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-José 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. This study is based on European local pig breeds genetically characterized using DNA-pool sequencing data and phenotypically characterized using breed level phenotypes related to stature, fatness, growth and reproductive performance traits. These data were analyzed 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 16 significant genomic regions associated with stature, 24 with fatness, 2 with growth and 192 with reproduction. 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 the *ANXA4* and *ANTXR1* genes, for fatness containing the *DNMT3A* and *POMC* genes and for reproductive performance containing the *HSD17B7* gene.

Conclusions

The present study on European local pig breeds used a dedicated 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). In parallel, most local breeds have not been subjected to such intensive management or genetic improvement and their use has declined. Local breeds are often raised in extensive farming systems, resulting in adaptation to specific environmental conditions and (often) 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 the Iberian or the 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 genetic biodiversity conservation (3). Being exposed to specific selection pressures in different local environments, local pig breeds also represent interesting genetic resources (8). They could become more important in the future as a reservoir of diversity useful to adapt to global change.

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

120 particular heterogeneous phenotypic traits that could reflect specific genetic potential and
121 adaptation to different production systems. As the measurement of phenotypic traits in pigs is to
122 some extent standardised, it is possible to compare different local breeds. As shown in the study of
123 Čandek-Potokar and Nieto (3) reviewing productive traits, not only do European local pig breeds
124 differ from modern breeds, but they also exhibit an extensive variation among themselves.

125 To better understand the genetic basis for variation in phenotypic traits in local pig breeds, several
126 genome-wide association studies focused on detecting possible associations of loci with different
127 phenotypic attributes, such as morphological, production or meat quality traits (11–14). However,
128 these studies worked with small sample sizes, increasing the risk to give false negative results due
129 to low statistical power (15). Another approach to search for the association between genetic
130 polymorphisms and phenotype is to look for genomic regions that have responded to selection (i.e.,
131 selection signatures). Several studies in local breeds have shown that the signals detected contain
132 gene variants/genes that may be associated with variations in phenotype, such as coat colour,
133 growth, reproduction, or fatness (8,9,16–19). The study by Muñoz et al. (9) determined the
134 signatures of selection using SNP-array data in 20 European local breeds. This study revealed
135 putative selection signals for regions containing genes involved in fatness, growth, reproduction,
136 development, behaviour and sensory perception. To increase precision on the localization of
137 selection signals, whole-genome sequencing was performed on DNA-pool samples from the same
138 breeds/animals, and new analyses detected several regions associated with coat colour, body size,
139 growth, reproduction and fat deposition (8). These two studies exploited methods based on the
140 differences in allele frequencies between breeds to detect selection signals. Other approaches have
141 been proposed, that jointly exploit information on allele frequencies and population-level
142 phenotypes (20,21). They can offer better power to detect adaptations specific to some traits or
143 environmental covariates. In the present study we adapt such an approach to combine the pool

144 sequencing data from 19 European local and 3 modern pig breeds from (8) and the database of
145 many phenotypic traits in 20 local pig breeds associated with stature, fatness, growth and
146 reproductive performance from (3), in order to identify further selection signatures linked to
147 specific breed level phenotypes and provide hypotheses on the physiological processes involved in
148 the genetic divergence between local breeds.

149 Methods

150 The aim of this study was to locate genomic regions associated with signatures of selection of local
151 pig breeds for production traits. This was performed by combining genetic and phenotypic data at
152 the breed (population) level. Most of the data used for this study have been previously published
153 (3,8,9). In this section, genetic and phenotypic datasets are presented, and then the methodological
154 approach to detect signatures of selection on phenotypic traits is described.

155 Genetic Datasets

156 The present study was based on genetic data collected from 19 populations of European local pig
157 breeds and 7 populations of industrial breeds. The data consisted of SNP genotyping performed
158 with a medium density array (9) for 20 local breeds and whole genome sequencing of the pooled
159 samples (8) for 19 of them (the Iberian breed was not included). To better describe the genetic
160 structure of the local pig breeds, samples from additional 7 populations of 4 modern pig breeds
161 were added. The final sample used for this study is shown in Table 1. Summary of whole-genome
162 sequencing statistics was previously described in (8) and is presented also in Additional file 1,
163 Table S1.

164

165 **Table 1. Name of pig breeds, country of origin of the breeds.**

Pig breed name	Country	Number of Individuals genotyped on SNP array	Sample size in sequenced 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	-

166

167 ## Quality control of genetic datasets

168 For the SNP array data, quality control of SNP genotypes was performed for the entire dataset using
169 standard filters: only autosomal SNPs with < 10% missing data were retained. Following this step,
170 10 individuals with > 3% missing genotypes were clear outliers in the sample and were therefore
171 discarded.

172 De novo SNV discovery was carried out on pool-sequencing data using CRISP (22) with default
173 parameters and yielded 34,751,691 variants. From these, we filtered out variants with more than
174 two alleles, variants with very low allele frequency that most likely resulted from sequencing errors
175 (field VP = 0 and AF=1), and variants with low mapping quality (QUAL < 1000, MQ < 20). After
176 filtering, 16,403,270 SNPs were kept. SNP allele frequencies were estimated for all discovered
177 variants.

178 **Genetic structure of pig populations**

179 The genetic structure of pig populations was assessed from individual SNP genotyping data using
180 principal component analysis (PCA), admixture analysis (23) and population tree reconstruction
181 with hapFLK (24) to reveal the genetic structure and covariance between populations. Admixture
182 analysis was performed for a number of clusters (K) between 2 to 40 to determine a value of K that
183 best explained the data. The decrease in cross-validation error was monotonous from 2 to 40
184 (Additional File 2: Figure S1), but showed a diminishing decrease after K=24, therefore this value
185 was used as a reference to describe the genetic structure of the breeds. Based on the admixture
186 results, individuals showing more than 80% of their genome assigned to the main cluster of their
187 assigned breed were selected for inclusion in the population tree analysis. The Nero Siciliano breed
188 did not show any individual matching to this criterion and was, therefore, not included in the
189 population tree reconstruction.

190 To verify the quality of the pool sequence data, a comparison with the SNP array genotyping results
191 was conducted. Allele frequencies of SNP array variants in the pooled sequence data were
192 estimated using allele counts extracted with Samtools mpileup (25) and PoPoolation2 (26).
193 Samtools was run with options -C 50 -q 20 (variants with mapping quality less than 20 were
194 discarded as recommended by the software documentation). PoPoolation2 was run with default

195 parameters on the resulting mpileup file. From the resulting pool allele counts, allele frequencies
196 and Fst for each pair of populations were then calculated using the approach of (27) implemented in
197 the R package poolfstat. The population tree was constructed applying the neighbour-joining
198 algorithm on the Fst matrix using the same procedure that was used on allele frequencies derived
199 from the SNP array genotypes.

200 Phenotypic characterization of local pig populations

201 A database of phenotypic traits of European local pig populations (3) was used to determine global
202 breed differences in phenotype. It gathers the results of different studies where the main
203 experimental unit for the phenotypic trait was a trial, experiment or part of the experiment in which
204 rearing conditions were sometimes very different from usual production conditions. However, all
205 traits/variables considered here are standard measurements that can be directly compared across
206 breeds and studies. Phenotypic variables were combined into four distinct groups summarizing
207 stature, fatness, growth, and reproductive performance. The growth performance group included
208 traits of average daily gain in 3 different growth periods, where the animals are fed *ad libitum* (i.e.,
209 from lactation to early fattening phase; up to approximately 60 kg). The stature group included traits
210 of body weight and height of adult male and female animals. The fatness group comprised backfat
211 thicknesses at different anatomical locations, fatty acid composition, carcass lean meat content and
212 intramuscular fat content. The reproductive performance group included traits related to number of
213 piglets per litter, number of litters per year, piglets/litter weights, and duration of lactation and
214 farrowing interval. A more detailed description of the traits included in the phenotypic
215 characterization can be found in Additional File 1: Table S2.

