# Discovering genomic regions associated with the phenotypic differentiation of European local pig breeds

Klavdija Poklukar[1,a], Camille Mestre[2,a], Martin Škrlep[1], Marjeta Čandek-Potokar[1], Cristina Ovilo[3], Luca Fontanesi[4], Juliette Riquet[2], Samuele Bovo[4], Giuseppina Schiavo[4], Anisa Ribani[4], Maria Muñoz[3], Maurizio Gallo[5], Ricardo Bozzi[6], Rui Charneca[7], Raquel Quintanilla[8], Goran Kušec[9], Marie J. Mercat[10], Christoph Zimmer[11], Violeta Razmaite[12], Jose P. Araujo[13], Čedomir Radović[14], Radomir Savić[15], Danijel Karolyi[16], and Bertrand Servin[2*]

[1]Agricultural Institute of Slovenia, Hacquetova ulica 17, 1000 Ljubljana, Slovenia

[2] GenPhySE, Université de Toulouse, INRAE, INP, ENVT, Castanet-Tolosan, 31320, France

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

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

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

[6]DAGRI - Animal Science Section, Università di Firenze, Via delle Cascine 5, 50144 Florence, Italy

[7]MED- Mediterranean Institute for Agriculture, Environment and Development, Universidade de Évora, Pólo da Mitra, Apartado 94, 7006-554 Évora, Portugal

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

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

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

[11]Bauerliche Erzeugergemeinschaft Schwäbisch Hall, Haller Str. 20, Wolpertshausen 74549, Germany

[12]Animal Science Institute, Lithuanian University of Health Sciences, Baisogala 82317, Lithuania

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

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

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

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

[a]Authors contributed equally to this work.

*Corresponding author

E-mail addresses:

KP: klavdija.poklukar@kis.si

CM: camille.mestre@inrae.fr or c.k.mestre@gmail.com

MŠ: martin.skrlep@kis.si

MČP: meta.candek-potokar@kis.si

CO: ovilo@inia.es

LF: luca.fontanesi@unibo.it

JR: juliette.riquet@inrae.fr

SB: samuele.bovo@unibo.it

GS: giuseppina.schiavo2@unibo.it

AR: anisa.ribani2@unibo.it

MM: mariamm@inia.es

MG: gallo@anas.it

RB: riccardo.bozzi@unifi.it

RC: rmcc@uevora.pt

RQ: raquel.quintanilla@irta.cat

GK: gkusec@fazos.hr

MJM: marie-jose.mercat@ifip.asso.fr

CZ: Christoph.Zimmer@besh.de

VR: violeta.razmaite@lsmuni.lt

JPA: pedropi@esa.ipvc.pt

ČR: cedomirradovic.izs@gmail.com

RS: savic@agrif.bg.ac.rs

DK: dkarolyi@agr.hr

BS: bertrand.servin@inrae.fr

# Abstract

## Background

Intensive selection of modern pig breeds resulted in genetic improvement of productive traits while local pig breeds remained less performant. As they have been bred in extensive systems, they have adapted to specifical environmental conditions resulting in a rich genotypic and phenotypic diversity. In this study, European local pig breeds were genotypically and phenotypically characterised using DNA-pool sequencing data and breed level phenotypes related to stature, fatness, growth and reproductive performance traits. These data were analysed using a dedicated approach to detect selection signatures linked to phenotypic traits in order to uncover potential candidate genes that may be under adaptation to specific environments.

## Results

Genetic data analysis of European pig breeds revealed four main axes of genetic variation represented by Iberian and modern breeds (i.e. Large White, Landrace, and Duroc). In addition, breeds clustered according to their geographical origin, for example French Gascon and Basque breeds, Italian Apulo Calabrese and Casertana breeds, Spanish Iberian and Portuguese Alentejano breeds. Principal component analysis of phenotypic data distinguished between larger and leaner breeds with better growth potential and reproductive performance on one hand and breeds that were smaller, fatter, and had low growth and reproductive efficiency on the other hand. Linking selection signatures with phenotype identified 561 significant genomic regions. Among them, several regions contained candidate genes with possible biological effect on stature, fatness, growth and reproduction performance traits. For example, strong associations were found for stature in two regions containing *ANXA4* and *ANTXR1* genes, for growth performance containing *TLL1* gene, for fatness containing *DNMT3A* and *POMC* genes and for reproductive performance containing *HSD17B7* gene.

## Conclusions

The present study on European local pig breeds used a newly developed approach for searching selection signatures supported by phenotypic data at the breed level to identify potential candidate genes that may have adapted to different living environments and production systems. Results can be useful to define conservation programs of local pig breeds.

# Background

In the last decades, pig breeding has focused mainly on improving growth rate, carcass leanness, and reproductive performances (1) of relatively few breeds (2). On the other hand, the use of many local breeds has decreased while these breeds have not been subjected to such intensive management and genetic improvement. They are often raised in extensive farming systems, resulting in adaptation to specific environmental conditions and (usually) poor feeding resources (3). However, this adaptation to seasonal fluctuations in feed availability may have resulted in low productivity (4). As a result, many local pig breeds have been abandoned or have even become extinct, while most of them have faced population bottlenecks and genetic drifts or introgression from other pig populations (5,6).

Today, many local breeds are used on a relatively limited scale and the available information on their phenotypic and genotypic traits is well known for only a few of them, such as Iberian or Meishan (7). Nonetheless, interest in local pig breeds has increased recently for several reasons, including their meat quality (allowing the production of high-quality meat products), their adaptation to local feeding resources, and society's awareness of phenotypic and genotypic biodiversity conservation (3). Being exposed to specific selection pressures in different local environments, local pig breeds also represent interesting genetic resources (8).

Recently, a genetic characterization of 20 European local pig breeds was carried out. It showed that some local breeds are clustered according to their geographical distribution (e.g. French Gascon and Basque breeds, Italian Apulo Calabrese and Casertana breeds, Spanish Iberian and Portuguese Alentejano breed), while some others suffer from introgressions or admixture with modern pig breeds (e.g. Lietuvos Baltosios Senojo Tipo and Lietuvos Vietiné with Large White and Landrace pigs; Mora Romagnola with Duroc pig) (8–10). Consequently, these breeds have developed particular heterogeneous phenotypic traits that could reflect specific genetic potential and adaptation to different

118 production systems. As the measurement of phenotypic traits in pigs is to some extent standardised,
119 it is possible to compare different local breeds. As shown in the study of Čandek-Potokar and Nieto
120 (3) reviewing productive traits, the studied European local pig breeds not only differ from modern
121 breeds, but are also quite variable and differ significantly among themselves.

122 To better understand the genetic basis for variation in phenotypic traits in local pig breeds, several
123 genome-wide association studies focused on detecting possible associations of loci with different
124 phenotypic attributes, such as morphological, production or meat quality traits (11–14). However,
125 these studies worked with small number of samples, increasing the risk to give false negative results
126 due to low statistical power (15). Another approach to search for the association between genetic
127 polymorphisms and phenotype is to look for genomic regions that have responded to selection (i.e.,
128 selection signatures). Several studies in local breeds have shown that the signals detected contain
129 gene variants/genes that may be associated with variations in phenotype, such as coat colour, growth,
130 reproduction, or fatness (8,9,16–19). The study by Muñoz et al. (9) determined the signatures of
131 selection using SNP-array data in 20 European local breeds. This study revealed putative selection
132 signals for regions containing genes involved in fatness, growth, reproduction, development,
133 behaviour and sensory perception. A subsequent study, in which whole-genome sequencing was
134 performed on DNA-pool samples from the same breeds/animals, detected several regions associated
135 with coat colour, body size, growth, reproduction and fat deposition (8). In the aforementioned
136 studies, the detected selection signatures did, however, not represent a direct association with a
137 specific phenotype. Therefore, the present study proposes a different approach that attempts to
138 determine selection signatures that are specifically linked to phenotype. It uses DNA-pool sequencing
139 data from 19 European local and 3 modern pig breeds, as well as a database of many phenotypic traits
140 in 20 local pig breeds associated with stature, fatness, growth and reproductive performance. The

141 discovered genomic regions could reveal candidate genes with potential biological effects on specific
142 phenotypic traits.

# 143 Methods

144 The aim of this study was to locate genomic regions associated with signatures of selection of local
145 pig breeds for production traits. This was performed by combining genetic and phenotypic data at the
146 breed (population) level. Most of the data used for this study have been previously published (3,8,9).
147 In this section, genetic and phenotypic datasets are presented, and then the methodological approach
148 to detect signatures of selection on phenotypic traits is described.

## 149 Genetic Datasets

150 The present study was based on genetic data collected from 19 populations of local European pig
151 breeds and 7 populations of industrial breeds. The data consisted of SNP genotyping performed with
152 a medium density array (9) for 20 local breeds and whole genome sequencing of the pooled samples
153 (8) for 19 of them (the Iberian breed was not included). To better describe the genetic structure of the
154 local pig breeds, samples from additional 7 populations of 4 modern pig breeds were added but only
155 3 of these samples were pool-sequenced. The final sample used for this study is shown in Table 1.

156

157 **Table 1. Name of pig breeds, country of origin and sample size.**

| Pig breed name | Country | # genotyped on SNP array | Sample size in sequencing pool |
|---|---|---|---|
| **Local breeds** | | | |
| Alentejano | Portugal | 48 | 35 |
| Apulo Calabrese | Italy | 53 | 35 |
| Basque | France | 39 | 30 |
| Bísaro | Portugal | 48 | 35 |
| Black Slavonian | Croatia | 52 | 35 |
| Cinta Senese | Italy | 51 | 35 |
| Gascon | France | 48 | 30 |
| Iberian | Spain | 48 | - |
| Krškopolje | Slovenia | 52 | 35 |
| Lietuvos Baltosios Senojo Tipo | Lithuania | 48 | 35 |
| Lietuvos Vietiné | Lithuania | 48 | 35 |
| Mangalitsa | Serbia | 50 | 35 |
| Mora Romagnola | Italy | 48 | 35 |
| Moravka | Serbia | 49 | 35 |
| Negre Mallorquí | Spain | 48 | 35 |
| Nero Casertano | Italy | 53 | 35 |
| Nero Siciliano | Italy | 48 | 35 |
| Sarda | Italy | 48 | 35 |
| Schwäbisch-Hällisches | Germany | 49 | 35 |
| Turopolje | Croatia | 49 | 35 |
| **Modern breeds** | | | |
| Italian Large White | Italy | 4 | 35 |
| Large White | France | 97 | - |
| Italian Duroc | Italy | 5 | 35 |
| Duroc | France | 33 | - |
| Italian Landrace | Italy | 4 | 35 |
| Landrace | France | 53 | - |
| Pietrain | France | 61 | - |

158

159 ## Quality control of genetic datasets

160 For the SNP array data, quality control of SNP genotypes was performed for the entire dataset using
161 standard filters: only autosomal SNPs with < 10% missing data were retained. Following this step,
162 10 individuals with > 3% missing genotypes were clear outliers in the sample and were therefore
163 discarded.

