ANIMAL GENETICS
Immunogenetics, Molecular Genetics and Functional Genomics

A genome-wide analyses reveals the regions involved in the phenotypic diversity between two local Sicilian pig populations

Journal:	*Animal Genetics*
Manuscript ID	AnGen-19-09-0287
Manuscript Type:	Short Communication
Date Submitted by the Author:	16-Sep-2019
Complete List of Authors:	D'Alessandro, Enrico; University of Messina, Dipartimento Scienze Veterinarie Sottile, Gianluca; University of Palermo, Scienze Economiche, Aziendali e Statistiche Sardina, Maria Teresa; University of Palermo, Scienze Agrarie, Alimentari e Forestali Criscione, Andrea; Università degli studi di Catania, DACPA - Dipartimento di Scienze Agronomiche Agrochimiche e delle Produzioni Animali - Sez. di Scienze delle Produzioni Animali Bordonaro, Salvatore; DACPA, Sez. di Scienze delle Produzioni Animali Sutera, Anna Maria; University of Messina, Dipartimento Scienze Veterinarie Zumbo, Alessandro; University of Messina, Dipartimento Scienze Veterinarie Portolano, Baldassare; University of Palermo, Scienze Agrarie e Forestali Mastrangelo, Salvatore; University of Palermo, Scienze Agrarie, Alimentari e Forestali
Keywords:	Local pig populations, genome-wide analyses, phenotype diversity, SNPs, candidate genes

A genome-wide analyses reveals the regions involved in the phenotypic diversity between two local Sicilian pig populations

E. D'Alessandro[1], G. Sottile[2], M.T. Sardina[3], A. Criscione[4], S. Bordonaro[4], A.M. Sutera[1], A. Zumbo[1], B. Portolano[3] and S. Mastrangelo[3]*

[1] Dipartimento Scienze Veterinarie, University of Messina, 98168 Messina, Italy

[2] Dipartimento Scienze Economiche, Aziendali e Statistiche, University of Palermo, 90128 Palermo, Italy

[3] Dipartimento Scienze Agrarie, Alimentari e Forestali, University of Palermo, 90128 Palermo, Italy

[4] Dipartimento di Agricoltura, Alimentazione, Ambiente, University of Catania, Catania, 95123, Italy

*Corresponding author:

Salvatore Mastrangelo

Tel: +39-09123896069

Fax: +39-09123860814

salvatore.mastrangelo@unipa.it

Summary

Phenotypic variability offers the opportunity to investigate the molecular basis involved in the differentiation. Nero Siciliano (Sicilian Black, SB) is a local pig breed reared in Sicily. The animals are usually completely black. Within this population, there are animals showing morphological characteristics resembling the SB genetic type but with grey hair (Sicilian Grey, SG). In this study, using the Illumina PorcineSNP60 BeadChip, we run a genome-wide analyses to identify regions that may explain the phenotypic differences between total black animals (SB, n=21) and pigs

classified as grey (SG, n=27). By combining the results obtained with the two case-control approaches (GWAS and FST), we identified two significant regions, one on SSC5 (95401083 bp) and one on SSC15 (55,051,435 bp), which contains several candidate genes related to growth traits in pig. The results of the Bayesian population differentiation approach highlighted five genomic regions. The locus with the highest FST value (0.315), (rs81386405 on SCC5), was also jointly identified by the two case-control approaches. Finally, scanning the genome for runs of homozygosity (ROH) islands, we found several candidate genes involved in coat color (ASIP, AHCY, EIF2S2, RALY, ITCH, PIGU, NCOA6 and GGT7) or related to different pig performance traits (CTSK and CTSS). In summary, the two analyzed populations differed for several phenotypic traits, and the genes that are involved in these traits (growth, meat traits and coat colour) were detected using different approaches. This study provide another contribution on the identification of regions involved in the phenotypic variability in local pig populations.

Keywords

Local pig populations, genome-wide analyses, phenotype diversity, SNPs, candidate genes.

Phenotypic variability offers the opportunity to investigate the molecular basis involved in the differentiation between populations. The availability of single nucleotide polymorphism (SNP) panels has greatly improved the power of genome-wide studies allowing the identification of highly differentiated genomic regions (e.g. Mastrangelo *et al.* 2019). In pig, several genome-wide studies have been conducted to identify the genomic regions involved in phenotypic differences among breeds (e.g., Fan *et al.* 2011; Wilkinson *et al.* 2013).

Nero Siciliano (Sicilian Black, SB), also known as "Nero dei Nebrodi", is a local pig breed reared on the island of Sicily mainly under extensive management (Guastella *et al.* 2010). The breed is well adapted to marginal conditions and is appreciated for its reproductive performance, disease resistance and production of tasty meat (Russo *et al.* 2004; D'Alessandro *et al*. 2019). The animals

are usually completely black (skin and hair), but a few present a white face or a face with white portions ("suino facciolo"). Its origin dates back to ancient times. However, the genetic pool of the SB breed seems to have been formed mainly during the last few centuries. In addition to the completely black and black with white animals, have been also identified animals showing morphological characteristics resembling the Nero Siciliano genetic type, but with grey hair on the body (Fontanesi *et al.* 2010a), here renamed Grigio Siciliano (Sicilian Grey, SG) (Fig. S1). This phenotype is different from that of wild boar. The SG, compared to the SB, shows a more compact structure with greater transverse diameters, higher average daily gains and lower thickness of the back fat; thanks to these characteristics, in the last century the SG morpho-type was often preferred for family consumption (S. Bordonaro, personal communication). The grey coat colour observed in this population was completely associated with a 4-bp deletion of intron 18 in a single copy *KIT* gene, providing evidence that this mutation characterizes the I^d allele described in the early genetic literature (Fontanesi *et al.* 2010a).

The main objective of the present study was to identify genomic regions that may explain the phenotypic differences observed between the two Sicilian pig populations using the Illumina PorcineSNP60 BeadChip v.2.

Blood or hair root samples were collected from 48 pigs corresponding to the Sicilian Grey (n=27) and Sicilian Black (n=21) (Fig. 1). The animals were genotyped with the Illumina PorcineSNP60 BeadChip v.2. Data filtering was performed using PLINK 1.9 (Chang *et al.* 2015). Chromosomal coordinates for each SNP were obtained from the Sscrofa 11.1 version genome. The dataset was filtered to remove animals with more than 2% of missing genotypes, non-autosomal markers, SNPs with a call rate < 90% and with a minor allele frequency (MAF) < 0.01. After filtering for quality, the final number of retained SNPs for the analysis was 36,301. All 48 animals had high quality genotyping and were included in the analysis.

PLINK v1.9 (Chang *et al.* 2015) was used to calculate pairwise identity-by-state distances between individuals, graphically represented by multidimensional scaling (MDS) analysis. The obtained

MDS plot showed a well-defined structure within groups and a clear separation of the two populations (Fig. S2).