216 Means for each variable and each breed were calculated and scaled. Since the phenotypic database
217 was composed of results from different studies, some variables were missing for some breeds

(Additional File 2: Figure S4). Missing data were imputed using regularized iterative PCA method using R package missMDA (28). All experimental units extracted from all studies were given equal weight regardless of the number of pigs involved. Principal component analyses were performed using the R package FactoMiner (29) for the growth performance, reproductive performance, stature and fatness traits. Uncertainty in the predictions of missing data was assessed by multiple imputation (MIPCA function (30)). Breed loadings on the PC1 of each PCA were used as "breed scores" for the analysis of selection signatures. All statistical analyses were performed using R statistical software.

Genome scan for selection on phenotypic breed scores

To identify genomic regions associated to selection on breed level phenotypes, we built upon the approach of (20) which we briefly describe here. Assuming data on allele frequencies measured in r populations at L loci, this association model at a locus l is:

$$p^l \sim N\left(\mathbf{1}p_0^l + \boldsymbol{x}\beta^l, \boldsymbol{V}p_0^l\left(1 - p_0^l\right)\right) \qquad (1)$$

where $\boldsymbol{p}^l$ is the vector of allele frequencies in the r populations, p_0^l is the (unknown) ancestral allele frequency of the locus, $\boldsymbol{x}$ is the vector of breed level phenotypes, β^l is the effect of phenotype on allele frequencies differences, and $\boldsymbol{V}$ is the genome-wide variance-covariance matrix of allele frequencies between populations. The idea behind this model is that the adaptation of populations to covariate x (here the breed level phenotypes) will drive allele frequency differences between populations away from their expectation under genetic drift. In (1), this expectation is modelled with the genome-wide covariance matrix **V** (see (20) for the detailed description). To perform statistical inference under this model the parameters **V**, β^l and p_0^l need to be estimated. Under the null hypothesis (selection associated with phenotype x has not affected allele frequencies), the parameter β^l is set to 0. To test for association with the covariate **x**, (20) use a MCMC algorithm

allowing to derive a statistic for association (a Bayes Factor). This was later extended by (21) to account for uncertainty in allele frequency estimation in pool sequencing experiments. This approach was tested on our dataset, but was found computationally inefficient due to the very high number of SNPs 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 shrunk 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 an

263 asymptotic p-value for each SNP for each phenotypic trait. Since model (1) is only valid for
264 intermediate ancestral allele frequencies, only SNPs where the ancestral allele frequency estimate
265 p_0^l lied between 0.05 and 0.95 were considered in the following. To reduce the possibility of false
266 positives, a permutation procedure was conducted to correct the asymptotic p-values. For each
267 phenotype tested, a new analysis was performed on a permutation of the phenotype, *i.e.* under the
268 null hypothesis of no association between the phenotype and the allele frequencies. This produced
269 an empirical distribution of the asymptotic p-values under the null that was used to obtain corrected
270 p-values at each SNP. Statistical significance was then established on the corrected p-values by
271 estimating the False Discovery Rates (FDR) with the approach of (31) using the qvalue R package.
272 The FDR threshold for detecting significant associations was set at 1%.

273 Based on association statistics for individual SNP, we defined association regions and extracted
274 candidate genes from the annotation of the reference genome (*Sscrofa11.1*). Regions containing at
275 least 4 significant SNPs less than 200Kb apart were further annotated. The detected genes within
276 these regions were reviewed in the literature for potential biological effects on phenotypic
277 differentiation.

278 Results

279 Genetic structure of European pig breeds

280 To characterize the genetic structure of populations, the individual genotypes on the SNP array
281 were used to perform a standard genetic structure analysis. PCA analyses of 24 breeds revealed
282 main genetic backgrounds in the dataset (i.e. Iberian, Duroc/Mora Romagnola, Large White and
283 Landrace/Pietrain backgrounds), visible in the first 3 principal components (Figure 1). PC1 and PC2
284 clearly distinguish between the Iberian pig, White pigs (Landrace, Large White and Pietrain) and
285 Duroc/Mora Romagnola backgrounds. PC3 shows that the White pigs' background separates into a

Large White specific background and a Landrace/Pietrain background. PC4 further separates the Turopolje breed from the Iberian group. The global pattern of differentiation between breeds in our dataset is thus strongly influenced by modern pig breeds. Local pig breeds are usually in intermediate positions, mostly within a triangle with summits corresponding to the Iberian breed, the Landrace/Pietrain and the Large White breeds (see Additional files 3). 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.

