Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE E TECNOLOGIE AGRARIE, AMBIENTALI E ALIMENTARI

Ciclo XXXIV

Settore Concorsuale: 07/G1 SCIENZE E TECNOLOGIE ANIMALI

Settore Scientifico Disciplinare: AGR/17 ZOOTECNICA GENERALE E MIGLIORAMENTO GENETICO

GENOMIC CHARACTERISATION OF PIGMENTATION RELATED TRAITS IN LIVESTOCK

Presentata da: Giulia Moscatelli

Coordinatore Dottorato — Prof. Massimiliano Petracci

Supervisore — Prof. Luca Fontanesi

Esame finale anno 2022

Abstract

With the intense use of high throughput genomic technologies, like high-density single nucleotide polymorphism (SNP) chip and next generation sequencing platforms, our knowledge about the cattle and pig genomes has rapidly evolved. Completing the genome sequences of the two species lead researchers to find genome variants that could characterize animal genetic resources and autochthonous breeds and that could be exploited in selection and conservation programs. Over the last years, many investigations have been already carried out to identify major genes and mutations underlying different morphological, productive, and reproductive traits in cattle and pigs.

This thesis is the result of research activities focused on the investigation of genomic features to find novel candidate gene markers associated with pigmentation in two Italian local cattle breeds (Reggiana and Modenese) and in an Italian heavy pig breed, the Italian Large White breed.

In the first study that we proposed, we detected signatures of selection in the genome of these two autochthonous cattle breeds using genome-wide SNP information in comparative F_{ST} analyses. Generally, these breeds are mainly distinguished at phenotypic level due to their exterior traits, particularly focused on coat colour: solid red in Reggiana and solid white with pale shades of grey in Modenese. Results show top F_{ST} values detected for the *melanocortin 1 receptor* (*MC1R*) gene region on BTA18, including the causative mutation for the red coat colour of the Reggiana breed, and for the *agouti signalling protein* (*ASIP*) gene region on BTA13, which emerged as a strong candidate affecting the white coat colour in the Modenese breed.

The second aim of this thesis was to investigate the pigmentation process of the iris in the Italian Large White pig breed. This is a white-coloured breed not affected by albinism. For this aim, we carried out several genome-wide association studies using high density SNP datasets and designed to contrast groups of pigs with different colour of iris (pigs were grouped into six different categories based on their eye pigmentation phenotypes). The results indicated that the eye pigmented patterns (different grades of brown pigmentation), the total absence of pigmentation in the both eyes, and heterochromia iridis defect were associated with SNPs close to the *SLC45A2* (on chromosome 16, SSC16), *EDNRB* (SSC11) and *KITLG* (SSC5) genes, respectively. In addition, other associated genomic regions with eye depigmented patterns were also identified on two SSC4 regions (including two candidate genes: *NOTCH2* and *PREX2*) and on SSC6, SSC8 and SSC14 (including *COL17A1* as candidate gene).

This thesis demonstrates how population genomic approaches designed to take advantage from the diversity between livestock genetic resources could provide interesting hints to explain pigmentation related traits not yet completely investigated in these species. These results could also integrate knowledge on the genetics of pigmentation in mammals.

Table of contents

Chapter 1: ***Genomics: a way to preserve biodiversity in livestock*** *p.5*

Chapter 2: ***Cattle genomics*** *p.8*

2.1 Cattle evolution and domestication

2.2 The cattle genome and genomic approaches

2.3 Dairy cattle: difference between commercial and autochthonous cattle

2.4 Reggiana and Modenese: two local breeds used for Parmigiano Reggiano cheese

Chapter 3: ***Pig genomics*** *p.23*

3.1 The domestic pig

3.2 The pig genome

3.3 The pig as animal model

Chapter 4: ***Pigmentation*** *p.30*

4.1 Molecular and development processes affecting pigmentation in mammals

4.2 Major genes affecting pigmentation

4.3 White coat colour in livestock

Aim *p.42*

Chapter 5: ***Research activity on cattle*** *p.43*

5.1 Signatures of selection are present in the genome of two close autochthonous cattle breeds raised in the North of Italy and mainly distinguished for their coat colours

Chapter 6: ***Research activity on pig*** *p.77*

6.1 Genome-wide association studies for iris pigmentation and heterochromia in Large White pigs identified several genomic regions affecting eye colours

Conclusions *p.110*

Acknowledgments

Chapter 1: Genomics: a way to preserve biodiversity in livestock

The era of animal genetics has led to spectacular increases in productivity in all major livestock species during the second half of past century.
Microsatellites were an abundant source of highly polymorphic and well-dispersed typed markers that had catalyzed the generation of primary livestock species' maps, especially for cattle and pig (Barendse et al., 1994; Ellegren et al., 1994; Rohrer et al., 1996).
Molecular markers and high-throughput SNP genotyping platforms led to the discovery of Quantitative Trait Loci (QTLs) and/or genomic regions with casual polymorphisms associated with economically important traits (Kim et al., 2007; Garrick, 2017).
SNPs (single nucleotide polymorphisms) are less informative about polymorphisms than other markers like microsatellites, for that reason SNPs have been widely utilized in SNP platform that allow thousands to millions of SNPs to be detected in a single reaction-test.
The availability of a large number of SNPs is connected to the recent development of Next Generation Sequencing (NGS) technology, in addition, to these technologies were developed strong bionformatic-tools to analyze the data (Ramos et al., 2009).
The globalization increases the demand of food, led to an intensive system breeding and results in a genetic erosion. Genetic improvement programs in meat quality, muscularity, reproductive attitude and age at slaughter, start to be always more suitable to satisfy humans' needs (Gandini and Oldenbroek, 2007).
The differences between breeds will have developed through a combination of four evolutionary forces: genetic drift, migration, selection and mutation (Falconer and Mackay, 1996). All these forces create genetic diversity in the genome of the breeds with gain and loss of alleles or different alleles frequency.
Biodiversity generally decreases when the intensity of farming increases (Biodiversity Brief 10, 2001).
Cosmopolitan breeds are directional selected for specific traits and for this reason, they differ from autochthonous breeds. Cosmopolitan breeds could be considered less rich in genetic diversity compared to autochthonous one. Loss in biodiversity derived by the employment of a restricted number of sires for reproduction (based on their estimated breeding value, EBV) and the increasing fixation of some important alleles through time (Ojeda et al., 2008; Ramirez et al., 2009).
Most of the local breeds, nearly 30%, are now considered endangered or in danger for extinction (FAO, 2007).

Until today, the merit for the survival of these breeds, is to be recognized to local farmers that still see value in local resources and try to conserve them instituting organizations of breeders at national level (Gandini and Villa, 2003; Gandini and Oldenbroek, 2007).

The available genomic tools are not specific for autochthonous breeds that were not initially included in the design as it was for the cosmopolitan breeds. Nonetheless, it is possible to explore biodiversity in local breeds, due to great efforts that have been carried on in last years with high-throughput genomic tools and NGS technologies (Herrero-Medrano et al., 2013; Mészáros et al., 2015).

Results obtain from these technologies, will help to defense biodiversity and enhance local breeding and their derived products.

References

Barendse, W., Armitage, S.M., Kossarek, L.M., Shalom, A., Kirkpatrick, B.W., Ryan, A.M., Hetzel, D.J.S. 1994. A genetic linkage map of the bovine genome. Nat. Genet. 6(3), 227-235. doi: 10.1038/ng0394-227.

Biodiversity Brief 10, 2001. Biodiversity in Development. Livestock and biodiversity. European Commision.

Ellegren, H., Chowdhary, B. P., Johansson, M., Marklund, L., Fredholm, M., 1994. A primary linkage map of the porcine genome reveals a low rate of genetic recombination. Genetics. 137, 1089–1100. doi: 10.1093/genetics/137.4.1089.

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to Quantitative Genetics. 4th edition, Longman, Harlow, Essex, UK.

FAO (2007) The State of the Worlds Animal Genetic Resources for Food and Agriculture. FAO, Rome.

Gandini, G. and Oldenbroek, K. in Oldenbroek, K. (Ed.), 2007. Utilisation and conservation of farm animal genetic resources. Wageningen Academic Publishers, The Netherlands. doi: 10.3920/978-90-8686-592-5.

Gandini, G.C., Villa, E., 2003. Analysis of the cultural value of local livestock breeds: a methodology. J. Anim. Breed. Genet. 120, 1–11. doi:1 0.1046/j.1439-0388.2003.00365.