A grey ancient genetic type has been already described in the Sicilian pig population (Chicoli, 1870), easy suggesting a genetic mixture with the more recently formed black breed. This aspect tends to minimize the confounding effects due to high genetic divergence and allowed us to adopt a genome-wide association study (GWAS) and F_{ST} analysis with a case–control model, generally applied to identify genomic regions affecting phenotypic variations between low divergent breeds (Mastrangelo *et al.* 2019). The GWAS was carried out using the univariate case–control model of the SNPASSOC R package (Gonzalez *et al.* 2007). The F_{ST} analysis was performed using PLINK 1.9 (Chang *et al.* 2015). The top 0.9998 SNPs of the percentile distribution were considered the most divergent across the comparison.

The GWAS analysis, at the $P<0.01$ Bonferroni corrected level, revealed a total of 262 significant markers (Fig. S3). In order to restrict the number of potential markers involved in the phenotypic diversity between SB and SG pig populations, we applied on these SNPs a new statistical method called induced smoothed lasso (IS-lasso) (Cilluffo *et al.* 2019). IS-lasso allows fitting generalized linear models with a l_1-penalty so returning, along with point estimates, the resulting standard errors that can be used to make inference in the lasso framework. The IS-lasso case-control analysis was carried out using the islasso R package (Sottile *et al.* 2019), and at $P<0.05$ level, revealed a total of four significant markers (Fig. 2) (Table S1) on different chromosomes (SSC). These markers are located within candidate genes (*RBKS* and *FLI1*) associated to the reproductive performance (de Melo *et al.* 2017; Xu *et al*. 2018). In the F_{ST} case-control analysis, nine SNPs were above the selected threshold (F_{ST} = 0.666) (Table S2). Four SNPs were mapped within known genes, such as rs80968587 which is located on *DENND1A*, a candidate gene associated with the number of vertebrae in pigs (Zhang *et al*. 2015). Combining the results obtained with the two case-control approaches, we identified two significant markers: one on SSC5 (95,401,083 bp) and one on SSC15 (55,051,435 bp). One (*MGAT4C*) and seven (*RBPMS, DCTN6, MBOAT4, LEPROTL1, SARAF,*

DUSP4 and *TNKS*) genes were annotated in the SSC5 and SSC15 regions, respectively (in a window of ± 500 kb from the significant marker). These genes are involved in several important biological process: *RBPMS* was reported as candidate genes in pig for growth trait (Puig-Oliveras *et al*. 2014); *MBOAT4* plays an important role in the control of energy balance and is essential for the survival of calorie-restricted mice by maintaining the levels of growth hormone (Zhao *et al*. 2010); *LEPROTL1*, is a growth hormone receptor highly expressed in porcine lung tissue (Demarchi *et al*. 2007); *SARAF* plays a role in calcium ion transport (Palty *et al.* 2012) and therefore on the pig traits which determine the postmortem characteristics (Piórkowska *et al*, 2017); *TNKS* gene has been shown to be related to energy expenditure, feed intake and adiposity in mice (Yeh *et al*. 2009), suggesting its possible role in adaptation to variation in local environment feed availability and quality (Yurchenko *et al*. 2018). These genes may very well explain the morphological difference between the pig populations.

The F_{ST} -outlier approach implemented in the Bayescan software (Foll & Gaggiotti 2008) was also adopted to identify loci under selection. Bayescan analyses comprised 20 pilot runs of 5,000 iterations, a burn-in of 50,000 iterations, a thinning interval of 10 (5,000 iterations were used for the estimation of posterior odds) with a resulting total number of 100,000 iterations. To control the number of false positives, significant SNPs were defined by applying a *q*-value threshold of 0.01. Results from the Bayesian population differentiation approach identified five significant markers (*q*-value ≤0.01) located on 4 autosomes (Table S3). The locus with the highest F_{ST} value (0.315) was rs81386405 on SCC5. This was also the most significant marker which was jointly identified by the two case-control approaches. The only closest gene in this genomic region was *MGAT4C* (62.7 Kb upstream of the marker). Few information are reported in the literature about the role of this gene. In a GWAS study dealing with average daily gain in pigs, *MGAT4C* was reported as the closest gene for one of the significant SNP associated with the studied phenotype (Fontanesi *et al.* 2014). Furthermore, the Bayescan analysis revealed a marker (rs81307769 on SSC6) mapping within a gene (*GPI*) which was also detected as significant in F_{ST} case-control analysis, and a SNP,

rs81453447 on SSC15, located within the *PLA2R1*, a fatty acid transporter associated with fat deposition and body weight in chicken (Gheyas *et al.* 2015).

Finally, we scanned the genome of the two pig populations to identify the genomic regions that were most commonly associated with runs of homozygosity (ROH). In fact, the genomic regions subjected to selection frequently show signatures, such as reduced nucleotide diversity, and tend to generate ROH islands (Metzger *et al*. 2015; Mastrangelo *et al*. 2017; Talenti *et al*. 2017). The genomic regions were defined according to Mastrangelo *et al*. (2019). In total, two ROH islands were identified. Table S4 provides the SSC position, the start and the end of these genomic regions. Among the annotated genes within ROH islands, some were worth mentioning because their function could plays important roles to explain the phenotypic differentiation between the SB and SG pigs. In the ROH island of SG, were mapped several genes involved in coat color (*ASIP, AHCY, EIF2S2, RALY, ITCH, PIGU, NCOA6* and *GGT7*) (Nazari-Ghadikolaei *et al*. 2018; Mendoza *et al.* 2019), as well as genes interested in lipid synthesis (Cai *et al.* 2015); the ROH island highlighted in SB mapped genes involved in the myogenesis and adipogenesis (*TMOD4*) (Zhao *et al*. 2013) or related to pig performance traits such as back fat thickness (*CTSK*) (Fontanesi *et al*. 2010b).

In summary, the two analyzed populations differed for several phenotypic traits, and the genes that are involved in some of these traits (stature and growth, meat traits and coat colour) were detected using different genome-wide approaches. We showed that the scan of the genome by comparing relatively close populations can be useful to identify genes that are likely responsible for the phenotypic diversity. This study provide another contribution on the identification of genomic regions involved in the phenotypic variability in local pig populations. Determination of the candidate genomic regions will help to protect and utilize these genetic resources. In future studies, the sequencing data and an increase in the number of genotyped animals would be particularly relevant to refine and validate these results.

Competing interests

The authors declare that they do not have competing interests.

Availability of data and materials

The dataset used and analyzed during this study is available on request from the corresponding author.

References

Cai Z., Zhang L., Jiang X., Sheng Y. & Xu, N. (2015) Differential miRNA expression profiles in the longissimus dorsi muscle between intact and castrated male pigs. *Research in Veterinary Science* **99**, 99-104.

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

Chicoli N. (1870). Riproduzione, allevamento e miglioramento degli animali domestici in Sicilia. Stamperia di G. 4. Lorsnaider, Palermo, Italy.

