ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Runs of homozygosity provide a genome landscape picture of inbreeding and genetic history of European autochthonous and commercial pig breeds

This is the final peer-reviewed author’s accepted manuscript (postprint) of the following publication:

Published Version:

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(2), 155-170 [10.1111/age.13045].

Availability:

This version is available at: https://hdl.handle.net/11585/847526 since: 2022-01-25

Published:

DOI: http://doi.org/10.1111/age.13045

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

06 May 2026

This is the final peer-reviewed accepted manuscript of:

Schiavo, G., Bovo, S., Muñoz, M., Ribani, A., Alves, E., Araújo, J.P., Bozzi, R., Čandek-Potokar, M., Charneca, R., Fernandez, A.I., Gallo, M., García, F., Karolyi, D., Kušec, G., Martins, J.M., Mercat, M.-.-J., Núñez, Y., Quintanilla, R., Radović, Č., Razmaite, V., Riquet, J., Savić, R., Usai, G., Utzeri, V.J., Zimmer, C., Ovilo, C. and Fontanesi, L. (2021), *Runs of homozygosity provide a genome landscape picture of inbreeding and genetic history of European autochthonous and commercial pig breeds*,. Anim. Genet., 52: 155-170.

The final published version is available online at:

https://doi.org/10.1111/age.13045

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

ANIMAL GENETICS
Immunogenetics, Molecular Genetics and Functional Genomics

Runs of homozygosity provide a genome landscape picture of inbreeding and genetic history of European autochthonous and commercial pig breeds

Journal:	*Animal Genetics*
Manuscript ID	Draft
Manuscript Type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Schiavo, Giuseppina; University of Bologna, Dept. of Agricultural and Food Sciences, Division of Animal Sciences Bovo, Samuele; University of Bologna, Department of Agricultural and Food Sciences Ribani, Anisa; University of Bologna, Department of Agricultural and Food Sciences Gallo, Maurizio; Associazione Nazionale Allevatori Suini, ANAS Utzeri, Valerio; University of Bologna, Dept. of Agricultural and Food Sciences, Division of Animal Sciences Óvilo, Cristina; INIA, Mejora Genetica Animal Fontanesi, Luca; University of Bologna, Department of Agricultural and Food Sciences
Keywords:	Autozygosity, Population genomics, Selection signature, SNP, Sus scrofa

SCHOLARONE™
Manuscripts

Runs of homozygosity provide a genome landscape picture of inbreeding and genetic history of European autochthonous and commercial pig breeds

Giuseppina Schiavo[1], Samuele Bovo[1], Maria Muñoz[2], Anisa Ribani[1], Estefania Alves[2], José P. Araújo[3], Riccardo Bozzi[4], Marjeta Čandek-Potokar[5], Rui Charneca[6], Ana I. Fernandez[2], Maurizio Gallo[7], Fabián García[2], Danijel Karolyi[8], Goran Kušec[9], José Manuel Martins[6], Marie-José. Mercat[10], Yolanda Núñez[2], Raquel Quintanilla[11], Čedomir Radović[12], Violeta Razmaite[13], Juliette Riquet[14], Radomir Savić[15], Graziano Usai[16], Valerio J. Utzeri[1], Christoph Zimmer[17], Cristina Ovilo[2], Luca Fontanesi[1]

[1]Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale Giuseppe Fanin 46, 40127 Bologna, Italy

[2]Departamento Mejora Genética Animal, INIA, Crta. de la Coruña, km. 7,5, 28040, Madrid, Spain

[3]Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Viana do Castelo, Escola Superior Agrária, Refóios do Lima, 4990-706 Ponte de Lima Portugal

[4]DAGRI – Animal Science Section, Università di Firenze, Via delle Cascine 5, 50144 Firenze, Italy

[5]Kmetijski Inštitut Slovenije, Hacquetova 17, SI-1000 Ljubljana

[6] Instituto de Ciências Agrárias e Ambientais Mediterrânicas (ICAAM), Universidade de Évora, Polo da Mitra, Apartado 94, 7006-554 Évora, Portugal

[7]Associazione Nazionale Allevatori Suini (ANAS), Via Nizza 53, 00198 Roma, Italy

[8]Department of Animal Science, Faculty of Agriculture, University of Zagreb, Svetošimunska c. 25, 10000 Zagreb, Croatia

[9]Faculty of Agrobiotechnical Sciences, University of Osijek, Vladimira Preloga 1, 31000, Osijek, Croatia

[10]IFIP Institut du porc, La Motte au Vicomte, BP 35104, 35651, Le Rheu Cedex, France

For Peer Review

[11]Programa de Genética y Mejora Animal, IRTA, Torre Marimon, 08140 Caldes de Montbui, Barcelona, Spain

[12]Department of Pig Breeding and Genetics, Institute for Animal Husbandry, 11080 Belgrade-Zemun, Serbia

[13]Animal Science Institute, Lithuanian University of Health Sciences, Baisogala, Lithuania

[14]GenPhySE, Université de Toulouse, INRA, Chemin de Borde-Rouge 24, Auzeville Tolosane, 31326 Castanet Tolosan, France

[15]Faculty of Agriculture, University of Belgrade, Nemanjina 6, 11080, Belgrade-Zemun, Serbia

[16]AGRIS SARDEGNA, Loc. Bonassai, 07100 Sassari, Italy

[17]Bäuerliche Erzeugergemeinschaft Schwäbisch Hall, Schwäbisch Hall, Germany

Corresponding author: Luca Fontanesi

Email: luca.fontanesi@unibo.it

For Peer Review

Running head: Runs of homozygosity in pigs

Summary

Runs of homozygosity (ROH) are long stretches of DNA homozygous at each polymorphic position. The proportion of genome covered by ROH and their length are indicators of the level and origin of inbreeding. Frequent common ROH within the same population define ROH islands and indicate hotspot of selection. In this work, we investigated ROH in a total of 1131 pigs, from 20 European local pig breeds and in three cosmopolitan breeds, genotyped with the GGP Porcine HD Genomic Profiler. PLINK software was used to identify ROH. Size classes and genomic inbreeding parameters were evaluated. ROH Islands were defined by evaluating different thresholds of homozygous SNP frequency. A functional overview of breed-specific ROH islands was obtained via over-representation analyses of Gene Ontology biological processes. Mora Romagnola and Turopolje breeds had the largest proportion of genome covered with ROH (~1003 and ~955 Mb, respectively) whereas Nero Siciliano and Sarda breeds had the lowest proportion (~207 and 247 Mb, respectively). The highest proportion of long ROH (>16 Mb) was in Apulo-Calabrese, Mora Romagnola e Casertana. The largest number of ROH islands was identified in the Italian Landrace (n. 32), Cinta Senese (n. 26) and Lithuanian White Old Type (n. 22) breeds. Several ROH islands were in regions encompassing genes known to affect morphological traits. Comparative ROH structure analysis among breeds indicted similar genetic structure of local breeds across Europe. This study contributed to understand the genetic history of the investigated pig breeds and provided information to manage these pig genetic resources.

Keywords: Autozygosity; Population genomics; Selection signature; SNP; *Sus scrofa*

For Peer Review

Introduction

Conservation programs of animal genetic resources, mainly constituted by numerous autochthonous breeds in all species, are usually challenged by their very small effective population size which, in turn, tends to increase inbreeding and to reduce genetic variability (Charlesworth & Willis 2009). Inbreeding depression is considered the result of the increased level of autozygosity. Pedigree information is traditionally used to calculate the inbreeding coefficient (F_{PED}), defined as the probability that in a diploid individual, the maternal and the paternal derived alleles at a randomly selected locus are identical by descent (Wright 1922). This definition is equivalent to consider F_{PED} as the proportion of autozygosity of an individual's genome. Then, the level of inbreeding of a population is expressed by averaging all F_{PED} individual values. Reliability of F_{PED} calculated in autochthonous breeds is in general lower than what is possible to obtain for animals in commercial selection nuclei. This is mainly due to incomplete registration and incorrect recording of all mating events derived by the extensive production systems in which local breeds are usually raised (Gomez-Raya *et al.* 2008; Kios *et al.* 2012). In addition, it is clear that a few assumptions used to calculate this pedigree-based coefficient are not correct and are used as approximations in the methods of calculations: i) all founder animals of the base population are expected to be unrelated, but this condition cannot be evaluated and it is usually not respected; ii) recombinant events occurring during meiosis mix equally the individual's paternal and maternal haploid genome copies, but this condition mimics only average events and not what actually happens in each specific meiosis; and iii) there is no selection biases on any parts of the genome, but this assumption is not respected considering that directional artificial selection or natural selection play important roles in shaping the genome of many domestic animal breeds.

For Peer Review

Genome wide analyses, usually based on single nucleotide polymorphism (SNP) arrays, can be used to estimate the level of autozygosity of an animal genome by directly interrogating the genotype status at thousands of polymorphic sites (e.g. Kristensen *et al.* 2010). The proportion of the genome covered by runs of homozygosity (ROH) of a certain minimal length has been considered one of the most precise estimation of the level of autozygosity, providing a measure of genomic inbreeding (F_{ROH}; Peripolli *et al.* 2017). Runs of homozygosity are defined as continuous chromosome stretches in which all loci have a homozygous genotype (Gibson *et al.* 2006). Some ROH characteristics in a population (the average length of ROH, the average proportion of the genome covered by ROH and the patterns of ROH distribution across the chromosomes) are considered indicators of the origin and genetic history of a population (Ceballos *et al.* 2018). The high frequency of ROH in some chromosome regions identifies selection signatures derived from a reduced haplotype variability around loci under natural or artificial selection (i.e. ROH island or ROH hotspots). By applying different strategies and methods, ROH islands have been used to detect signatures of selection in several livestock species (Purfield *et al.* 2017; Bertolini *et al.* 2018; Grilz-Seger *et al.* 2018; Mastrangelo *et al.* 2018; Peripolli *et al.* 2018), including the pig (Zhang *et al.* 2018; Gorssen et al. 2020; Schiavo *et al.* 2020b).