164 Two analyses were performed for the pooled sequence data: estimation of SNP allele frequencies for
165 variants in the SNP array and de novo variant discovery. De novo SNV discovery was carried out
166 using CRISP (20) with default parameters and yielded 34,751,691 variants. From these, we filtered
167 out variants with more than two alleles, variants with very low allele frequency that most likely
168 resulted from sequencing errors (field VP = 0 and AF=1), and variants with low mapping quality
169 (QUAL < 1000, MQ < 20). After filtering, 16,403,270 SNPs were kept.

## 170 Genetic structure of pig populations

171 The genetic structure of pig populations was assessed from individual SNP genotyping data using
172 PCA, admixture analysis (21) and population tree reconstruction with hapFLK. Admixture analysis
173 was performed for a number of clusters (K) between 2 to 40 to determine a value of K that best
174 explained the data. The decrease in cross-validation error was monotonous from 2 to 40 (Additional
175 File 1: Figure S1), but showed a diminishing decrease after K=24, therefore this value was used as a
176 reference to describe the genetic structure of the breeds. Based on the admixture results, individuals
177 showing more than 80% of their genome assigned to the main cluster of their assigned breed were
178 selected for inclusion in the population tree analysis. The Nero Siciliano breed did not show any
179 individual matching to this criterion and was, therefore, not included in the population tree
180 reconstruction.

181 To verify the quality of the pool sequence data, a comparison with the SNP array genotyping results
182 was conducted. Allele frequencies of SNP array variants in the pooled sequence data were estimated
183 using allele counts extracted with Samtools mpileup (22) and PoPoolation2 (23). Samtools was run
184 with options -C 50 -q 20 (variants with mapping quality less than 20 were discarded as recommended
185 by the software documentation). PoPoolation2 was run with default parameters on the resulting
186 mpileup file. From the resulting pool allele counts, allele frequencies and Fst for each pair of

187 populations were then calculated using the approach of (24) implemented in the R package poolfstat.
188 The population tree was constructed applying the neighbour-joining algorithm on the Fst matrix using
189 the same procedure that was used on allele frequencies derived from the SNP array genotypes. PCA
190 was performed on the pool-seq SNPs extracted from the SNP-chip data and showed high correlation
191 across breeds (Additional File 1: Figure S2).

## 192 Phenotypic characterization of local pig populations

193 A database of phenotypic traits of European local pig populations was used to determine global breed
194 differences in phenotype (3). Phenotypic variables were combined into four distinct groups
195 summarizing stature, fatness, growth, and reproductive performance. The growth performance group
196 included traits of average daily gain in 3 different growth periods (i.e. from lactation to early fattening
197 phase; up to approximately 60 kg). The stature group included traits of body weight and height of
198 adult male and female animals. The fatness group comprised backfat thicknesses at different
199 anatomical locations, fatty acid composition, carcass lean meat content and intramuscular fat content.
200 To standardize the value of backfat thickness at the level of the last rib, an adjustment was made to
201 the final body weight of 120 kg (see Additional file 2 for a more detailed description). The adjusted
202 values were included in the PCA. The reproductive performance group included traits related to
203 number of piglets per litter, number of litters per year, piglets/litter weights, duration of lactation and
204 farrowing interval. A more detailed description of the traits included in the phenotypic
205 characterization can be found in Additional File 3: Table S1.

206 All statistical analyses were performed using R statistical software. Means for each variable and each
207 breed were calculated and scaled. Since the phenotypic database was composed of results from
208 different studies, some variables were missing for some breeds (Additional File 1: Figure S3).
209 Missing data were imputed using regularized iterative PCA method using R package missMDA (25).

210 Principal component analyses were performed using the R package FactoMiner (26) for the growth
211 performance, reproductive performance, stature and fatness traits. Uncertainty in the predictions of
212 missing data was assessed by multiple imputation (MIPCA function (27)). Breed loadings on the PC1
213 of each PCA were used as "breed scores" for the analysis of selection signatures.

## 214 Genome scan for selection on phenotypic breed scores

215 To identify genomic regions associated to selection on breed level phenotypes, we built upon the
216 approach of (28) which we briefly describe here. Assuming data on allele frequencies measured in r
217 populations at L loci, this association model at a locus l is:

218 $$p^l \sim N\left(1p_0^l + x\beta^l, Vp_0^l\left(1 - p_0^l\right)\right) \quad (1)$$

219 where $p^l$ is the vector of allele frequencies in the r populations, $p_0^l$ is the (unknown) ancestral allele
220 frequency of the locus, $x$ is the vector of breed level phenotypes, $\beta^l$ is the effect of phenotype on
221 allele frequencies differences, and $V$ is the genome-wide variance-covariance matrix of allele
222 frequencies between populations. The idea behind this model is that the adaptation of populations to
223 covariate $x$ (here the breed level phenotypes) will drive allele frequency differences between
224 populations away from their expectation under genetic drift. In (1), this expectation is modelled with
225 the genome-wide covariance matrix V (see (28) for the detailed description). To perform statistical
226 inference under this model the parameters V, $\beta^l$ and $p_0^l$ need to be estimated. Under the null
227 hypothesis (selection associated with phenotype x has not affected allele frequencies), the parameter
228 $\beta^l$ is set to 0. To test for association with the covariate x, (28) use a MCMC algorithm allowing to
229 derive a statistic for association (a Bayes Factor). This was later extended by (29) to account for
230 uncertainty in allele frequency estimation in pool sequencing experiments. This approach was tested
231 on our dataset, but was found computationally inefficient due to the very high number of SNPs
232 considered. We therefore used a frequentist treatment of the model that consists in maximizing the

likelihood of model (1) under the null and the alternative hypothesis and performing a likelihood ratio test. One deviation from this approach is that we account for the variation in sequencing depth in different pools by using regularized allele frequencies $\tilde{p}_r^l$ (see below) rather than fitting the model on the usual allele frequency estimates $\hat{p}_r^l$. The maximum likelihood estimator of allele frequencies is $\hat{p}_r^l = \frac{c_r^l}{n_r^l}$ , where $c_r^l$ is the number of alternative allele counts and $n_r^l$ the sequencing depth for population r at locus l. This estimator has good properties provided the sequencing depth $n_r^l$ at the locus is large. However, this depth is highly variable along the genome and can be quite low (even 0) in some genomic regions. Moreover, it will vary between populations at a given locus. To regularize allele frequencies estimates, we used shrank allele frequencies estimates in the form:

$$\tilde{p}_r^l = \frac{a^l + c_r^l}{b^l + n_r^l}$$

where $a^l$and $b^l$are regularizing (prior) parameters set so that $a^l/b^l$ equals the alternative allele frequency among all populations at locus l. Doing so, if a population has no observed data at a SNP, its allele frequency will be similar to other populations which reduces the risk of false positives due to uneven sequencing coverage.

The association of breed scores to allele frequencies was tested for all SNPs for all breed scores (stature, growth performance, fatness traits, reproductive performance). The result of the test was a p-value for each SNP for each phenotypic trait. Since model (1) is only valid for intermediate ancestral allele frequencies, only SNPs where the ancestral allele frequency estimate $p_0^l$ lied between 0.05 and 0.95 were considered in the following. Statistical significance was established by estimating False Discovery Rates (FDR) with the approach of (30) using the qvalue R package. The FDR

254 threshold for detecting significant associations was set at 1%. The number of significant variants with
255 corresponding q-values for all phenotypic groups can be found in Additional File 3: Table S2.

256 Based on association statistics for individual SNP, we defined association regions by assuming that
257 two SNPs belonged to the same region if they were less than 200 kb apart. For each significant region,
258 candidate genes were extracted from the annotation of the reference genome (Sscrofa11.1). We chose
259 to focus on the top 20 significant regions of each phenotypic group (i.e. those with the lowest P-value
260 and with at least 3 significant SNPs) and genes within these regions were reviewed in the literature
261 for potential biological effects on phenotypic differentiation. The top 20 significant regions for each
262 phenotypic trait were summarized using the R package circlize (31). The allele frequencies of each
263 breed at each SNP were extracted and plotted to highlight which breed most likely adapted for specific
264 phenotypic traits.

# 265 Results

## 266 Genetic structure of European pig breeds

267 To characterize the genetic structure of populations, the individual genotypes on the SNP array were
268 used to perform a standard genetic structure analysis. PCA analyses of 24 breeds revealed four main
269 genetic backgrounds in the dataset (i.e. Iberian, Duroc/Mora Romagnola, Large White and
270 Landrace/Pietrain backgrounds), visible in the first 3 principal components (Figure 1). PC1 and PC2
271 clearly distinguish between the Iberian pig, White pigs (Landrace, Large White and Pietrain) and
272 Duroc/Mora Romagnola backgrounds. PC3 shows that the White pigs' background separates into a
273 Large White specific background and a Landrace/Pietrain background. PC4 further separates the
274 Turopolje breed from the Iberian group. The global pattern of differentiation between breeds in our
275 dataset is thus strongly influenced by modern pig breeds. Local pig breeds are usually in intermediate
276 positions, mostly within a triangle with summits corresponding to the Iberian breed, the

Landrace/Pietrain and the Large White breeds (see PC1 vs. PC3 in Figure 1). The two exceptions are the Mora Romagnola and Turopolje breeds, which appear much more differentiated. In order to interpret these patterns, further analyses using admixture clustering and population tree reconstruction were performed.

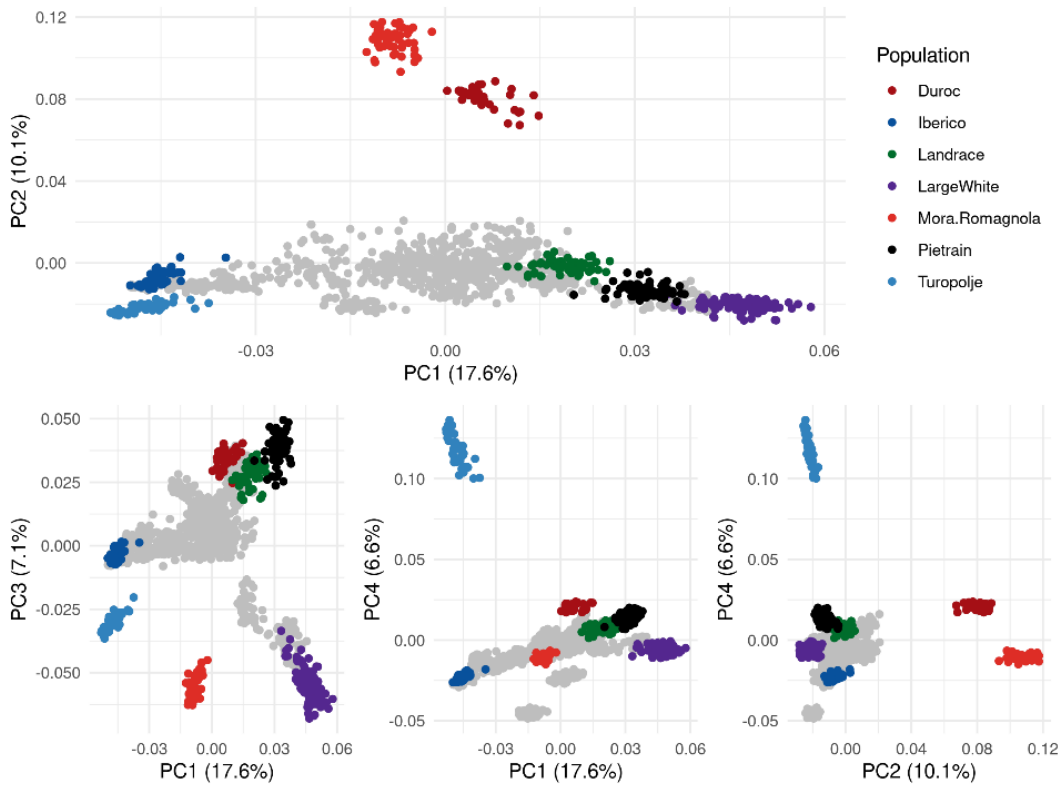


**Figure 1**. Principal component analysis of 24 pig breeds genotyped on a medium density SNP array. The most differentiated breeds are highlighted with different colours.