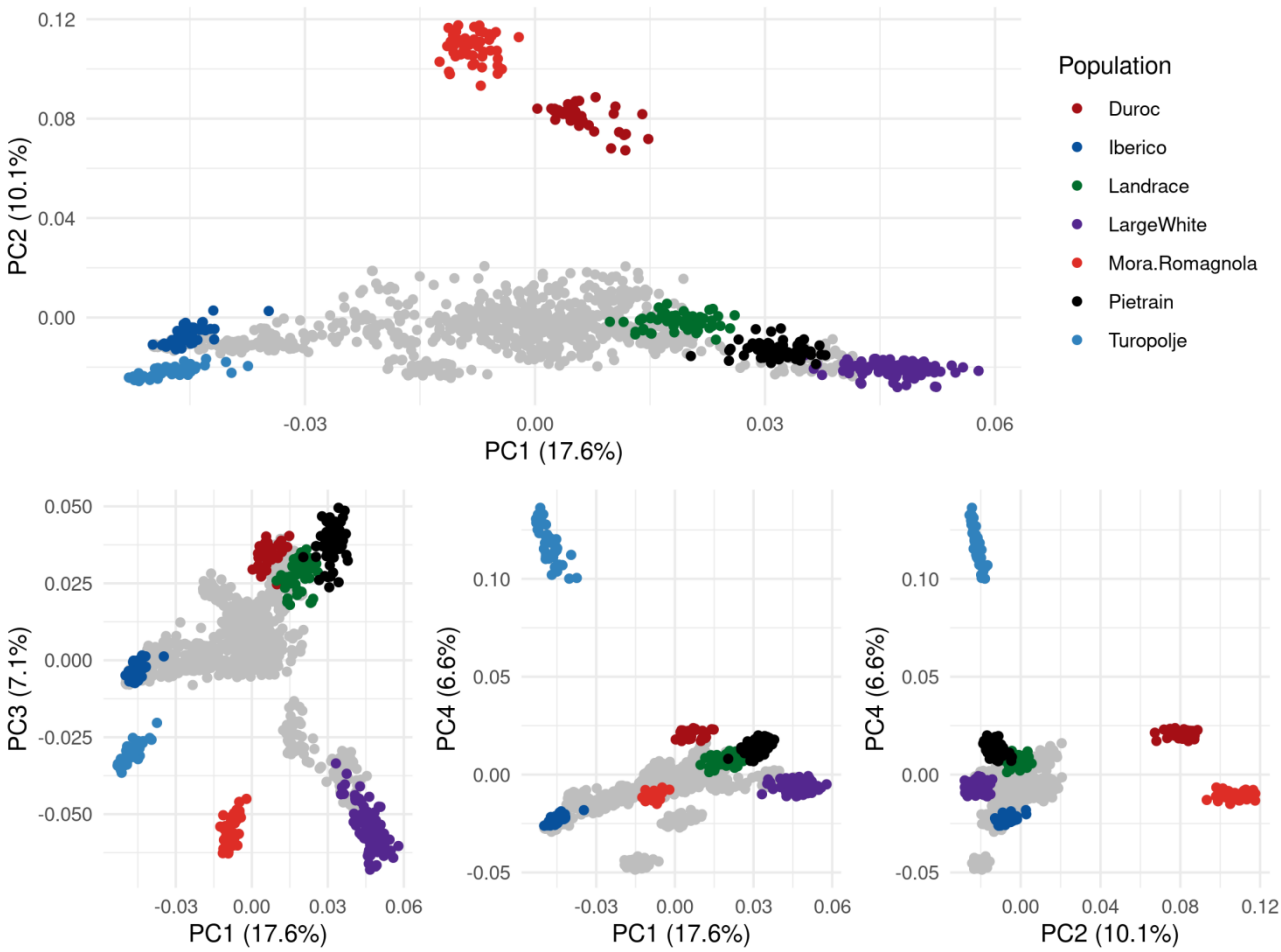


Figure 1. Principal component analysis of 24 pig breeds genotyped on a medium density SNP array.

297 The results of the admixture analysis results show that the number of homogeneous clusters in the
298 dataset is difficult to determine. The cross-validation procedure was conducted for up to 40 clusters
299 and resulted in a general decrease in cross-validation error with increasing number of clusters
300 (Additional File 2: Figure S1). However, the decrease slowed down after K=24 clusters, a number
301 that corresponds to the number of named breeds (Table 1). The further decrease is due to the fact
302 that some breeds are sub-structured and require additional clusters to be well fitted. This is typical
303 of local breeds (e.g Gama et al., 2013; (32)). In the following, we present the results obtained with
304 K = 24 corresponding to the inflexion point in the cross-validation curve. Figure 2 shows side-by-
305 side the result of the population tree reconstruction and the admixture analysis. At K=24, most
306 breed exhibit a homogeneous pattern of admixture and belong to a specific cluster. Exceptions are
307 the Alentejano and Iberian breeds (which belong to the same cluster, in consistence with their
308 common origin), the Casertana and Apulo Calabrese breeds (each of them further splits into two
309 groups) and Sarda, Moravka, Bisaro and Nero Siciliano breeds that show high heterogeneity. In the
310 latter breeds, admixed individuals do not appear to be recent hybrids with other breeds. In addition,
311 admixture plots with K = 20, K = 15, K = 10 and K = 6 are shown using circular plot in Additional
312 file S2, Figure S2.

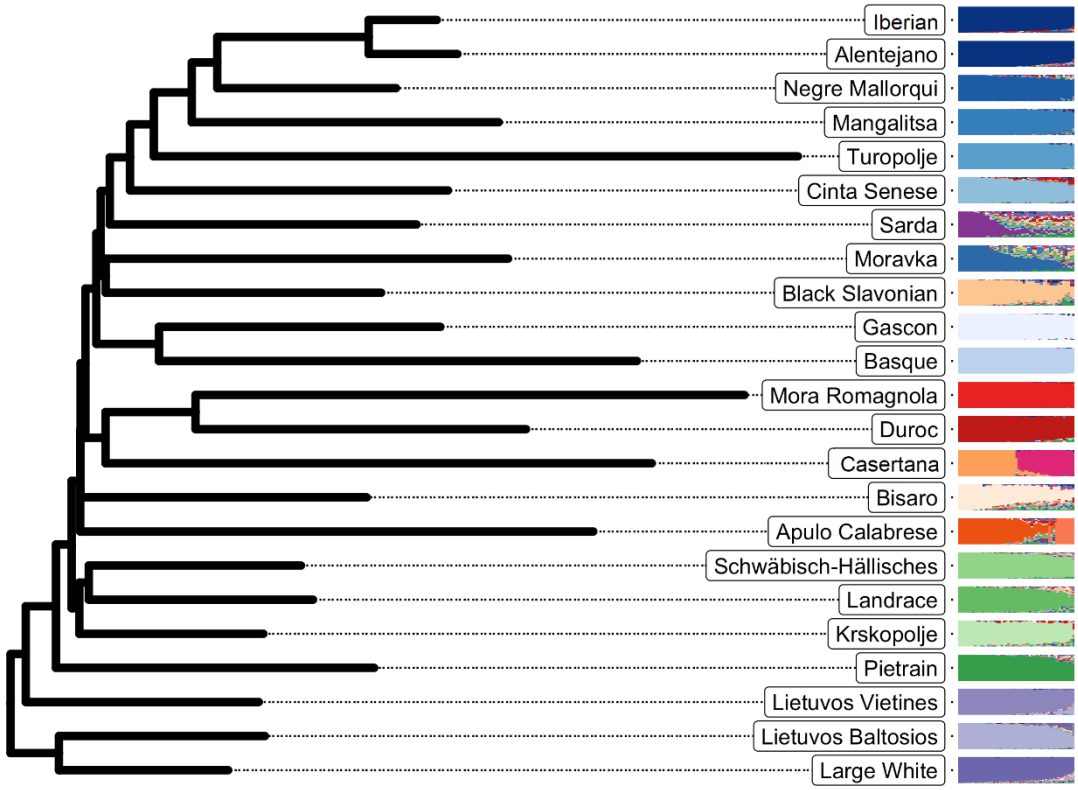


Figure 2. Population tree and admixture analysis of 23 pig breeds genotyped on a medium density SNP array. The population tree reconstructed from pairwise genetic distances (Fst) is shown on the left, and the admixture component for all individuals a priori belonging to the breed is shown on the right. 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.

The population tree is structured consistently with the main axes of variation identified in the PCA analysis with main genetic backgrounds (Figure 2). Turopolje and Mora Romagnola breeds differentiated early in the PCA analysis, which could be explained with the population analysis and the fact that they exhibit long branches corresponding to low heterozygosity. This analysis also reveals a general clustering of local breeds according to their geographical origin. Breeds closer to the Iberian group mostly originate from the Iberian peninsula or close geographical areas (South West of France with Basque and Gascon breeds, Balearic Islands with Negre Mallorqui breed). Interestingly, some other breeds from different geographical area appear to be related to this

330 background such as Mangalitsa and Moravka from Serbia, the Black Slavonian from Croatia or the
331 Cinta Senese from Italy. Breeds from Central Europe, such as Schwabisch-Hällisches and
332 Krškopolje pig, showed genetic proximity to the Landrace/Pietrain background and breeds from
333 Lithuania in Northern Europe with the Large White component. Finally, some breeds such as the
334 Bísaro from Portugal or Apulo Calabrese from Italy cannot be considered related to any other breed
335 in the dataset.

336 Sequences from the investigated breeds were aligned to the Sscrofa11.1 reference genome and
337 SNPs were discovered with the pool-seq variant caller CRISP. For SNPs present on the genotyping
338 array, allele frequencies were consistent with those derived from individual genotyping (9)
339 (Additional file 2, Figure S3). This confirmed the quality of SNPs obtained by pool-sequencing.

340 Phenotypic differentiation of European local pig breeds

341 A database of available published results on phenotypic traits of 20 local pig breeds (3) was used in
342 order to distribute local pig populations according to phenotype including stature, growth, fatness
343 traits, and reproductive performance traits. The representation of relationships between breeds and
344 variables was evaluated using PCA analyses. The first two principal components of the PCA for
345 stature, growth, fatness, and reproduction group accounted for 98.2%, 83.2%, 84.3% and 70.7% of
346 the total variance, respectively (Figure 3, Additional File 1: Figures S5, S6, S7). Scores for each
347 breed from PC of all phenotypic groups were extracted from PCA.