A lot of different pig breeds have been developed through the combined action of artificial directional selection and natural pressures that contributed to shape a large reservoir of genetic diversity within the *Sus scrofa* species (Porter 1993). A large fraction of these genetic resources is however constituted by autochthonous breeds of small population size, usually well adapted to their local agro-climatic and environmental conditions but less productive, compared to cosmopolitan breeds or lines. Conservation programmes for these breeds, some of which considered unexplored genetic resources, have different levels of managing actions that range from advanced Herd Book structures with specific breeding and selection plans to preliminary voluntary farmer-based herd books or primitive conservation programmes (Čandek-Potokar & Nieto 2019). We recently analysed major and candidate gene markers in 20 autochthonous European pig breeds from several different countries and obtained preliminary population structure results (Muñoz *et al.* 2018) that were refined using SNP array information (Muñoz *et al.* 2019) and whole genome resequencing data (Bovo *et al.* 2020a, 2020b). Genome wide data indicated that average persistence and strength of linkage disequilibrium between markers and SNP based effective population size varied among breeds

For Peer Review

depending by the genetic structures and history of these breeds that experienced different genetic events (e.g. admixture, bottlenecks and genetic drift). Selection signatures were also obtained using F_{ST} statistics by analyzing SNP chip genotyping and sequencing data (Muñoz *et al.* 2019; Bovo *et al.* 2020a). Genomic inbreeding analyses in these breeds could add other information to refine their conservation programmes and identify appropriate strategies to control inbreeding level and infer other population structures or features.

In this study we analysed the same 20 European autochthonous pig breeds from nine different countries (Croatia, France, Germany, Italy, Lithuania, Portugal, Serbia, Slovenia and Spain) and other three cosmopolitan-derived breeds to obtain genomic inbreeding information from whole genotyping datasets by using ROH and other genomic approaches. We then evaluated the distribution of ROH in the genome of these breeds and identified putative selection hotspot regions that might be originated by different selection histories and structures of these pig genetic resources.

For Peer Review

Materials and methods

Animals

Pigs included in this study were from 20 autochthonous breeds distributed in nine European countries (Alentejana and Bísara from Portugal; Iberian and Majorcan Black from Spain; Basque and Gascon from France; Apulo-Calabrese, Casertana, Cinta Senese, Mora Romagnola, Nero Siciliano and Sarda from Italy; Krškopolje from Slovenia; Black Slavonian and Turopolje from Croatia; Moravka and Swallow-Bellied Mangalitsa from Serbia; Schwäbisch-Hällisches Schwein from Germany; Lithuanian indigenous wattle and Lithuanian White old type from Lithuania) and three commercial breeds (Italian large White, Italian Landrace and Italian Duroc). Analysed pigs were selected by avoiding highly related animals (no full- or half-sibs). All animals had standard breed characteristics and were registered to their respective Herd Books. Table S1 reports detailed descriptions of the investigated breeds and selected animals (Čandek-Potokar & Nieto 2019). Pictures of animals of the autochthonous breeds are reported in Muñoz *et al.* (2018, 2019) and Bovo et al. (2020a).

Genotyping of single nucleotide polymorphisms

All pigs (39-55 for each breed; Table S2) were genotyped with the GeneSeek ® GGP Porcine HD Genomic Profiler v1 (Illumina Inc, USA), which includes 68,516 SNPs evenly distributed with a median of 25 kb gap spacing. The average genotyping call rate was 0.94. Single nucleotide polymorphisms were mapped on the Sscrofa11.1 genome version, following the procedure already described (Fontanesi *et al.* 2012, 2014). Only autosomal SNPs located in unique positions were considered. Genotyping data were then filtered using PLINK software version 1.9 (Chang *et al.* 2015). Call rate of 0.90 and Hardy Weinberg equilibrium *P* of 0.001 were set as thresholds to keep SNPs. Although filtering for minor allele frequency (MAF) is necessary as best practice in most SNP chip analyses, this approach excludes the SNPs that are homozygous for the whole breed, therefore it could bring to an underestimation of the coverage in ROH (Meyermans *et al.* 2020). For this reason, we analysed ROH without applying any MAF pruning. For comparison with other studies that applied a MAF threshold and to evaluate the impact of MAF on the calculated ROH parameters, we also used a MAF threshold of 0.01 (indicated as method based on MAF > 0.01) and results are included in the Supplementary material. All analyses in the text are derived without MAF pruning (indicated as method based on MAF ≥ 0.00), if not stated otherwise. Animals were discarded if their call rate was <0.90. Table S2 reports the number of SNPs and animals considered for further analyses after filtering.

Multidimentional-plot analysis of pig breeds and effective population size

The first three dimensions for a multidimensional (MDS)-plot have been obtained with PLINK software version 1.9 and plotted with the R package "Scatterplot3d" (Ligges & Mächler 2003) to graphically visualize the genetic distances between the 23 pig breeds. Effective population size at

For Peer Review

recent and remote generations was computed using SNP data with the software *SNeP* (Barbato *et al.* 2015) with default parameters, except for the maximum distance in bp between SNPs to be analysed, that has been set to 10 Mb, and the binwidth for the calculation of linkage disequilibrium that was set to 100 kb.

Identification of runs of homozygosity

Runs of homozygosity (ROH) were identified using PLINK software version 1.9 (Chang *et al.* 2015). No pruning was performed based on linkage disequilibrium to avoid biases that could be derived by this practice (Marras *et al.* 2015; Meyermans *et al.* 2020) but a minimum length of 1 Mb was set to detect ROH. This threshold may exclude short and common ROH determined by markers in linkage disequilibrium, as previously demonstrated (e.g. Ferencakovic *et al.* 2013; Marras *et al.* 2015). The following parameters, already used by Schiavo *et al.* (2020a), were considered to call ROH: i) the minimum number of consecutive homozygous SNPs included in the ROH was 15; ii) the minimum length that constituted the ROH was 1 Mb; iii) the number of heterozygous SNPs that were allowed in the ROH was 0; iv) the minimum density of SNP in a genome window was 1 SNP every 100 kb; v) the maximum gap between consecutive SNPs was 1000 kb. ROH were placed into five size classes (Kirin *et al.* 2010; Ferenčaković *et al.* 2013a; Schiavo *et al.* 2020a): 1–2, 2–4, 4–8, 8–16 and >16 Mb, identified as ROH1–2 Mb , ROH2–4 Mb , ROH4–8 Mb , ROH8–16 Mb and ROH>16 Mb, respectively. The total number of ROH (nROH) was then obtained for each individual and for each length class. The average length of ROH (L_{ROH}, in Mb) and the sum of all ROH segments by animals (S_{ROH}, in Mb) were calculated. These parameters were also calculated for each breed by averaging individual data.

Genomic inbreeding measures

F_{ROH} was calculated for each pig as the proportion of the autosomal genome covered by ROH. F_{ROH} was calculated using all the detected ROH with length >1 Mb (F_{ROH1}) and also considering

For Peer Review

higher thresholds of length, namely >4 Mb, >8 Mb, >16 Mb to obtain, respectively, F_{ROH4}, F_{ROH8} and F_{ROH16} inbreeding coefficients. Averaged F_{ROH} values were calculated for each breed. In addition, chromosome (SSC) F_{ROH} (F_{ROHSSC}) values were also estimated for each breed: $F_{ROHSSC} = L_{ROHSSC}/L_{SSC}$ (Silió *et al.* 2013), in which L_{ROHSSC} is the total length of an individual's ROH in each SSC and L_{SSC} is the length of each chromosome covered by the involved SNPs.

Other genomic inbreeding coefficients were calculated: i) the variance-standardized relationship minus 1 (F_{hat1}); ii) the excess of homozygosity-based inbreeding estimate (F_{hat2}); iii) the estimate based on correlation between uniting gametes (F_{hat3}); iv) the values of the diagonal elements of the genomic relationship matrix, GRM (F_{GRM}; Van Raden *et al.* 2011); v) the difference between observed and expected number of homozygous genotypes (F_{HOM}). F_{hat1}, F_{hat2}, F_{hat3} and $F_{GRM.}$ GRM coefficients were calculated using PLINK1.9 with the ported functions of GCTA software v. 1.92 (Yang *et al.* 2011). F_{HOM} was computed with PLINK software version 1.9 (Chang *et al.* 2015). Pearson correlation coefficients (*r*) between all evaluated inbreeding coefficients were calculated.

For Peer Review

Identification of runs of homozygosity islands and annotation of genome regions

First, the proportion of SNPs residing within a ROH was calculated for a given breed by counting the amount of times a SNP appeared in a ROH within the given breed divided by the total number of genotyped pigs of that breed. Then, to call ROH islands a threshold of frequency should be defined. A few methods have been proposed for this purpose, each with pros and cons (e.g. Purfield *et al.* 2017; Grilz-Seger *et al.* 2018, Gorssen *et al.* 2020). However, there is no general agreement on their use in different contexts and populations. In this study, we used three methods to identify ROH islands that differed on the threshold that was applied.

One method already reported in other studies (Grilz-Seger *et al.* 2018, 2019a, 2019b) uses an empirical threshold defined as the percentage of animals (usually 50%), whiting a population, positive for a ROH at each tested SNP (hereinafter called 50% of animals-based threshold). When the level of inbreeding is high, the identification of islands due to signature of selection based on a fixed percentage of animals having ROH at each position of the genome might increase the number of false positive ROH islands that indicate the presence of signature of selection. This method could increase the risk of type II errors when the level of inbreeding in the population is low. Another method, frequently applied for this aim (e.g. Szmatoła *et al.* 2016; Purfield *et al.* 2017; Bertolini *et al.* 2018; Mastrangelo *et al.* 2018; Zhang *et al.* 2018), defines a percentile threshold (99th percentile) based on the top 1% of SNPs observed in a ROH in each breed (hereinafter called percentile-based threshold). Adjacent SNPs over this threshold are then merged into genomic regions corresponding to ROH hotspots. This method identifies always ROH islands as the threshold is defined on a percentile within the breed dataset and does not consider the structure of the population or its level of inbreeding.

Considering the problems that these two methods could have, we developed a third method where the identification of the threshold was chosen using a linear model in which the number of animals having SNPs in a ROH was a function of the average S_{ROH} level of the breed, which approximate the genomic inbreeding level of a population (hereinafter called S_{ROH} based-threshold). ROH islands were then considered in the text and annotated based on the results derived by this latter method. Results obtained with the other two methods were used for a comparative analysis. ROH co-occurrence between different breeds were investigated by comparing the average homozygosity level in each breed at each island region. For this evaluation, each ROH island identified in at least one breed was considered.