The results of the admixture analysis results show that the number of homogeneous clusters in the dataset is difficult to determine. The cross-validation procedure was conducted for up to 40 clusters and resulted in a general decrease in cross-validation error with increasing number of clusters (Additional File 1: Figure S1). However, the decrease slowed down after K=24 clusters, a number that corresponds to the number of named breeds (Table 1). The further decrease is due to the fact that some breeds are sub-structured and require additional clusters to be well fitted. In the following, we present the results obtained with K = 24 corresponding to the inflexion point in the cross-validation

curve. Figure 2 shows side-by-side the result of the population tree reconstruction and the admixture analysis. At K=24, most breed exhibit a homogeneous pattern of admixture and belong to a specific cluster. Exceptions are the Alentejano and Iberian breeds (which belong to the same cluster, in consistence with their common origin), the Casertana and Apulo Calabrese breeds (each of them further split into two groups) and Sarda, Moravka, Bisaro and Nero Siciliano breeds that show high heterogeneity. In the latter breeds, admixed individuals do not appear to be recent hybrids with other breeds.

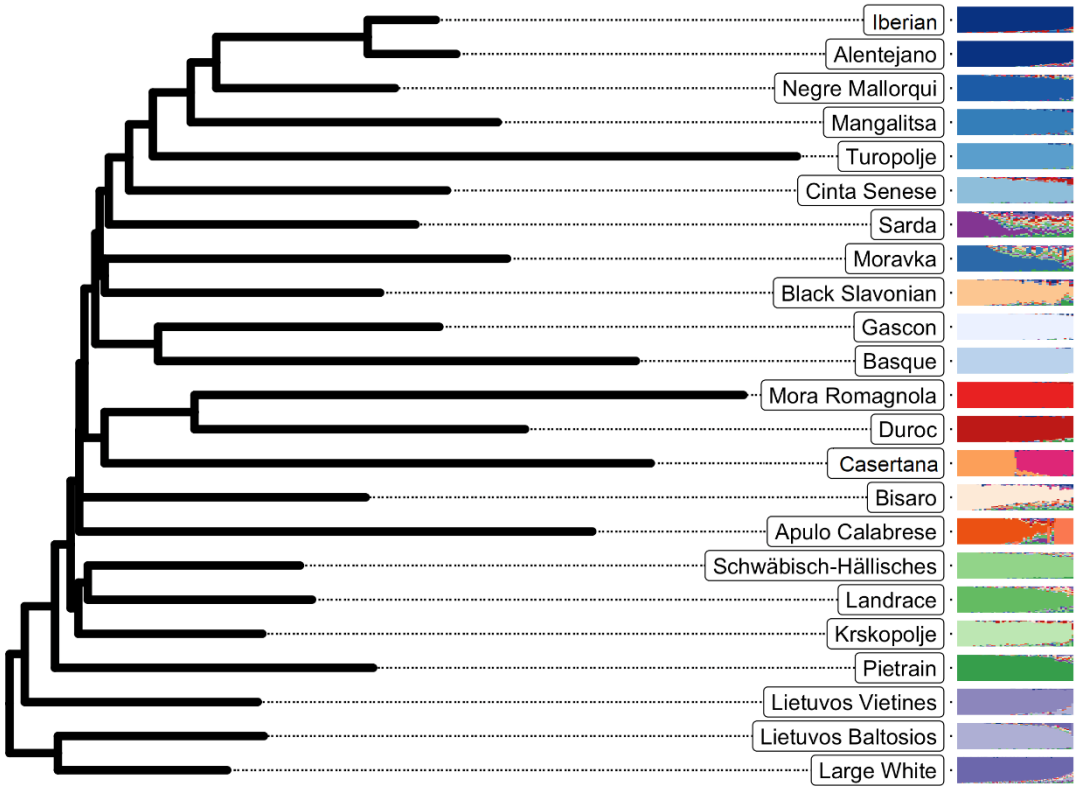


**Figure 2**. Population tree and admixture analysis of 23 pig breeds genotyped on a medium density SNP array. The population tree reconstructed from pairwise genetic distances (Fst) is shown on the left, and the admixture component for all individuals a priori belonging to the breed is shown on the right. The colour levels follow the global axes of genetic variation, with populations most closely related to the Iberian type shown in shades of blue, to the Duroc in shades of red, to the Large White in shades of purple, and to the Landrace in shades of green. Heterogeneous populations, or those equidistant from these four clusters, are shown in orange. The Nero Siciliano breed exhibited extremely high heterogeneity, and was therefore not included in the reconstruction of the population tree and is not shown in this figure.

306 The population tree is structured consistently with the main axes of variation identified in the PCA
307 analysis with four main genetic backgrounds, highlighted with similar colours in Figure 2. Turopolje
308 and Mora Romagnola breeds differentiated early in the PCA analysis, which could be explained with
309 the population analysis and the fact that they exhibit long branches corresponding to low
310 heterozygosity. This analysis also reveals a general clustering of breeds according to their
311 geographical origin. Breeds closer to the Iberian group mostly originate from the Iberian peninsula
312 or close geographical areas (South West of France with Basque and Gascon breeds, Balearic Islands
313 with Negre Mallorqui breed). Interestingly, some other breeds from different geographical area
314 appear to be related to this background such as Mangalitsa and Moravka from Serbia, the Black
315 Slavonian from Croatia or the Cinta Senese from Italy. Breeds from Central Europe, such as
316 Schwabisch-Hällisches and Krškopolje pig, showed genetic proximity to the Landrace/Pietrain
317 background and breeds from Lithuania in Northern Europe with the Large White component. Finally,
318 some breeds such as the Bísaro from Portugal or Apulo Calabrese from Italy cannot be considered
319 related to any other breed in the dataset.

320 Sequences from the investigated breeds were aligned to the Sscrofa11.1 reference genome and SNPs
321 were discovered with the pool-seq variant caller CRISP. The detected variants were verified with
322 SNP-chip results, which were published in a previous study of (9).

## 323 Phenotypic differentiation of European local pig breeds

324 A database of available published results on phenotypic traits of 20 local pig breeds (3) was used in
325 order to distribute local pig populations according to phenotype including stature, growth, fatness
326 traits, and reproductive performance traits. The representation of relationships between breeds and
327 variables was demonstrated by PCA analyses. The first two principal components of the PCA for
328 stature, growth, fatness, and reproduction group accounted for 98.2%, 83.2%, 84.3% and 70.7% of

329 the total variance, respectively (Figure 3, Additional File 1: Figures S4, S5, S6). Scores for each breed
330 from PC of all phenotypic groups were extracted from PCA.

331 The first principal component of the stature group (Figure 3) represented the majority of the total
332 variability (77.3%) and clearly distributed breeds according to average height and weight. Thus,
333 breeds with lower breed scores for stature (e.g. Moravka, Turopolje, Nero Siciliano, Negre Mallorqui)
334 were smaller and lighter, while breeds with higher breed scores were taller and heavier breeds (e.g.
335 Schwäbisch-Hällisches, Lietuvos Baltosios Senojo Tipo, Krškopolje pig, Lietuvos Vietiné).

336 The growth group distributed local breeds according to their growth capacity, including data on
337 average daily gain for the rearing period from the lactation period to the early fattening phase (i.e. up
338 to 60 kg body weight) (Additional File 1: Figure S4). The first PC explained 56.1% of the total
339 variability and distributed the breeds with the highest (i.e. Schwäbisch-Hällisches, Lietuvos Baltosios
340 Senojo Tipo, Lietuvos Vietiné, Bisaro, Krškopolje pig) and the lowest (i.e. Alentejano, Moravka,
341 Black Slavonian, Mangalitsa and Turopolje) growth potential.

342 Fatness traits comprised variables associated with the fatty phenotype (Additional File 1: Figure S5).
343 Principal component 1 (59.3% of total variability) was positively correlated with backfat thicknesses
344 at different anatomical locations and intramuscular fat content. Conversely, lean meat and PUFA
345 content (i.e., lean phenotype) were negatively correlated with PC1. Because intramuscular fat content
346 is a trait of particular interest, the distribution of breeds by intramuscular fat content is shown in
347 Additional File 1: Figure S7. Local breeds were divided into fatter (e.g. Moravka, Iberian, Mangalitsa,
348 Negre Mallorqui) or leaner phenotypes (e.g. Schwäbisch-Hällisches, Lietuvos Baltosios Senojo Tipo,
349 Bisaro).