Garrick, D.J., 2017. The role of genomics in pig improvement. Anim. Prod. Sci. 57(12), 2360-2365. doi: 10.1071/AN17277.

Herrero-Medrano, J.M., Megens, H.J., Groenen, M.A. 2013. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. BMC Genet. 14, 106 doi: 10.1186/1471-2156-14-106.

Kim, S.C., Adesogan, A.T., Arthington, J.D., 2007. Optimizing nitrogen utilization in growing steers fed forage diets supplemented with dried citrus pulp. J. Anim. Sci. 85(10), 2548-2555. doi: 10.1523/JNEUROSCI.1167-07.2007.

Mészáros, G., Boison, S.A., O'Brien, A.M.P., Ferencakovic, M., Curik, I., Da Silva, M.V.B. 2015. Genomic analysis for managing small and endangered populations: a case study in Tyrol Grey cattle. Front Genet. 6, 173. doi: 10.3389/fgene.2015.00173.

Ojeda, A., Huang, L.S., Ren, J., Angiolillo, A., Cho, I.C., Soto, H., Lemús-Flores, C., Makuza, S.M., Folch, J.M., Pérez-Enciso, M., 2008. Selection in the making: a worldwide survey of haplotypic diversity around a causative mutation in porcine IGF2. Genetics. 178(3)1639-1652. doi: 10.1534/genetics.107.084269 .

Ramìrez, O., Ojeda, A., Tomas, A., Gallardo, D., Huang, L.S., Folch, J.M., 2009. Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. Mol. Biol. Evol. 26, 2061–2072. doi: 10.1093/molbev/msp118.

Ramos, A.M., Crooijmans, R.P.M.A., Affara, N.A., Amaral, A.J., Archibald, A.L., Beever, J.E., Bendixen, C., Churcher, C., Clark, R., Dehais, P., Hansen, M.S., Hedegaard, J., Hu, Z.-L., Kerstens, H.H., Law, A.S., Megens, H.-J., Milan, D., Nonneman, D.J., Rohrer, G.A., Rothschild, M.F., Smith, T.P.L., Schnabel, R.D., Van Tassell, C.P., Taylor, J.F., Wiedmann, R.T., Schook, L.B., Groenen, M.A.M., 2009. Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. PLoS One. 4, e6524. doi: 10.1371/journal.pone.0006524.

Rohrer, G.A., Alexander, L.J., Hu, Z., Smith, T.P., Keele, J.W., Beattie, C.W. 1996. A comprehensive map of the porcine genome. Genome Res. 6(5), 371-391. doi: 10.1101/gr.6.5.371.

Chapter 2: Cattle genomics

2.1 Cattle evolution and domestication

Domestication of animals was an essential step in human demographic and cultural development (Bruford et al., 2003; Groeneveld et al., 2010). During the subsequent history of livestock, the main evolutionary forces of mutation, selective breeding, adaptation, isolation and genetic drift have created an enormous diversity of local populations. This has culminated in the establishment of several well-defined breeds utilized for a number of reasons with varying levels of performance in recent decades. Artificial insemination, embryo transfer and application of Genomic Selection (GS) have facilitated the dissemination of genetic material (Van der Werf, 2013; Food and Agriculture Organization of the United Nations, FAO 2007).

Several bovine species have been domesticated although, practically, all cattle breeds derived from taurine cattle (*Bos taurus*) and zebu (*Bos indicus*) (Lenstra and Bradley, 1999; Lenstra et al., 2014; Ajmone-Marsan et al., 2010).

Taurine cattle (*Bos taurus*) and zebu (*Bos indicus*) both descend from the extinct wild ox or aurochs (*Bos primigenius*), which, during the Pleistocene and Holocene (10,000 years B.P.), ranged from the Atlantic to the Pacific coasts and from the northern tundra to India and Africa (Zeuner, 1963; Bradley et al., 1996). A deep bifurcation in bovine mitochondrial DNA (mtDNA) phylogeny has been described and is indicative of a pre-domestic divergence well in excess of 100,000 years between the two cattle taxa, *Bos indicus* and *Bos taurus* (Loftus et al., 1994).

In the past, all modern cattle were considered to have the same roots in captured aurochs from the primary domestication centers, however, this opinion was an artifact of the history of archaeology (Mannen et al., 1998).

Previous study of mitochondrial DNA (mtDNA) D-loop show the diversity among European, African, and Indian cattle, which suggests that multiple strains of ancestral aurochs were ancestors of modern day cattle, at geographically and temporally separate stages of the domestication process (Bradley et al., 1996; Mannen et al., 1998).

Archaeological data indicate that taurine cattle have been domesticated between 10,300–10,800 years ago in the Fertile Crescent, most probably on the western Turkish-Syrian border (Helmer et al., 2005; Vigne, 2011). MtDNA analysis initially supported separate domestications in Africa and eastern Asia.

MtDNA shows that *Bos taurus* has the high (T, T1,T2 and T3) haplotypes (Bradley et al., 1996; Mannen et al., 2004; Troy et al., 2001; Zeder et al., 2006). One of these, T3, which is the dominating haplogroup in European cattle has also been found in ancient cattle DNA from the European Neolithicum (Kuhn et al., 2005; Anderung et al., 2005). Northern and central European aurochs carry a different P haplotype (Bollongino et al., 2006; Edwards et al., 2007) and thus are excluded as maternal ancestors of European cattle (Ajmone-Marsan et al., 2010).
In contrast, southwestern Asian aurochs carried the mitochondria now found in taurine domestic cattle, which is consistent with the notion that taurine cattle originate from the Fertile Crescent. The auroch haplotype P mtDNA has also been found sporadically in domestic cattle, first in a late Neolithic sample and along the Danube.

2.2 The cattle genome and genomic approaches

In 1977, the invention of DNA sequencing started a new era of data generation in the field of genetics (Sanger et al., 1977).
Through sequencing, the order of DNA bases in an organism's genome could be established. Early versions of the technology were time intensive and confined to small sequences (Ewing and Green, 1998). Furthermore, at the beginning of these technologies, no specific methods for computational analysis to overlapping short sequence reads into longer contiguous sequences have been established (contigs).
In 1986, the Human Genome Project (HGP) was iniziated. The project developed new technologies and method in sequencing platforms and analysis algorithms to analyse the DNA sequence to achieve the goal to realize the first complete genome reference (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001).
This project inspired researchers that such techniques would be applied to mammalian livestock species.
In 2020 the bovine species was proposed as the first livestock species to be sequenced for its strong interest in the world food production (Gibbs et al., 2002).
As such a reference genome was released in 2009 (MacNeil, 2009) paving the way to the use of use of genomics in livestock selection as postulated by Meuwissen in 2001 (Meuwissen et al., 2001).
Cattle genome, as mammalian’s genome, presents fragmentation of the assembled sequence, due to the high presence of repetitive elements (Nagarajan and Pop, 2013; Bentley et al., 2008; Koren et al., 2012).

Many resources like microsatellite and genetic or radiation hybrid maps were been developed to validate the sequence of DNA in assembled contigs and to identify the chromosomes (Barendse et al., 1994).
The first public release of the cattle reference genome (Btau1.0) consisted of 795,212 contigs generated from 3X WGS sequence data released to NCBI's GenBank in 2004. Many other releases were followed later until when in the four release (Btau4.0) contigs were placed and oriented onto chromosome scaffold (Liu et al., 2009; Bovine Genome Sequencing and Analysis Consortium et al., 2009).
Simultaneously, researchers at the University of Maryland published the second iteration of their assembly of the cattle genome (UMD2) which used an alternative assembly approach, a graph-based approach, that improved assembly continuity (Zimin et al., 2009).
The UMD2 assembly was supplanted by the polished release of the UMD3.1 reference in 2010, which resulted in several gap closures and the placement of more contigs on chromosome scaffolds.
The Baylor College of Medicine released the Btau4.6.1 composite assembly in 2012, including substitutions of WGS contigs with the finished BAC sequence as well as several scaffolds predicted to originate from the Y chromosome. In the following years, the cattle genomics community would use the Btau4.0, UMD3.1, and Btau4.6.1 assemblies extensively for basic research and practical applications.
By the mid 2000s, several high- throughput genotyping technologies have been developed (Kwok, 2001) in humans, and of these genotyping technologies were suitable for use in cattle (Sachidanandam et al., 2001).
Thanks to the recent advances in high-throughput DNA sequencing, machine learning, and bioinformatics the identification of SNP markers have been easier than using microsatellites (Ajmone-Marsan et al., 1997; Heyen et al., 1997).
An initial panel of 26 highly informative SNP was released by USDA Meat Animal Research Center (MARC) scientists, for use as parentage markers. Seven of them were found to contain at least one SNP that match with criteria for a highly informative diallelic marker (Heaton et al., 2002).
Even though the costs were initially higher, SNP have marked advantages over microsatellites, such as lower mutation rates (Krawczak, 1999), more streamlined laboratory methodologies, and more refined data interpretation (Kruglyak, 1997). They are also more readily standardized than microsatellites (Fries and Durstewitz, 2001).