Similarity among breeds was investigated by computing a first matrix **A** (*n* breeds × *m* ROH islands regions identified across all the analyzed breeds) whose generic entry *a* is the average breed-specific frequency value of a given ROH island computed as follows: $a = \frac{\sum_i AF_i}{n}$, where AF_i is the allele frequency of the i^{th} SNP belonging to the ROH island and including *n* SNPs. This matrix was used to compute a similarity matrix D ($n \times n$), whose generic entry *d* is the Euclidean distance between pairs of breeds with values scaled in the range 0 to 1. A final dissimilarity matrix (1-D) was obtained

For Peer Review

and used to produce a heatmap in R (package *corrplot*; Wei and Simko, 2007) showing similarity among breeds.

Genes annotated in the Sscrofa11.1 pig genome version that mapped in the identified ROH islands were retrieved using the Ensembl Biomart tool (http://www.ensembl.org/biomart/martview/) and from NCBI Sscrofa11.1 GFF file. Functional enrichment analysis was carried out with *Enrichr* (Chen *et al.* 2013) via Fisher's exact test. Analyses run over the Biological Process (BP) branch of the Gene Ontology (GO) (Ashburner *et al.* 2000), by interrogating a total of 5103 functional terms covering 14433 human genes. Breed-specific analyses were run by using as input set the list of genes included in ROH islands. We considered as statistically over-represented terms those having: i) at least two input genes from two or more different ROH islands and ii) an adjusted *P* lower than 0.10.

Results

Genomic relationships among breeds and effective population size

Genomic information on the analysed breeds based on SNP data was graphically presented in a tri-dimensional MDS-plot (Figure S1). This plot showed that distinct groups of individuals were usually from the same breed. Several breeds were well separated from other groups. These distinct groups included breeds from several countries: Gascon and Basque from France; Italian Large White, Italian Duroc and Mora Romagnola from Italy; Iberian from Spain; Turopolje from Croatia. Most of the other breeds formed a continuous large cluster showing a general geographical distribution gradient as already reported in principal component analyses that included the same autochthonous breeds (Muñoz *et al.* 2019).

Effective population size (*Ne*) estimated with software *SNeP* for the 23 breeds is reported in Table S3. For all 20 autochthonous breeds, results confirmed the general low *Ne* for most breeds as already reported by Muñoz *et al.* (2019) who applied a similar estimation method. At 5 generations ago, breeds with the lowest *Ne* values were Turopolje, Mora Romagnola, Apulo-Calabrese and Casertana (*Ne* = 15, 16, 22 and 22, respectively). These breeds had the lowest estimated *Ne* also in

For Peer Review

the study of Muñoz *et al.* (2019) even if in different order. The autochthonous breeds with the largest *Ne* were Iberian, Nero Siciliano, Alentejana, Majorcan Black, Sarda and Bísara (*Ne* = 69, 68, 61, 58, 57 and 55, respectively). The commercial breeds had a higher *Ne* than all other remaining autochthonous breeds. In Italian Duroc, Italian Landrace and Italian Large White N_e at 5 generation ago was equal to 53, 59 and 61, respectively.

Runs of homozygosity in the investigated breeds

Table 1 (MAF ≥ 0.00) and Table S4 (MAF > 0.01) show the average size and average number of ROH (considering all ROH>1 Mb) per pig (average L_{ROH} and average nROH, respectively) and the average S_{ROH} values per animal in the 23 breeds. Minimum and maximum values for these three parameters are reported in Table S5. As expected, the parameters calculated without any MAF pruning were always higher than the parameters calculated using MAF >0.01. The breeds that had the highest mean nROH were Basque, Italian Duroc and Turopolje (n. 107, n. 104 and n. 80, respectively) and the breeds with the lowest mean nROH were Nero Siciliano (n. 24) Sarda (n. 27) and Moravka (n. 30). The mean L_{ROH} in all autochthonous breeds was larger than that of all three commercial breeds. Three Italian local breeds (Mora Romagnola, Apulo-Calabrese, and Casertana had the largest L_{ROH} (14.38, 14.21 and 12.63 Mb, respectively). Among the autochthonous breeds, the lowest L_{ROH} was observed in Alentejana (6.49 Mb), Iberian (6.50 Mb) and Majorcan Black (6.58 Mb). The maximum ROH length was observed in the largest chromosomes and reached 24.34 Mb in Mora Romagnola (SSC1), 23.36 Mb in Nero Siciliano (SSC1), 22.64 Mb in Moravka (SSC1) and 21.55 Mb in Apulo-Calabrese (SSC13). Mora Romagnola and Turopolje breeds had the largest mean S_{ROH} (a total of ~1003 and ~955 Mb, respectively) whereas Nero Siciliano and Sarda breeds had the lowest mean values for this parameter (~207 and ~247 Mb, respectively). The maximum S_{ROH} value was observed in one Mora Romagnola and one Black Slavonian pig that had about half of their genome covered by ROH (Table S5).

For Peer Review

Figure 1 shows the correlation plots between the S_{ROH} and the nROH values over the individual pigs in the 23 breeds. Basque and Gascon showed very homogeneous plots, indicating that most pigs of these two breeds had very similar within individual ROH parameters (nROH, L_{ROH} and S_{ROH}). The opposite was the heterogeneous distribution observed in the Apulo-Calabrese, Bísara, Casertana and Turopolje breeds (Figure 1).

Figure 2 reports the proportion of ROH of the five different length classes in each breed. Table S6 lists the corresponding values. The highest proportion of long ROH (>16 Mb) was in Apulo-Calabrese, Mora Romagnola e Casertana (about 25%, 23% and 23%, respectively). Apulo-Calabrese, Casertana, Mora Romagnola and Turopolje had the lowest proportion of short-medium ROH (ROH1-8). All three commercial breeds, Alentejana, Gascon, Iberian, Majorcan Black, Nero Siciliano, Lithuanian indigenous wattle, Lithuanian White Old Type and Schwäbisch-Hällisches had more than 50% of short ROH (ROH1-2 and ROH2-4).

For Peer Review

Genomic inbreeding parameters based on runs of homozygosity

Table 2 reports the mean and standard deviation of genomic inbreeding parameters calculated using ROH from different size classes in the 23 breeds. Mora Romagnola, Turopolje and Apulo Calabrese and Casertana were the autochthonous breeds with the highest F_{ROH} values, considering all ROH classes. For example, among these breeds F_{ROH1} ranged from 0.409 (Mora Romagnola) to 0.243 (Casertana). Among the commercial breeds, Italian Duroc had the highest F_{ROH} values. The lowest F_{ROH1} levels were observed in Nero Siciliano (0.085), Sarda (0.101) and Moravka (0.118).

When considering only medium-long ROH to calculate other ROH based inbreeding parameters (i.e. F_{ROH4}, F_{ROH8} and F_{ROH16}), the values decreased in all breeds, as expected. Among those with high F_{ROH1}, this drop was more evident in the breeds that had a high percentage of short ROH than in breeds that had many long ROH. For example, the Italian Duroc F_{ROH16} value was about 2.5 times lower than that of F_{ROH1} value whereas in Mora Romagnola, Turopolje, Apulo-Calabrese and Casertana their F_{ROH16} values decreased only 1.4-1.6 times compared to their respective F_{ROH1}

values. The distribution of the F_{ROH} values in the analysed breeds is shown in the boxplots of Figure 3.

The genome wide F_{ROH} information was also dissected by considering the average proportion of all ROH covering the different autosomes (F_{ROHSSC}). Among all breeds, Mora Romagnola and Turopolje had the highest F_{ROHSSC} values for 10 (SSC1, SSC4, SSC8, SSC9, SSC10, SSC13, SSC14, SSC15, SSC16 and SSC17) and 5 (SSC2, SSC3, SSC5, SSC6 and SSC11) chromosomes, respectively. Apulo-Calabrese had the highest F_{ROHSSC} values for SSC7 and SSC18 whereas Basque had the highest F_{ROHSSC} value for SSC12 (Figure S2).

Mean F_{ROH1}, F_{ROH4}, F_{ROH8} and F_{ROH16} breed values were negatively correlated with the estimated breed *Ne* values at 5 generation ago, defined as reported above (r = -0.685, -0.722, -0.737 and -0.716, respectively; P <0.0001).

Other genomic inbreeding parameters and their correlations with F_{ROH}

Other parameters that have been proposed as estimators of the level of genomic inbreeding were calculated in the 23 breeds (Table S8). The average F_{hat1} value was positive in only two breeds (Apulo-Calabrese and Sarda) and ranged from -0.320 (Mora Romagnola) to 0.010 (Sarda), with large within breed variability (the largest standard deviation was in Turopolje) and among breeds variability. These considerations could be also applied for the F_{GRM} parameter which is equivalent to F_{hat1} (even if scaled in a different way). The average F_{hat2} and F_{hat3} parameters had both the extreme values for the same breeds (Lithuanian indigenous wattle with the lowest values and Apulo-Calabrese with the highest values) with similar within and among breed variability (Table S8). The average F_{HOM} values were negative in 11 out or 23 breeds and ranged from -0.070 in Lithuanian Indigenous Wattle to 0.124 in Apulo-Calabrese. Turopolje had the largest standard deviation for this parameter (0.24). Distribution plots of the F_{hat1}, F_{hat2}, F_{hat3} and F_{HOM}, parameters in the analysed breeds are reported in Figure S3 and Figure S4.

For Peer Review

Correlations between all F_{ROH} parameters and all other genomic inbreeding measures for each breed are reported in Table S9. F_{HOM} had always very high and consistent correlations with the ROH based measures over all breeds. For example, correlations with F_{ROH1} and F_{ROH4} ranged from 0.819 and 0.814 for the Nero Siciliano breed to 0.987 and 0.982 for the Bísara breed. Correlations between F_{hat2} and F_{ROH1} and F_{ROH4} had some lower values even if again very high and consistent across breeds (they ranged from 0.447 or 0.450 in Swallow-Bellied Mangalitsa to 0.909 and 0.906 in Casertana). F_{hat1} and F_{hat3} showed inconsistent correlations compared to those of the other measures, including also negative values (Table S9). All these other genomic inbreeding measures had low negative correlations with *Ne* (from -0.11 to -0.18).

Run of homozygosity islands

Table 3 summarizes the number of ROH islands and the fraction of the genome covered by ROH islands identified using the S_{ROH} based-threshold in the 23 pig breeds. Figure 4 includes the Manhattan plots of a few breeds with extreme numbers of ROH islands. Figure 5 reports the pairwise similarities between breeds when overlapping ROH islands across breeds were considered. Some common features across breeds were evident.