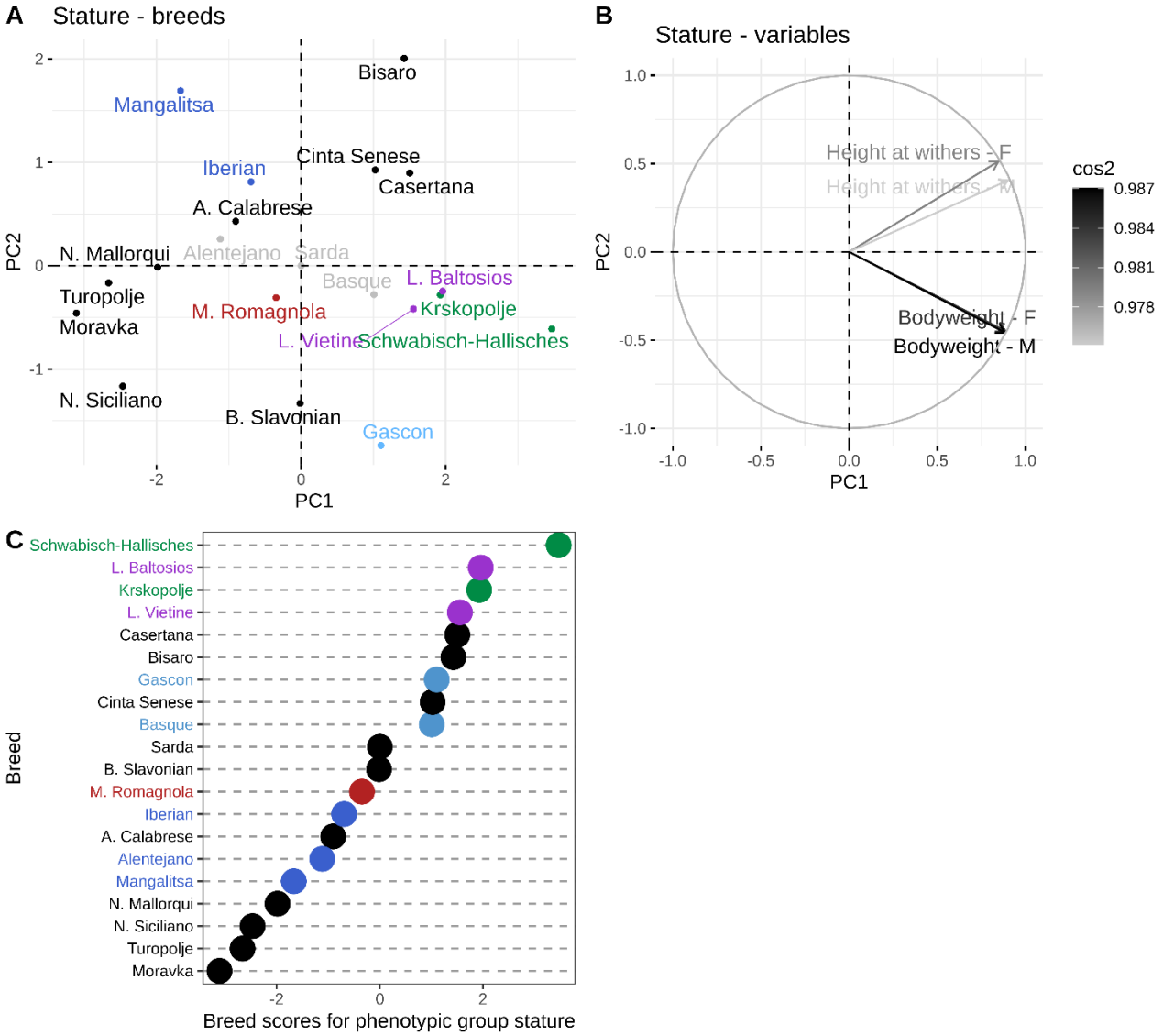


**Figure 3.** Principal component analysis showing the relationship between breeds (A) and the traits associated with stature (B) and the corresponding phenotypic breed scores (C).
Breeds (A) coloured in grey are the breeds with more than 50% of missing variables, thus, their position on the PCA must be interpreted carefully. The variables (B) are coloured according to quality of the representation, which is measured by squared cosine between the vector originating from the element and its projection on the axis. The variables that contribute most to the separation of the trait into PC1 and PC2 are coloured black. Breeds (A, C) are coloured according to genetic similarity. Breeds (A, C) in green are genetically Landrace-like breeds, in purple are Large White-like breeds, in blue are Iberian-like breeds, in red are Duroc-like breeds and in light blue are Gascon and Basque.

The final characterization of local pig breeds was done according to reproductive performance (Additional File 1: Figure S6). Herein, the PC1 clearly distinguished breeds with larger litter sizes and piglet birth weights (i.e. Schwäbisch-Hällisches, Lietuvos Baltosios Senojo Tipo, Lietuvos

Vietiné, Bisaro, Krškopolje pig) from breeds that had lower reproductive performance with smaller litters and lighter piglet birth weights (i.e. Nero Siciliano, Turopolje, Casertana, Mora Romagnola).

Lastly, breed scores of phenotypic trait groups were plotted to examine the global distribution of breeds and production traits in local pig populations (Figure 4). Figure 4 shows that breeds characterised by larger size and higher growth potential were also more reproductively efficient than smaller breeds with lower growth rate. In addition, fatter breeds were smaller and lighter with smaller growth rate than leaner breeds that are larger.

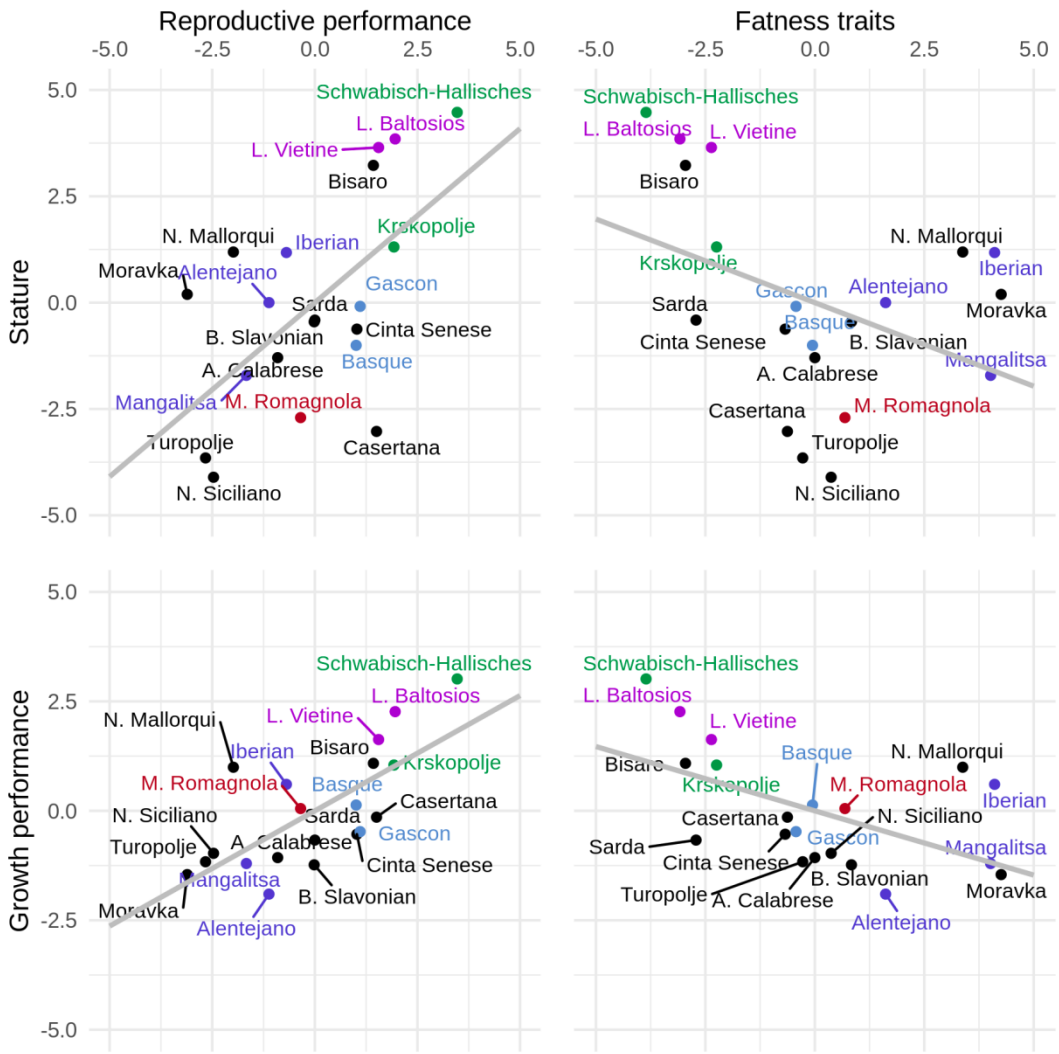

**Figure 4.** Global differences in production traits for 20 European local pig populations. Breeds are distributed according to breed scores for phenotypic traits and colored according to genetic similarity. Low values for breed score growth are representing lower average daily gain from lactation up to 60 kg of live weight, while higher values represent higher average daily gain in the same growth period. Low values for breed score reproduction represent low reproductive performance, while high values represent higher reproductive performance of breeds. Low values for breed score stature represent lighter and smaller breeds, while high values represent heavier and larger breeds. Low values for the breed score fat represent leaner breeds, while high values represent fatter breeds. Breeds in green are genetically Landrace-like breeds, in purple are Large White-like breeds, in blue are Iberian-like breeds, in red are Duroc-like breeds and in light blue are Gascon and Basque.

## Detection of genomic regions associated with phenotypic differentiation

In the present study, the approach proposed by Coop et al. (28) was extended to breed-level phenotypes to find genomic regions potentially influencing phenotypic differentiation in local European pig breeds. Therefore, a selection signature scan was performed for each phenotypic breed score, resulting in p-values for each SNP. Figure 5 shows an example of the genome-wide distribution of selection signatures for phenotypic breed score stature.

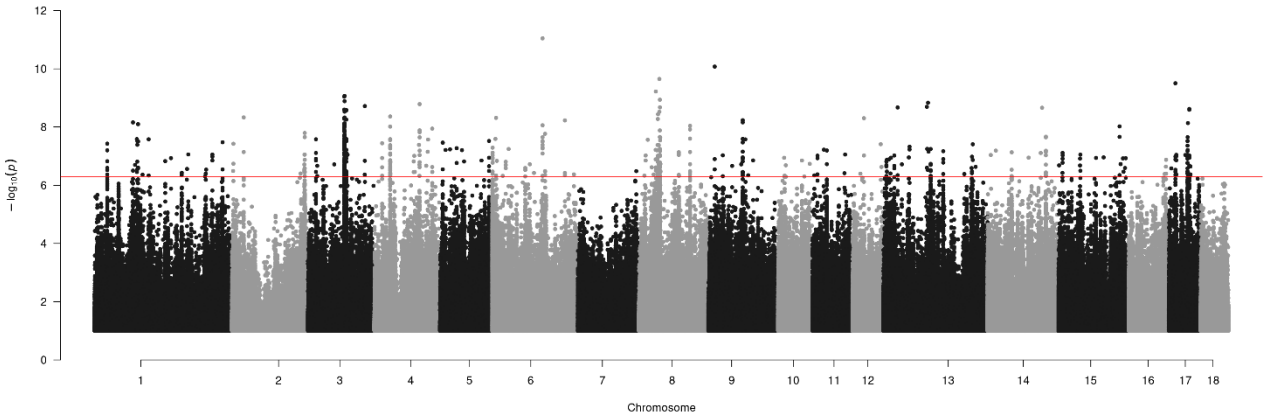


Figure 5. Manhattan plot for phenotypic group of traits stature with significance threshold 0.01. Each change of color is a new chromosome. Each peak is a possibly under selection region.

The enrichment of significant variants associated with phenotype was tested (Figure 6). The significant SNPs in the phenotypic group fatness were significantly enriched in the intergenic, 5' prime UTR and coding region. The growth group had significantly enriched SNPs in the 5' prime UTR and the stature group has significantly enriched SNPs in the intronic region.