Cilluffo G., Sottile G., La Grutta S. & Muggeo V.M.R. (2019) The Induced Smoothed lasso: A practical framework for hypothesis testing in high dimensional regression. *Statistical Methods in Medical Research*, doi: 10.1177/0962280219842890.

D'Alessandro E., Giosa D., Sapienza I., Giuffrè L., Aiese Cigliano R., Romeo O. & Zumbo A. (2019) Whole genome SNPs discovery in Nero Siciliano pig. *Genetics and Molecular Biology*, doi: 10.1590/1678-4685-GMB-2018-0169

de Melo T.P., De Camargo G.M.F., De Albuquerque L.G. & Carvalheiro R. (2017) Genome-wide association study provides strong evidence of genes affecting the reproductive performance of Nellore beef cows. *PloS one* 12(5), e0178551.

Demarchi F., Bertoli C., Copetti T., Eskelinen E.L. & Schneider C. (2007) Calpain as a novel regulator of autophagosome formation. *Autophagy* **3**(3), 235–237

Fan B., Onteru S.K., Du Z.Q., Garrick D.J., Stalder K.J. & Rothschild M.F. (2011) Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *PloS ONE* **6**(2), e14726.

Foll M. & Gaggiotti O.E. (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**, 977–993

Fontanesi L., D'Alessandro E., Scotti E., Liotta L., Crovetti A., Chiofalo V. & Russo V. (2010a) Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. *Animal Genetics* **41**(5), 478-492.

Fontanesi L., Scotti E., Buttazzoni L., Dall'Olio S., Davoli R. & Russo V. (2010b) A single nucleotide polymorphism in the porcine cathepsin K (CTSK) gene is associated with back fat thickness and production traits in Italian Duroc pigs. *Molecular Biology Reports* **37**(1), 491-495.

Fontanesi L., Schiavo G., Galimberti G., Calo D.G. & Russo V. (2014) A genomewide association study for average daily gain in Italian Large White pigs. *Journal of Animal Science* **92**(4), 1385-1394.

Gheyas A.A., Boschiero C., Eory L., Ralph H., Kuo R., Woolliams J.A. & Burt D.W. (2015) Functional classification of 15 million SNPs detected from diverse chicken populations. *DNA Research* **22**(3), 205-217.

Guastella A.M., Criscione A., Marletta D., Zuccaro A., Chies L., & Bordonaro S. (2010). Molecular characterization and genetic structure of the Nero Siciliano pig breed. *Genetics and Molecular Biology*, **33**(4), 650-656.

Mastrangelo S., Ben Jemaa S., Sottile G., Casu S., Portolano B., Ciani E. & Pilla F. (2019) Combined approaches to identify genomic regions involved in phenotypic differentiation between low divergent breeds: Application in Sardinian sheep populations. *Journal of Animal Breeding and Genetics* doi: 10.1111/jbg.12422

Mastrangelo S., Tolone M., Sardina M.T., Sottile G., Sutera A.M., Di Gerlando R. & Portolano B. (2017) Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genetics Selection Evolution* **49**(1),84.

Mendoza M.N., Raudsepp T., Alshanbari F., Gutierrez Reynoso G.A. & Ponce de León F.A. (2019) Chromosomal localization of candidate genes for fiber growth and color in alpaca (Vicugna pacos). *Frontiers in Genetics* **10**, 583.

Metzger J., Karwath M., Tonda R, Beltran S., Agueda L., Gut M., Gut I.G. & Distl O. (2015) Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics* **16**, 764.

Nazari-Ghadikolaei A., Mehrabani-Yeganeh H., Miarei-Aashtiani S.R., Staiger E.A., Rashidi A. & Huson H.J. (2018) Genome-wide association studies identify candidate genes for coat color and mohair traits in the iranian markhoz goat. *Frontiers in Genetics* **9**, 105.

Palty R., Raveh A., Kaminsky I., Meller R. & Reuveny E. (2012) SARAF inactivates the store operated calcium entry machinery to prevent excess calcium refilling. *Cell* **149**, 425–438.

Piórkowska K., Żukowski K., Szmatoła T., Ropka-Molik K. & Tyra M. (2017) Transcript variants of a region on SSC15 rich in QTLs associated with meat quality in pigs. *Annals of Animal Science* **17**(3), 703-715.

Puig-Oliveras A., Ballester M., Corominas J., Revilla M., Estellé J., Fernández A.I., Ramayo-Caldas Y. & Folch J.M. (2014) A Co-Association Network Analysis of the Genetic Determination of Pig Conformation, Growth and Fatness. *PLoS ONE* **9**(12), e114862.

Russo V., Fontanesi L., Davoli R., Chiofalo L., Liotta L. & Zumbo, A. (2004). Analysis of single nucleotide polymorphisms in major and candidate genes for production traits in Nero Siciliano pig breed. *Italian Journal of Animal Science* **3**(1), 19-29.

Sottile G., Cilluffo G. & Muggeo V.M.R. (2019) The R package islasso: estimation and hypothesis testing in lasso regression. *Technical Report on ResearchGate*.

Talenti A., Bertolini F., Pagnacco G., Pilla F., Ajmone-Marsan P., Rothschild M.F. & Crepaldi P. (2017) The Valdostana goat: a genome-wide investigation of the distinctiveness of its selective sweep regions. *Mammalian Genome* **28**, 114–128.

Wilkinson S., Lu Z.H., Megens H.J., Archibald A.L., Haley C., Jackson I.J., Groenen M.A.M., Crooijmans R.P.M.A., Ogden R. & Wiener P. (2013). Signatures of diversifying selection in European pig breeds. *PLoS Genetics* **9**(4),e1003453.

Woolliams J.A., Gonzalez J.R., Armengol L., Sole X., Guino E., Mercader J.M., Estivill X. & Moreno V. (2007) SNPASSOC: an R package to perform whole genome association studies. *Bioinformatics* **23**, 654–5.

Xu S.S., Gao L., Xie X.L., Ren Y.L., Shen Z.Q., Wang F., Shen M., Eyþórsdóttir E., Hallsson J.H., Kiseleva T., Kantanen J. & Li M-H. (2018) Genome-wide association analyses highlight the potential for different genetic mechanisms for litter size among sheep breeds. *Frontiers in Genetics* **9**, 118.

Yeh T.Y.J., Beiswenger K.K., Li P., Bolin K.E., Lee R.M., Tsao, T.S., Murphy A.N., Hevener A.L. & Chi N.W. (2009) Hypermetabolism, hyperphagia, and reduced adiposity in tankyrase-deficient mice. *Diabetes* **58**(11), 2476-2485.

Yurchenko A.A., Daetwyler H.D., Yudin N., Schnabel R.D., Vander Jagt C.J., Soloshenko V., Lhasaranov B., Popov R., Taylor J.F. & Larkin D.M. (2018) Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Scientific Reports* 8.