The largest number of ROH islands was identified in the Italian Landrace (n. 34), Cinta Senese (n. 26) and Lithuanian White Old Type (n. 22) breeds. The largest covered fraction of the genome was observed in the Italian Duroc (92.85 Mb), Turopolje (80.82 Mb, with the largest averaged size of ROH islands) and Italian Landrace (75.03 Mb). No ROH islands were observed in Apulo-Calabrese and in Sarda breeds.

Table S10 compares the results obtained using the S_{ROH} based-threshold method with the results obtained using the other two methods considered in this study (the 50% of animals-based threshold and the percentile-based threshold methods, see Materials and methods). The Manhattan plots for all breeds and including the thresholds derived by the three methods is reported in Figure S5. Breeds with the highest level of genomic inbreeding estimated using F_{ROH} measures, like Mora Romagnola,

For Peer Review

Turopolje and Basque (Table 2), showed the highest number of ROH islands and the largest fraction of genome covered by ROH islands with the 50% of animals-based threshold method (n. 91 with 756 Mb in Mora Romagnola, n. 129 with 747 Mb in Turopolje and n. 93 in Basque with 312.9 Mb). Using the percentile-based threshold method, the number of ROH islands and the total length of the genome fractions covered by these regions were similar in all breeds and ranged from n. 7 (Mora Romagnola) to n. 20 (Italian Landrace) and from 19.83 Mb (Casertana) to 44.51 Mb (Turopolje). These methods could capture different information from the analysed populations. It seems however, that these two latter methods are, to some extent, biased by the genetic structure of the analysed populations and by the methodologies that are applied.

For Peer Review

The complete list of ROH islands identified in the investigated breeds, using the S_{ROH} based-threshold method, including the genes annotated in these regions, is reported in Table S11. Several breeds had ROH islands encompassing genes that are well known to affect exterior traits, that might contribute to differentiate these pig breeds. For example, Gascon and Turopolje had a ROH island on SSC6 which includes the *melanocortin 1 receptor* (*MC1R*) gene and Krškopolje and Turopolje had another ROH island on SSC8 which includes the *v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog* (*KIT*) gene. These two genes are well known to affect coat colour and colour patterns (Fontanesi & Russo 2013). Two genes that are known to affect vertebral number (*nuclear receptor subfamily 6 group A member 1*, *NR6A1* on SSC1; and *vertnin*, *VRTN* on SSC7; Mikawa *et al.* 2007, 2011) were in two ROH islands observed in Italian Landrace and in Schwäbisch-Hällisches breeds, respectively. Moravka and Schwäbisch-Hällisches breeds had a ROH island on SSC5 including the *methionine sulfoxide reductase B3* (*MSRB3*) gene whose variants have been associated with ear size in pigs (Chen *et al.* 2018; Bovo *et al*. 2020a). Cinta Senese and Italian Duroc had a ROH island including other genes that have been shown to affect body size (*caspase 10*, *CASP10*; and *non-SMC condensin I complex subunit G*, *NCAPG*; Rubin *et al.* 2012).

A functional overview of breed-specific ROH islands identified using the S_{ROH} based-threshold method was obtained via over-representation analyses of GO biological processes (Table S12). Few terms characterizing ROH islands were detected in two breeds (Krškopolje and Swallow-Bellied Mangalitsa) only. Terms were general and included pattern recognition receptor signaling pathway, toll-like receptor signaling pathway, zymogen activation, cellular response to radiation and negative regulation of cell differentiation.

Discussion

The demographic history of a population can be inferred using information from the average distribution, coverage, size and patterns of ROH that can be identified in the individuals belonging to the population using high density SNPs data (Ceballos *et al.* 2018). In this study we detected ROH in the genome of pigs from 20 autochthonous and three commercial breeds and compared the obtained ROH genome landscapes patterns. These breeds represent populations that derived from several countries and originated in different production systems that largely contributed to shape their genetic structures.

Combining different population genomic parameters calculated in this study it could be possible to reconstruct, to some extent, the genetic events and history that contributed to define the current genetic pools of the investigated breeds. ROH based fingerprinting are left in the analysed breeds and can be used to divide the 23 breeds in a few macro-groups that could have independently experienced similar genetic trajectories.

The ROH complement of recently inbred populations is defined by a large number of ROH with large size and a large fraction of the genome covered by ROH (high S_{ROH}), owing to very recent pedigree inbreeding loops, accompanied by a small *Ne*. The large S_{ROH} standard deviation indicates a low uniformity of the animals, that means that there might be different substructures or heterogeneity in the population or that an original bottleneck or founder effect could have increased the range of ROH size. Recent inbreeding features accompanied by a constituting bottleneck series of events can be clearly evidenced in a few Italian local breeds, i.e. Apulo-Calabrese, Casertana, Mora Romagnola, and in Turopolje. The high level of inbreeding could have masked regions that harbor

For Peer Review

selection of signatures as most of these breeds showed a low number of ROH islands (from zero to 7, considering the S_{ROH} based method; Table 3) apart Turopolje that seems to maintain a quite high level of ROH specific regions (n. 17; Table 3). These breeds need to be carefully managed to reduce or control the high level of inbreeding. Programmes in this direction are currently under way in the Italian breeds (ANAS, 2020).

Other breeds have a quite high S_{ROH} level but with short ROH indicating the occurrence of a past bottleneck and then a quite good isolation of the genetic pool. This is a case that can be observed in the two French breeds, Basque and Gascon, and in the Italian breed Cinta Senese. Differences in the three breeds are evident in the number of ROH islands that might indicate a low-medium level of specific signatures of selection in the French breeds (7 in the Basque that also had the largest number of nROH among the three - and 12 in the Gascon) and a high level of characterizing signatures in the Cinta Senese (26 ROH islands) probably due to different levels of selection pressures and adaptation of the three considered populations. A similar genetic history seems evident in the Italian Duroc breed (which however had a larger *Ne*; Table S3), reflecting deeper parental relatedness and consistent with an original strong bottleneck that occurred at the beginning of the 1990’ when the heavy pig selection programme was defined and differentiated the Italian Duroc breed from other Duroc lines (Bosi & Russo 2004).

Breeds that experienced recent admixtures had, in general, a low nROH and as a proportion, had a higher frequency of short-medium ROH than long ROH, with high *Ne*. This group included the two breeds that had nROH <30, S_{ROH} <300.00 Mb, and Ne >55, i.e. Nero Siciliano and Sarda for which the ROH derived landscape was in agreement with the large variability observed in candidate gene markers and SNP chip data (Muñoz *et al.* 2018, 2019). Other breeds (i.e. Alentejana, Black Slavonian, Krskopolje, Lithuanian indigenous wattle and Moravka) had similar ROH patterns with that described for these two Italian breeds even if not so extreme (nROH <40, S_{ROH} <350.00 Mb). They constitute a heterogeneous group of populations that might have experienced some moderate introgression over the period of their constitution or that these events occurred in the past and at

For Peer Review

present they maintain a moderate level of variability. The low-medium number of ROH islands (from 3, Moravka, to 15 Krskopolje) indicates a low-medium level of differentiation in terms of specific ROH features. Another group of intermediate breeds (which some features partially overlapping with those of the previous group) with medium nROH and, in general, with a medium level of inbreeding (nROH>40 and S_{ROH}>300) includes Bísara, Lithuanian White Old Type, Majorcan Black, Schwäbisch-Hällisches and Swallow-Bellied Mangalitsa.

Three other breeds, i.e. Iberian, Italian Landrace and Italian Large White, had characteristic ROH derived feature of commercial breeds or large populations, as expected from their large population size (consistent with the large *Ne*). The two Italian breeds had some indicators of more specific differentiations and signatures of selection with a higher number of nROH, lower *Ne* and larger fraction of the genome included in ROH islands than the Iberian breed. This fact could be also due to the high level of genetic diversity observed within the Iberian breed, sometime higher than in some European pig breeds (Fabuel *et al.* 2004). This is consistent with the structure of these three populations, with the two Italian breeds being derived by small selection nuclei specifically addressing a selection programme for heavy pigs. The presence of common features among breeds raised in different countries suggests that a few ROH islands might capture some adaptive features that are shared across populations and production systems.

The general picture depicted by the ROH profiles was able to summarize the main elements that characterize the population structure of the analysed breeds. For a few of them the potential burden derived by the ROH should be evaluated with attention. An increased homozygosity for (partially) recessive detrimental mutations maintained at low frequency in populations by mutation–selection balance has been suggested to be one of the main causes of inbreeding depression. Genomic inbreeding measures can help to manage all these pig populations. F_{ROH} based measures seems more appropriate than all other calculated parameters and are highly correlated with *Ne* indicating that they better reflect the population structure and then the effective inbreeding level of the animals, as we

For Peer Review

already reported comparing these measures with pedigree based inbreeding estimations (Schiavo *et al.* 2020a).

The method that we considered to identify ROH islands considers the level of inbreeding of the breeds to reduce the biases derived by the large fraction of the genome covered by ROH in highly inbreed populations and to increase the probability to capture signatures of selection able to explain morphological or adaptative features that characterize the uniqueness of these genetic resources. Some of the ROH islands contained genes responsible for domestication signatures related to exterior traits and morphological adaptation (i.e. coat colour genes: *MC1R* and *KIT*; Fontanesi & Russo 2013; vertebral number: *NR6A1* and *VRTN*, Mikawa *et al.* 2007, 2011; parts of the body and body size: *CASP10*, *MSRB3* and *NCAPG*; Rubin *et al.* 2012; Chen *et al.* 2018) indicating that fixation or increased frequency for some haplotypes containing breed specific alleles or features differentiating the domestic pool from wild boars could be captured by ROH.

Runs of homozygosity can complement other methods that have been applied to extract signatures of selection in these pig breeds (Muñoz *et al.* 2018, 2019; Bovo *et al.* 2020a, 2020b) and can provide additional information useful to design conservation plans and mating strategies to maintain the diversity of these pig genetic resources.

Acknowledgments

This work has been funded by University of Bologna RFO 2016-2019 programmes and by the European Union's Horizon 2020 research and innovation programme (TREASURE project - grant agreement No. 634476). The content of this work reflects only the authors' view and the European Union Agency is not responsible for any use that may be made of the information it contains.

Conflict of interests

The authors declare they do not have any competing interests.