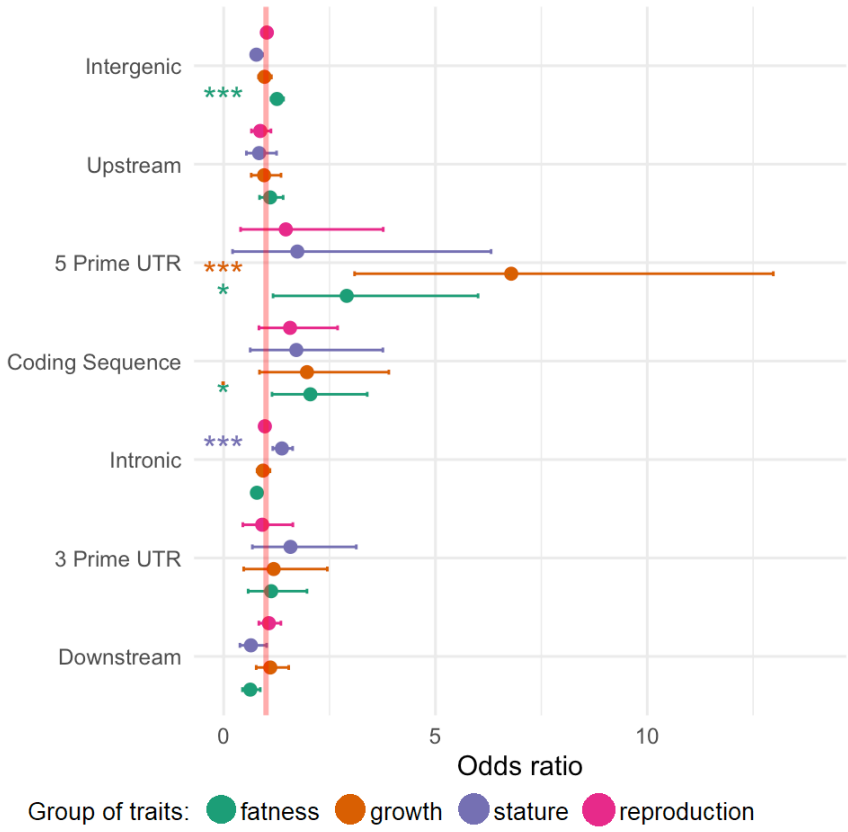


**Figure 6**. Enrichment of SNPs associated with different traits into functional categories.

Across the genome, windows/regions with at least 3 significant SNPs with less than 200 kbp apart were considered as selection signals for phenotypes. Overall, a total of 561 regions/windows ranging in length from 58 bp to 1.282 Mbp were discovered. Among the detected regions, 51 regions were in the stature group, 47 regions were in the growth group, 221 regions were in the fatness group, and 242 regions were in the reproduction group. For a list of all selection signatures for phenotypes, see Additional File 4.

Within the detected regions in each phenotypic group, the 20 most significant regions were used for interpretation and more detailed description (Figure 7, Additional File 5: Table S1-S4). Regions in the phenotypic group stature contained regions between 1.6 kbp and 565.7 kbp with 3 to 63 SNPs and included from none to five genes. The strongest selection signal for stature was detected on

406 chromosome SSC6, which contains the *ARHGAP28* gene. The phenotypic group growth contained
407 regions from 6.3 to 709.1 kbp with 3 to 124 significant SNPs and comprised between zero and 20
408 genes. The region with the strongest selection signal was on SSC18 and did not contain any gene.
409 The phenotypic group fatness contained regions ranging from 3.4 kbp to 980.9 kbp in length with 4
410 to 240 SNPs and included between none and 30 genes. The region with the strongest selection signal
411 was on SSC3 and contained 7 genes, including *POMC*, *DNMT3A*, *EFR3B*, and *NCOA1*. The
412 phenotypic group reproduction contained regions from 5.3 kbp to 1,282.0 kbp in length with 3 to 115
413 SNPs and zero to 48 genes. The region with the strongest signals was located on chromosome SSC6
414 and contained 4 genes (i.e. *YES1*, *ENOSF1*, *TYMS*, and *CLUL1*). The detected selection signals for
415 production traits were further examined to find overlaps between groups (Figure 7). In total, there
416 were 12 overlaps.

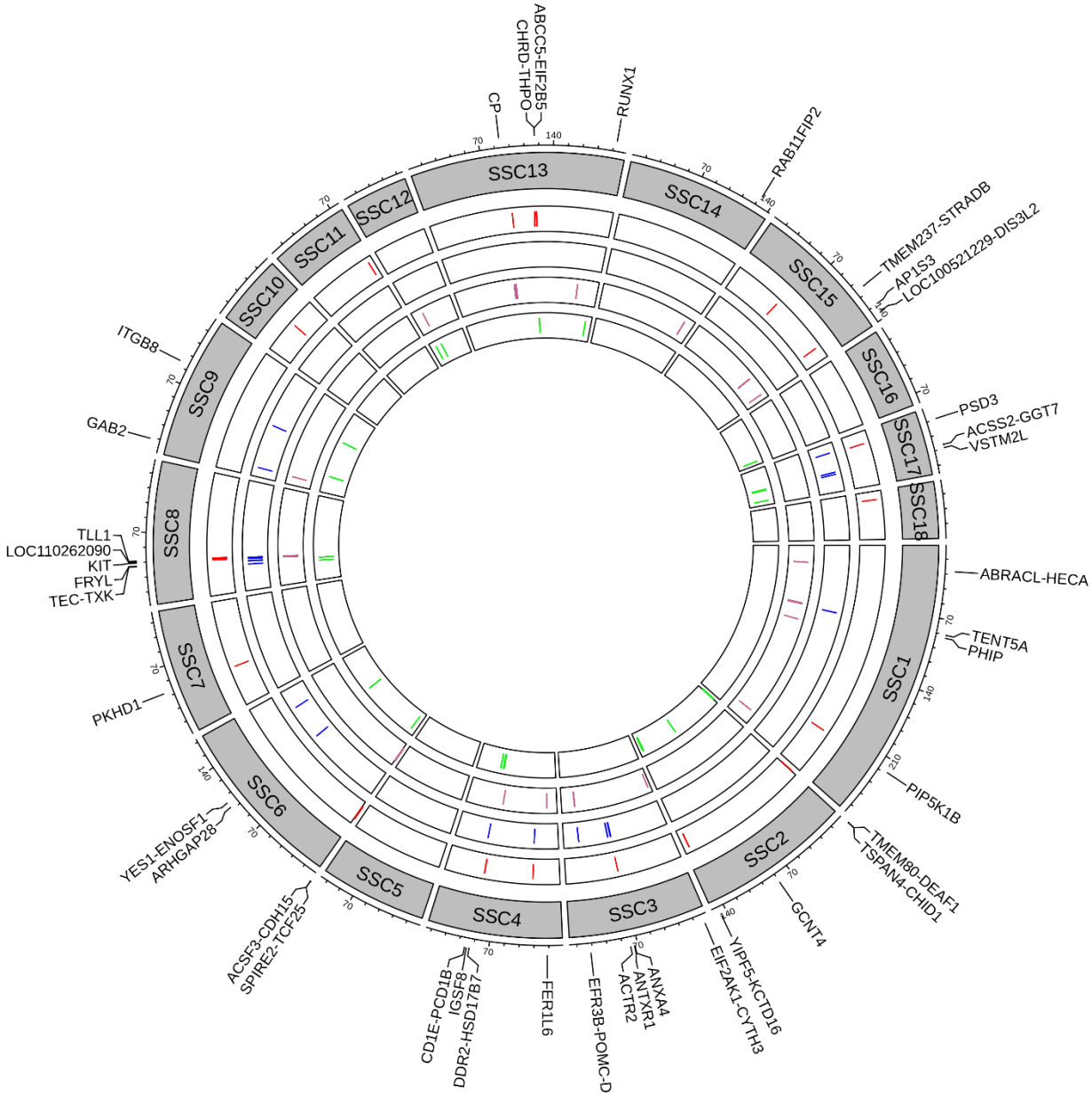


**Figure 7**. Circular plot summarizing top 20 regions with candidate genes with the highest P-value for each phenotypic group. Discovered regions associated with growth are coloured in red, with stature are coloured in blue, with fatness are coloured in purple and with reproductive performance in green.

# Discussion

Deciphering the genetic basis of variation in complex traits can help to understand their biology and evolution, and improve breeding, selection programs and conservation plans of animal genetic resources. One approach to address this question is to link genetic and phenotypic variation by

425 identifying genomic regions that have evolved in response to selection on complex traits. Establishing
426 such links requires a collection of data on genetic groups (e.g. populations, lines, breeds) that share a
427 common origin and have information on adaptive traits that can be phenotyped in a standardized
428 fashion. Livestock species generally meet these criteria and therefore provide good models for
429 mapping genomic regions associated with selective constraints. Here, we studied local pig
430 populations that were genetically characterized using individual genotyping and pool sequencing for
431 which a recent common phenotypic database was constituted. The genetic relationships between
432 breeds confirmed genome scans for response to selection for a wide range of traits measured as part
433 of their breeding management.

## 434 Genetic diversity of European local pig breeds

435 Using individual genotyping of a large set of European breeds, we have found that their genetic
436 variation is attracted to four main genetic groups: the Duroc-like, the Large White pigs, the Landrace
437 and the Iberian-like. The origin of the Duroc is unclear, as it originated from American continent and
438 was then brought back to Europe. This history implies that it was separated from the European breeds
439 for a long time, which is evident in our genetic analysis by its rather large genetic distance with most
440 European breeds, except for Mora Romagnola breed from Italy, which has a history of crossbreeding
441 with Duroc breed (32). In addition, the present study showed that Mora Romagnola (and also
442 Turopolje pig breed) differentiated early in the PCA analysis, which can be explained with the high
443 levels of inbreeding (33). The White pig group is composed of the Lietuvos breeds and the Large
444 White, all of which originate from Northern Europe and are clearly separated from the Landrace,
445 which includes populations originating from more central countries such as Germany (Landrace,
446 Schwäbisch-Hällisches), Belgium (Pietrain) and Slovenia (Krškopolje). The Large White and
447 Landrace groups are close to the root of the population tree, which could be explained by the well-
448 documented influence (34) of Asian pigs on these genetic pools. Finally, the Iberian group includes

449 a set of breeds that are quite geographically dispersed; while many of the Iberian-type breeds are
450 found around southwestern Europe (Spain, Portugal, and France), some populations from
451 Eastern/Central Europe clearly belong to the same genetic pool (e.g. Mangalitsa from Serbia,
452 Turopolje from Croatia, Cinta Senese from Italy) as it was already reported in Muñoz et al. (9). A
453 possible interpretation is that the genetic background responsible for variation in the Iberian pig breed
454 was historically widespread in Europe and can still be found in some local pig breeds. Overall,
455 analysis of the genetic structure of European local pig breeds reveals extensive genetic variation
456 clustered in differentiated breeds with continuously distributed genetic distances. The historical
457 events that led to the present genetic variation are most likely complex, and involve differentiation
458 from an ancestral pool followed by or accompanied by outcrossing between populations, including
459 wild boars, and cannot be reconstructed from the data used here. However, the sample of available
460 breeds is well adapted to our main objective, as we can interrogate a diverse set of polymorphisms
461 for adaptation to contrasting environments or production systems in different genetic backgrounds.

462 While SNP genotyping is sufficient to characterize the genetic relationships and population structure
463 of local pig breeds, the density of the SNP array used limits the resolution and power to detect
464 potential associations with breed phenotypes especially in local breeds where a big proportion of
465 SNPs included in the commercial chips are not segregating. To alleviate this issue, a cost-effective
466 alternative to SNP genotyping consists of pool sequencing of populations. In the dataset used here,
467 we confirmed that allele frequency and genetic structure estimates obtained from pool sequencing are
468 consistent with those obtained from SNP array genotyping as it was already reported in Bovo et al.
469 (8), while providing comprehensive information on the genetic diversity of local pig breeds.