Zhang L., Liu X., Liang J., Zhao K., Yan H., Li, N., Pu L., Zhang Y., Shi H., Wang L. & Wang L. (2015) Genome-wide Association Study for Number of Vertebrae in an F2 Large White× Minzhu Population of Pigs. *bioRxiv* 016956.

Zhao T.J, Liang G., Li R.L., Xie X., Sleeman M.W., Murphy A.J., Valenzuela D.M., Yancopoulos G.D., Goldstein J.L. & Brown M.S. (2010) Ghrelin O-acyltransferase (GOAT) is essential for growth hormone-mediated survival of calorie-restricted mice. *Proceedings of the National Academy of Sciences* **107**,7467–7472.

Zhao X., Huang Z., Liu X., Chen Y., Gong W., Yu K. & Mo D. (2013) The switch role of the Tmod4 in the regulation of balanced development between myogenesis and adipogenesis. *Gene* **532**(2), 263-271.

Figures

Figure 1 Examples of (a) Sicilian Black and (b) Sicilian Grey pigs.

Figure 2 Manhattan plot of the *P*-values in the induced smoothed lasso (IS-lasso) case-control analysis. The horizontal lines represent the genome-wide significance (red; $P < 0.05$) and suggestively significant (green; $P < 0.10$).

Supporting information

Table S1 List of significant single nucleotide polymorphisms (SNPs) obtained in the genome-wide association study (GWAS). When markers were located within genes, these are indicated.

SSC	SNP	Position	Annotated genes
3	rs81375874	111,354,650	*RBKS*
5	rs81386405	95,401,083	-
9	rs81411669	55,687,607	*FLI1*
15	rs80807225	55,051,435	-

SSC, porcine chromosome number

Table S2 List of significant single nucleotide polymorphisms (SNPs) obtained in the case-control F_{ST} analysis. When markers were located within genes, these are indicated.

SSC	SNP	Position	F_{ST} value	Annotated genes
1	rs80968587	264,309,057	0.705	*DENND1A*
2	rs81362513	12,438,091	0.678	*LOC*
2	rs81361469	94,633,673	0.686	
5	rs81386405	95,401,083	0.927	
6	rs81307769	44,059,341	0.666	*GPI*
6	rs81312099	121,412,908	0.672	
11	rs80804521	56,512,042	0.733	*SLITRK6*
15	rs80807225	55,051,435	0.686	
16	rs81462450	71,769,147	0.666	*GM2A*

SSC, porcine chromosome number

Table S3 List of significant single nucleotide polymorphisms (SNPs) identified using the F_{ST}-outlier method implemented in BayeScan. When markers were located within genes, these are indicated.

SSC	SNP	Position	*q*value	F_{ST}	Annotated genes
2	rs81361469	94,633,673	0.0029	0.287	-
5	rs81386405	95,401,083	0.0001	0.315	-
6	rs81307769	44,059,341	0.0052	0.283	*GPI*
15	rs81331202	58,833,925	0.0094	0.280	-
15	rs81453447	67,008,603	0.0075	0.278	*PLA2R1*

Table S4 Run of homozygosity (ROH) islands identified within the two pig population.

Pop	SSC	SNP	Start and end (bp)

Sicilian Grey	17	38	37,273,372–39,597,192
Sicilian Black	4	37	97,587,537–99,233,113

SSC, porcine chromosome number; SNP, single nucleotide polymorphisms.

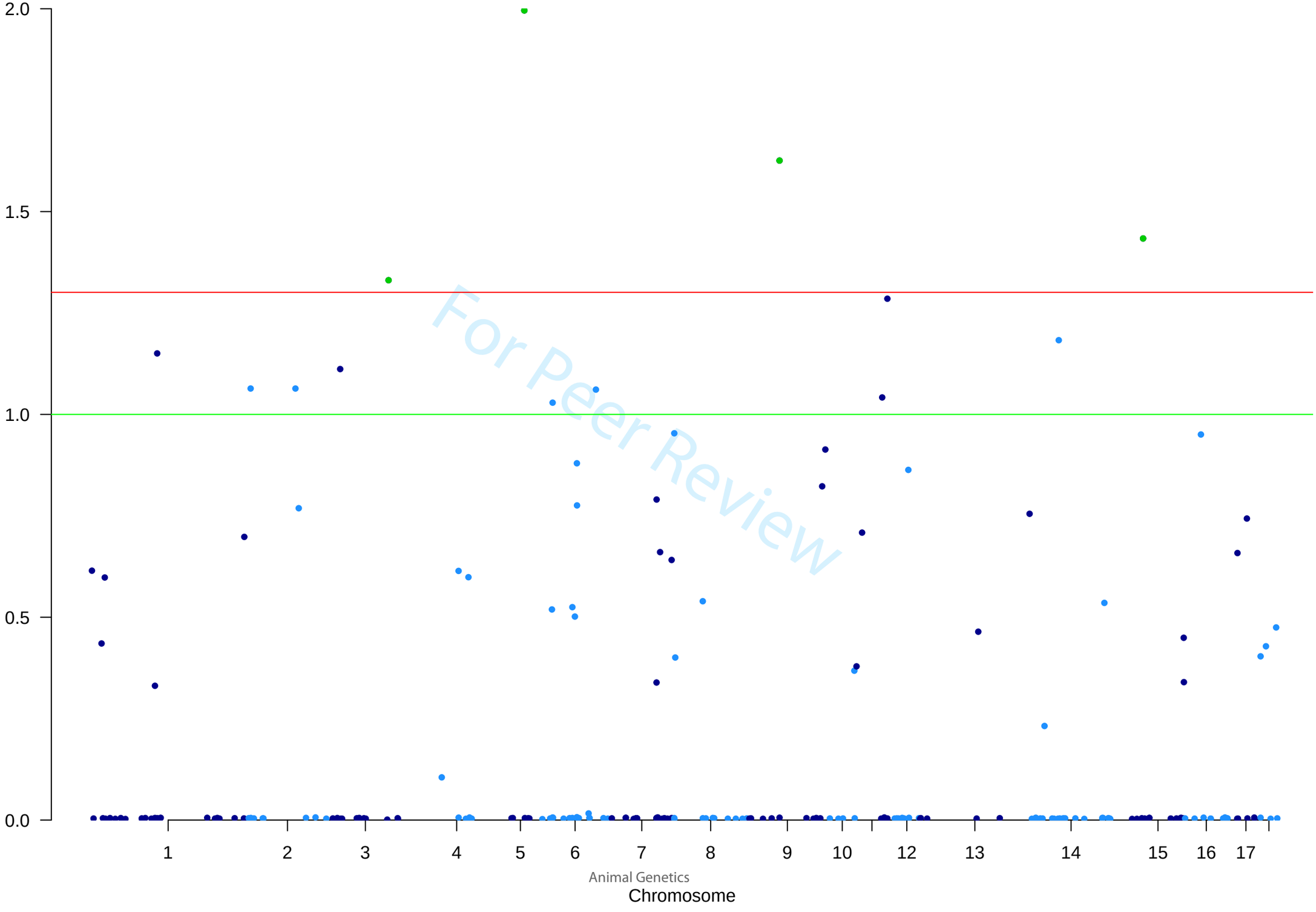
Figure S1 Examples of Sicilian Black and Sicilian Grey pigs reared in the same flock.