For Peer Review

Data availability

Genotyping data of the autochthonous breeds can be shared after the signature of an agreement on their use with the TREASURE Consortium. Genotyping data of the commercial breeds can be shared after the signature of an agreement on their use with the University of Bologna.

References

ANAS 2019. Registro Anagrafico. Retrieved on 6th October 2020, from http://www.anas.it/.

Ashburner M., Ball C.A., Blake J. A., Botstein D., Butler H., Cherry J. M., Davis A.P., Dolinski K., Dwight S. S., Eppig J.T., Harris M. A., Hill D. P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J.C., Richardson J. E., Ringwald M., Rubin G.M. & Sherlock G. (2000) Gene ontology: Tool for the unification of biology. *Nature Genetics* **25**, 25–9.

Barbato M., Orozco-terWengel P., Tapio M. & Bruford M.W. (2015) SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics* **6**, 109.

Bertolini F., Cardoso T.F., Marras G., Nicolazzi E.L., Rothschild M.F. & Amills M. (2018) Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genetics Selection Evolution* **50**, 59.

Bosi P. & Russo V. (2004) The production of the heavy pig for high quality processed products. *Italian Journal of Animal Science* **3**, 309–21.

Bovo S., Ribani A., Muñoz M., Alves E., Araújo J.P., Bozzi R., Čandek-Potokar M., Charneca R., Di Palma F., Etherington G., Fernandez A.I., García F., García-Casco J., Karolyi D., Gallo M., Margeta V., Martins J.M., Mercat M.J., Moscatelli G., Núñez Y., Quintanilla R., Radović Č., Razmaite V., Riquet J., Savić R., Schiavo G., Usai G., Utzeri V.J., Zimmer C., Ovilo C., Fontanesi L. (2020a) 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.

For Peer Review

Bovo S., Ribani A., Muñoz M., Alves E., Araújo J.P., Bozzi R., Charneca R., Di Palma F., Etherington G., Fernandez A.I., García F., García-Casco J., Karolyi D., Gallo M., Gvozdanović K., Martins J.M., Mercat M.J., Núñez Y., Quintanilla R., Radović, Razmaite V., Riquet J., Savić R., Schiavo G., Škrlep M., Usai G., Utzeri V.J., Zimmer C., Ovilo C. & Fontanesi L. (2020b) 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**, 541–56.

Čandek-Potokar M. & Nieto L.R.M. (2019) European Local Pig Breeds - Diversity and Performance. A study of project TREASURE. IntechOpen. https://www.intechopen.com/books/european-local-pig-breeds-diversity-and-performance-a-study-of-project-treasure

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. *Nature Review Genetics* **19**, 220–34.

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

Charlesworth D. & Willis J.H. (2009) The genetics of inbreeding depression. *Nature Review Genetics* **10**, 793–6.

Chen E.Y., Tan C.M., Kou Y., Duan Q., Wang Z., Meirelles G.V., Clark N.R. & Ma'ayan A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 1–14.

Chen C., Liu C., Xiong X., Fang S., Yang H., Zhang Z., Ren J., Guo Y. & Huang L. (2018) Copy number variation in the MSRB3 gene enlarges porcine ear size through a mechanism involving miR-584-5p. *Genetics Selection Evolution* **50**, 72.

Fabuel E., Barragán C., Silió L., Rodríguez M.C. & Toro M.A. (2004) Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. *Heredity* **93**, 104–13.

For Peer Review

Ferenčaković M., Sölkner J. & Curik I. (2013) Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genetics Selection Evolution* **45**, 42.

Fontanesi L. & Russo V. (2013) Molecular genetics of coat colour in pigs. *Acta Agriculturae Slovenica* **Suppl. 4**, 15–20.

Fontanesi L., Galimberti G., Calò D.G., Fronza R., Martelli P.L., Scotti E., Colombo M., Schiavo G., Casadio R., Buttazzoni L. & Russo V. (2012) Identification and association analysis of several hundred single nucleotide polymorphisms within candidate genes for back fat thickness in Italian large white pigs using a selective genotyping approach. *Journal of Animal Science* **90**, 2450–64.

Fontanesi L., Schiavo G., Galimberti G., Calò D.G. & Russo V. (2014) A genomewide association study for average daily gain in Italian large white pigs. *Journal of Animal Science* **92**, 1385–94.

Gibson J., Morton N.E. & Collins A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15**, 789–95.

Gomez-Raya L., Priest K., Rauw W.M., Okomo-Adhiambo M., Thain D., Bruce B., Rink A., Torell R., Grellman L., Narayanan R. & Beattie C.W. (2008) The value of DNA paternity identification in beef cattle: Examples from Nevada’s free-range ranches. *Journal of Animal Science* **86**, 17–24.

Gorssen W., Meyermans R., Buys N. & Janssens S. (2020) SNP genotypes reveal breed substructure, selection signatures and highly inbred regions in Piétrain pigs. *Animal Genetics* **51**, 32–42.

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**, 174.

Grilz-Seger G., Druml T., Neuditschko M., Mesarič 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. *Animal Genetics* **50**, 334–46.

For Peer Review

Grilz-Seger G., Mesarič M., Cotman M., Neuditschko M., Druml T. & Brem G. (2018) Runs of homozygosity and population history of three horse breeds with small population size. *Journal of Equine Veterinary Science* **71**, 27–34.

Kios D., van Marle-Köster E. & Visser C. (2012) Application of DNA markers in parentage verification of Boran cattle in Kenya. *Tropical Animals and Health Prodution* **44**, 471–6.

Kirin M., McQuillan R., Franklin C.S., Campbell H., Mckeigue P.M. & Wilson J.F. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* **5**, e13996.

Kristensen T.N., Pedersen K.S., Vermeulen C.J. & Loeschcke V. (2010) Research on inbreeding in the "omic" era. *Trends in Ecology and Evolution* **25**, 44-52.

Ligges U. & Mächler M. (2003) Scatterplot3d - An R package for visualizing multivariate data. *Journal of Statistical Software* **8**, 1–20.

Marras G., Gaspa G., Sorbolini S., Dimauro C., Ajmone-Marsan P., Valentini A., Williams J.L. & Macciotta N.P.P. (2015) Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Animal Genetics* **46**, 110–21.

Mastrangelo S., Sardina M.T., Tolone M., Di Gerlando R., Sutera A.M., Fontanesi L. & Portolano B. (2018) Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. *Animal* **12**, 2480–8.

Meyermans R., Gorssen W., Buys N. & Janssens S. (2020) How to study runs of homozygosity using plink? A guide for analyzing medium density snp data in livestock and pet species. *BMC Genomics* **21**, 94.

Mikawa S., Morozumi T., Shimanuki S.I., Hayashi T., Uenishi H., Domukai M., Okumura N. & Awata T. (2007) Fine mapping of a swine quantitative trait locus for number of vertebrae and analysis of an orphan nuclear receptor, germ cell nuclear factor (NR6A1). *Genome Research* **17**, 586–93.

For Peer Review

Mikawa S., Sato S., Nii M., Morozumi T., Yoshioka G., Imaeda N., Yamaguchi T., Hayashi T. & Awata T. (2011) Identification of a second gene associated with variation in vertebral number in domestic pigs. *BMC Genetics* **12**, 5.

Muñoz M., Bozzi R., García F., Núñez Y., Geraci C., Crovetti A., García-Casco J., Alves E., Škrlep M., Charneca R., Martins J.M., Quintanilla R., Tibau J., Kušec G., Djurkin-Kušec I., Mercat M.J., Riquet J., Estellé J., Zimmer C., Razmaite V., Araújo J.P., Radović Č., Savić R., Karolyi D., Gallo M., Čandek-Potokar M., Fontanesi L., Fernández A.I. & Óvilo C. (2018) Diversity across major and candidate genes in European local pig breeds. *PLoS One* **13**, e0207475.

Muñoz M., Bozzi R., García-Casco J., Núñez Y., Ribani A., Franci O., García F., Škrlep M., Schiavo G., Bovo S., Utzeri V.J., Charneca R., Martins J.M., Quintanilla R., Tibau J., Margeta V., Djurkin-Kušec I., Mercat M.J., Riquet J., Estellé J., Zimmer C., Razmaite V., Araújo J.P., Radović Č., Savić R., Karolyi D., Gallo M., Čandek-Potokar M., Fernández A.I., Fontanesi L. & Óvilo C. (2019) Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip. *Scientific Reports* **9**, 13546.

Peripolli E., Metzger J., De Lemos M.V.A., Stafuzza N.B., Kluska S., Olivieri B.F., Feitosa F.L.B., Berton M.P., Lopes F.B., Munari D.P., Lôbo R.B., Magnabosco C.D.U., Di Croce F., Osterstock J., Denise S., Pereira A.S.C. & Baldi F. (2018) Autozygosity islands and ROH patterns in Nellore lineages: Evidence of selection for functionally important traits *BMC Genomics* **19**, 680.

Peripolli E., Munari D.P., Silva M.V.G.B., Lima A.L.F., Irgang R. & Baldi F. (2017) Runs of homozygosity: current knowledge and applications in livestock. *Animal Genetics* 48, 255–71

Porter V. (1993) *Pigs: A handbook to the breeds of the world.* Comstock Publishing Associates, Cornell University Press, Ithaca, New York, 256 pp., ISBN 0-8014-2920-X.

Purfield D.C., McParland S., Wall E. & Berry D.P. (2017) The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One* **12**, e0176780.

Rubin C.J., Megens H.J., Barrio A.M., Maqbool K., Sayyab S., Schwochow D., Wang C., Carlborg Ö., Jern P., Jørgensen C.B., Archibald A.L., Fredholm M., Groenen M.A.M. & Andersson L.

For Peer Review

(2012) Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 19529–36.

Schiavo G., Bovo S., Bertolini F., Tinarelli S., Dall'Olio S., Nanni Costa L., Gallo M. & Fontanesi L. (2020a) Comparative evaluation of genomic inbreeding parameters in seven commercial and autochthonous pig breeds. *Animal* **14**, 915–20.

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

Silió L., Rodríguez M.C., Fernández A., Barragán C., Benítez R., Óvilo C. & Fernández A.I. (2013) Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *Journal of Animal Breeding and Genetics* **130**, 349–60.

Szmatoła T., Gurgul A., Ropka-Molik K., Jasielczuk I., Zabek T. & Bugno-Poniewierska M. (2016) Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. *Livestock Science* **188**, 72–80.