470 **Phenotypic diversity of European local pig breeds**

471 In parallel with the comprehensive information on the genetic diversity of local pig breeds, we
472 developed an original approach based on PCA to characterize their phenotypic diversity in traits
473 related to stature, growth, fatness and reproduction, which have been described in a comprehensive
474 analysis (3). In addition, local breeds could be phenotypically characterized in regard to meat quality
475 traits (pH24, pH48, intramuscular fat content and *longissimus dorsi* muscle colour, Additional File
476 S1: Figure S8). However, meat quality traits (with the exception of IMF) are strongly influenced by
477 pre-slaughter handling stress (35). Breed score for meat quality was therefore not included in the
478 phenotype-genotype selection scan.

479 Local pig breeds exhibit a wide variety of exterior phenotypic characteristics, including body size or
480 body weight. This could be due to genetic factors or could reflect the management of the breed. For
481 example, male and female animals of the Schwäbisch-Hällisches breed are medium to large in size
482 and also have well-developed management systems. In contrast, more untapped breeds (e.g.
483 Moravka) are smaller and lighter (3). Moreover, the distribution of local pig breeds in stature traits
484 was similar to that in the growth performance group.

485 Regarding the growth performance of local pig breeds, existing knowledge is limited. Moreover, the
486 production systems may not be sufficiently adapted to the needs of the breeds which could be reflected
487 in the phenotype of the breeds. For example, the study of Brossard et al. (36) argues that in many
488 studies on local pig breeds, animals are fed below ad libitum level and, therefore, do not reach their
489 full production potential. This is consistent with an analytical literature review (37) on European local
490 pig breeds, which demonstrated that local pig breeds are generally fed in ad libitum conditions only
491 earlier in life (during the growing and early fattening phase). In the late fattening phase, the feeding
492 is likely to be restricted. To limit the influence of such effects in the present study and to use a better
493 representation of breeds' growth potential, only average daily gain values from the beginning of the

494 lactation to the end of the early fattening phase (i.e. up to 60 kg) were used for the phenotypic score
495 on growth. Despite this, better growth performance might still partly reflect breeding management,
496 as shown for the Schwäbisch-Hällisches pig, which has a growth potential comparable to genetically
497 improved modern pig breeds (38,39).

498 It has been reported that local pig breeds have a lower protein and higher lipid deposition capacity
499 compared to modern pig breeds (36). The higher genetic capacity of local pig breeds for backfat and
500 intramuscular fat deposition, and lower muscle accretion has been demonstrated in several studies
501 comparing fat characteristics between local and modern breeds (40–44). Although the local pig breeds
502 are recognised as fatty, they still exhibit large phenotypic variability in the dataset used here.

503 Regarding reproduction traits, sows of local breeds typically exhibit relatively high age at first
504 parturition, few litters per sow per year, long lactation periods, small litter sizes and high piglet
505 mortality (37). In our study, PCA of reproductive performance distinguished breeds with better
506 reproductive efficiency, which are usually reared in more intensive systems within more developed
507 pork chains (e.g. Schwäbisch-Hällisches pig) from breeds with lower reproductive efficiency (e.g.
508 Nero Siciliano, Turopolje, Casertana, Mora Romagnola, Mangalitsa) reared in extensive or semi-
509 extensive systems (3).

510 While some of this variation in phenotypes can certainly be explained by differences in production
511 systems among local breeds, our genetic association results suggest that some of it may be linked to
512 genetic variation. In order to be able to test for these genetic associations, the methods used require
513 all breeds to have an associated phenotype. To do so, we imputed using PCA the missing phenotypes
514 of some breeds (see Methods). The resulting imputed phenotypes are likely to be biased toward the
515 average of all breeds. This is for example most likely the case for the Sarda breed, considered to be a

516 very small breed (in term of body size) but lying at an intermediate value after imputation. This
517 certainly limits our power to detect genetic associations but is unlikely to create false positives.

## 518 Genome scan of genomic regions associated with phenotypic differentiation in 519 European local pig breeds

520 The functions of some candidate genes identified within the signatures of selection indicated that
521 some local pig breeds may have adapted to specific phenotypic traits such as stature, growth, fatness
522 or reproductive performance. For these regions, it is believed that genetic variations contributed to
523 the development of different phenotypic traits in European local pig breeds. For example, a strong
524 association was found for stature in a region on chromosome SSC3 containing the *ANXA4* gene.
525 Annexin A4 encodes a calcium-dependent phospholipid-binding protein involved in various
526 membrane processes. This candidate gene or region has been previously proposed as a QTL for
527 stature in cattle (45,46). Another genomic region associated with the signature of selection for stature
528 and growth was discovered on SSC3 containing also candidate gene *ANTXR1*. This gene is involved
529 in cell morphogenesis, cellular development process and cytoskeleton organisation. Part of this region
530 has also been previously proposed as a QTL for bovine stature (47). On chromosome SSC8, a region
531 containing the *TLL1* gene was found, which encodes a tolloid member of metalloprotease that
532 converts latent to biologically active myostatin, and thus inhibits skeletal muscle development (48).
533 This gene has previously been linked to average daily gain in cattle (49). In the fatness and also stature
534 group, an overlapping region was found on chromosome SSC3. This region contained common
535 candidate genes; namely *DNMT3A* (responsible for CpG methylation), *POMC* (prohormone) and
536 *EFR3B* (localises phosphatidylinositol 4-kinase to the plasma membrane). *DNMT3A* has been
537 previously shown to affect stature and body weight in cattle and humans (47,50,51). In addition, the
538 *DNMT3A* gene has previously been implicated to the regulation of adipose tissue development (52).
539 Another candidate gene found in the same region is the *POMC* gene, which encodes the precursor of

540 several peptide hormones that contribute to the regulation of feed intake and energy balance via the
541 leptin/melanocortin pathway (53). Polymorphisms in *POMC* have previously been associated with
542 *longissimus dorsi* muscle area and backfat thickness in cattle (54–56) and with obesity and body mass
543 index in humans (57,58). Additionally, in the fatness group, the *NCOA1* gene was detected in the
544 same region on SSC3, which plays an important role in adipogenesis. An interesting signature of
545 selection on reproductive performance was found on chromosome SSC4, which contains the
546 *HSD17B7* gene. This gene encodes an enzyme involved in the biosynthesis of sex steroids and
547 cholesterol (59,60). Due to the clear role of the *HSD17B7* gene in steroidogenesis, this gene has been
548 proposed as a good candidate gene for reproductive performance in local pig breeds.

549 Several other regions have also been identified as selection signatures for stature, growth, and fatness
550 with not so direct connection to their biological function in production traits, but still with some
551 association to phenotype. These regions could be useful for further/future studies on selection
552 signatures. For example, chromosome SSC6 contained a region with possible adaptation on stature
553 with the gene *ARHGAP28*, which has previously been associated with the number of vertebrae in
554 pigs, thus affecting carcass length (61). On chromosome SSC7, a region under adaptation on growth
555 containing the gene *PKHD1* has been found, which has been previously associated with carcass
556 conformation score in cattle (62). In the fatness group, the gene *EIF2AK1* (located on chromosome
557 SSC3) was identified, with a role of inhibiting of the protein synthesis in response to stress. This gene
558 has previously been associated to body mass index in pigs (63).

559 Some of regions that have been discovered must be carefully interpreted. For instance, the conducted
560 studies differ in production conditions. Therefore, correlations between phenotypic groups could
561 create non-causal signal with genes. An example of a non-causal signal due to phenotypic correlation
562 found in the growth and fatness groups is a region on SSC8 containing the gene *KIT*. This gene

563 encodes the tyrosine kinase receptor and has been associated with coat colour in pigs. Another
564 selection signature associated with coat colour was found on chromosome SSC6 in the growth and
565 fatness groups. It contains the *MC1R* gene, which plays a major role in controlling the transition from
566 eumelanin (black or brown) to pheomelanin (yellow to red) (64). The same selection signatures (i.e.
567 *KIT* and *MC1R* gene) were also observed using the same pool-sequencing data in the study of Bovo
568 et al. (8). Since the coat colour was favoured by breeders, it was strongly selected in several local
569 breeds (65) (e.g. White breeds were under selection for leanness and better growth performance).
570 Interestingly, the study performed on SNP-chip data on the same animals/breeds (9) did not detect
571 any signal near *MC1R* or *KIT* genes probably due to different SNP-chip informativity and statistical
572 approach used in their study.

573 **Comparison of the detected regions with other studies**

574 Detected selection signatures were compared with previous study performed on European local pig
575 breeds using the same whole genome re-sequencing data (8). Selection signals in individual breeds,
576 different groups of breeds (i.e. local vs commercial pig breeds and domestic vs wild boar) and on
577 breeds with distinct exterior traits (i.e. coat colour, body size) were searched using Fst approach. The
578 comparison of both approaches revealed few common regions, probably due to the different
579 methodologies applied. For instance, an overlapping region was found on chromosome SSC8 in the
580 stature group containing *FRYL* gene, which is responsible for maintaining the integrity of polarized
581 cell extensions during morphogenesis. In the fatness group, two overlapping selection signals were
582 detected. The first one on chromosome SSC14 contained *RAB11FIP2* gene (found in Alentejano
583 breed) and the second one on chromosome SSC15 contained several candidate genes (e.g.,
584 *TMEM237*, *STRADB*, *MPP4*, *CASP10*). This selection signal was in the study of Bovo et al. (8)
585 detected in two individual breeds (i.e. Casertana and Negre Mallorqui), in the comparison of breeds
586 with small and large body sizes and in the comparison of local and commercial breeds.

# Conclusions

In this study, DNA-pool sequencing data of European local pig breeds were used to identify genomic regions associated with phenotypic differences in groups of traits related to stature, growth, fatness and reproductive performance. The genome scan for selection revealed several strong candidate genes with potential implication in adaptation of European local pig breeds to different production systems. This study will be helpful for future conservation, association or selection approaches in these European local pig breeds.

# Declarations

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of data and materials**

Data used in this study are available from the original sources as detailed in the Materials and Methods section.

# Competing interests

The authors declare that they have no competing interests.

**Funding**

This research was funded by the Horizon 2020 Framework Programme, grant number No 634476 and Slovenian Research Agency (grants P4-0133, J4-3094, PhD scholarship for K.P.).

# Authors' contributions

BS, MČP and MŠ conceived the study. MČP and MŠ obtained project funding and were responsible for project administration. BS supervised the study and designed the methodology. CO, LF, SB, GS,

AR, MM, MG, RB. RC, RQ, GK, MJM, CZ, VR, JPA, ČR, RS, DK provided data and samples. KP, CM, BS performed data analysis. KP, CM, BS, MČP, MŠ, LF, CO and JR contributed to data interpretation. KP, CM and BS wrote the initial draft of the paper. All authors reviewed and edited paper. All authors read and approved the final manuscript.

# Acknowledgements

This study has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 634476 (project acronym TREASURE). Project TREASURE consortium is acknowledged for providing the samples. The content of this paper reflects only the author's view and the European Union Agency is not responsible for any use that may be made of the information it contains. Core financing of the Slovenian Research Agency (grant P4-0133, J4-3094, PhD scholarship for KP) for KP, MČP and MŠ is acknowledged.