Figure S2 Multidimensional scaling (MDS) plot of Sicilian Black (blu) and Sicilian Grey (red) individuals.

Figure S3 Manhattan plot of the *P*-values in the genome-wide association study (GWAS) analysis. The horizontal lines represent the genome-wide significance (red; $P < 0.01$) and suggestively significant (green; $P < 0.05$).

Figure 1

80x102mm (300 x 300 DPI)



A genome-wide analyses reveals the regions involved in the phenotypic diversity between two local Sicilian pig populations

E. D'Alessandro[1], G. Sottile[2], M.T. Sardina[3], A. Criscione[4], S. Bordonaro[4], A.M. Sutera[1], A. Zumbo[1], B. Portolano[3] and S. Mastrangelo[3]*

[1] Dipartimento Scienze Veterinarie, University of Messina, 98168 Messina, Italy

[2] Dipartimento Scienze Economiche, Aziendali e Statistiche, University of Palermo, 90128 Palermo, Italy

[3] Dipartimento Scienze Agrarie, Alimentari e Forestali, University of Palermo, 90128 Palermo, Italy

[4] Dipartimento di Agricoltura, Alimentazione, Ambiente, University of Catania, Catania, 95123, Italy

*Corresponding author:

Salvatore Mastrangelo

Tel: +39-09123896069

Fax: +39-09123860814

salvatore.mastrangelo@unipa.it

Summary

Phenotypic variability offers the opportunity to investigate the molecular basis involved in the differentiation. Nero Siciliano (Sicilian Black, SB) is a local pig breed reared in Sicily. The animals are usually completely black. Within this population, there are animals showing morphological characteristics resembling the SB genetic type but with grey hair (Sicilian Grey, SG). In this study, using the Illumina PorcineSNP60 BeadChip, we run a genome-wide analyses to identify regions that may explain the phenotypic differences between total black animals (SB, n=21) and pigs

classified as grey (SG, n=27). By combining the results obtained with the two case-control approaches (GWAS and FST), we identified two significant regions, one on SSC5 (95401083 bp) and one on SSC15 (55,051,435 bp), which contains several candidate genes related to growth traits in pig. The results of the Bayesian population differentiation approach highlighted five genomic regions. The locus with the highest FST value (0.315), (rs81386405 on SCC5), was also jointly identified by the two case-control approaches. Finally, scanning the genome for runs of homozygosity (ROH) islands, we found several candidate genes involved in coat color (ASIP, AHCY, EIF2S2, RALY, ITCH, PIGU, NCOA6 and GGT7) or related to different pig performance traits (CTSK and CTSS). In summary, the two analyzed populations differed for several phenotypic traits, and the genes that are involved in these traits (growth, meat traits and coat colour) were detected using different approaches. This study provide another contribution on the identification of regions involved in the phenotypic variability in local pig populations.

Keywords

Local pig populations, genome-wide analyses, phenotype diversity, SNPs, candidate genes.

Phenotypic variability offers the opportunity to investigate the molecular basis involved in the differentiation between populations. The availability of single nucleotide polymorphism (SNP) panels has greatly improved the power of genome-wide studies allowing the identification of highly differentiated genomic regions (e.g. Mastrangelo *et al.* 2019). In pig, several genome-wide studies have been conducted to identify the genomic regions involved in phenotypic differences among breeds (e.g., Fan *et al.* 2011; Wilkinson *et al.* 2013).

Nero Siciliano (Sicilian Black, SB), also known as “Nero dei Nebrodi”, is a local pig breed reared on the island of Sicily mainly under extensive management (Guastella *et al.* 2010). The breed is well adapted to marginal conditions and is appreciated for its reproductive performance, disease resistance and production of tasty meat (Russo *et al.* 2004; D’Alessandro *et al*. 2019). The animals are usually completely black (skin and hair), but a few present a white face or a face with white

portions ("suino facciolo"). Its origin dates back to ancient times. However, the genetic pool of the SB breed seems to have been formed mainly during the last few centuries. In addition to the completely black and black with white animals, have been also identified animals showing morphological characteristics resembling the Nero Siciliano genetic type, but with grey hair on the body (Fontanesi *et al.* 2010a), here renamed Grigio Siciliano (Sicilian Grey, SG) (Fig. S1). This phenotype is different from that of wild boar. The SG, compared to the SB, shows a more compact structure with greater transverse diameters, higher average daily gains and lower thickness of the back fat; thanks to these characteristics, in the last century the SG morpho-type was often preferred for family consumption (S. Bordonaro, personal communication). The grey coat colour observed in this population was completely associated with a 4-bp deletion of intron 18 in a single copy *KIT* gene, providing evidence that this mutation characterizes the I^d allele described in the early genetic literature (Fontanesi *et al.* 2010a).

The main objective of the present study was to identify genomic regions that may explain the phenotypic differences observed between the two Sicilian pig populations using the Illumina PorcineSNP60 BeadChip v.2.

Blood or hair root samples were collected from 48 pigs corresponding to the Sicilian Grey (n=27) and Sicilian Black (n=21) (Fig. 1). The animals were genotyped with the Illumina PorcineSNP60 BeadChip v.2. Data filtering was performed using PLINK 1.9 (Chang *et al.* 2015). Chromosomal coordinates for each SNP were obtained from the Sscrofa 11.1 version genome. The dataset was filtered to remove animals with more than 2% of missing genotypes, non-autosomal markers, SNPs with a call rate < 90% and with a minor allele frequency (MAF) < 0.01. After filtering for quality, the final number of retained SNPs for the analysis was 36,301. All 48 animals had high quality genotyping and were included in the analysis.

PLINK v1.9 (Chang *et al.* 2015) was used to calculate pairwise identity-by-state distances between individuals, graphically represented by multidimensional scaling (MDS) analysis. The obtained

MDS plot showed a well-defined structure within groups and a clear separation of the two populations (Fig. S2).

A grey ancient genetic type has been already described in the Sicilian pig population (Chicoli, 1870), easy suggesting a genetic mixture with the more recently formed black breed. This aspect tends to minimize the confounding effects due to high genetic divergence and allowed us to adopt a genome-wide association study (GWAS) and F_{ST} analysis with a case–control model, generally applied to identify genomic regions affecting phenotypic variations between low divergent breeds (Mastrangelo *et al.* 2019). The GWAS was carried out using the univariate case–control model of the SNPASSOC R package (Gonzalez *et al.* 2007). The F_{ST} analysis was performed using PLINK 1.9 (Chang *et al.* 2015). The top 0.9998 SNPs of the percentile distribution were considered the most divergent across the comparison.

