*	3.44	1.00
	ecb04976:Bile secretion	0.056	*ABCG8, ABCG5, ADCY9*	7.65	1.00
	ecb05200:Pathways in cancer	0.058	*CREBBP, ADCY9, TPR, TRAF2, NFKB1, GLI2*	2.77	1.00
	ecb05032:Morphine addiction	0.087	*ADCY9, ADORA1, CACNA1B*	5.91	1.00

Sanfratellano horse annotated gene list

In the SAN horses sample GO analysis revealed significant enrichment for *KAT7* and *CCT4* genes involved in biological process related to regulation of protein localization.

Category	Term	p-value	Genes	Fold Enrichment	Bonferroni p-value
Biological process	GO:1900180~regulation of protein localization to nucleus	0.027	*KAT7, CCT4*	48.47	0.85
	GO:1903829~positive regulation of cellular protein localization	0.042	*KAT7, CCT4*	31.18	0.95
	GO:1903827~regulation of cellular protein localization	0.066	*KAT7, CCT4*	19.82	0.99
	GO:0033365~protein localization to organelle	0.100	*KAT7, CCT4*	13.02	1.00

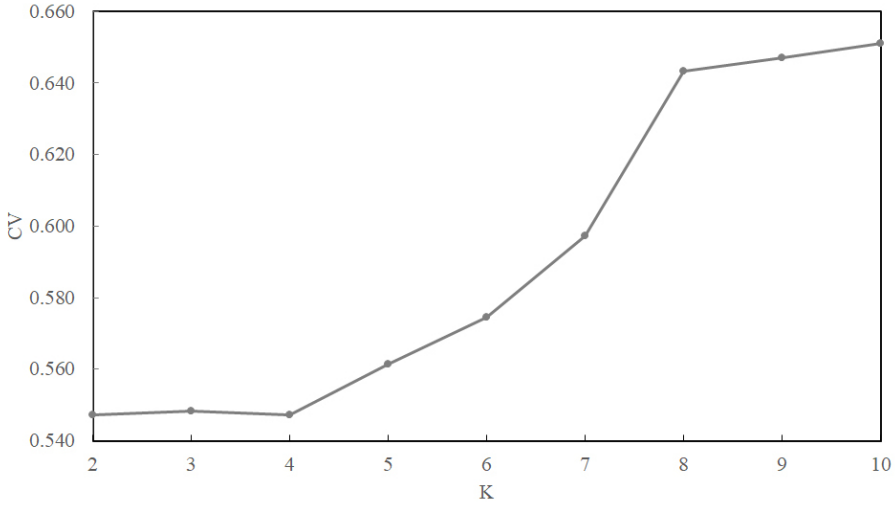

Figure S1. Distribution of the mean cross validation errors (y-axis) for each inferred K genomic cluster (x-axis).

199x121mm (133 x 133 DPI)

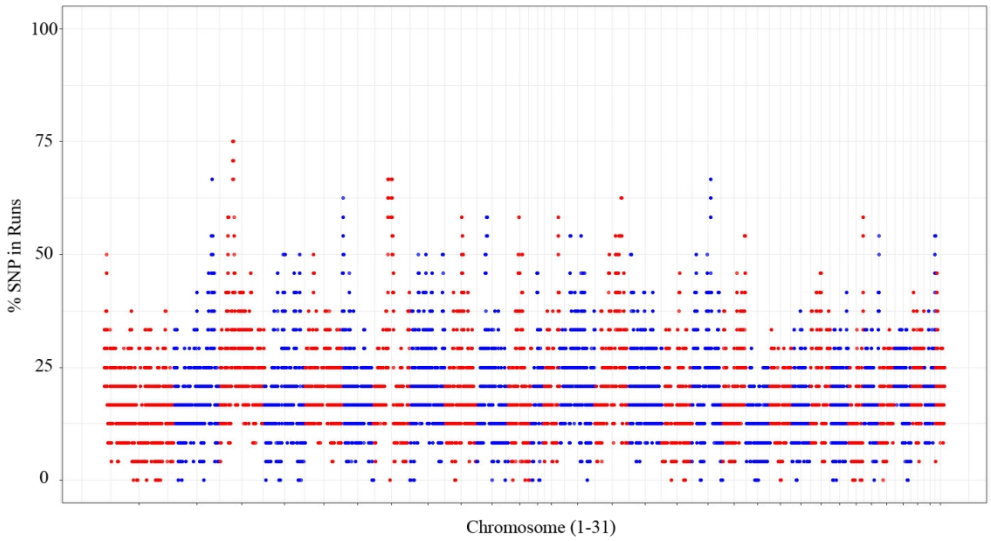

Figure S2. Manhattan plot of the proportion of time each SNP falls within a ROH (y-axis), over the 31 chromosome (x-axis), in the Arab horse (ARR). Dots, representing markers, are alternatively coloured in red and blue from chr1 to chr31.