By 2009, the cost of SNP genotyping had declined enough that it was less than half the cost of microsatellite genotyping (Tokarska et al., 2009).
Many companies in the years that followed provided different panels of cattle genotyping chip, e.g., the Affymetrix 10K cattle genotyping chip (Affy 10K) was developed by Affymetrix to provide initial parentage verification and consisted of approximately 10,000 SNP marker sites (with an average spacing of 1.71 SNP per centimorgan) (Daetwyler et al., 2008). Of the markers included on this chip, 92% were identified from the WGS sequence data of 4 cattle breeds, Holstein, Angus, Limousin and Hereford), and 8% contributed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) were identified as being within genes (though not necessarily coding genes; http://tools.thermofisher.com/content/sfs/brochures/bovine10k_snp_datasheet.pdf).
Consequently, to the release by the Bovine HapMap Consortium of a dataset comprising WGS reads from dozens of individuals for 24 distinct breeds and 2 subspecies of cattle, a new cattle SNP chip was product: the BovineSNP50 (Matukumalli et al., 2009; Bovine Hap- Map Consortium, 2009).
The success of the BovineSNP50 was slightly over- shadowed by questions regarding its density of markers. There was substantial debate in the research community about whether a higher-density assay (>100,000 genetic markers) would greatly improve breeding value estimations (Meuwissen and Goddard, 2010) or not provide much benefit at all (VanRaden, et al., 2011) within dairy breeds of cattle.
In 2011, the BovineHD genotyping array was developed.
With the advance of NGS technologies, in 2012, it was possible to create long DNA sequence reads in addition with improvements in genome assembly algorithms.
Despite having read error rates approaching 17% false incorporation of insertions and deletions (Eid et al., 2009), longer reads were still found to be suitable for consensus overlap assembly methods (Koren et al., 2017) due to the random distribution of errors.
So, it was decided that was useful improving the cattle reference genome with long-read sequences. To made this was used the same Hereford animal used for the original draft genome, based on the idea, that existing genomics studies would be more easily integrated if the same animal or breed were used as the reference.
Thus, L1 Dominette was sequenced using long-read technologies and was create a reference assembly for cattle with minimal polishing (ARS-UCD1.2, accession number GCA_002263795.2; Rosen et al., 2020). This assembly consists of only 2,597 contigs and just 315 gaps in chromosome scaffolds. These methodologies allow to identifying and correcting

thousands of previously unconfirmed inversions of scaffolded contigs present in the UMD3.1 reference.

This resulted in a 98% decrease in artifact inversion calls compared with the UMD3.1 assembly. Furthermore, the X-chromosome pseudoautosomal region (PAR), which was noncontiguous in previous cattle reference assemblies, was resolved into a contiguous block in ARS-UCD1.2 (Rosen et al., 2020).

Using a strategy previously adopted (Koren et al., 2018), SNP marker probe sequences from the Affy 10K, BovineSNP50, and BovineHD SNP chips were aligned to the ARS-UCD1.2 reference. The proportion of discrepancies was highest in the Affy 10K array with 12% of markers with chromosome assignment predicted to exist on a different chromosome than originally assigned.

The BovineHD chip had the highest count of probe sequences that did not remap to ARS-UCD1.2, suggesting that these markers may target genomic sequence that was not present in Dominette's genome. Out of a combined set of 794,731 SNP probe sequences from all 3 chips, 3,107 markers map to a different chromosome in ARS-UCD1.2 and 1,043 SNP probe sequences did not align to the reference sequence at all. This represents a remapping rate of 99.87%, suggesting that ARS-UCD1.2 is still highly representative of legacy genotyping data in cattle.

Due to scaffolding errors, manual edits to contig order needed to be performed between scaffolds on chromosomes 7 and 10, as well as on chromosomes 21 and 27 (Rosen et al., 2020). Alignment of BovineSNP50 markers to ARS-UCD1.2 show 11 and 18 markers for the 7–10 and 21–27 chromosome pairs, respectively, that are still mapped on the wrong chromosomes.

In September 2019, the Council on Dairy Cattle Breeding reports that a total of 3,541,090 SNP genotypes have been collected from dairy cattle alone (https://queries.uscdcb.com/Genotype/cur_freq .html).

2.3 Dairy cattle: difference between cosmopolitan and autochthonous cattle

Genetic diversity and population structure, is important to study, not only for understanding the evolutionary history of breed origins, but also for providing crucial information for local biodiversity conservation and management (Boettcher et al., 2010; Williams et al., 2016).

The variability within and among livestock populations is the result of natural and artificial selection, genetic drift and admixture events that have contributed to shape the genetic

uniqueness and diversity of many different breeds (Andersson and Georges, 2004; Decker et al., 2014).
In many regions of the world, the amount of the demand changed breeds raised: to obtain high productions autochthonous cows were abandoned in favor of more productive cosmopolitan breeds (Pastorino et al., 2018).
Among livestock species, *Bos taurus* to have the greatest number of breeds at risk of extinction, resulting in a steady loss of genetic diversity. Indeed, FAO reported that the extinction of many local breeds occurred over the last 15 years (Ramljak et al., 2018; Scherf, 2000).
Autochthonous breeds are a major source of revenue for the entire regional economy, since they are fully adapted to the environment and carry important traits for livestock production (Battaglini et al., 2014).
In comparison to local breeds from other European nations, or, in general, other worldwide cow breeds, the genome-wide genetic diversity and population structure of Italian cattle breeds is still poorly investigated. Indeed, Mastrangelo et al., report that only a few Italian breeds have been characterized using medium-density single nucleotide polymorphism (SNP) arrays (Mastrangelo et al., 2014; Mastrangelo et al., 2018). With more then 30 officially recognized cattle breeds, i.e. Agerolese, Burlina, Cabannina, Calvana, Cinisara, Garfagnina, Modenese, Modicana, Mucca Pisana, Pezzata Rossa d'Oropa, Pontremolese, Reggiana, Sarda, Sardo-Bruna, Sardo-Modicana and Ottonese-Varzese, Italy can be considered as one of the most important centre of cattle diversity (Mastrangelo et al., 2018). Generally, a breed is acknowledged by the Italian Breeders Association of that breed. The Breeders Association can manage more than one breed, when these have low consistency, i.e. due to the low number of Modenese individuals, the Herd Book of this breed is managed by the National Association of Reggiana Cattle Breeders (ANABORARE). The National Association Breeders with an official genealogical register are responsible for the safeguard and preservation of the recorded animals. In addition, the National Breeders Association defines the goals of the breeding program and the mating strategies.
To date, the genetic variability and the relationships between commercial and autochthonous breeds at the genomic level have been investigated by several studies (Medugorac et al., 2009; Mastrangelo et al., 2014; Mastrangelo et al., 2018).
Commercial SNP genotyping tools have been recently developed for several species, including cattle, providing information from many polymorphic sites to conduct detailed characterizations of the genetic diversity and population structure (Matukumalli et al., 2009; Senczuk et al., 2020).

Multiplex SNP genotyping allows the simultaneous high-throughput interrogation of hundreds of thousands of loci with high measurement precision at a cost that enables large-scale studies. SNP genotyping technology has created a surge in the number of genome-wide association (GWA) studies (Matukumalli et al., 2009). In addition, GWAS conducted on local breeds might provide valuable insights into the genetic determinism of the production-related traits by capturing genetic variants that are no longer detectable in cosmopolitan breeds (Sorbolini et al., 2017).

Over the past centuries, milk production and composition were the main selection goals in dairy cattle breeding programs (Miglior et al., 2017).

Currently, ~95% of the high-yielding dairy cows raised in the main dairy producing regions around the globe are represented by only three breeds: Holstein (or Holstein-Friesian), Jersey, Brown Swiss, and their crosses (Brito et al., 2021). In the past years, until the establishment of the herd book, bulls of cosmopolitan breed (especially Holstein) were used to improve milk production levels and responsiveness to high-input production systems in autochthonous breeds.