The GWAS analysis, at the $P<0.01$ Bonferroni corrected level, revealed a total of 262 significant markers (Fig. S3). In order to restrict the number of potential markers involved in the phenotypic diversity between SB and SG pig populations, we applied on these SNPs a new statistical method called induced smoothed lasso (IS-lasso) (Cilluffo *et al.* 2019). IS-lasso allows fitting generalized linear models with a l_1-penalty so returning, along with point estimates, the resulting standard errors that can be used to make inference in the lasso framework. The IS-lasso case-control analysis was carried out using the islasso R package (Sottile *et al.* 2019), and at $P<0.05$ level, revealed a total of four significant markers (Fig. 2) (Table S1) on different chromosomes (SSC). These markers are located within candidate genes (*RBKS* and *FLI1*) associated to the reproductive performance (de Melo *et al.* 2017; Xu *et al*. 2018). In the F_{ST} case-control analysis, nine SNPs were above the selected threshold (F_{ST} = 0.666) (Table S2). Four SNPs were mapped within known genes, such as rs80968587 which is located on *DENND1A*, a candidate gene associated with the number of vertebrae in pigs (Zhang *et al*. 2015). Combining the results obtained with the two case-control approaches, we identified two significant markers: one on SSC5 (95,401,083 bp) and one on SSC15 (55,051,435 bp). One (*MGAT4C*) and seven (*RBPMS, DCTN6, MBOAT4, LEPROTL1, SARAF,*

DUSP4 and *TNKS*) genes were annotated in the SSC5 and SSC15 regions, respectively (in a window of ± 500 kb from the significant marker). These genes are involved in several important biological process: *RBPMS* was reported as candidate genes in pig for growth trait (Puig-Oliveras *et al*. 2014); *MBOAT4* plays an important role in the control of energy balance and is essential for the survival of calorie-restricted mice by maintaining the levels of growth hormone (Zhao *et al*. 2010); *LEPROTL1*, is a growth hormone receptor highly expressed in porcine lung tissue (Demarchi *et al*. 2007); *SARAF* plays a role in calcium ion transport (Palty *et al.* 2012) and therefore on the pig traits which determine the postmortem characteristics (Piórkowska *et al*, 2017); *TNKS* gene has been shown to be related to energy expenditure, feed intake and adiposity in mice (Yeh *et al*. 2009), suggesting its possible role in adaptation to variation in local environment feed availability and quality (Yurchenko *et al*. 2018). These genes may very well explain the morphological difference between the pig populations.

The F_{ST} -outlier approach implemented in the Bayescan software (Foll & Gaggiotti 2008) was also adopted to identify loci under selection. Bayescan analyses comprised 20 pilot runs of 5,000 iterations, a burn-in of 50,000 iterations, a thinning interval of 10 (5,000 iterations were used for the estimation of posterior odds) with a resulting total number of 100,000 iterations. To control the number of false positives, significant SNPs were defined by applying a *q*-value threshold of 0.01. Results from the Bayesian population differentiation approach identified five significant markers (*q*-value ≤0.01) located on 4 autosomes (Table S3). The locus with the highest F_{ST} value (0.315) was rs81386405 on SCC5. This was also the most significant marker which was jointly identified by the two case-control approaches. The only closest gene in this genomic region was *MGAT4C* (62.7 Kb upstream of the marker). Few information are reported in the literature about the role of this gene. In a GWAS study dealing with average daily gain in pigs, *MGAT4C* was reported as the closest gene for one of the significant SNP associated with the studied phenotype (Fontanesi *et al.* 2014). Furthermore, the Bayescan analysis revealed a marker (rs81307769 on SSC6) mapping within a gene (*GPI*) which was also detected as significant in F_{ST} case-control analysis, and a SNP,

rs81453447 on SSC15, located within the *PLA2R1*, a fatty acid transporter associated with fat deposition and body weight in chicken (Gheyas *et al.* 2015).

Finally, we scanned the genome of the two pig populations to identify the genomic regions that were most commonly associated with runs of homozygosity (ROH). In fact, the genomic regions subjected to selection frequently show signatures, such as reduced nucleotide diversity, and tend to generate ROH islands (Metzger *et al*. 2015; Mastrangelo *et al*. 2017; Talenti *et al*. 2017). The genomic regions were defined according to Mastrangelo *et al*. (2019). In total, two ROH islands were identified. Table S4 provides the SSC position, the start and the end of these genomic regions. Among the annotated genes within ROH islands, some were worth mentioning because their function could plays important roles to explain the phenotypic differentiation between the SB and SG pigs. In the ROH island of SG, were mapped several genes involved in coat color (*ASIP, AHCY, EIF2S2, RALY, ITCH, PIGU, NCOA6* and *GGT7*) (Nazari-Ghadikolaei *et al*. 2018; Mendoza *et al.* 2019), as well as genes interested in lipid synthesis (Cai *et al.* 2015); the ROH island highlighted in SB mapped genes involved in the myogenesis and adipogenesis (*TMOD4*) (Zhao *et al*. 2013) or related to pig performance traits such as back fat thickness (*CTSK*) (Fontanesi *et al*. 2010b).

In summary, the two analyzed populations differed for several phenotypic traits, and the genes that are involved in some of these traits (stature and growth, meat traits and coat colour) were detected using different genome-wide approaches. We showed that the scan of the genome by comparing relatively close populations can be useful to identify genes that are likely responsible for the phenotypic diversity. This study provide another contribution on the identification of genomic regions involved in the phenotypic variability in local pig populations. Determination of the candidate genomic regions will help to protect and utilize these genetic resources. In future studies, the sequencing data and an increase in the number of genotyped animals would be particularly relevant to refine and validate these results.

Competing interests

The authors declare that they do not have competing interests.

Availability of data and materials

The dataset used and analyzed during this study is available on request from the corresponding author.

References

Cai Z., Zhang L., Jiang X., Sheng Y. & Xu, N. (2015) Differential miRNA expression profiles in the longissimus dorsi muscle between intact and castrated male pigs. *Research in Veterinary Science* **99**, 99-104.

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

Chicoli N. (1870). Riproduzione, allevamento e miglioramento degli animali domestici in Sicilia. Stamperia di G. 4. Lorsnaider, Palermo, Italy.

Cilluffo G., Sottile G., La Grutta S. & Muggeo V.M.R. (2019) The Induced Smoothed lasso: A practical framework for hypothesis testing in high dimensional regression. *Statistical Methods in Medical Research*, doi: 10.1177/0962280219842890.