# References

1. Tribout T, Caritez JC, Gruand J, Bouffaud M, Guillouet P, Billon Y, et al. Estimation of genetic trends in French Large White pigs from 1977 to 1998 for growth and carcass traits using frozen semen. J Anim Sci. 2010;88(9):2856–67.

2. Rauw WM, Kanis E, Noordhuizen-Stassen EN, Grommers FJ. Undesirable side effects of selection for high production efficiency in farm animals: a review. Livest Prod Sci. 1998;56(1):15–33.

3. Čandek-Potokar M, Nieto R. European Local Pig Breeds - Diversity and performance. A study of project TREASURE. London: Intech Open; 2019. 303 p.

4. Pugliese C, Sirtori F. Quality of meat and meat products produced from southern European pig breeds. Meat Sci. 2012;90(3):511–8.

5. Porter V. Pigs: A Handbook to the breeds of the world. New York: Cornell University Press; 1993. 256 p.

6. Ribani A, Utzeri VJ, Geraci C, Tinarelli S, Djan M, Veličković N, et al. Signatures of de-domestication in autochthonous pig breeds and of domestication in wild boar populations from MC1R and NR6A1 allele distribution. Anim Genet. 2019;50(2):166–71.

7. Poklukar K, Čandek-Potokar M, Batorek Lukač N, Tomažin U, Škrlep M. Lipid deposition and metabolism in local and modern pig breeds: A review. Animals. 2020;10(3):424.

8. Bovo S, Ribani A, Muñoz M, Alves E, Araujo JP, Bozzi R, et al. 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. Gen Sel Evol. 2020;52(1):33.

9. Muñoz M, Bozzi R, García-Casco J, Núñez Y, Ribani A, Franci O, et al. Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip. Sci Rep. 2019;9(1):1–14.

646 10. Muñoz M, Bozzi R, García F, Núñez Y, Geraci C, Crovetti A, et al. Diversity across major and
647 candidate genes in European local pig breeds. PLOS ONE. 2018 Nov 20;13(11):e0207475.

648 11. Amaral AJ, Bressan MC, Almeida J, Bettencourt C, Moreira O, Sá J, et al. Combining
649 genome-wide association analyses and gene interaction networks to reveal new genes
650 associated with carcass traits, meat quality and fatty acid profiles in pigs. Livest Sci.
651 2019;220:180–9.

652 12. Schiavo G, Bovo S, Tinarelli S, Bertolini F, Dall'Olio S, Gallo M, et al. Genome-wide
653 association analyses for several exterior traits in the autochthonous Casertana pig breed. Livest
654 Sci. 2019;230:103842.

655 13. Jiang Y, Tang S, Wang C, Wang Y, Qin Y, Wang Y, et al. A genome-wide association study
656 of growth and fatness traits in two pig populations with different genetic backgrounds. J Anim
657 Sci. 2018;96(3):806–16.

658 14. Wang Y, Ning C, Wang C, Guo J, Wang J, Wu Y. Genome-wide association study for
659 intramuscular fat content in Chinese Lulai black pigs. Asian-Australas J Anim Sci.
660 2019;32(5):607–13.

661 15. Strucken EM, Schmitt AO, Bergfeld U, Jurke I, Reissmann M, Brockmann GA. Genomewide
662 study and validation of markers associated with production traits in German Landrace boars. J
663 Anim Sci. 2014;92(5):1939–44.

664 16. Wilkinson S, Lu ZH, Megens H-J, Archibald AL, Haley C, Jackson IJ, et al. Signatures of
665 Diversifying Selection in European Pig Breeds. PLOS Genet. 2013;9(4):e1003453.

666 17. Yang S, Li X, Li K, Fan B, Tang Z. A genome-wide scan for signatures of selection in
667 Chinese indigenous and commercial pig breeds. BMC Genet. 2014;15(1):7.

668 18. Choi J-W, Choi B-H, Lee S-H, Lee S-S, Kim H-C, Yu D, et al. Whole-Genome Resequencing
669 Analysis of Hanwoo and Yanbian Cattle to Identify Genome-Wide SNPs and Signatures of
670 Selection. Mol Cells. 2015;38(5):466–73.

671 19. Gurgul A, Jasielczuk I, Ropka-Molik K, Semik-Gurgul E, Pawlina-Tyszko K, Szmatoła T, et
672 al. A genome-wide detection of selection signatures in conserved and commercial pig breeds
673 maintained in Poland. BMC Genet. 2018;19:95.

674 20. Bansal V. A statistical method for the detection of variants from next-generation resequencing
675 of DNA pools. Bioinform. 2010;26(12):i318–24.

676 21. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated
677 individuals. Genome Res. 2009;19(9):1655–64.

678 22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence
679 alignment/map format and SAMtools. Bioinform. 2009;25(16):2078–9.

680 23. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between
681 populations using sequencing of pooled DNA samples (Pool-Seq). Bioinform.
682 2011;27(24):3435–6.

683 24. Hivert V, Leblois R, Petit EJ, Gautier M, Vitalis R. Measuring Genetic Differentiation from
684 Pool-seq Data. Genetics. 2018;210(1):315–30.

685 25. Josse J, Husson F. missMDA: A Package for Handling Missing Values in Multivariate Data
686 Analysis. Journal of Statistical Software. 2016 Apr 4;70(1):1–31.

687 26. Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. J Stat Softw.
688 2008;25(1):1–18.

689 27. Josse J, Pagès J, Husson F. Multiple imputation in principal component analysis. Adv Data
690 Anal Classif. 2011;5(3):231–46.

691 28. Coop G, Witonsky D, Di Rienzo A, Pritchard JK. Using environmental correlations to identify
692 loci underlying local adaptation. Genetics. 2010;185(4):1411–23.

693 29. Gautier M. Genome-Wide Scan for Adaptive Divergence and Association with Population-
694 Specific Covariates. Genetics. 2015;201(4):1555–79.

695 30. Storey JD, Tibshirani R. Statistical significance for genomewide studies. PNAS.
696 2003;100(16):9440–5.

697 31. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular
698 visualization in R. Bioinform. 2014;30(19):2811–2.

699 32. Tinarelli S, Ribani A, Utzeri VJ, Taurisano V, Bovo C, Dall'Olio S, et al. Redefinition of the
700 Mora Romagnola Pig Breed Herd Book Standard Based on DNA Markers Useful to
701 Authenticate Its "Mono-Breed" Products: An Example of Sustainable Conservation of a
702 Livestock Genetic Resource. Animals. 2021 Feb;11(2):526.

703 33. Schiavo G, Bovo S, Muñoz M, Ribani A, Alves E, Araújo JP, et al. Runs of homozygosity
704 provide a genome landscape picture of inbreeding and genetic history of European
705 autochthonous and commercial pig breeds. Animal Genetics. 2021;52(2):155–70.

706 34. Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Cagan A, Bosse M, et al. Evidence of
707 long-term gene flow and selection during domestication from analyses of Eurasian wild and
708 domestic pig genomes. Nat Genet. 2015;47(10):1141–8.

709 35. Dokmanovic M, Baltic MZ, Duric J, Ivanovic J, Popovic L, Todorovic M, et al. Correlations
710 among stress parameters, meat and carcass quality parameters in pigs. Asian-Australas J Anim
711 Sci. 2015;28(3):435–41.

712 36. Brossard L, Nieto R, Charneca R, Araujo JP, Pugliese C, Radović Č, et al. Modelling
713 nutritional requirements of growing pigs from local breeds using InraPorc. Animals.
714 2019;9(4):169.

715 37. Čandek-Potokar M, Batorek-Lukač N, Tomažin U, Škrlep M, Nieto R. Analytical review of
716 productive performance of local pig breeds. In: European local pig breeds - diversity and
717 Performance A study of project TREASURE. London: Intech Open; 2019. p. 281–303.

718 38. Petig M, Zimmer C, Bühler R, Batorek-Lukač N. Schwäbisch-Hällisches Pig. In: European
719 local pig breeds - diversity and performance A study of project TREASURE. London: Intech
720 Open; 2019. p. 258–66.

721 39. Bozzi R, Škrlep M, Lenoir H, Lebret B, Gasco G, Petig M, et al. Survey of demographic and
722 phenotypic data of local pig breeds of TREASURE project. In: Archivos de Zootecnia. 2018.
723 p. 4.

724 40. Serra X, Gil F, Perez-Enciso M, Oliver MA, Vazquez JM, Gispert M, et al. A comparison of
725 carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and
726 Landrace pigs. Livest Prod Sci. 1998;56(3):215–23.

727 41. Lebret B, Dourmad JY, Mourot J, Pollet PY, Gondret F. Production performance, carcass
728 composition, and adipose tissue traits of heavy pigs: influence of breed and production system.
729 J Anim Sci. 2014;92(8):3543–56.

730 42. Madeira MS, Pires VMR, Alfaia CM, Costa ASH, Luxton R, Doran O, et al. Differential
731 effects of reduced protein diets on fatty acid composition and gene expression in muscle and
732 subcutaneous adipose tissue of Alentejana purebred and Large White × Landrace × Pietrain
733 crossbred pigs. Br J Nutri. 2013;110(2):216–29.

734 43. Palma-Granados P, Haro A, Seiquer I, Lara L, Aguilera JF, Nieto R. Similar effects of lysine
735 deficiency in muscle biochemical characteristics of fatty and lean piglets. J Anim Sci.
736 2017;95(7):3025–36.

737 44. Parunović N, Petrović M, Matekalo-Sverak V, Radović Č, Stanišić N. Carcass properties,
738 chemical content and fatty acid composition of the musculus longissimus of different pig
739 genotypes. S Afr j anim sci. 2013;43(2):123–36.

740 45. Kolbehdari D, Wang Z, Grant JR, Murdoch B, Prasad A, Xiu Z, et al. A whole-genome scan to
741 map quantitative trait loci for conformation and functional traits in Canadian Holstein bulls. J
742 Dairy Sci. 2008;91(7):2844–56.

743 46. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, et al. Genome-wide
744 association analysis of thirty one production, health, reproduction and body conformation
745 traits in contemporary U.S. Holstein cows. BMC Genom. 2011;12(1):408.

746 47. Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, et al.
747 Meta-analysis of genome-wide association studies for cattle stature identifies common genes
748 that regulate body size in mammals. Nat Genet. 2018;50(3):362–7.

749 48. Kostyunina DS, Ivanova AD, Smirnova OV. Myostatin: Twenty Years Later. Hum Physiol.
750 20;44(1):88–101.

751 49. Higgins MG, Fitzsimons C, McClure MC, McKenna C, Conroy S, Kenny DA, et al. GWAS
752 and eQTL analysis identifies a SNP associated with both residual feed intake and GFRA2
753 expression in beef cattle. Sci Rep. 2018;8(1):1–12.

754 50. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of
755 common variation in the genomic and biological architecture of adult human height. Nat
756 Genet. 2014;46(11):1173–86.

757 51. Liu X, Usman T, Wang Y, Wang Z, Xu X, Wu M, et al. Polymorphisms in epigenetic and
758 meat quality related genes in fourteen cattle breeds and association with beef quality and
759 carcass traits. Asian Australas J Anim Sci. 2015;28(4):467–75.