VanRaden P.M., Olson K.M., Null D.J. & Hutchison J.L. (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of Dairy Science* **94**, 6153–61.

VanRaden P.M. (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414–23.

Wei T. & Simko V. (2017) R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot.

Wright S. (1922) Coefficients of Inbreeding and Relationship. *American Naturalist* **56**, 330–8.

Yang J., Lee S.H., Goddard M.E. & Visscher P.M. (2011) GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**, 76–82.

Zhang Z., Zhang Q., Xiao Q., Sun H., Gao H., Yang Y., Chen J., Li Z., Xue M., Ma P., Yang H., Xu N., Wang Q. & Pan Y. (2018) Distribution of runs of homozygosity in Chinese and Western pig breeds evaluated by reduced-representation sequencing data. *Animal Genetics* **49**, 579–91.

For Peer Review

653

For Peer Review

Tables

Table 1. Runs of homozygosity (ROH) parameters calculated in the 23 pig breeds obtained without any pruning for minor allele frequency (MAF), i.e. MAF ≥0.00. Parameters calculated using MAF >0.01 are reported in Table S4.

Breed	**Acronym**	**nROH (SD)**[1]	**L_{ROH} (SD)**[2]	**S_{ROH} (SD)**[3]
Alentejana	AL	50.90 (10.67)	6.49 (2.48)	339.97 (167.31)
Apulo-Calabrese	AC	56.74 (11.67)	14.21 (3.60)	813.75 (266.55)
Basque	BA	106.62 (9.36)	7.21 (1.13)	764.56 (105.38)
Bísara	BI	43.88 (12.93)	7.59 (2.67)	352.18 (211.11)
Black Slavonian	BS	36.61 (14.72)	8.75 (3.29)	336.98 (230.97)
Casertana	CA	45.34 (11.20)	12.63 (4.04)	595.06 (268.90)
CintaSenese	CS	55.62 (15.47)	7.75 (2.28)	424.32 (144.99)
Gascon	GA	75.08 (8.52)	6.97 (1.06)	522.14 (89.18)
Iberian	IB	51.38 (11.97)	6.50 (2.25)	341.52 (148.95)
Krškopolje	KR	34.96 (7.36)	8.62 (2.72)	306.47 (138.31)
Lithuanian indigenous wattle	LIW	42.69 (7.07)	7.69 (1.74)	330.44 (98.97)
Lithuanian White Old Type	LWOT	56.27 (10.16)	6.59 (1.82)	373.55 (133.34)
Majorcan Black	MB	48.50 (10.47)	6.58 (1.95)	327.89 (147.08)
Mora Romagnola	MR	70.35 (7.37)	14.38 (2.48)	1003.13 (139.75)
Moravka	MO	30.14 (12.34)	8.48 (4.36)	289.36 (220.73)
Nero Siciliano	NS	24.15 (10.00)	7.30 (4.91)	207.33 (208.19)
Sarda	SA	27.46 (10.26)	7.77 (4.70)	246.77 (221.24)
Schwäbisch-Hällisches	SHS	49.14 (6.63)	7.28 (2.13)	360.16 (123.64)
Swallow-Bellied Mangalitsa	SBMA	49.96 (8.11)	9.75 (2.04)	483.27 (115.50)
Turopolje	TU	79.76 (15.31)	11.91 (1.78)	955.04 (242.37)
Italian Duroc	IDU	104.00 (10.49)	6.33 (1.03)	655.35 (106.75)
Italian Landrace	ILA	65.56 (8.86)	5.27 (1.08)	347.80 (92.75)
Italian Large White	ILW	62.46 (12.90)	5.52 (1.00)	349.22 (107.11)

[1] nROH: the average total number of ROH and the standard deviation (SD) calculated for each breed.

[2] L_{ROH}: the average length of ROH (in Mb) considering all length classes and the standard deviation (SD) calculated for each breed.

[3] S_{ROH}: the average sum of all ROH segments (in Mb) by animals considering all length classes and the standard deviation (SD) calculated for each breed.

For Peer Review

Table 2. Mean F_{ROH} values calculated in the 23 pig breeds using all ROH >1 (F_{ROH1}), >4 (F_{ROH4}), >8 (F_{ROH8}) and >16 (F_{ROH16}) Mb. Standard deviation is in parenthesis.

Breed	F_{ROH1}	F_{ROH4}	F_{ROH8}	F_{ROH16}
Alentejana	0.139 (0.072)	0.110 (0.071)	0.084 (0.062)	0.059 (0.061)
Apulo-Calabrese	0.332 (0.111)	0.314 (0.110)	0.281 (0.102)	0.229 (0.101)
Basque	0.312 (0.042)	0.261 (0.052)	0.194 (0.053)	0.120 (0.042)
Bísara	0.144 (0.093)	0.122 (0.082)	0.098 (0.081)	0.071 (0.062)
Black Slavonian	0.138 (0.091)	0.121 (0.091)	0.101 (0.092)	0.072 (0.071)
Casertana	0.243 (0.112)	0.226 (0.110)	0.202 (0.110)	0.162 (0.100)
Cinta Senese	0.173 (0.064)	0.147 (0.063)	0.111 (0.052)	0.075 (0.050)
Gascon	0.213 (0.042)	0.175 (0.042)	0.132 (0.041)	0.087 (0.031)
Iberian	0.139 (0.063)	0.111 (0.061)	0.082 (0.060)	0.056 (0.050)
Krškopolje	0.125 (0.061)	0.109 (0.060)	0.089 (0.063)	0.065 (0.052)
Lithuanian indigenous wattle	0.135 (0.042)	0.114 (0.040)	0.089 (0.044)	0.060 (0.032)
Lithuanian White Old Type	0.152 (0.052)	0.122 (0.050)	0.093 (0.051)	0.063 (0.050)
Majorcan Black	0.134 (0.061)	0.108 (0.060)	0.081 (0.051)	0.055 (0.052)
Mora Romagnola	0.409 (0.062)	0.386 (0.062)	0.345 (0.060)	0.286 (0.061)
Moravka	0.118 (0.092)	0.103 (0.091)	0.087 (0.080)	0.068 (0.071)
Nero Siciliano	0.085 (0.084)	0.073 (0.082)	0.059 (0.081)	0.043 (0.072)
Sarda	0.101 (0.092)	0.088 (0.094)	0.073 (0.092)	0.053 (0.070)
Schwäbisch-Hällisches	0.147 (0.051)	0.120 (0.052)	0.093 (0.052)	0.065 (0.051)
Swallow-Bellied Mangalitsa	0.197 (0.052)	0.175 (0.050)	0.146 (0.050)	0.107 (0.042)
Turopolje	0.390 (0.101)	0.362 (0.101)	0.311 (0.093)	0.238 (0.081)
Italian Duroc	0.267 (0.043)	0.211 (0.041)	0.157 (0.041)	0.104 (0.042)
Italian Landrace	0.142 (0.042)	0.104 (0.040)	0.069 (0.031)	0.041 (0.031)
Italian Large White	0.143 (0.041)	0.106 (0.042)	0.075 (0.040)	0.046 (0.030)

For Peer Review

Table 3. The number of runs of homozygosity (ROH) islands and information on the genome covered by ROH islands identified in the 23 pig breeds with the method that used the S_{ROH} based-threshold.

Breed	Frequency[1]	N. of ROH islands	Genome covered (Mb)[2]	Average length (Mb)[3]
Alentejana	19/48 (40%)	12	35.88	2.99 (2.25)
Apulo-Calabrese	38/53 (72%)	0	-	-
Basque	36/39 (92%)	7	16.58	2.37 (1.84)
Bísara	20/48 (42%)	7	13.32	1.90 (1.36)
Black Slavonian	19/49 (39%)	3	2.64	0.88 (0.44)
Casertana	29/53 (55%)	7	10.23	1.46 (1.52)
Cinta Senese	23/53 (43%)	26	69.37	2.67 (2.42)
Gascon	27/48 (56%)	12	27.99	2.33 (2.00)
Iberian	19/48 (40%)	15	36.74	2.45 (1.49)
Krškopolje	18/52 (35%)	15	34.89	2.33 (2.14)
Lithuanian indigenous wattle	19/48 (40%)	15	41.81	2.79 (2.00)
Lithuanian White Old Type	21/48 (44%)	22	44.84	2.04 (2.19)
Majorcan Black	19/48 (40%)	12	27.23	2.27 (1.87)
Mora Romagnola	46/48 (96%)	4	12.34	3.09 (3.41)
Moravka	17/49 (35%)	9	19.11	2.12 (2.65)
Nero Siciliano	14/48 (29%)	4	7.41	1.85 (1.83)
Sarda	16/48 (33%)	0	-	-
Schwäbisch-Hällisches	20/49 (41%)	17	36.40	2.14 (1.76)
Swallow-Bellied Mangalitsa	25/50 (50%)	8	23.41	2.93 (1.89)
Turopolje	44/50 (88%)	17	80.82	4.75 (3.50)
Italian Duroc	32/48 (67%)	19	92.85	4.89 (6.48)
Italian Landrace	20/48 (42%)	32	75.03	2.34 (2.48)
Italian Large White	20/48 (42%)	12	46.51	3.88 (2.57)

[1] Frequency of the SNPs in a ROH that identifies the threshold to declare a ROH island. The frequency has been calculated dividing the number of animals needed to reach the define level by the number of animals retained after genotyping (see Table S2).

[2] Sum of the length of the chromosome regions in the genome covered by ROH islands in Mb.

[3] Average length of the ROH islands (standard deviation) in Mb.

The three blocs indicate the two different thresholds that can be used to define an island. For each block, there is information about: the number of animals that is used as threshold to define ad Island, the number of islands Identified, the total length of genome that is covered by islands, the average length of islands.

For Peer Review

Figure 1. Correlation plots between nROH (y axis) and S_{ROH} (x axis) for the 23 pig breeds including all animals. Acronyms of the breeds and are defined in Table 1 and Table S1. Pearson correlation coefficient is reported beside the acronym of each breed.