Using bulls from cosmopolitan breeds led to a decrease in number of bulls from autochthonous one.

The low number of available animals lead to an increase of inbreeding. For some breeds their low genetic diversity is confirmed by high genomic inbreeding coefficients and a small value of contemporary N_e. Other factors that could led to the small cN_e inferred, are the geographic isolation of some farms and the reduced interest of breeders these kinds of animals. Generally, the main causes of small N_e in livestock is due to selection pressures and use of artificial insemination. In local cattle breeds, the small N_e is related to inbreeding and low genetic diversity (Mastrangelo et al., 2017).

In the other hand, comparing the Italian local breeds with the cosmopolitan breeds, the cosmopolitan one had lower genetic diversity indices and N_e: between them: Holstein breeds had the smallest Ne. In the case of this breed, the low Ne is not directly connected to a high inbreeding value. The results of the ROH analysis also confirmed the above findings, i.e. F_{ROH} values were high for the highly selected cosmopolitan breeds (Holstein, Jersey and Brown Swiss) (Senczuk et al., 2020).

2.4 Reggiana and Modenese: two local breeds used for Parmigiano Reggiano cheese

Milk from local breeds can be used to develop branded dairy products to exploit the uniqueness of breeds. Numerous examples exist of branded products from local breeds that have been successfully promoted and marketed. For example, is well known, the Parmigiano Reggiano cheese product with only milk of Reggiana is marked with two additional brands: “Vacche Rosse” and “ANaBoRaRe" (https://www.razzareggiana.it/).

This increasing interest in marketing mono-breed labelled lines of meat as well as diary products, has lead to the necessity of detailed characterization at genomic level of local breeds. Analysis of the DNA present in all animal products (including dairy products as the milk contains the somatic cells of the cow) could be used to trace back its origin to the individual animals and to infer their breed.

The Reggiana cattle genome was shown to still contain several signatures that are reminiscent of its past of un-specialized purpose (until the 1960s), before it was redirected towards a dairy specialization, partially through some early and mostly undeclared introgression events from other more productive dairy breeds (Bertolini et al., 2020).

Since 1992, milk of this breed is almost exclusively used to produce a mono-breed Parmigiano–Reggiano cheese, that is labelled with a brand name (*Vacche Rosse*) created to refer to the characteristic typical red coat colour of the Reggiana cattle (known as *fromentino*; Russo et al., 2007).

The branded cheese is sold at premium prices compared with standard Parmigiano Reggiano cheese and is revitalizing interest in Reggiana cows.

A few studies have been carried out to characterise this breed at the genetic level, and compared the breed with cosmopolitan one to find the main differences. Most of these investigated traits using a GWAS approach on, for example coat colour (like *MC1R* gene), and other exterior trais (Bovo et al., 2021), exterior traits were also studied with genome-wide analyses of signature of selection and population genomic parameters (Bertolini et al., 2015, 2018). Considerable effort was expended to find candidate gene markers used in association studies which compare the frequency of relevant alleles affecting milk production traits (Fontanesi et al., 2007; Scotti et al., 2010).

Mariani and Russo (1971) first investigated the frequency distribution of k-casein protein variants in the Reggiana population.

Milk from Reggiana cows has higher milk solids than milk from Holsteins; and milk from Reggiana cows is particularly well-suited for cheese production because of its high percentage of CN and superb properties for rennet coagulation (Mariani and Pecorari, 1987).
Reggiana milk, compared with Holstein, has dairy technological properties more suitable for Parmigiano Reggiano cheese production, including better fermentative aptitude and curds with more favorable cooking and syneresis properties (Gandini et al., 2007).
Considering the mean milk yield production, the Reggiana breed produces milk about 30% less than that of the Holstein breed (Gandini et al., 2007), which brought up an important concern for the conservation of the Reggiana breed related to profitability (Fontanesi et al., 2015).
Caroli et al. (2004) monitored the presence of polymorphisms in three caseins and in *β*-lactoglobulin by isoelectrofocusing on the milk of Reggiana cows.
Fontanesi et al. (2015) reported SNPs in several candidate genes affecting milk production traits, about the 22, the *DGAT1* p.K232A mutation was the most important.
Modenese milk is sold to local market or industry, linked to the production of Parmigiano Reggiano cheese Modenese branded.
The Modenese cow milk contains 0.4 percent more units of crude protein (3.48 *vs* 3.01 g/100g) and casein (2.75 *vs* 2.32 g/100g) compared to milk from the Italian Friesian. Casein is clearly higher proportion in Modenese milk (79.05 *vs* 76.92%). The Italian Friesian cow milk shows a casein number similar to the conventional value (77%), while the Modenese milk is 2 percent units higher (Mariani, 1975; Summer et al., 2002).
The percentage repartition of caseins of Modenese cow differ to that of Italian Friesian cow (Mariani et al., 1998), especially concerning the k-casein and α_{S2}-casein. Casein of Modenese cow contains more k-casein (12.28%) than that of Italian Friesian (11.25%).
In any case, the rennet-coagulation properties of Modenese milk are not technologically suitable for Parmigiano-Reggiano cheese production, even if it is better than those of the Italian Friesian: probably because of the lower content of colloidal calcium phosphate per casein unit (lower mineralisation degree of the micelle): the Modenese milk tends to form a curd of mealy type, less elastic and considerably different from those “gelatinous” typical of the milk of other breeds, as Italian Brown and Reggiana (Pecorari et al., 1987; Mariani and Battistotti, 1999).

References

Ajmone-Marsan, P., Garcia, J.F., Lenstra, J.A., 2010. On the origin of cattle: how aurochs became cattle and colonized the world. Evolutionary Anthropology: Issues, News, and Reviews, 19(4), 148- 157. doi: 10.1002/evan.20267.

Ajmone-Marsan, P., Vecchiotti-Antaldi, G., Bertoni, G., Valentini, A., Cassandro, M., Kuiper, M., 1997. AFLP™ markers for DNA fingerprinting in cattle. Anim. Genet. 28(6), 418-426. doi: 10.1111/j.1365-2052.1997.00204.x.

Andersson, L., Georges, M., 2004. Domestic animal genomics: deciphering the genetics of complex traits. Nat. Rev. Genet. 5, 202–212. doi: 10.1038/nrg1294.

Anderung, C., Bouwman, A., Persson, P., Carretero, J.M., Ortega, A.I., Elburg, R., Smith, C., Arsuaga, J.L., Ellegren, H., Gotherstrom, A., 2005. Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. Proc. Natl. Acad. Sci. USA. 102, 8431–8435. doi: 10.1073/pnas.0503396102.

Barendse, W., Armitage, S.M., Kossarek, L.M., Shalom, A., Kirkpatrick, B.W., Ryan, A.M., Hetzel, D.J.S., 1994. A genetic linkage map of the bovine genome. Nat. Genet. *6*(3), 227-235. doi: 10.1038/ng0394-227.

Battaglini, L., Bovolenta, S., Gusmeroli, F., Salvador, S., Sturaro, E., 2014. Environmental sustainability of Alpine livestock farms. Ital. J. Anim. Sci. 13,3155. doi:10.4081/ijas.2014.3155.

Bentley, D., Balasubramanian, S., Swerdlow, H. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59. doi: 10.1038/nature07517.

Bentley, D., Balasubramanian, S., Swerdlow, H., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 456, 53–59. doi: 10.1038/nature07517.

Bertolini, F., Galimberti, G., Calò, D.G., Schiavo, G., Matassino, D., Fontanesi, L., 2015. Combined use of principal component analysis and random forests identify population-informative single nucleotide polymorphisms: application in cattle breeds. J. Anim. Breed. Genet. 132(5), 346-356. doi:10.1111/jbg.12155.

Bertolini, F., Galimberti, G., Schiavo, G., Mastrangelo, S., Di Gerlando, R., Strillacci, M.G., Fontanesi, L., 2018. Preselection statistics and Random Forest classification identify population informative single nucleotide polymorphisms in cosmopolitan and autochthonous cattle breeds. Animal. 12(1), 12-19. doi:10.1017/S1751731117001355.

Bertolini, F., Schiavo, G., Bovo, S., Sardina, M.T., Mastrangelo, S., Dall’Olio, S., Fontanesi, L., 2020. Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano-Reggiano cheese production system. Animal. 14(5), 921-932. doi:10.1017/S1751731119003318.