760 52. Ma X, Kang S. Functional implications of DNA methylation in adipose biology. Diabetes.
761 2019;68(5):871–8.

762 53. Switonski M, Stachowiak M, Cieslak J, Bartz M, Grzes M. Genetics of fat tissue accumulation
763 in pigs: a comparative approach. J Appl Genet. 2010;51(2):153–68.

764 54. Gill JL, Bishop SC, McCorquodale C, Williams JL, Wiener P. Associations between single
765 nucleotide polymorphisms in multiple candidate genes and carcass and meat quality traits in a
766 commercial Angus-cross population. Meat Sci. 2010;86(4):985–93.

767 55. Seong J, Kong HS. Association between polymorphisms of the CRH and POMC genes with
768 economic traits in Korean cattle (Hanwoo). Genet Mol Res. 2015;14(3):10415–21.

769 56. Liu Y, Zan L, Li L, Xin Y. Proopiomelanocortin gene polymorphisms and its association with
770 meat quality traits by ultrasound measurement in Chinese cattle. Gene. 2013;529(1):138–43.

771 57. Singh RK, Kumar P, Mahalingam K. Molecular genetics of human obesity: A comprehensive
772 review. Comptes Rendus Biologies. 2017;340(2):87–108.

773 58. Harno E, Gali Ramamoorthy T, Coll AP, White A. POMC: The physiological power of
774 hormone processing. Physiol Rev. 2018;98(4):2381–430.

775 59. LaVoie HA. Transcriptional control of genes mediating ovarian follicular growth,
776 differentiation, and steroidogenesis in pigs. Mol Reprod Dev. 2017;84(9):788–801.

777 60. Robic A, Feve K, Louveau I, Riquet J, Prunier A. Exploration of steroidogenesis-related genes
778 in testes, ovaries, adrenals, liver and adipose tissue in pigs. Anim Sci J. 2016;87(8):1041–7.

779 61. Rohrer GA, Nonneman DJ, Wiedmann RT, Schneider JF. A study of vertebra number in pigs
780 confirms the association of vertnin and reveals additional QTL. BMC Genet. 2015;16(1):129.

781 62. Mao X, Sahana G, De Koning D-J, Guldbrandtsen B. Genome-wide association studies of
782 growth traits in three dairy cattle breeds using whole-genome sequence data. J Anim Sci.
783 2016;94(4):1426–37.

784 63. Gong H, Xiao S, Li W, Huang T, Huang X, Yan G, et al. Unravelling the genetic loci for
785 growth and carcass traits in Chinese Bamaxiang pigs based on a 1.4 million SNP array. J Anim
786 Breed Genet. 2019;136(1):3–14.

787 64. Kijas JMH, Wales R, Törnsten A, Chardon P, Moller M, Andersson L. Melanocortin receptor
788 1 (MC1R) mutations and coat color in pigs. Genetics. 1998;150(3):1177–85.

789 65. Groenen MAM. A decade of pig genome sequencing: a window on pig domestication and
790 evolution. Genet Sel Evol. 2016;48(1):23.

791

792

# Figures

**Figure 1 Principal component analysis of 24 pig breeds genotyped on a medium density SNP array.**

The most differentiated breeds are highlighted with different colours.

**Figure 2**. **Population tree and admixture analysis of 23 pig breeds genotyped on a medium density SNP array**.

The population tree reconstructed from pairwise genetic distances (Fst) is shown on the left, and the admixture component for all individuals a priori belonging to the breed is shown on the right. The colour levels follow the global axes of genetic variation, with populations most closely related to the Iberian type shown in shades of blue, to the Duroc in shades of red, to the Large White in shades of purple, and to the Landrace in shades of green. Heterogeneous populations, or those equidistant from these four clusters, are shown in orange. The Nero Siciliano breed exhibited extremely high heterogeneity, and was therefore not included in the reconstruction of the population tree and is not shown in this figure.

**Figure 3. Principal component analysis showing the relationship between breeds (A) and the traits associated with stature (B) and the corresponding phenotypic breed scores (C).**

Breeds (A) coloured in grey are the breeds with more than 50% of missing variables, thus, their position on the PCA must be interpreted carefully. The variables (B) are coloured according to quality of the representation, which is measured by squared cosine between the vector originating from the element and its projection on the axis. The variables that contribute most to the separation of the trait into PC1 and PC2 are coloured black. Breeds (A, C) are coloured according to genetic similarity. Breeds (A, C) in green are genetically Landrace-like breeds, in purple are Large White-like breeds, in blue are Iberian-like breeds, in red are Duroc-like breeds and in light blue are Gascon and Basque.

**Figure 4. Global differences in production traits for 20 European local pig populations.**

Breeds are distributed according to breed scores for phenotypic traits and colored according to genetic similarity. Lower values for breed score growth are representing lower average daily gain from lactation up to 60 kg of live weight, while higher values represent higher average daily gain in the same growth period. Low values for breed score reproduction represent low reproductive performance, while high values represent higher reproductive performance of breeds. Low values for breed score stature represent lighter and smaller breeds, while high values represent heavier and larger breeds. Low values for the breed score fat represent leaner breeds, while high values represent fatter breeds. Breeds in green are genetically Landrace-like breeds, in purple are Large White-like breeds, in blue are Iberian-like breeds, in red are Duroc-like breeds and in light blue are Gascon and Basque.

**Figure 5. Manhattan plot for phenotypic group of traits stature with significance threshold 0.01.**

828 Each change of color is a new chromosome. Each peak is a possibly under selection region.

829 **Figure 6. Enrichment of SNPs associated with different traits into functional categories.**

830 **Figure 7. Circular plot summarizing top 20 regions with candidate genes with the highest P-**
831 **value for each phenotypic group.**

832 Discovered regions associated with growth are coloured in red, with stature are coloured in blue, with
833 fatness are coloured in purple and with reproductive performance in green.

834

# Additional files

**Additional file 1 Figure S1**

Title: Plot od admixture cross-validation error from K=2 to K=40.

**Additional file 1 Figure S2**

Title: Principal component analysis of pool-sequencing SNPs extracted from SNP-chip data.

**Additional file 1 Figure S3**

Title: A matrix of missing phenotypic variables in a database of European local pig breeds. The yellow colour is representing missing variables.

Description: SFA = saturated fatty acid content, PUFA = polyunsaturated fatty acid content, Piglets W weight = piglets weaning weight, MUFA = monounsaturated fatty acid content, Litter W birth = litter weaning weight, LD IMF = *longissimus dorsi* intramuscular fat content, M = male, F = female, Death WN = death rate to weaning, BFT = backfat thickness, GM = *gluteus medius* muscle, ADG1 = Average daily gain during the lactation period, ADG2 = Average daily gain in the growing period from weaning to 30 kg of weight, ADG3 = Average daily gain during fattening period from 30 kg to 60 kg.

**Additional file 1 Figure S4**

Title: Principal component analysis showing the relationship between breeds (A) and the traits associated with growth (B) and the corresponding phenotypic breed scores (C).

Description: Breeds (A) coloured in grey are the breeds with more than 50% of missing variables, thus, their position on the PCA must be interpreted carefully. The variables (B) are coloured according to quality of the representation, which is measured by squared cosine between the vector originating from the element and its projection on the axis. The variables that contribute most to the separation of the trait into PC1 and PC2 are coloured black. Breeds (A, C) are coloured according to genetic similarity. Breeds (A, C) in green are genetically Landrace-like breeds, in purple are Large White-like breeds, in blue are Iberian-like breeds, in red are Duroc-like breeds and in light blue are Gascon and Basque.

**Additional file 1 Figure S5**

Title: Principal component analysis showing the relationship between breeds (A) and the traits associated with fatness (B) and the corresponding phenotypic breed scores (C).

Description: Breeds (A) coloured in grey are the breeds with more than 50% of missing variables, thus, their position on the PCA must be interpreted carefully. The variables (B) are coloured according to quality of the representation, which is measured by squared cosine between the vector originating

from the element and its projection on the axis. The variables that contribute most to the separation of the trait into PC1 and PC2 are coloured black. Breeds (A, C) are coloured according to genetic similarity. Breeds (A, C) in green are genetically Landrace-like breeds, in purple are Large White-like breeds, in blue are Iberian-like breeds, in red are Duroc-like breeds and in light blue are Gascon and Basque.

**Additional file 1 Figure S6**

Title: Principal component analysis showing the relationship between breeds (A) and the traits associated with reproduction performance (B) and the corresponding phenotypic breed scores (C).

Description: Breeds (A) coloured in grey are the breeds with more than 50% of missing variables, thus, their position on the PCA must be interpreted carefully. The variables (B) are coloured according to quality of the representation, which is measured by squared cosine between the vector originating from the element and its projection on the axis. The variables that contribute most to the separation of the trait into PC1 and PC2 are coloured black. Breeds (A, C) are coloured according to genetic similarity. Breeds (A, C) in green are genetically Landrace-like breeds, in purple are Large White-like breeds, in blue are Iberian-like breeds, in red are Duroc-like breeds and in light blue are Gascon and Basque.

**Additional file 1 Figure S7**

Title: Distribution of the European local pig breeds according to intramuscular fat content in longissimus dorsi (LD) muscle.

Description: Higher values on x-axis are representing higher intramuscular fat content, while the lower values are representing lower intramuscular fat content.

**Additional file 1 Figure S8**

Title: Principal component analysis showing the relationship between breeds (A) and the traits associated with meat quality (B) and the corresponding phenotypic breed scores (C).

Description: Breeds (A) coloured in grey are the breeds with more than 50% of missing variables, thus, their position on the PCA must be interpreted carefully. The variables (B) are coloured according to quality of the representation, which is measured by squared cosine between the vector originating from the element and its projection on the axis. The variables that contribute most to the separation of the trait into PC1 and PC2 are coloured black. Breeds (A, C) are coloured according to genetic similarity. Breeds (A, C) in green are genetically Landrace-like breeds, in purple are Large White-like breeds, in blue are Iberian-like breeds, in red are Duroc-like breeds and in light blue are Gascon and Basque.

**Additional file 2**

Title: Description of the statistical analysis used for adjustment of backfat thickness at the level of the last rib to final body weight of 120 kg.

903 **Additional file 3 Table 1**

904 Title: Description of phenotypic traits used for phenotypic characterisation of European local pig
905 breeds.

906 **Additional file 3 Table 2**

907 Title: Number of significant variants with corresponding q-values at the 1% level of false discovery
908 rate for phenotypic covariables.

909 **Additional file 4**

910 Title: A list of detected genomic regions associated with phenotypic differentiation in production
911 traits in European local pig breeds.

912 Format: xls

913 **Additional file 5 Table S1**

914 Title: Best 20 significant regions under adaptation on stature in European local pig breeds.

915 **Additional file 5 Table S2**

916 Title: Best 20 significant regions under adaptation on growth in European local pig breeds.

917 **Additional file 5 Table S3**

918 Title: Best 20 significant regions under adaptation on fatness in 20 European local pig breeds.

919 **Additional file 5 Table S4**

920 Title: Best 20 significant regions under adaptation on reproductive performance in 20 European local
921 pig breeds.