348 The first principal component of the stature group (Figure 3) represented the majority of the total
349 variability (77.3%) and clearly distributed breeds according to average height and weight. Thus,
350 breeds with lower breed scores for stature (e.g. Moravka, Turopolje, Nero Siciliano, Negre
351 Mallorqui) were smaller and lighter, while breeds with higher breed scores were taller and heavier

352 breeds (e.g. Schwäbisch-Hällisches, Lietuvos Baltosios Senojo Tipo, Krškopolje pig, Lietuvos
353 Vietiné).

354 The growth group distributed local breeds according to their growth capacity, including data on
355 average daily gain for the rearing period from the lactation period to the early fattening phase (i.e.
356 up to 60 kg body weight) (Additional File 2: Figure S5). The first PC explained 56.1% of the total
357 variability and distributed the breeds with the highest (i.e. Schwäbisch-Hällisches, Lietuvos
358 Baltosios Senojo Tipo, Lietuvos Vietiné, Bisaro, Krškopolje pig) and the lowest (i.e. Alentejano,
359 Moravka, Black Slavonian, Mangalitsa and Turopolje) growth potential.

360 Fatness traits comprised variables associated with the fatty phenotype (Additional File 2: Figure
361 S6). Principal component 1 (59.3% of total variability) was positively correlated with backfat
362 thicknesses at different anatomical locations and intramuscular fat content. Conversely, lean meat
363 and PUFA content (i.e., lean phenotype) were negatively correlated with PC1. Because
364 intramuscular fat content is a trait of particular interest, the distribution of breeds by intramuscular
365 fat content is shown in Additional File 2: Figure S8. Local breeds were divided into fatter (e.g.
366 Moravka, Iberian, Mangalitsa, Negre Mallorqui) or leaner phenotypes (e.g. Schwäbisch-Hällisches,
367 Lietuvos Baltosios Senojo Tipo, Bisaro).

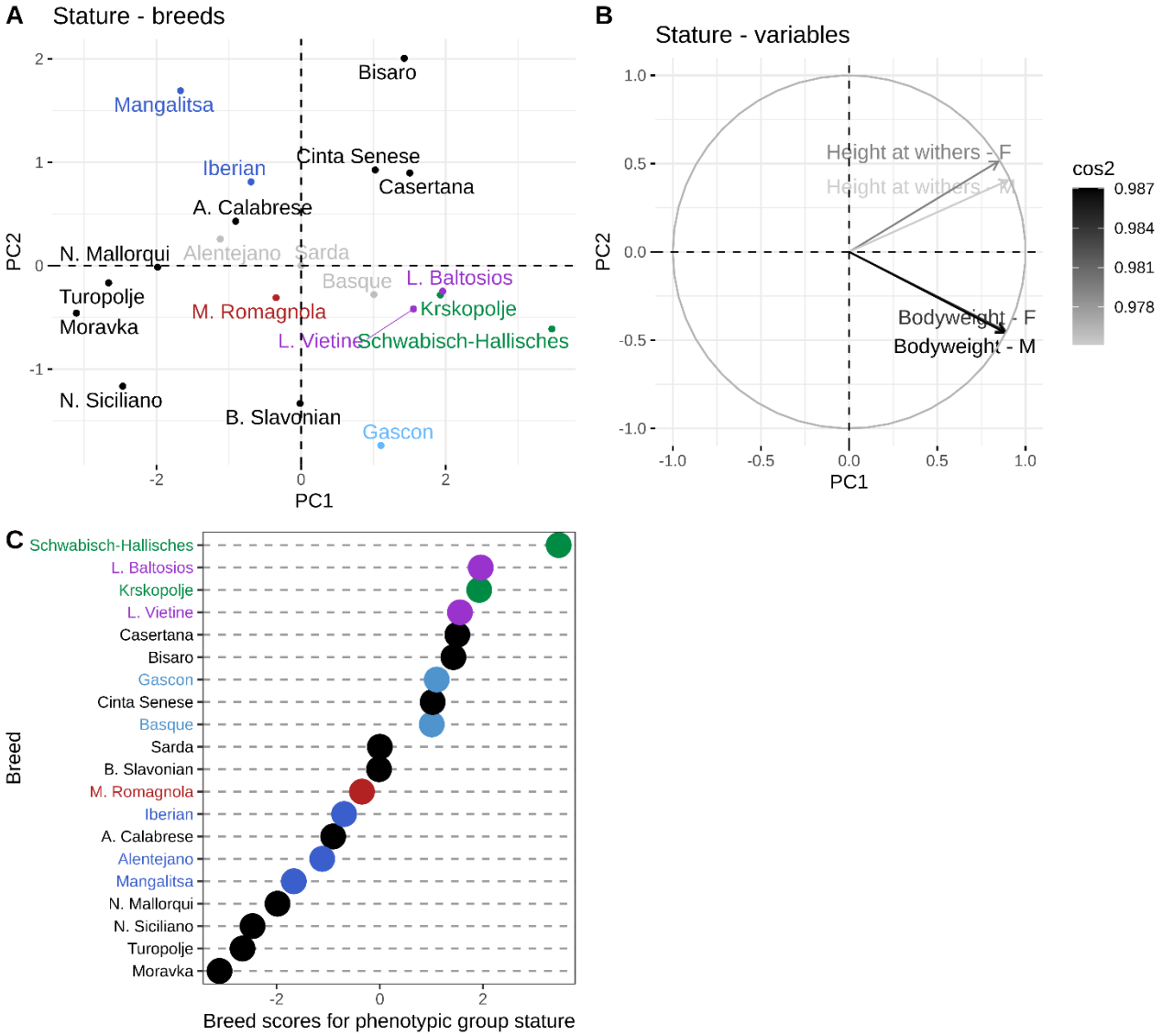


Figure 3. Principal component analysis showing (A) the relationship between breeds on the first tow PCs (B) the contribution of the traits associated with stature to the PCs and (C) the corresponding phenotypic breed scores (PC1 loadings). (A) Breeds colored in grey are the breeds with more than 50% of missing variables, thus, their position on the PCA must be interpreted carefully. (B) The variables are colored 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 colored in black. Breeds (A, C) are colored according to their genetic similarity: 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 2: Figure S7). 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.