Boettcher, P.J., Tixier-Boichard, M., Toro, M.A., Simianer, H., Eding, H., Gandini, G., 2010. Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. Anim. Genet. 41, 64–77; doi:10.1111/j.1365-2052.2010.02050.

Bollongino, R., Edwards, C.J., Alt, K.W., Burger, J., Bradley, D.G., 2006. Early history of European domestic cattle as revealed by ancient DNA. Biol. Lett. 2, 155–159. doi:10.1098/rsbl.2005.0404.

Bovine Genome Sequencing and Analysis Consortium, Elsik, C.G., Tellam, R.L., Worley, K. C., Gibbs, R.A., Muzny, D.M., Hitchens, M.E. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science. 324(5926), 522-528. doi: 10.1126/science.1169588.

Bovo, S., Schiavo, G., Kazemi, H., Moscatelli, G., Ribani, A., Ballan, M., Bonacini, M., Prandi, M., Dall'Olio, S., Fontanesi, L., 2021. Exploiting within-breed variability in the autochthonous Reggiana breed identified several candidate genes affecting pigmentation-related traits, stature and udder defects in cattle. Anim. Genet. 52,(5)579-597. doi:10.1111/age.13109.

Bradley, D.G., MacHugh, D.E., Cunningham, P., Loftus, R.T., 1996. Mitochondrial diversity and the origins of African and European cattle. Proc. Natl. Acad. Sci. U. S. A. 93(10), 5131-5135. doi:10.1073/pnas.93.10.5131

Bruford M.W., Bradley D.G., Luikart, G., 2003. DNA markers reveal the complexity of livestock domestication. Nature Reviews Genetics. 4, 900–10. doi: 10.1038/nrg1203.

Caroli, A., Chessa, S., Bolla, P., Budelli, E., Gandini, G.C., 2004. Genetic structure of milk protein polymorphisms and effects on milk production traits in a local dairy cattle. J. Anim. Breed. Genet. 121(2), 119-127. doi:10.1111/j.1439-0388.2003.00443.

Daetwyler, H.D., Villanueva, B., Woolliams, J.A., 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PloS one. 3(10), e3395. doi:10.1371/journal.pone.0003395.

de Brito, J., Kurda, R., 2021. The past and future of sustainable concrete: A critical review and new strategies on cement-based materials. J. Clean. Prod. 281, 123558. doi:10.1016/j.jclepro.2020.123558.

Decker, J.E., McKay, S.D., Rolf, M.M., Kim, J., Molina Alcalá, A., Sonstegard, T.S.,Hanotte, O., Götherström, A., Seabury, C.M., Praharani, L., Babar, M.E., Correia de Almeida Regitano, L., Yildiz, M.A., Heaton, M.P., Liu, W.S., Lei, C.Z., Reecy, J.M., Saif-Ur-Rehman, M., Schnabel, R.D., Taylor, J.F., 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLoS Genetics. 10,e1004254. doi:10.1371/journal.pgen.1004254.

Edwards, C.J., Bollongino, R., Scheu, A., Cham-berlain, A., Tresset, A., Vigne, J.D., Baird, J.F., Larson, G., Ho, S.Y.W., Heupink, T.H., 2007. Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. Proc. Biol. Sci. 274, 1377–1385. doi:10.1098/rspb.2007.0020.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Turner, S., 2009. Real-time DNA sequencing from single polymerase molecules. Science, 323(5910), 133-138. doi:10.1126/science.1162986.

Ewing, B., Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8(3), 186-194.

FAO (2007) The State of the Worlds Animal Genetic Resources for Food and Agriculture. FAO, Rome.

Fontanesi, L. 2017. Metabolomics and genomics in livestock: taking advantages from the next phenotyping frontiers in animal breeding and selection. In A new view of animal science: challenges and perspectives. Proceedings of the 54th Annual Meeting of the Brazilian Society of Animal Science, Foz do Iguaçu, Brazil, 24-28 July 2017 (pp. 177-184). Brazilian Society of Animal Science (Sociedade Brasileira de Zootecnia-SBZ) & Department of Animal Science, Universidade Federal do Paraná-UFPR.

Fontanesi, L., Scotti, E., Samorè, A.B., Bagnato, A., Russo, V., 2015. Association of 20 candidate gene markers with milk production and composition traits in sires of Reggiana

breed, a local dairy cattle population. Livest. Sci. 176, 14–21. doi:10.1016/j.livsci.2015.03.022.

Fries, R., Durstewitz, G., 2001. Digital DNA signatures for animal tagging. Nature Biotechnology. 19(6), 508-508.

Gandini, G., Maltecca, C., Pizzi, F., Bagnato, A., Rizzi, R., 2007. Comparing local and commercial breeds on functional traits and profitability: the case of Reggiana dairy cattle. Int. J. Dairy Sci. 90, 2004–2011. doi:10.3168/jds.2006-204.

Gibbs, G., 2002. Qualitative data analysis: Explorations with NVivo. Open University Press, Buckingham.

Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., Groeneveld, E., Weigend, S., 2010. Genetic diversity in farm animals – a review. Anim. Genet. 41, 6–31. doi:10.1111/j.1365-2052.2010.02038.

Heaton, M.P., Harhay, G.P., Bennett, G.L., Stone, R.T., Grosse, W.M., Casas, E., Laegreid, W. W., 2002. Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle. Mamm. Genome. 13(5), 272-281. doi:10.1007/s00335-001-2146-3.

Helmer, D., Gourichon, L., Monchot, H., Peters, J., Saña Segui, M., 2005. Identifying early domestic cattle from prepottery Neolithic sites on the middle Euphrates using sexual dimorphism. In *The First Steps of Animal Domestication*; Vigne, J.D., Peters, J., Helmer, D., Eds.; Oxbow Books: Oxford, UK, pp. 86–95.

Heyen, D.W., Beever, J.E., Da, Y., Everts, R.E., Green, C., Lewin, H.A., Ziegle, J.S., 1997. Exclusion probabilities of 22 bovine microsatellite markers in fluorescent multiplexes for semiautomated parentage testing. Anim. Genet. 28(1), 21-27. doi:10.1111/j.1365-2052.1997.t01-1-00057.

International Human Genome Mapping Consortium. 2001. A physical map of the human genome. Nature. 409(6822), 934-941.

Koren, I., 2018. Computer arithmetic algorithms. AK Peters/CRC Press.

Koren, S., Schatz, M., Walenz, B. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol 30, 693–700. doi: 10.1038/nbt.2280

Koren, S., Schatz, M., Walenz, B., 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat. Biotechnol. 30, 693–700. doi:10.1038/nbt.2280.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27(5), 722-736. doi:10.1101/gr.215087.116.

Krawczak, M., 1999. Informativity assessment for biallelic single nucleotide polymorphisms. ELECTROPHORESIS: An International Journal. 20(8), 1676-1681.

Kruglyak, L., 1997. The use of a genetic map of biallelic markers in linkage studies. Nat. Genet. 17(1), 21-24.

Kuhn, R., Ludt, C., Manhart, H., Peters, J., Neumair, E., Rottmann, O., 2005. Close genetic relationship of early neolithic cattle from Ziegelberg (Freising, Germany) with modern breeds. J. Anim. Breed. Genet. 122, 36 – 44. doi:10.1111/j.1439-0388.2005.00507.

Kwok, P.Y., 2001. Methods for genotyping single nucleotide polymorphisms. Annual review of genomics and human genetics. 2(1), 235-258. doi:10.1146/annurev.genom.2.1.235.

Lenstra, J.A., Ajmone-Marsan, P., Beja-Pereira, A., Bollongino, R., Bradley, D.G., Colli, L., Radoslavov, G.A., 2014. Meta-analysis of mitochondrial DNA reveals several population bottlenecks during worldwide migrations of cattle. Diversity. 6(1), 178-187. doi:10.3390/d6010178.

Lenstra, J.A., Bradley, D.G., 1999. Systematics and phylogeny of cattle. The genetics of cattle. 1-14.

Liu, L., Yu, L., Pearl, D.K., Edwards, S.V., 2009. Estimating Species Phylogenies Using Coalescence Times among Sequences. Systematic Biology. 58(5), 468–477. doi:10.1093/sysbio/syp031.

Mannen, H., Kohno, M., Nagata, Y., 2004. Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle. Molecular Phylogenetics and Evolution. 32, 539–44. doi: 10.1016/j.ympev.2004.01.010.