383x238mm (96 x 96 DPI)

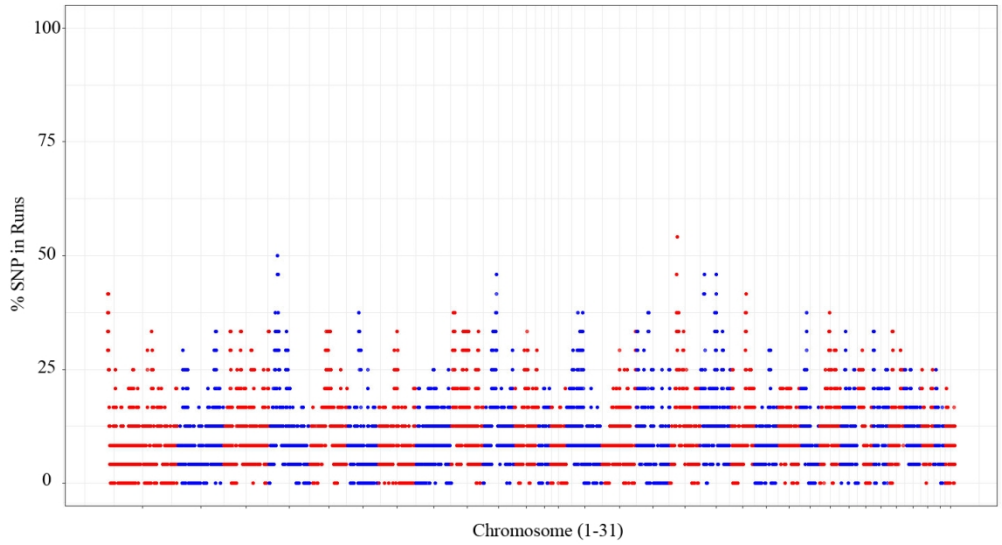

Figure S3. Manhattan plot of the proportion of time each SNP falls within a ROH (y-axis), over the 31 chromosome (x-axis), in the Maremmano horse (MARM). Dots, representing markers, are alternatively coloured in red and blue from chr1 to chr31.

199x123mm (183 x 183 DPI)

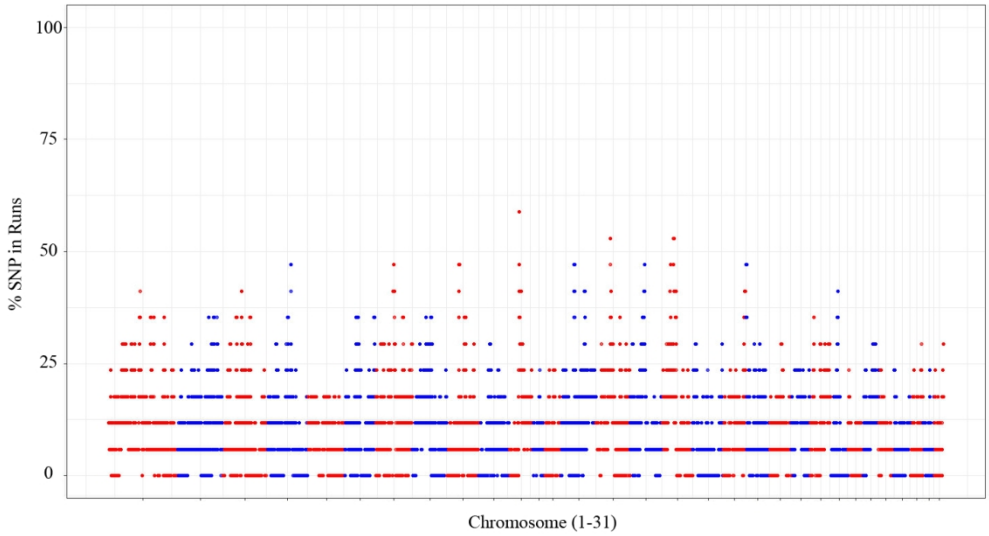

Figure S4. Manhattan plot of the proportion of time each SNP falls within a ROH (y-axis), over the 31 chromosome (x-axis), in the Sanfratellano horse (SAN). Dots, representing markers, are alternatively coloured in red and blue from chr1 to chr31.