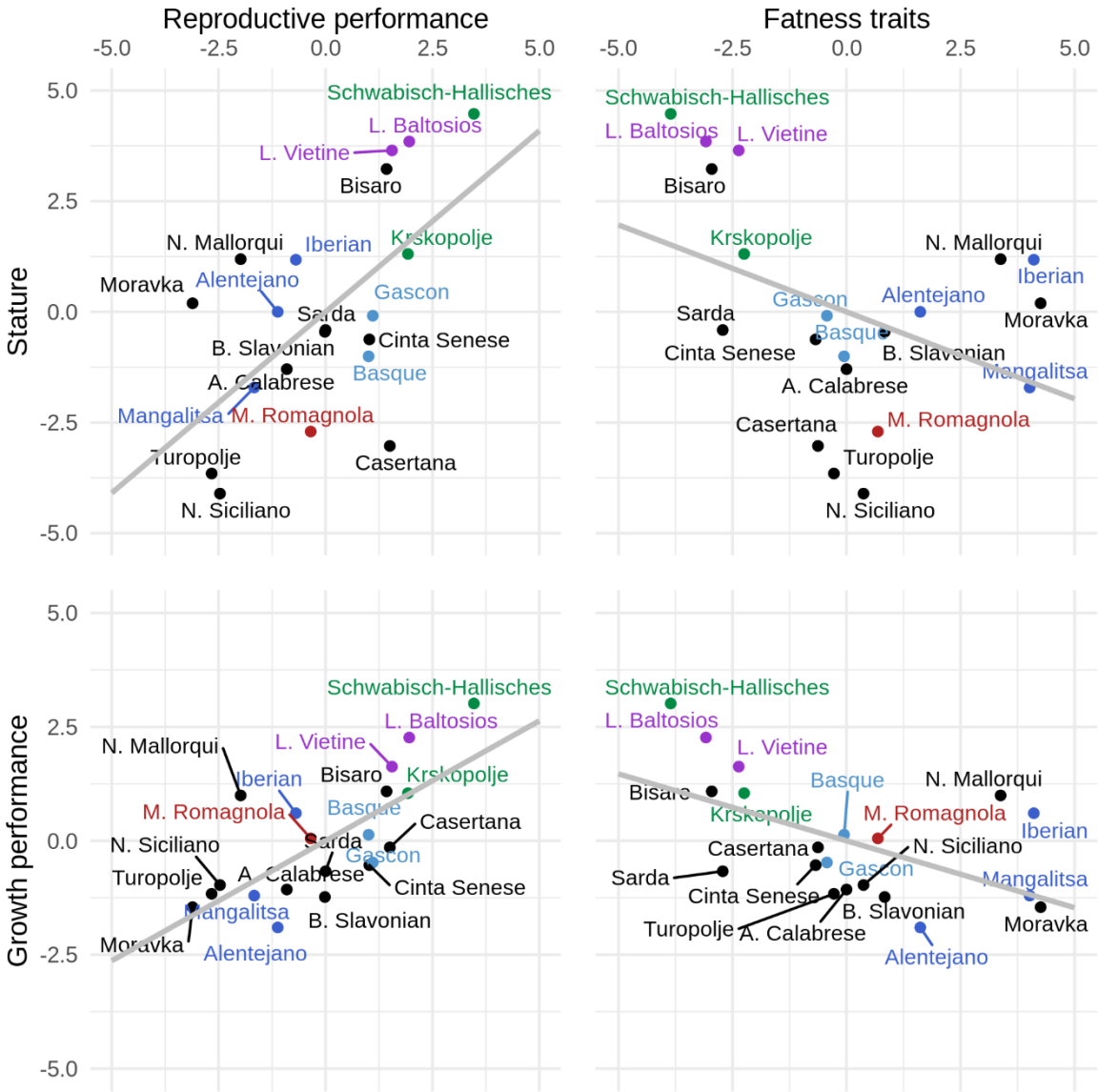


Figure 4. Global differences in production traits for 20 European local pig populations. Breeds are distributed according to breed scores for phenotypic traits and colored according to genetic similarity. Low values for breed score growth are representing lower average daily gain from lactation up to 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

392 reproductive performance, while high values represent higher reproductive performance of breeds. Low values for
393 breed score stature represent lighter and smaller breeds, while high values represent heavier and larger breeds. Low
394 values for the breed score fat represent leaner breeds, while high values represent fatter breeds. Breeds in green are
395 genetically Landrace-like breeds, in purple are Large White-like breeds, in blue are Iberian-like breeds, in red are
396 Duroc-like breeds and in light blue are Gascon and Basque.

397 Detection of genomic regions associated with phenotypic differentiation

398 The approach proposed by Coop et al. (20) was extended to breed-level phenotypes to find genomic
399 regions potentially influencing phenotypic differentiation in local European pig breeds (see
400 Methods). A selection signature scan was performed for each phenotypic breed score, resulting in p-
401 values for each SNP. Figure 5 shows Manhattan plots for stature, fatness, growth and reproduction
402 along with a pp-plot constrasting the expected distribution of (-log10) p-values to the observed one.

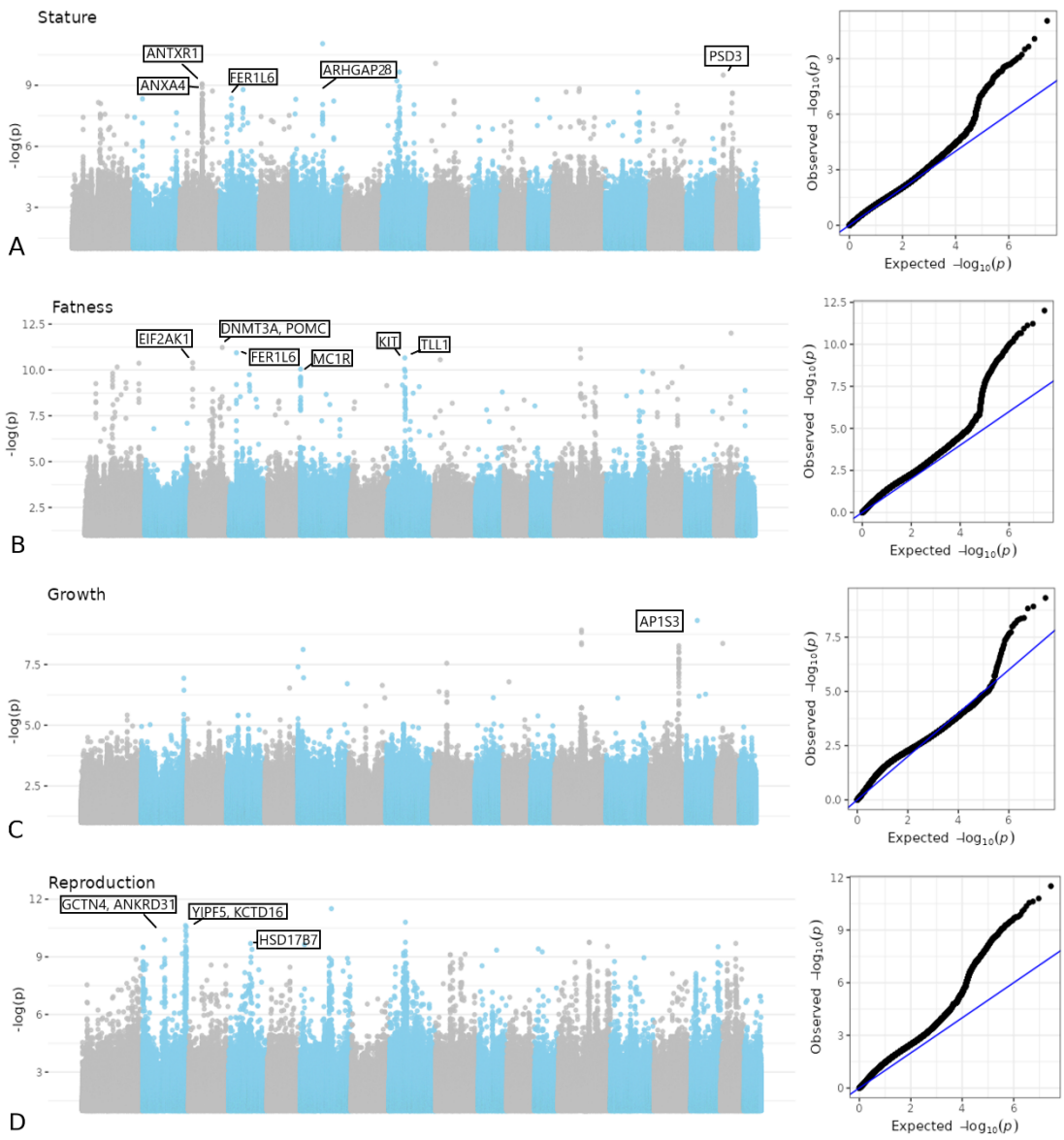
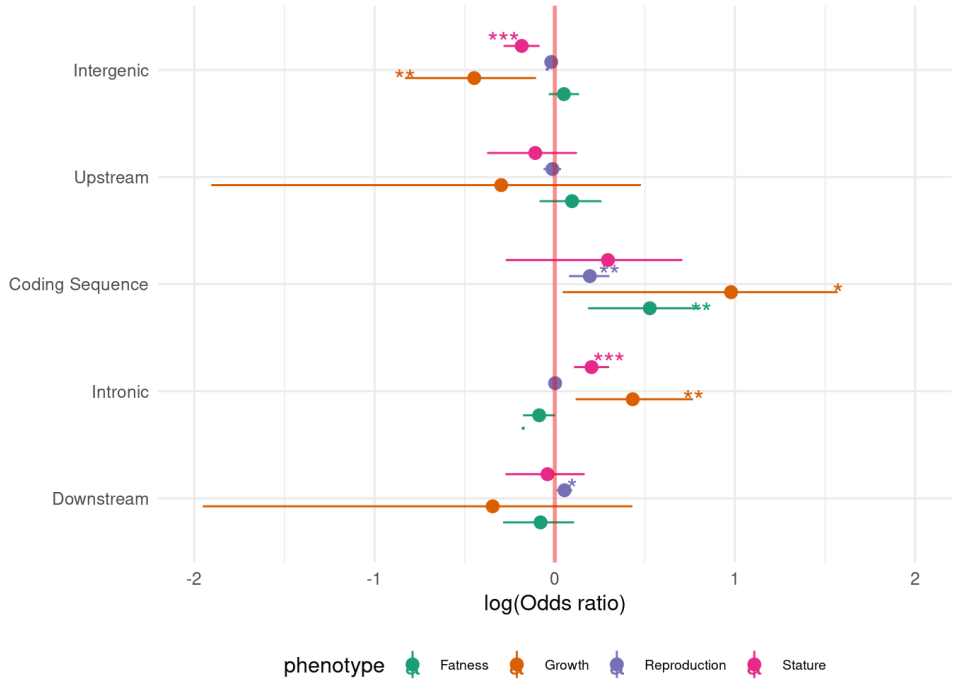