Mannen, H., Tsuji, S., Loftus, R.T., Bradley, D.G., 1998. Mitochondrial DNA Variation and Evolution of Japanese Black Cattle (Bos taurus). Genetics. 150, 3,(1),1169–1175. doi: 10.1093/genetics/150.3.1169.

Mariani, P. and Pecorari, M., 1987. Fattori genetici, attitudine alla caseificazione e resa del latte in formaggio. Scienza e Tecnica Lattiero-Casearia, 38, pp.286-326.

Mariani, P., 1975. Ripartizione delle proteine del latte nelle razze Frisona, Bruna alpina, Reggiana e Modenese. Riv. Zoot. Vet., 3, 13-35.

Mariani, P., Battistotti, B., 1999. Milk quality for cheesemaking. Proceedings XIII° ASPA Congress, 499-516.

Mariani, P., Russo, V., 1971. Distribuzione delle varianti genetiche delle caseine e della β-lattoglobulina nelle vacche di razza Reggiana. Riv. Zoot. 44: 310–322.

Mariani, P., Summer, A., Franchetti, M., Vecchia, P., Fossa, E., 1998. Ripartizione percentuale delle caseine in latti di massa delle vacche di razza Frisona, Bruna, Reggiana e Modenese. Sci. Tecn. Latt.-Cas., 49, 181-192.

Mastrangelo, D., Pelosi, E., Castelli, G., Lo-Coco, F., Testa, U., 2018. Mechanisms of anti-cancer effects of ascorbate: Cytotoxic activity and epigenetic modulation. Blood Cells Mol Dis. 69, 57-64. doi: 10.1016/j.bcmd.2017.09.005.

Mastrangelo, M.E., Weyland, F., Villarino, S.H., Barral, M.P., Nahuelhual, L., Laterra, P., 2014. Concepts and methods for landscape multifunctionality and a unifying framework based on ecosystem services. Landscape Ecology. 29(2), 345-358. doi:10.1007/s10980-013-9959-9.

Mastrangelo, S., Ciani, E., Ajmone Marsan, P., Bagnato, A., Battaglini, L., Bozzi, R., et al. (2018). Conservation status and historical relatedness of Italian cattle breeds. Genet. Sel. Evol. 50:35. doi: 10.1186/s12711-018-0406-x

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. Genet. 49(1), 1-10. doi:10.1186/s12711-017-0360.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PloS one. 4(4), e5350. doi:10.1371/journal.pone.0005350.

Medugorac, I., Medugorac, A., Russ, I., Veit-Kensch, C.E., Taberlet, P., Luntz, B., Foerster, M., 2009. Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. Mol Ecol. 18(16), 3394-3410. doi:10.1111/j.1365-294X.2009.04286.x

Meuwissen, T., Goddard, M., 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics. 185(2), 623-631. doi:10.1534/genetics.110.116590.

Meuwissen, T.H., Hayes, B.J., Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157(4):1819-29. doi: 10.1093/genetics/157.4.1819.

Miglior, F., Fleming, A., Malchiodi, F., Brito, L.F., Martin, P., Baes, C.F., 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. Int. J. Dairy Sci. 100(12), 10251-10271. doi: 0.3168/jds.2017-12968.

Nagarajan, N., Pop, M. 2013. Sequence assembly demystified. Nat Rev Genet. 14(3):157-67. doi: 10.1038/nrg3367.

Pastorino, G., Faustini, M., Vitali, F., Mazzola, S., Curone, G., 2018. What Are We Losing? Are the Personality Traits of Italian Autochthonous Cows Different from Those of Cosmopolitan Breeds? J. Adv. Vet. Anim. Res. *5*, 315. doi:10.5455/javar.2018.e282.

Pecorari, M., Sandri, S., Mariani, P., 1987. Attitudine alla coagulazione dei latti delle razze Frisona, Bruna, Reggiana e Modenese. Sci. Tecn. Latt.-Cas. 38, 376-384.

Ramljak, J., Bunevski, G., Bytyqi, H., Marković, B., Brka, M., Ivanković, A., 2018. Conservation of a domestic metapopulation structured into related and partly admixed strains. Mol Ecol. 27:1633–50. doi: 10.1111/mec.14555.

Rosen, B.D., Bickhart, D.M., Schnabel, R.D., Koren, S., Elsik, C.G., Tseng, E., ..., Medrano, J.F., 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience. 9(3), giaa021. doi:10.1093/gigascience/giaa021.

Russo, V., Fontanesi, L., Scotti, E., Tazzoli, M., Dall'Olio, S., Davoli, R., 2007. Analysis of melanocortin 1 receptor (MC1R) gene polymorphisms in some cattle breeds: Their usefulness and application for breed traceability and authentication of Parmigiano Reggiano cheese. Ital. J. Anim. Sci. 6, 257–272. doi:10.4081/ijas.2007.257.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Altshuler, D., 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 409(6822), 928-934.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74(12), 5463-5467. doi:10.1073/pnas.74.12.5463.

Scherf, B.D., 2000. World watch list for domestic animal diversity. 3rd ed. Rome: Food and Agriculture Organization (FAO).

Scotti, E., Fontanesi, L., Schiavini, F., La Mattina, V., Bagnato, A., Russo, V., 2010. DGAT1 p. K232A polymorphism in dairy and dual purpose Italian cattle breeds. Ital. J. Anim. Sci. 9, e16. doi:10.4081/ijas.2010.e16.

Senczuk, G., Mastrangelo, S., Ciani, E., Battaglini, L., Cendron, F., Ciampolini, R., Cassandro, M. 2020. The genetic heritage of Alpine local cattle breeds using genomic SNP data. Genet. 52(1), 1-12. doi:10.1186/s12711-020-00559-1.

Snelling, W.M., Chiu, R., Schein, J.E., Hobbs, M., Abbey, C.A., Adelson, D.L., Zhao, S., 2007. A physical map of the bovine genome. Genome Biology. 8(8), 1-17. doi: 10.1186/gb-2007-8-8-r165

Sorbolini, S., Bongiorni, S., Cellesi, M., Gaspa, G., Dimauro, C., Valentini, A., Macciotta, N.P.P., 2017. Genome wide association study on beef production traits in Marchigiana cattle breed. J. Anim. Breed. Genet. 134(1), 43-48. doi:10.1111/jbg.12227.

Summer, A., Malacarne, M., Martuzzi, F., Mariani, P., 2002. Structural and func-tional characteristics of Modenese cow milk in Parmigiano-Reggiano cheeseproduction. Annali Facolta`Medicina Veterinaria, Università Parma, 23, 163–174.

Tokarska, M., Marshall, T., Kowalczyk, R., 2009. Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. Heredity. 103, 326–332. doi:10.1038/hdy.2009.73.

Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C., Bradley, D.G., 2001. Genetic evidence for Near-Eastern origins of European cattle. Nature. 410, 1088–1091. doi:10.1038/35074088.

Van der Werf, J., 2013. Genomic selection in animal breeding programs. In Genome-Wide Association Studies and Genomic Prediction. Humana Press, Totowa, NJ. 543-561. doi:10.3389/fgene.2019.01389.

VanRaden, P.M., Olson, K.M., Null, D.J., Hutchison, J.L., 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. Int. J. Dairy Sci.. 94(12), 6153-6161. doi:10.3168/jds.2011-4624.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., et al., 2001. The sequence of the human genome. Science. 291(5507):1304-51. doi:10.1126/science.1058040.

Vigne, J.D., 2011. The origins of animal domestication and husbandry: A major change in the history of humanity and the biosphere. C. R. Biol. 334, 171–181. doi: 10.1016/j.crvi.2010.12.009.

Williams, J.L., Hall, S.J., Del Corvo, M., Ballingall, K.T., Colli, L., Ajmone-Marsan, P., 2016. Inbreeding and purging at the genomic level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. Anim. Genet. 47,19–27. doi: 10.1111/age.12376.

Zeder, M.A., Emshwiller, E., Smith, B.D., Bradley, D.G., 2006. Documenting domestication: the intersection of genetics and archaeology. Trends. Genet. 22, 139–155. doi:10.1016/J.TIG.2006.01.007.

Zeuner, F.E., 1963. A history of domesticated animals. London: Hutchinson.

Zimin, A.V., Delcher, A.L., Florea, L., 2009. A whole-genome assembly of the domestic cow, Bos taurus.Genome. Biol. 10, R42. doi:10.1186/gb-2009-10-4-r42.