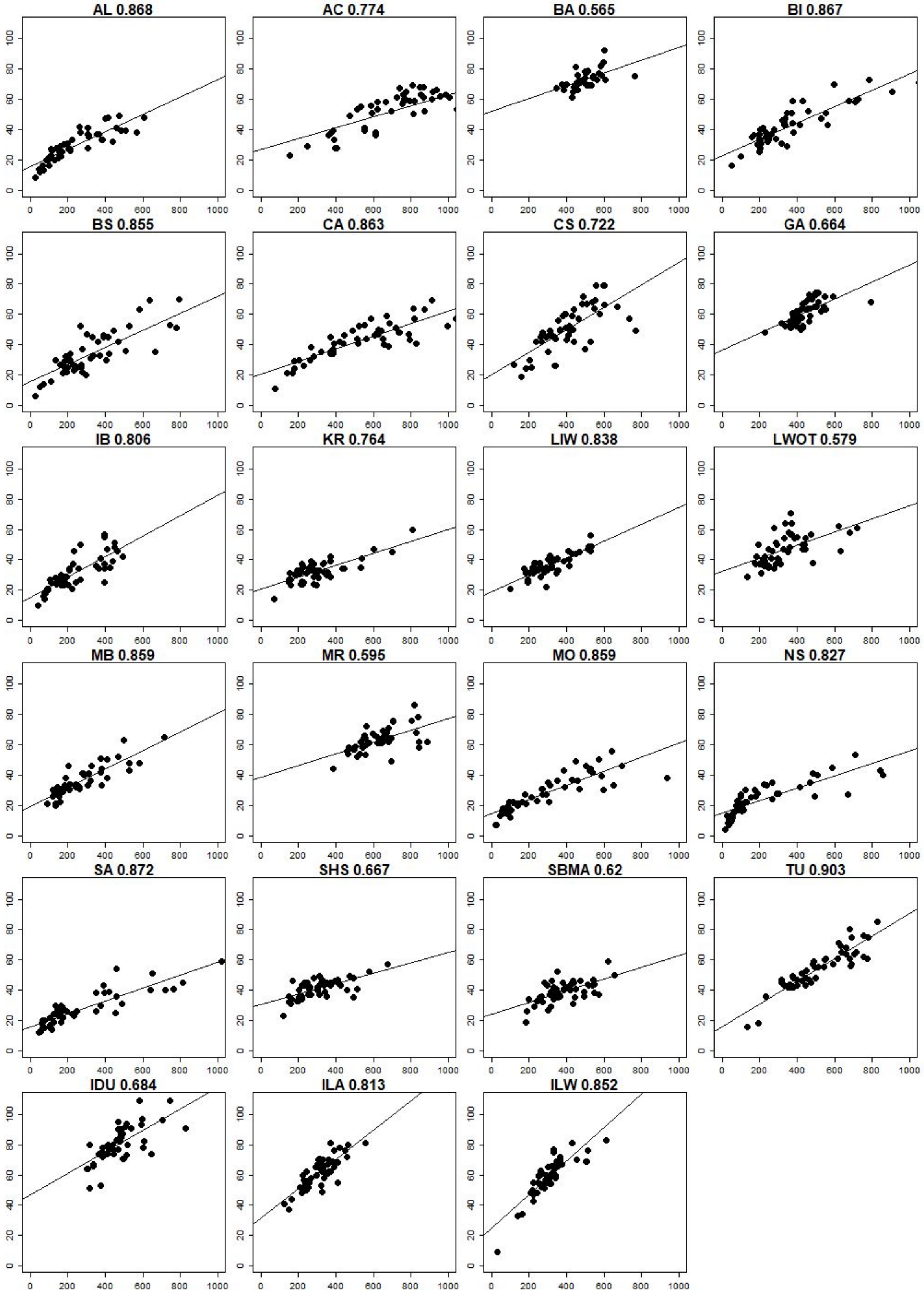


Figure 2. Proportion of runs of homozygosity of different class size in the 23 pig breeds. ROH classes were defined according to their size: 1–2, 2–4, 4–8, 8–16 and >16 Mb, identified as ROH1–2, ROH2–4, ROH4–8, ROH8–16 and ROH>16, respectively.

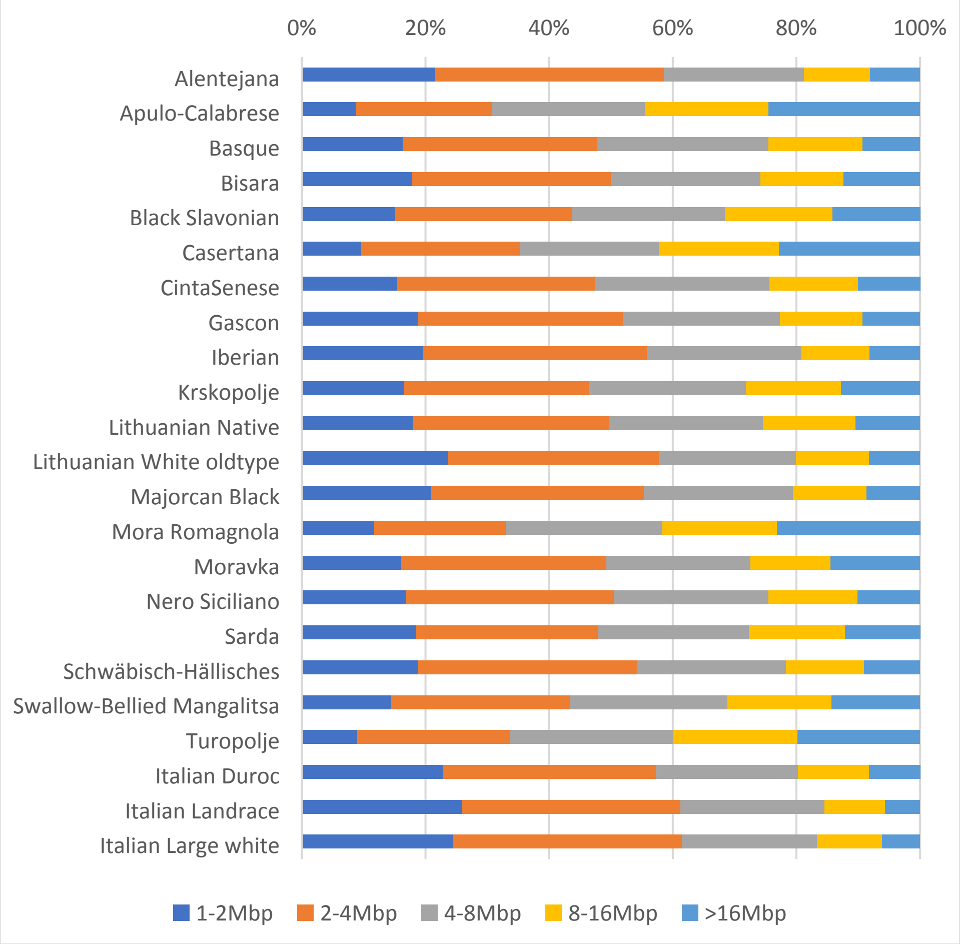


Figure 3. Boxplots of the F_{ROH} distribution in the 23 breeds: a) F_{ROH1}; b) F_{ROH4}; c) F_{ROH8}; d) F_{ROH16}. Acronyms of the breeds are explained in Table 1 and Table S1.

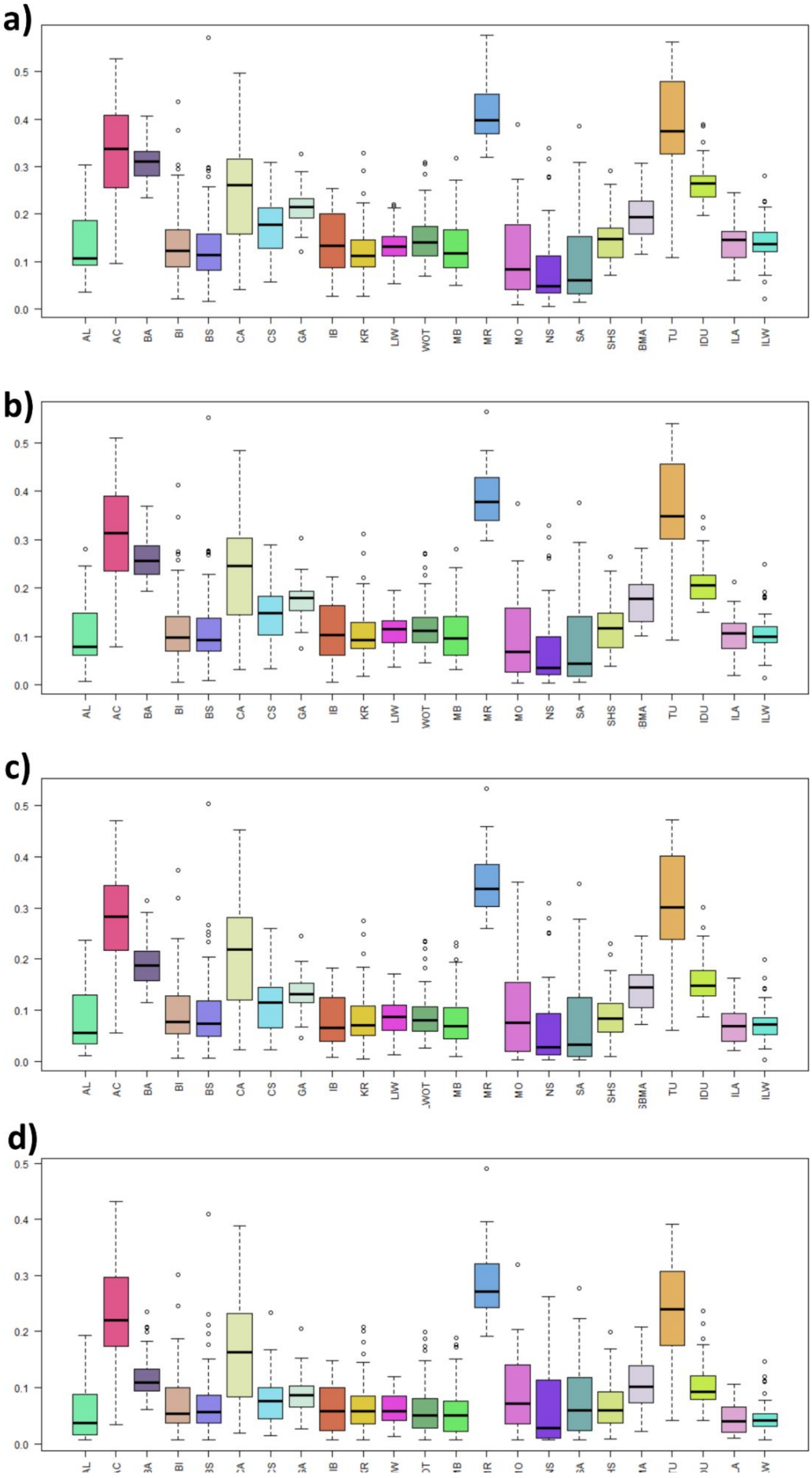


Figure 4. Manhattan plots showing ROH islands in a few analysed pig breeds with extreme patterns. The red line indicates the S_{ROH}-based threshold, the blue line indicates the frequency corresponding to the top 1% most frequent SNP in the population, the green line indicates the 50% of individuals within the population. The y axes indicate the number of animals carrying that SNP in a ROH.