Figure 5. Manhattan plots for phenotypes (A) stature, (B) fatness, (C) growth and (D) reproduction and associated pp-plots.

Across the genome, windows/regions with 4 significant SNPs with significance threshold 0.01 were considered as selection signals for phenotypes. Overall, a total of 234 regions/windows ranging in length from 3.1 kbp to 1.282 Mbp were discovered. Among the detected regions, 16 regions were in the stature group, 2 regions were in the growth group, 24 regions were in the fatness group, and 192 regions were in the reproduction group. For a list of selection signatures for phenotypes, see Additional File 4.

412 Regions in the stature phenotypic group spanned between 3.1 kbp and 362.2 kbp with a maximum
413 of 35 SNPs and included from none to four genes. The region with the strongest signal included the
414 *ARHGAP28* gene. The growth phenotypic group contained 2 regions with a maximum of 18
415 significant SNPs and included 1 gene (i.e. *AP1S3*). The fatness phenotypic group spanned from 3.1
416 kbp to 590.5 kbp in length with maximum of 30 SNPs and included between up to 12 genes. The
417 region with the strongest signal included the *DNMT3A, POMC* and *EFR3B* genes. The reproduction
418 phenotypic group contained regions spanning from 12.8 kbp to 1,282.0 kbp in length with a
419 maximum of 170 SNPs. The region with the strongest signals was located on chromosome SSC6
420 and contained the *YES1*, *ENOSF1*, *TYMS*, and *CLUL1* genes. The detected selection signals for
421 production traits were further examined to find overlaps between groups (Additional file 2: Figure
422 S9).



423

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

425 For each phenotypic group, we estimated the enrichment/depletion of significant SNPs in different
426 functional categories: Intergenic, upstream of genes, genic (coding regions or introns) and
427 downstream of genes (Figure 6). We observed a general trend for an enrichment of significant SNPs
428 in genic regions (exons and introns) and some depletion in intergenic regions.

429 Discussion

430 Deciphering the genetic basis of variation in complex traits can help to understand their biology and
431 evolution, and improve breeding, selection programs and conservation plans of animal genetic
432 resources. One approach to address this question is to link genetic and phenotypic variation by
433 identifying genomic regions that have evolved in response to selection on complex traits.
434 Establishing such links requires a collection of data on genetic groups (e.g. populations, lines,
435 breeds) that share a common origin and have information on adaptive traits that can be phenotyped
436 in a standardized fashion. Livestock species generally meet these criteria and therefore provide
437 good models for mapping genomic regions associated with selective constraints. Here, we studied
438 local pig populations that were genetically characterized using individual genotyping and pool
439 sequencing for which a recent common phenotypic database was constituted.

440 Genetic diversity of European local pig breeds

441 Using individual genotyping of a large set of European breeds, we have found that their genetic
442 variation is structured around four main genetic groups: the Duroc, the Large White, the Landrace
443 and the Iberian. The origin of the Duroc is unclear, as it originated from American continent and
444 was then brought back to Europe. This history implies that it was separated from the European
445 breeds for a long time, which is evident in our genetic analysis by its rather large genetic distance
446 with most European breeds, except for Mora Romagnola breed from Italy, which has a history of
447 crossbreeding with Duroc breed (33). In addition, the present study showed that Mora Romagnola

448 (and also Turopolje pig breed) differentiated early in the PCA analysis, which can be explained by
449 their high degrees of inbreeding (34). The White pig group is composed of the Lietuvos breeds and
450 the Large White, all of which originate from Northern Europe and are clearly separated from the
451 Landrace, which includes populations originating from more central countries such as Germany
452 (Landrace, Schwäbisch-Hällisches), Belgium (Pietrain) and Slovenia (Krškopolje). The Large
453 White and Landrace groups are close to the root of the population tree, which could be explained by
454 the well-documented influence of Asian pigs on these genetic pools (35). Finally, the Iberian group
455 includes a set of breeds that are quite geographically dispersed; while many of the Iberian-type
456 breeds are found around southwestern Europe (Spain, Portugal, and France), some populations from
457 Eastern/Central Europe clearly belong to the same genetic pool (e.g. Mangalitsa from Serbia,
458 Turopolje from Croatia, Cinta Senese from Italy). A possible interpretation is that the genetic
459 background responsible for variation in the Iberian pig breed was historically widespread in Europe
460 and can still be found in some local pig breeds. Overall, analysis of the genetic structure of
461 European local pig breeds reveals extensive genetic variation clustered in differentiated breeds with
462 continuously distributed genetic distances. The historical events that led to the present genetic
463 variation are most likely complex, and involve differentiation from an ancestral pool followed by or
464 accompanied by outcrossing between populations, including wild boars, and cannot be
465 reconstructed from the data used here. However, the sample of available breeds is well adapted to
466 our main objective, as we can interrogate a diverse set of polymorphisms for adaptation to
467 contrasting environments or production systems in different genetic backgrounds.

468 While SNP genotyping is sufficient to characterize the genetic relationships and population
469 structure of local pig breeds, the density of the SNP array used limits the resolution and power to
470 detect potential associations with breed phenotypes especially in local breeds where a big
471 proportion of SNPs included in the commercial chips are not segregating. To alleviate this issue, a

472 cost-effective alternative to SNP genotyping consists of pool sequencing of populations. In the
473 dataset used here, we confirmed that allele frequency and genetic structure estimates obtained from
474 pool sequencing are consistent with those obtained from SNP array, while providing comprehensive
475 information on the genetic diversity of local pig breeds.