Chapter 3: Pig genomics

3.1 The domestic pig

The domestic pig (*Sus scrofa domesticus*) is a member of the Suidae family and was one of the first animal species in the world to be domesticated. The domestication process can be defined as a long-term mutualistic interaction between humans and animal species that benefits both parties. (Zeder et al., 2006). This evolutionary process resulted in modifications in animal genetic, morphophysiological (for example, brain size and skull, teeth shape and increased growth and prolificacy), and behavioural (for example, decreased aggression, stress and watchfulness) characteristics, in order to meet human needs (Ramos-Onsins et al., 2014). In other words, pig domestication was a genetic adaptation to better satisfy and meet the human needs involving allelic frequency changes (Wang et al., 2017). There is a general consensus that the *Sus scrofa* and other sister species, emerged in Southeast Asia in the early Pliocene, approximately 5.3–3.5 Mya (Larson et al., 2007a, 2011; Frantz et al., 2013), and approximately 0.8 Mya reaching Europe (Frantz et al., 2013).

Genetic analyses showing a much higher level of diversity in *Sus scrofa* populations from Asia than in those from Europe (Larson et al., 2005; Ramìrez et al., 2009). There is genetic divergence between Near East and European *Sus scrofa* as demonstrated by Manunza et al. (2013).

The migration of wild boars to Europe, was likely followed by a long period of geographic isolation, roughly corresponding to the Calabrian stage, which resulted in the formation of two distinct *Sus scrofa* gene pools, one representing eastern populations (Asia) and the other gathering western pig groups (Europe, Near East and North Africa) (Ramos-Onsins et al., 2014).

In a pioneering study, Giuffra et al. (2000) sequenced several mitochondrial and nuclear loci and concluded that European and Asian pigs diverged long before domestication, 500,000 years ago (YBP). Is possible to identify two distinct Asian and European lineages that probably diverged during the mid-Pleistocene, around 0.8-1 Mya, between wild board and domestic pigs (Giuffra et al., 2000; Fang and Andersson, 2006; Groenen et al., 2012). According to the most ancient archaeological evidence of pig found in Anatolia, the process of domestication from *Sus scrofa* wild boars started approximately 9,000 years ago with two main domestication events occurring independently in multiple regions of the world (Giuffra et al., 2000; Larson et

al., 2005). European pigs and wild boars share mitochondrial haplotypes (Larson et al., 2005), a feature that would suggest that Europe was a primary domestication centre for pigs.
China, is considered the main domestication centre with multiple native breeds starting their domestication process around 8,000 years ago, for what regarding the pig domestication in Asia. The largest distinction between the two clades can be traced back to 4,500–7,000 years ago, with the two populations reporting differing population sizes, indicating primary and separate domestication in Europe and China, respectively (Wang et al., 2017).

3.2 The pig genome

Over the past decade, tremendous progress has been made in mapping and characterizing the swine genome.
The porcine research community has a long history in quantitative genetics, and more recently in genomics research.
Sequencing of the pig genome was initiated with the establishment of the Swine Genome Sequencing Consortium (SGSC) in September 2003 (Schook et al., 2005), following the successful generation of genetic and physical (Raudsepp et al., 2011) maps for the pig (Humphray et al., 2007a ; Humphray et al., 2007b; Archibald et al., 2010).
The porcine BAC map was mad with two libraries (RPCI-44 and CHORI-242) made by Pieter J. de Jong, after that data has been merged (Rogel-Gaillard et al., 1999). These BAC resources have facilitated the production of high-resolution physical maps in specific chromosomal regions (Rogel-Gaillard et al., 1999) and support the construction of sequence-ready mapping resources for the porcine genome.
These efforts resulted in the assembly and publication of a draft reference genome sequence of *Sus scrofa* in 2012 (Groenen et al., 2012).
The pig genome is of similar size, complexity and chromosomal organization (2n = 38, including meta and acrocentric chromosomes) as the human genome, and is approximately 2.6 GB.
Currently the last version release Sscrofa11.1 (GCA_000003025.6) was made using the use of Pacific Biosciences (PacBio) long read technology. The assembly, Sscrofa11.1, had a final contig N50 of 48.2 Mb, only 103 gaps in the sequences assigned to chromosomes, and only 583 remaining unplaced contigs (Warr et al., 2020).

The previously published draft pig reference genome sequence (Sscrofa10.2), has a number of significant deficiencies: about 10% of the pig genome, including some important genes, were not represented (i.e. CD163), or incompletely represented (i.e., IGF2) in the assembly (Robert et al., 2014).

Later, in addition to this reference genome sequence, which was derived from a female Duroc pig, were sequence another 48 pigs; since then, the genomes of many individual pigs have been re-sequenced, and currently around 350 complete genomes are publicly available.

To date, >3000 mapped loci are catalogued for the pig genome (http://www.thearkdb.org), using different methodologies and snip chip panels of different density.

The need for more genetic markers is also supported by the extent of linkage disequilibrium (LD) in the pig genome, which has been estimated to extend from as little as 40– 60 kb up to 400 kb in the commonly used commercial pig breeds, such as Duroc, Landrace and Large White (Jungerius et al., 2005; Amaral et al., 2008).

A major limitation to the development of highly parallel genotyping assays for swine is a lack of suitable SNPs for genotyping. To date there are a little over 8,400 SNPs for swine in dbSNP, but many of these are clustered into a small number of sequences that do not effectively cover the genome (Wieldmann et al., 2008).

One of the first SNP panel commercially available on the market, is the PorcineSNP60 BeadChip v2 (Illumina, San Diego, CA), that contains about 60K SNPs that cover all autosomal and X chromosomes (Steemers, 2007).

Ramos et al. (2009) conducted a study to evaluate the efficiency of the panels genotyping 158 individuals from five pools (include the four main breeds used in worldwide pig production (Durock, Pietrain, Landrace and Large White, as well as the wild boar); finding that the PorcineSNP60 Beadchip will be highly efficient for genomic selection with 97.5% animals genotyped successfully, while the 2.5% could not be reliably genotyped. Of the 62,621 validated loci, 58,994 were polymorphic which indicates that the SNP conversion success rate was 94%.

A commercial LD SNP chip was developed by GeneSeek/Neogen (Lincoln, NE) to face the need of the market (GeneSeek/Neogen GPP-Porcine LD Illumina Bead Chip panel). GeneSeek/Neogen prepared also a higher density SNP panel including about 70K SNPs.

Recently Affimetrix (Santa Clara, CA) release a HD SNP panel, containing ~650,000 SNPs in which are included all SNPs of the Illumina PorcineSNP60 BeadChip v2 array.

3.3 The pig as animal model

Domestic pigs also serve an important role in biomedical research and translational medicine. An animal model is a non-human species choose to investigate humans disease or infection because it can simulate aspect of the disease (https://www.genome.gov/genetics-glossary/Animal-Model). Several studies have been conducted, and more are being conducted, to examine the pig as an animal model for human illnesses. This is feasible because pigs and humans have extremely similar physiologies, anatomy and genetics (Meurens et al., 2012).
Thanks to the large number of breeds with different size or features, is very easy find breeds of pigs that allow various surgical and non-surgical procedures typically used in human medicine In addition to the farm breeds, many studies involved also miniature breeds, which develop adult human-sized organs between 6 and 8 months of age. Miniature are largely used thanks to their growth rate and size at sexual maturity, their size and the reduction in amounts of drug needed for testing (Swindle et al., 2012).
In general, examining animal models of human disease aids scientists in better understanding the mechanisms involved in disease pathogenesis and, as a result, gives tools for the development of gene therapy to treat the disease/condition in people.
Comparative maps have indicated that the porcine and human genomes are more similarly organized than when either is compared to the mouse.
During the Plant and Animal Genome meeting, it was reported that a 1.0 Mb human – pig comparative map has been completed (Meyers et al., 2005).
For genomics, it is an asset that the pig genome has high sequence and chromosome structure homology with humans.
Some studies already employ genomics approaches, such as the heart, transplantation and melanoma models; others are still in the early stages; and studies on atherosclerosis and diabetes largely used swine physiological parameters and utility as a human biomedical model. In addition, they share disease-causing mutations in genes that cause serious diseases like Parkinson's or Alzheimer's disease, as well as multifactorial characteristics like obesity and diabetes (Lunney, 2007; Groenen et al., 2012).
Swine skin studies have been very important. Pig and human skin are very similar, among similarities the most important is that the skin is tightly attached to the subcutaneous connective tissue in both species (Sullivan et al., 2001). Pig was used to studies several diseases in human skin as wound healing, burns, radiation and UVB impact, and dermatitis (Ansell et al., 2012; Sheu et al., 2014; Smirnova et al., 2014).