D'Alessandro E., Giosa D., Sapienza I., Giuffrè L., Aiese Cigliano R., Romeo O. & Zumbo A. (2019) Whole genome SNPs discovery in Nero Siciliano pig. *Genetics and Molecular Biology*, doi: 10.1590/1678-4685-GMB-2018-0169

de Melo T.P., De Camargo G.M.F., De Albuquerque L.G. & Carvalheiro R. (2017) Genome-wide association study provides strong evidence of genes affecting the reproductive performance of Nellore beef cows. *PloS one* 12(5), e0178551.

Demarchi F., Bertoli C., Copetti T., Eskelinen E.L. & Schneider C. (2007) Calpain as a novel regulator of autophagosome formation. *Autophagy* **3**(3), 235–237

Fan B., Onteru S.K., Du Z.Q., Garrick D.J., Stalder K.J. & Rothschild M.F. (2011) Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *PloS ONE* **6**(2), e14726.

Foll M. & Gaggiotti O.E. (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**, 977–993

Fontanesi L., D'Alessandro E., Scotti E., Liotta L., Crovetti A., Chiofalo V. & Russo V. (2010a) Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. *Animal Genetics* **41**(5), 478-492.

Fontanesi L., Scotti E., Buttazzoni L., Dall'Olio S., Davoli R. & Russo V. (2010b) A single nucleotide polymorphism in the porcine cathepsin K (CTSK) gene is associated with back fat thickness and production traits in Italian Duroc pigs. *Molecular Biology Reports* **37**(1), 491-495.

Fontanesi L., Schiavo G., Galimberti G., Calo D.G. & Russo V. (2014) A genomewide association study for average daily gain in Italian Large White pigs. *Journal of Animal Science* **92**(4), 1385-1394.

Gheyas A.A., Boschiero C., Eory L., Ralph H., Kuo R., Woolliams J.A. & Burt D.W. (2015) Functional classification of 15 million SNPs detected from diverse chicken populations. *DNA Research* **22**(3), 205-217.

Guastella A.M., Criscione A., Marletta D., Zuccaro A., Chies L., & Bordonaro S. (2010). Molecular characterization and genetic structure of the Nero Siciliano pig breed. *Genetics and Molecular Biology*, **33**(4), 650-656.

Mastrangelo S., Ben Jemaa S., Sottile G., Casu S., Portolano B., Ciani E. & Pilla F. (2019) Combined approaches to identify genomic regions involved in phenotypic differentiation between low divergent breeds: Application in Sardinian sheep populations. *Journal of Animal Breeding and Genetics* doi: 10.1111/jbg.12422

Mastrangelo S., Tolone M., Sardina M.T., Sottile G., Sutera A.M., Di Gerlando R. & Portolano B. (2017) Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genetics Selection Evolution* **49**(1),84.

Mendoza M.N., Raudsepp T., Alshanbari F., Gutierrez Reynoso G.A. & Ponce de León F.A. (2019) Chromosomal localization of candidate genes for fiber growth and color in alpaca (Vicugna pacos). *Frontiers in Genetics* **10**, 583.

Metzger J., Karwath M., Tonda R, Beltran S., Agueda L., Gut M., Gut I.G. & Distl O. (2015) Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics* **16**, 764.

Nazari-Ghadikolaei A., Mehrabani-Yeganeh H., Miarei-Aashtiani S.R., Staiger E.A., Rashidi A. & Huson H.J. (2018) Genome-wide association studies identify candidate genes for coat color and mohair traits in the iranian markhoz goat. *Frontiers in Genetics* **9**, 105.

Palty R., Raveh A., Kaminsky I., Meller R. & Reuveny E. (2012) SARAF inactivates the store operated calcium entry machinery to prevent excess calcium refilling. *Cell* **149**, 425–438.

Piórkowska K., Żukowski K., Szmatoła T., Ropka-Molik K. & Tyra M. (2017) Transcript variants of a region on SSC15 rich in QTLs associated with meat quality in pigs. *Annals of Animal Science* **17**(3), 703-715.

Puig-Oliveras A., Ballester M., Corominas J., Revilla M., Estellé J., Fernández A.I., Ramayo-Caldas Y. & Folch J.M. (2014) A Co-Association Network Analysis of the Genetic Determination of Pig Conformation, Growth and Fatness. *PLoS ONE* **9**(12), e114862.

Russo V., Fontanesi L., Davoli R., Chiofalo L., Liotta L. & Zumbo, A. (2004). Analysis of single nucleotide polymorphisms in major and candidate genes for production traits in Nero Siciliano pig breed. *Italian Journal of Animal Science* **3**(1), 19-29.

Sottile G., Cilluffo G. & Muggeo V.M.R. (2019) The R package islasso: estimation and hypothesis testing in lasso regression. *Technical Report on ResearchGate*.

Talenti A., Bertolini F., Pagnacco G., Pilla F., Ajmone-Marsan P., Rothschild M.F. & Crepaldi P. (2017) The Valdostana goat: a genome-wide investigation of the distinctiveness of its selective sweep regions. *Mammalian Genome* **28**, 114–128.

Wilkinson S., Lu Z.H., Megens H.J., Archibald A.L., Haley C., Jackson I.J., Groenen M.A.M., Crooijmans R.P.M.A., Ogden R. & Wiener P. (2013). Signatures of diversifying selection in European pig breeds. *PLoS Genetics* **9**(4),e1003453.

Woolliams J.A., Gonzalez J.R., Armengol L., Sole X., Guino E., Mercader J.M., Estivill X. & Moreno V. (2007) SNPASSOC: an R package to perform whole genome association studies. *Bioinformatics* **23**, 654–5.

Xu S.S., Gao L., Xie X.L., Ren Y.L., Shen Z.Q., Wang F., Shen M., Eyþórsdóttir E., Hallsson J.H., Kiseleva T., Kantanen J. & Li M-H. (2018) Genome-wide association analyses highlight the potential for different genetic mechanisms for litter size among sheep breeds. *Frontiers in Genetics* **9**, 118.

Yeh T.Y.J., Beiswenger K.K., Li P., Bolin K.E., Lee R.M., Tsao, T.S., Murphy A.N., Hevener A.L. & Chi N.W. (2009) Hypermetabolism, hyperphagia, and reduced adiposity in tankyrase-deficient mice. *Diabetes* **58**(11), 2476-2485.

Yurchenko A.A., Daetwyler H.D., Yudin N., Schnabel R.D., Vander Jagt C.J., Soloshenko V., Lhasaranov B., Popov R., Taylor J.F. & Larkin D.M. (2018) Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Scientific Reports* 8.

Zhang L., Liu X., Liang J., Zhao K., Yan H., Li, N., Pu L., Zhang Y., Shi H., Wang L. & Wang L. (2015) Genome-wide Association Study for Number of Vertebrae in an F2 Large White× Minzhu Population of Pigs. *bioRxiv* 016956.