476 Phenotypic diversity of European local pig breeds

477 In parallel to the comprehensive information on the genetic diversity of local pig breeds, we
478 developed an original approach based on PCA to characterize their phenotypic diversity in traits
479 related to stature, growth, fatness and reproduction, which have been described in a comprehensive
480 analysis (3). While local breeds can be phenotypically characterized on meat quality traits (pH24,
481 pH48, intramuscular fat content and *longissimus dorsi* muscle colour, Additional File S2: Figure
482 S10)., these traits (with the exception of IMF) are strongly influenced by pre-slaughter handling
483 stress (36). Breed score for meat quality was therefore not included in the phenotype-genotype
484 selection scan.

485 Local pig breeds exhibit a wide variety of exterior phenotypic characteristics, including body size or
486 body weight. This could be due to genetic factors or could reflect the management of the breed. For
487 example, male and female animals of the Schwäbisch-Hällisches breed are medium to large in size
488 and also have well-developed management systems. In contrast, more untapped breeds (e.g.
489 Moravka) are smaller and lighter (3).

490 Regarding the growth performance of local pig breeds, existing knowledge is limited. Moreover, the
491 production systems may not be sufficiently adapted to the needs of the breeds which could be
492 reflected in the phenotype of the breeds. For example, the study of Brossard et al. (37) argues that in
493 many studies on local pig breeds, animals are fed below ad libitum level and, therefore, do not reach
494 their full production potential. This is consistent with an analytical literature review (38) on

495 European local pig breeds, which demonstrated that local pig breeds are generally fed in ad libitum
496 conditions only earlier in life (during the growing and early fattening phase). In the late fattening
497 phase, the feeding is likely to be restricted. To limit the influence of such effects in the present
498 study and to use a better representation of breeds' growth potential, only average daily gain values
499 from the beginning of the lactation to the end of the early fattening phase (i.e. up to 60 kg) were
500 used for the phenotypic score on growth. Despite this, better growth performance might still partly
501 reflect breeding management, as shown for the Schwäbisch-Hällisches pig, which has a growth
502 potential comparable to genetically improved modern pig breeds (39,40).

503 It has been reported that local pig breeds have a lower protein and higher lipid deposition capacity
504 compared to modern pig breeds (37). The higher genetic capacity of local pig breeds for backfat and
505 intramuscular fat deposition, and lower muscle accretion has been demonstrated in several studies
506 comparing fat characteristics between local and modern breeds (41–45). Although the local pig
507 breeds are recognised as fatty, they still exhibit a large phenotypic variability in the dataset used
508 here.

509 Regarding reproduction traits, sows of local breeds typically exhibit relatively high age at first
510 parturition, few litters per sow per year, long lactation periods, small litter sizes and high piglet
511 mortality (38). In our study, PCA of reproductive performance distinguished breeds with better
512 reproductive efficiency, which are usually reared in more intensive systems within more developed
513 pork chains (e.g. Schwäbisch-Hällisches pig) from breeds with lower reproductive efficiency (e.g.
514 Nero Siciliano, Turopolje, Casertana, Mora Romagnola, Mangalitsa) reared in extensive or semi-
515 extensive systems (3).

516 While some of this variation in phenotypes can certainly be explained by differences in production
517 systems among local breeds, our genetic association results suggest that some of it may be linked to

genetic variation. In order to be able to test for these genetic associations, the methods used require all breeds to have an associated phenotype. To do so, we imputed using PCA the missing phenotypes of some breeds (see Methods). The resulting imputed phenotypes are likely to be biased toward the average of all breeds. This is for example most likely the case for the Sarda breed, considered to be a very small breed (in term of body size) but lying at an intermediate value after imputation. This certainly limits our power to detect genetic associations but is unlikely to create false positives.

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

The genome scans for selection associated to broad phenotypic groups revealed 234 regions associated with stature, fatness, growth or reproduction traits. However, the distribution of the number of regions across traits is highly imbalanced, the reproduction phenotypic group containing 192 regions, while the growth traits only two. While these numbers could change slightly depending on the definition of a "significant region", it is clear that the reproduction group exhibits more significant regions than the others. We could explain this result by different, non-exclusive reasons. First, it seems that our permutation procedure to correct p-values is less efficient for this breed score than for the others. This can be seen in the slightly biased distribution of p-values on the pp-plot of Figure 5 (Panel D-right). But another possibility is that the reproductive traits are more influenced by the production systems than the other traits. Therefore, testing for an association with the production system actually captures complex adaptations to other, non-measured traits than those linked to reproduction. The consequence is that by using this reproductive score, we are actually capturing adaption to many other traits linked to adaptation to the rearing conditions. Therefore, a bigger portion of the genome can be contributing to the response of local breeds to selection which could explain why p-values are generally smaller for this trait group.

542 Relatively little overlap was found between the significant regions discovered with the trait-
543 association approach here and the trait agnostic one applied to the same pool sequencing data (8).
544 Only two regions were found clearly overlapping, one on chromosome 8, encompassing the *MAP9*
545 gene but also close to the *KIT* gene, associated notably to coat color, and the other on chromosome
546 15, containing genes *TMEM237* and *MPP4*. This illustrates the complementarity of the two
547 approaches, and highlights empirically the gain in power that can be obtained by testing specific
548 hypotheses (*e.g.* the association of allele frequencies to phenotypic variation) as was shown in
549 simulations studies (20,21).

550 A strong association was found for stature and fatness in a region on chromosome 3 containing the
551 *ANXA4* gene. Annexin A4 encodes a calcium-dependent phospholipid-binding protein involved in
552 various membrane processes. This candidate gene or region has been previously proposed as a QTL
553 for stature in cattle (46,47). Another genomic region associated with the signature of selection for
554 stature and growth was discovered on chromosome 3 containing also the candidate gene *ANTXR1*.
555 This gene is involved in cell morphogenesis, cellular development process and cytoskeleton
556 organisation. Part of this region has also been previously proposed as a QTL for bovine stature (48).
557 In the fatness and also stature group, two overlapping regions were found on chromosome 3. These
558 regions contained common candidate genes; namely *DNMT3A* (responsible for CpG methylation),
559 *POMC* (prohormone) and *EFR3B* (localises phosphatidylinositol 4-kinase to the plasma
560 membrane). *DNMT3A* has been previously shown to affect stature and body weight in cattle and
561 humans (48–50). In addition, the *DNMT3A* gene has previously been implicated to the regulation of
562 adipose tissue development (51). Another candidate gene found in the same region is the *POMC*
563 gene, which encodes the precursor of several peptide hormones that contribute to the regulation of
564 feed intake and energy balance via the leptin/melanocortin pathway (52). Polymorphisms in *POMC*
565 have previously been associated with *longissimus dorsi* muscle area and backfat thickness in cattle

566 (53–55) and with obesity and body mass index in humans (56,57). An interesting signature of
567 selection on reproductive performance was found on chromosome 4, which contains the *HSD17B7*
568 gene. This gene encodes an enzyme involved in the biosynthesis of sex steroids and cholesterol
569 (58,59). Due to the clear role of the *HSD17B7* gene in steroidogenesis, it is a good candidate gene
570 for reproductive performance in local pig breeds.

571 Several other regions have also been identified as selection signatures for stature and fatness with
572 not so direct connection to their biological function in production traits, but still with some
573 association to phenotype. These regions could be useful for further/future studies on selection
574 signatures. For example, chromosome 6 contained a region with possible adaptation on stature with
575 the gene *ARHGAP28*, which has previously been associated with the number of vertebrae in pigs,
576 thus affecting carcass length (60). In the fatness group, the gene *EIF2AK1* (located on chromosome
577 3) was identified, with a role of inhibiting of the protein synthesis in response to stress. This gene
578 has previously been associated to body mass index in pigs (61).