386x234mm (96 x 96 DPI)

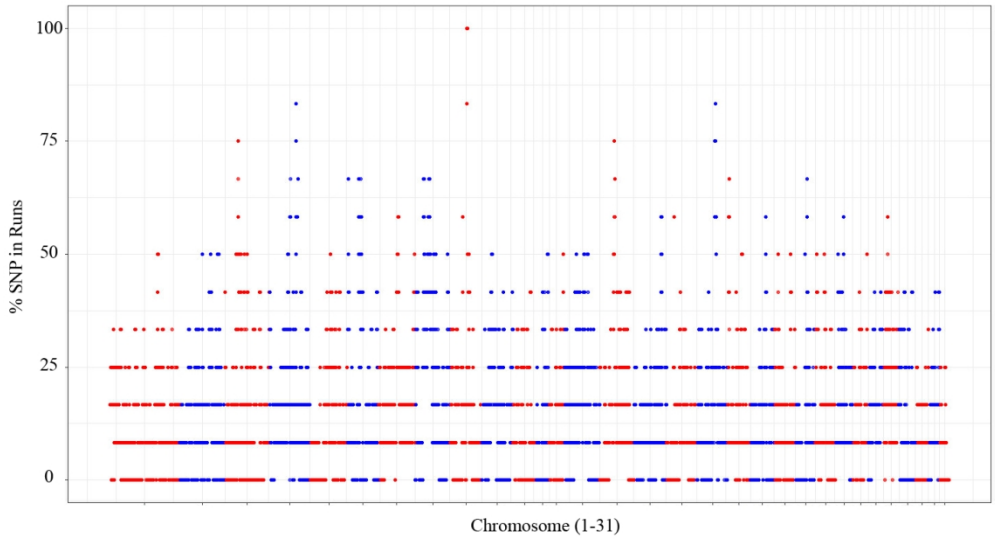

Figure S5. Manhattan plot of the proportion of time each SNP falls within a ROH (y-axis), over the 31 chromosome (x-axis), in the Purosangue Orientale Siciliano horse (SOP). Dots, representing markers, are alternatively coloured in red and blue from chr1 to chr31.

200x122mm (184 x 184 DPI)

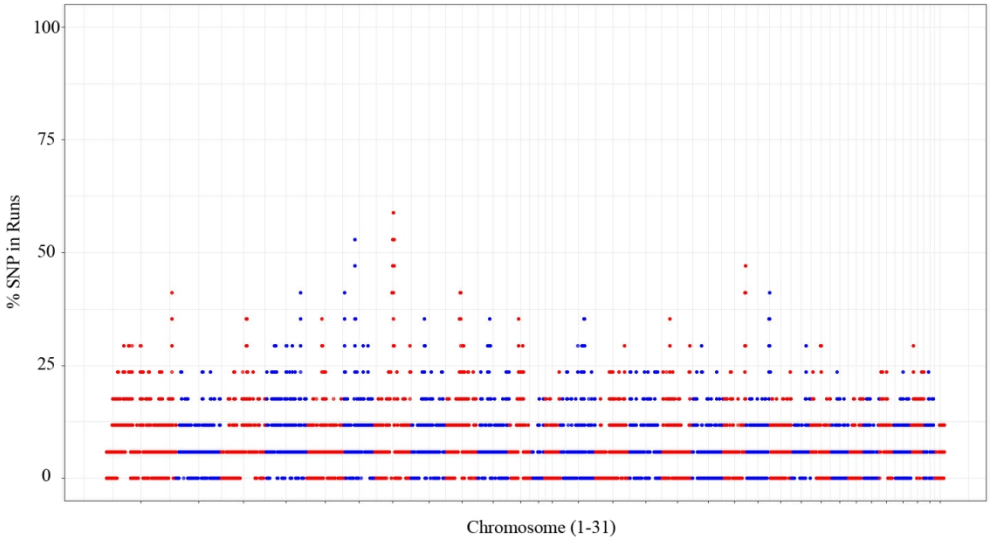

Figure S6. Manhattan plot of the proportion of time each SNP falls within a ROH (y-axis), over the 31 chromosome (x-axis), in the Siciliano horse (SIC). Dots, representing markers, are alternatively coloured in red and blue from chr1 to chr31.

200x122mm (184 x 184 DPI)