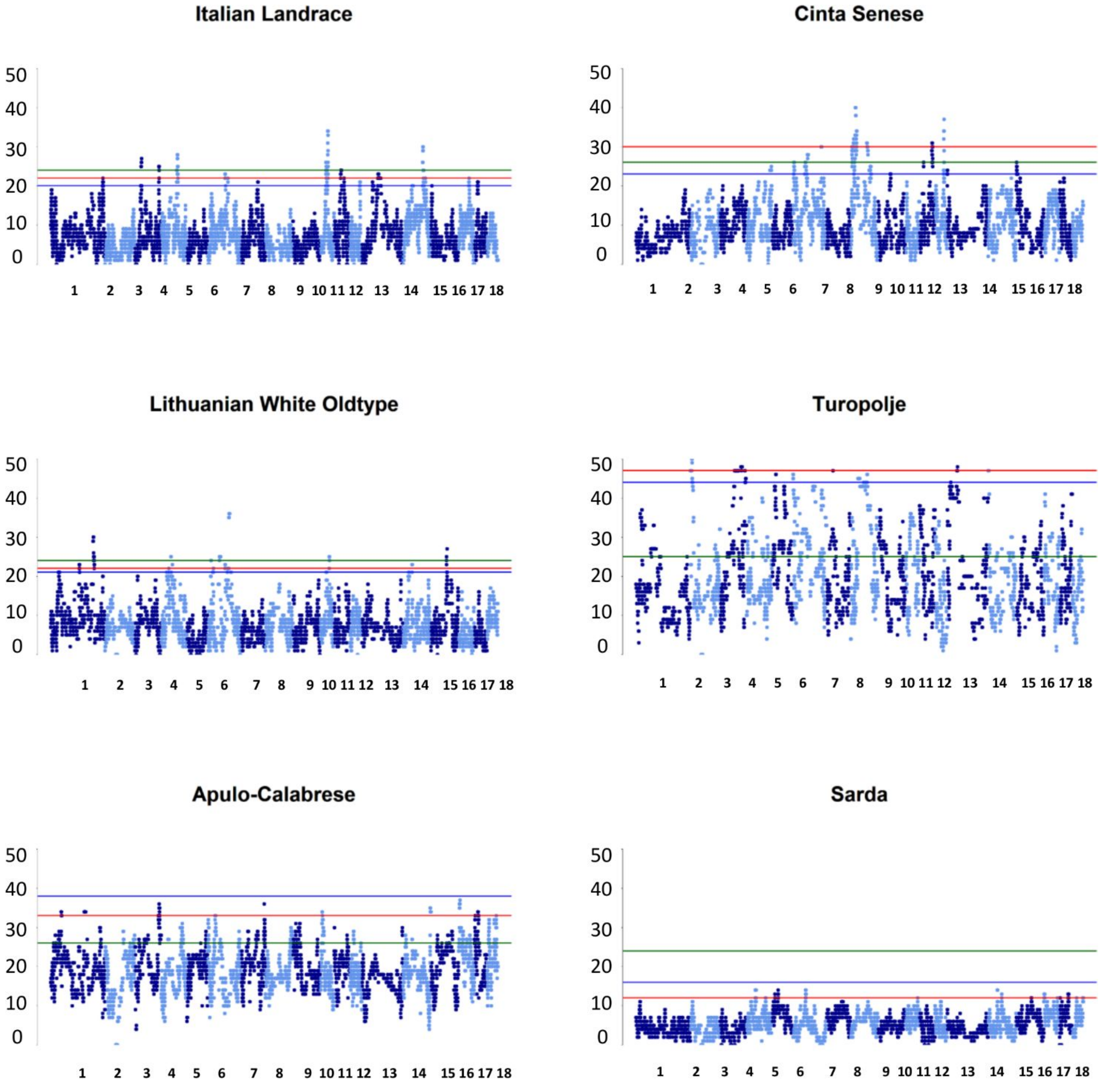
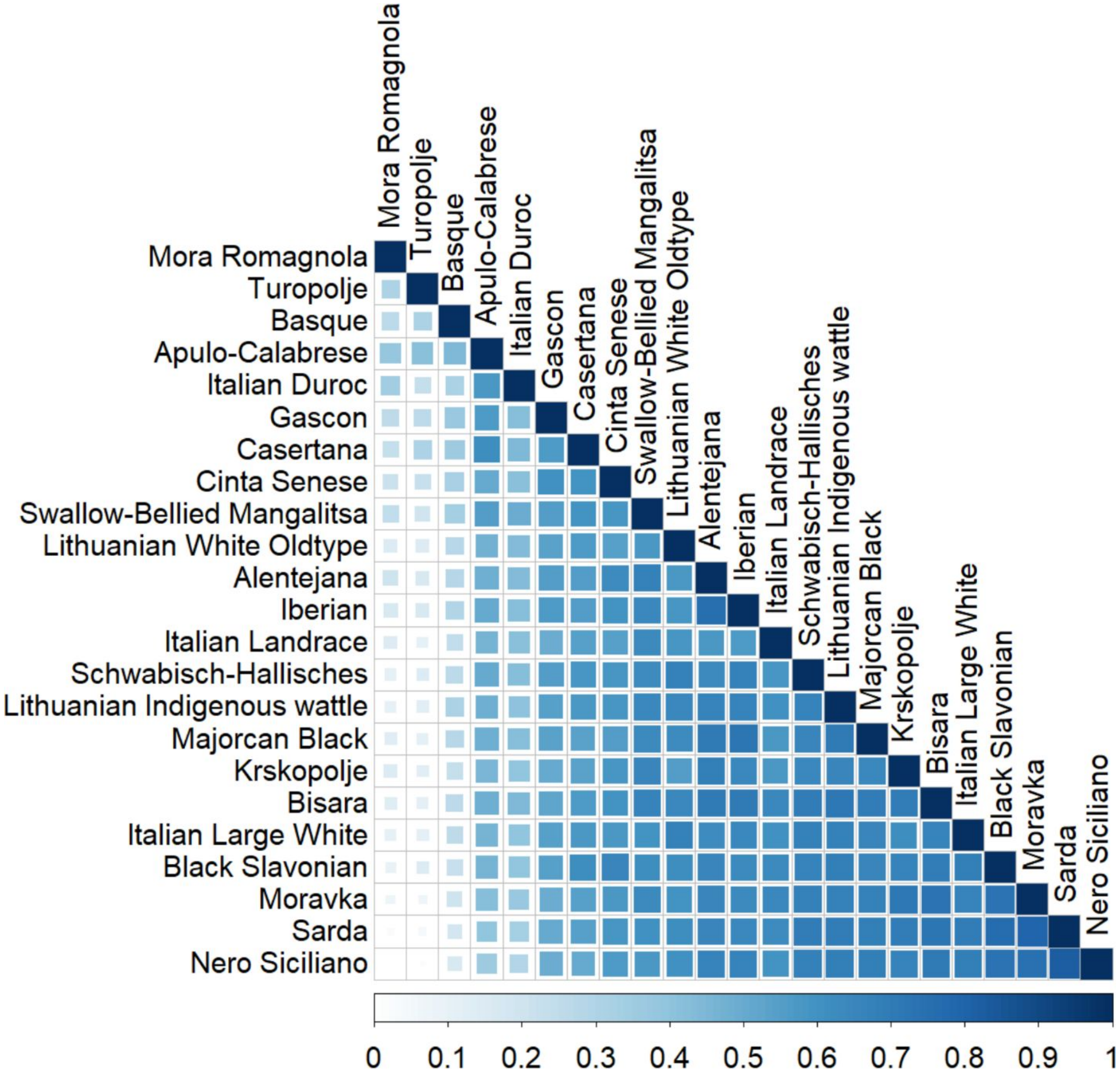


Figure 5. Similarity plot between patterns of homozygosity between pairs breeds. Color intensity and size of the squares are proportional to the similarity values.



Supporting information legend

Table S1: Analysed breeds, their country and region of origin and other information useful to describe the breeds.

Table S2: Number of animals and analysed SNP before and after the filtering steps.

Table S3. Effective population size (*Ne*) calculated for each breed.

Table S4. Runs of homozygosity (ROH) parameters using minor allele frequency (MAF) ≥0.01

Table S5. Minimum and maximum values for the number and size of ROH (nROH and L_{ROH}, respectively) and for the sum of all ROH segments by animals.

Table S6. Proportion of the five different runs of homozygosity (ROH) classes for each breed.

Table S7. Mean F_{ROH} values calculated using different ROH lengths and MAF >0.01.

Table S8. Average values for several genomic inbreeding measures.

Table S9. Correlation between all genomic inbreeding parameters in all breeds.

Table S10. The number ROH islands and information on the genome covered.

Table S11. ROH Islands and annotations (Excel file).

Table S12. Results of the gene enrichment analysis on all ROH Islands.

Table S13. Results of the gene enrichment analysis on ROH Islands that overlapped previous work regions identifying selection signature.

Figure S1. Multidimensional scaling (MDS) plot of the 23 pig breeds.

Figure S2. Genomic inbreeding based on F_{ROH} across chromosomes (F_{ROHSS}).

Figure S3. Boxplot of the Inbreeding Coefficients estimated with all the different methods.

Figure S4. Boxplot of the Inbreeding Coefficients estimated with all the different methods.

Figure S5. Manhattan plots showing ROH island patterns in all investigated pig breeds. The red line indicates the S_{ROH}-based threshold, the blue line indicates the frequency corresponding to the top 1% most frequent SNP in the population, the green line indicates the 50% of individuals within the population. The y axes indicate the number of animals carrying that SNP in a ROH.

For Peer Review