579 Some of the regions that have been discovered must be carefully interpreted. For instance, the
580 conducted studies differ in production conditions. Therefore, correlations between phenotypic
581 groups could create non-causal signal with genes. An example of a non-causal signal due to
582 phenotypic correlation found in the growth and fatness groups is a region on chromosome 8
583 containing the gene *KIT*. This gene encodes the tyrosine kinase receptor and is associated with coat
584 colour in pigs. Another selection signature associated with coat colour was found on chromosome
585 chromosome 6 in the growth and fatness groups. It contains the *MC1R* gene, which plays a major
586 role in controlling the transition from eumelanin (black or brown) to pheomelanin (yellow to red)
587 (62). Since the coat colour was favoured by breeders, it was strongly selected in several local breeds
588 (63) (e.g. White breeds were under selection for leanness and better growth performance).

589 Interestingly, the study performed on SNP-chip data on the same animals/breeds (9) did not detect
590 any signal near *MC1R* or *KIT* genes probably due to different SNP-chip informativity and statistical
591 approach used in their study.

592 Conclusions

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

599 Declarations

600 **Ethics approval and consent to participate**

601 Not applicable.

602 **Consent for publication**

603 Not applicable.

604 **Availability of data and materials**

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

607 Competing interests

608 The authors declare that they have no competing interests.

609 **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. The authors also wish to thank Nina Batorek-Lukač and Urška Tomažin for their contribution in phenotypic data collection.

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.

645 8. Bovo S, Ribani A, Muñoz M, Alves E, Araujo JP, Bozzi R, et al. Whole-genome sequencing of European
646 autochthonous and commercial pig breeds allows the detection of signatures of selection for
647 adaptation of genetic resources to different breeding and production systems. Gen Sel Evol.
648 2020;52(1):33.

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

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

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

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

659 13. Jiang Y, Tang S, Wang C, Wang Y, Qin Y, Wang Y, et al. A genome-wide association study of growth and
660 fatness traits in two pig populations with different genetic backgrounds. J Anim Sci. 2018;96(3):806–
661 16.

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

664 15. Strucken EM, Schmitt AO, Bergfeld U, Jurke I, Reissmann M, Brockmann GA. Genomewide study and
665 validation of markers associated with production traits in German Landrace boars. J Anim Sci.
666 2014;92(5):1939–44.

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

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

671 18. Choi JW, Choi BH, Lee SH, Lee SS, Kim HC, Yu D, et al. Whole-genome resequencing analysis of
672 Hanwoo and Yanbian cattle to identify genome-wide SNPs and signatures of selection. Mol Cells.
673 2015;38(5):466–73.

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

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

679 21. Gautier M. Genome-wide acan for adaptive divergence and association with population-specific
680 covariates. Genetics. 2015;201(4):1555–79.

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

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

685 24. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. Detecting signatures of selection through
686 haplotype differentiation among hierarchically structured populations. Genetics. 2013;193(3):929–41.

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

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

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

693 28. Josse J, Husson F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis.
694 Journal of Statistical Software. 2016;70(1):1–31.

695 29. Lê S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. J Stat Softw. 2008;25(1):1–
696 18.

697 30. Josse J, Pagès J, Husson F. Multiple imputation in principal component analysis. Adv Data Anal Classif.
698 2011;5(3):231–46.

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

700 32. Gama LT, Martínez AM, Carolino I, Landi V, Delgado JV, Vicente AA, et al. Genetic structure,
701 relationships and admixture with wild relatives in native pig breeds from Iberia and its islands.
702 Genetics Selection Evolution. 2013 Jun 14;45(1):18.

703 33. Tinarelli S, Ribani A, Utzeri VJ, Taurisano V, Bovo C, Dall'Olio S, et al. Redefinition of the Mora
704 Romagnola Pig Breed Herd Book Standard Based on DNA Markers Useful to Authenticate Its "Mono-
705 Breed" Products: An Example of Sustainable Conservation of a Livestock Genetic Resource. Animals.
706 2021;11(2):526.

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

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

713 36. Dokmanovic M, Baltic MZ, Duric J, Ivanovic J, Popovic L, Todorovic M, et al. Correlations among stress
714 parameters, meat and carcass quality parameters in pigs. Asian-Australas J Anim Sci. 2015;28(3):435–
715 41.

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

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

721 39. Petig M, Zimmer C, Bühler R, Batorek-Lukač N. Schwäbisch-Hällisches pig. In: European local pig
722 breeds - diversity and performance A study of project TREASURE. London: Intech Open; 2019. p. 258–
723 66.

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

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

729 42. Lebret B, Dourmad JY, Mourot J, Pollet PY, Gondret F. Production performance, carcass composition,
730 and adipose tissue traits of heavy pigs: influence of breed and production system. J Anim Sci.
731 2014;92(8):3543–56.

732 43. Madeira MS, Pires VMR, Alfaia CM, Costa ASH, Luxton R, Doran O, et al. Differential effects of reduced
733 protein diets on fatty acid composition and gene expression in muscle and subcutaneous adipose
734 tissue of Alentejana purebred and Large White × Landrace × Pietrain crossbred pigs. Br J Nutri.
735 2013;110(2):216–29.

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

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

741 46. Kolbehdari D, Wang Z, Grant JR, Murdoch B, Prasad A, Xiu Z, et al. A whole-genome scan to map
742 quantitative trait loci for conformation and functional traits in Canadian Holstein bulls. J Dairy Sci.
743 2008;91(7):2844–56.

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

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

750 49. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common
751 variation in the genomic and biological architecture of adult human height. Nat Genet.
752 2014;46(11):1173–86.

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

756 51. Ma X, Kang S. Functional implications of DNA methylation in adipose biology. Diabetes.
757 2019;68(5):871–8.

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

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

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

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

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

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

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

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

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

777 61. Gong H, Xiao S, Li W, Huang T, Huang X, Yan G, et al. Unravelling the genetic loci for growth and
778 carcass traits in Chinese Bamaxiang pigs based on a 1.4 million SNP array. J Anim Breed Genet.
779 2019;136(1):3–14.

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

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

784

785

Figures

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

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.

821 **Figure 5.** Manhattan plots for phenotypes (A) stature, (B) fatness, (C) growth and (D) reproduction
822 and associated pp-plots.

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

824

Additional files

Additional file 1 Table S1

Title: Summary of sequencing statistics (adapted from Bovo et al., 2020).

Additional file 1 Table S2

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

Additional file 2 Figure S1

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

Additional file 2 Figure S2

Title: Admixture plots of European local pig breeds for K=24, K=20, K=15, K=10, and K=6 (from the outside to the inside of the circle).

Additional file 2 Figure S3

Title: Projection of the pool-sequencing data (allele frequencies) on the PCA established with individual genotypes on the SNP genotyping array. The x-axis is the average coordinate of individuals of each population on each of the PC of the genotyped-based PCA, and the y-axis is the coordinate of the allele frequencies of each population projected on each PCA.

Additional file 2 Figure S4

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 2 Figure S5

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 2 Figure S6

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 2 Figure S7

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 2 Figure S8

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 2 Figure S9

Title: Circular plot summarizing regions with candidate genes for phenotypic groups growth, stature, fatness and reproduction.

Description: 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.

Additional file 2 Figure S10

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 3

Title: The genetic structure of European local pig populations assessed from individual SNP genotyping data using principal component analysis.

Additional file 4

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

Format: xls