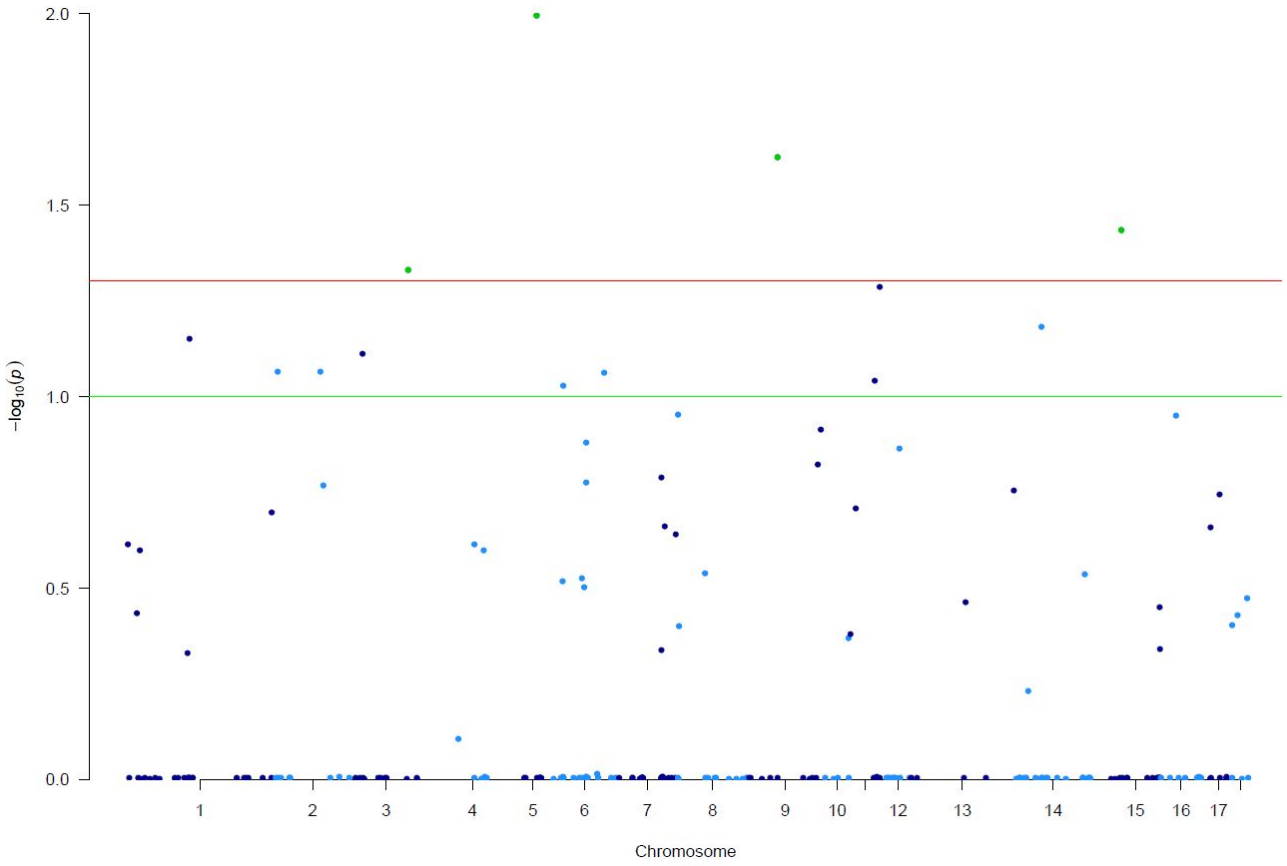
Zhao T.J, Liang G., Li R.L., Xie X., Sleeman M.W., Murphy A.J., Valenzuela D.M., Yancopoulos G.D., Goldstein J.L. & Brown M.S. (2010) Ghrelin O-acyltransferase (GOAT) is essential for growth hormone-mediated survival of calorie-restricted mice. *Proceedings of the National Academy of Sciences* **107**,7467–7472.

Zhao X., Huang Z., Liu X., Chen Y., Gong W., Yu K. & Mo D. (2013) The switch role of the Tmod4 in the regulation of balanced development between myogenesis and adipogenesis. *Gene* **532**(2), 263-271.

Figures

Figure 1 Examples of (a) Sicilian Black and (b) Sicilian Grey pigs.

Figure 2 Manhattan plot of the *P*-values in the induced smoothed lasso (IS-lasso) case-control analysis. The horizontal lines represent the genome-wide significance (red; $P < 0.05$) and suggestively significant (green; $P < 0.10$).



Supporting information

Table S1 List of significant single nucleotide polymorphisms (SNPs) obtained in the genome-wide association study (GWAS). When markers were located within genes, these are indicated.

SSC	SNP	Position	Annotated genes
3	rs81375874	111,354,650	*RBKS*
5	rs81386405	95,401,083	-
9	rs81411669	55,687,607	*FLI1*
15	rs80807225	55,051,435	-

SSC, porcine chromosome number

Table S2 List of significant single nucleotide polymorphisms (SNPs) obtained in the case-control F_{ST} analysis. When markers were located within genes, these are indicated.

SSC	SNP	Position	F_{ST} value	Annotated genes
1	rs80968587	264,309,057	0.705	*DENND1A*
2	rs81362513	12,438,091	0.678	*LOC*
2	rs81361469	94,633,673	0.686	
5	rs81386405	95,401,083	0.927	
6	rs81307769	44,059,341	0.666	*GPI*
6	rs81312099	121,412,908	0.672	
11	rs80804521	56,512,042	0.733	*SLITRK6*
15	rs80807225	55,051,435	0.686	
16	rs81462450	71,769,147	0.666	*GM2A*

SSC, porcine chromosome number

Table S3 List of significant single nucleotide polymorphisms (SNPs) identified using the F_{ST}-outlier method implemented in BayeScan. When markers were located within genes, these are indicated.

SSC	SNP	Position	*q*value	F_{ST}	Annotated genes
2	rs81361469	94,633,673	0.0029	0.287	-
5	rs81386405	95,401,083	0.0001	0.315	-
6	rs81307769	44,059,341	0.0052	0.283	*GPI*
15	rs81331202	58,833,925	0.0094	0.280	-
15	rs81453447	67,008,603	0.0075	0.278	*PLA2R1*

Table S4 Run of homozygosity (ROH) islands identified within the two pig population.

Pop	SSC	SNP	Start and end (bp)

Sicilian Grey	17	38	37,273,372–39,597,192
Sicilian Black	4	37	97,587,537–99,233,113

SSC, porcine chromosome number; SNP, single nucleotide polymorphisms.

Figure S1 Examples of Sicilian Black and Sicilian Grey pigs reared in the same flock.

Figure S2 Multidimensional scaling (MDS) plot of Sicilian Black (blu) and Sicilian Grey (red) individuals.

Figure S3 Manhattan plot of the *P*-values in the genome-wide association study (GWAS) analysis. The horizontal lines represent the genome-wide significance (red; $P < 0.01$) and suggestively significant (green; $P < 0.05$).