In particular, the swine melanoma model has been particularly informative. The swine melanoma model is a well established spontaneous melanoma model and one of the best developed for genomic approaches (Perez et al., 2002; Apiou et al., 2004).

Each model will be impacted by the availability of the functional genomic tools and swine genome sequence and maps outlined (Tuggle et al., 2007; Rothschild et al., 2007).

References

Amaral, A.J., Megens, H.J., Crooijmans, R.P., Heuven, H.C., Groenen, M.A., 2008. Linkage disequilibrium decay and haplotype block structure in the pig. Genetics. 179(1), 569–579. doi: 10.1534/genetics.107.084277.

Ansell, D. M., Holden, K. A., Hardman, M. J. 2012. Animal models of wound repair: Are they cutting it?. Experimental dermatology. 21(8), 581-585. doi: 10.1111/j.1600-0625.2012.01540.

Apiou, F., Vincent-Naulleau, S., Spatz, A., Vielh, P., Geffrotin, C., Frelat, G., Dutrillaux, B., Le Chalony, C., 2004. Comparative genomic hybridization analysis of hereditary swine cutaneous melanoma revealed loss of the swine 13q36-49 chromosomal region in the nodular melanoma subtype. Int J Cancer. 110, 232-8. doi: 10.1002/ijc.20108.

Archibald, A.L., Bolund, L., Churcher, C. 2010. Pig genome sequence - analysis and publication strategy. BMC Genomics 11, 438. doi: 10.1186/1471-2164-11-438.

Archibald, A.L., Bolund, L., Churcher, C., Fredholm, M., Groenen, M.A., Harlizius, B., 2010. Pig genome sequence–analysis and publication strategy. BMC Genomics. 11:438. doi: 10.1186/1471-2164-11-438.

Debeer, S., Le Luduec, J. B., Kaiserlian, D., Laurent, P., Nicolas, J. F., Dubois, B., Kanitakis, J. 2013. Comparative histology and immunohistochemistry of porcine versus human skin. European journal of dermatology. 23(4), 456-466. doi: 10.1684/ejd.2013.2060.

Fang, M., Andersson, L., 2006. Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. Proc. R. Soc. B Biol. Sci. 273, 1803–1810. doi: 10.1098/rspb.2006.3514.

Frantz, L.A., Schraiber, J.G., Madsen, O., Megens, H.J., Bosse, M., Paudel, Y., 2013. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. Genome Biol. 14: R107. doi: 10.1186/gb-2013-14-9-r107.

Giuffra, E., Kijas, J.M., Amarger, V., Carlborg, O., Jeon, J.T., Andersson, L., 2000. The origin of the domestic pig: independent domestication and subsequent introgression. Genetics. 154, 1785–91. doi: 10.1093/genetics/154.4.1785.

Groenen, M.A., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 491, 393–8. doi: 10.1038/nature11622.

Groenen, M.A.M., Schook, L.B., Archibald, A.L., 2011. Pig genomics. In: Rothschild MF, Ruvinsky A, editors. The genetics of the pig. 2nd ed. Wallingford: CABI. p. 179–99.

Humphray, S.J., Scott, C.E., Clark, R. 2007a. A high utility integrated map of the pig genome. Genome Biol 8, R139. doi:10.1186/gb-2007-8-7-r139.

Humphray, S.J., Scott, C.E., Clark, R., Marron, B., Bender, C., Camm, N., 2007b. A high utility integrated map of the pig genome. Genome Biol. 8, R139. doi: 10.1186/gb-2007-8-7-r139.

Jungerius, B.J., Gu, J., Crooijmans, R.P.M.A., Poel, J.J., Groenen, M.A.M., 2005. Estimation of the extent of linkage disequilibrium in seven regions of the porcine genome. Anim Biotechnol. 16, 41–54. doi: 10.1081/ABIO-200053402.

Kong, R., Bhargava, R. 2011. Characterization of porcine skin as a model for human skin studies using infrared spectroscopic imaging. Analyst. 136(11), 2359-2366. doi: 10.1039/C1AN15111H

Larson, G., Cucchi, T., Dobney, K., 2011. Genetic aspects of pig domestication. In: Rothschild MF, Ruvinsky A (eds) The Genetics of the Pig. CABI Publishing: Oxfordshire, pp 14–37.

Larson, G., Cucchi, T., Fujita, M., Matisoo-Smith, E., Robins, J., Anderson, A., 2007a. Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in Island Southeast Asia and Oceania. Proc. Natl. Acad. Sci. USA.. 104, 4834–4839. doi: 10.1073/pnas.0607753104.

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science. 307, 1618–1621. doi: 10.1126/science.1106927.

Lunney, J.K., 2007. Advances in swine biomedical model genomics. Int. J. Biol. Sci. 3, 179–84. doi: 10.7150/ijbs.3.179.

Manunza, A., Zidi, A., Yeghoyan, S., Balteanu, V.A., Carsai, T.C., Scherbakov, O., Ramírez, O., Eghbalsaied, S., Castelló, A., Mercadé, A., Amills, M., 2013. A high throughput genotyping approach reveals distinctive autosomal genetic signatures for European and Near Eastern wild boar. PLoS One 8, e55891. doi: 10.1371/journal.pone.0055891.

Meurens, F., Summerfield, A., Nauwynck, H., Saif, L., Gerdts, V. 2012. The pig: a model for human infectious diseases. Trends in microbiology, 20(1), 50-57. doi: 10.1016/j.tim.2011.11.002

Myers, S., Bottolo, L., Freeman, C., McVean, G., Donnelly, P., 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science. 310(5746), 321-324. doi: 10.1126/science.1117196.

Perez, J., Garcia, P.M., Bautista, M.J., Millan, Y., Ordas, J., Martin de las Mulas, J., 2002. Immunohistochemical characterization of tumor cells and inflammatory infiltrate associated with cutaneous melanocytic tumors of Duroc and Iberian swine. Vet Pathol. 39: 445-51. doi: 10.1354/vp.39-4-445.

Ramìrez, O., Ojeda, A., Tomas, A., Gallardo, D., Huang, L.S., Folch, J.M., 2009. Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. Mol Biol Evol 26: 2061–2072. doi: 10.1093/molbev/msp118.

Ramos-Onsins, S.E., Burgos-Paz, W., Manunza, A., Amills, M., 2014. Mining the pig genome to investigate the domestication process. Heredity (Edinb). 113, 471–484. doi: 10.1038/hdy.2014.68.

Raudsepp, T., Chowdhary, B. P., 2011. Cytogenetics and chromosome maps. The genetics of the pig. 2nd ed. Wallingford: CABI, 134-78.

Robert, C., Fuentes-Utrilla, P., Troup, K., Loecherbach, J., Turner, F., Talbot, R., 2014. Design and development of exome capture sequencing for the domestic pig (Sus scrofa). BMC Genomics, 15,550. doi: 10.1186/1471-2164-15-550.

Rogel-Gaillard, C., Bourgeaux, N., Billault, A., Vaiman, M., Chardon, P., 1999. Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements. Cytogenet Cell Genet, 85(3-4), 205-211. doi: 10.1159/000015294.

Rothschild, M.F., Hu, Z.L., Jiang, Z., 2007. Advances in QTL Mapping in Pigs. Int J Biol Sci. 3, 192-197. doi: 10.7150/ijbs.3.192.

Schook, L.B., Beever, J.E., Rogers, J., Humphray, S., Archibald, A., Chardon, P., 2005. Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome. Comp. Funct. Genomics. 6, 251–5. doi: 10.1002/cfg.479.

Steemers F.J., Gunderson K.L., 2007. Whole genome genotyping technologies on the BeadArray platform. Biotechnol. J 2(1): 41–9. doi: 10.1002/biot.200600213.

Summerfield, A., Meurens, F., & Ricklin, M. E. 2015. The immunology of the porcine skin and its value as a model for human skin. Molecular immunology. 66(1), 14-21. doi:10.1016/j.molimm.2014.10.023

Swindle, M. M., Makin, A., Herron, A. J., Clubb Jr, F. J., Frazier, K. S. 2012. Swine as models in biomedical research and toxicology testing. Veterinary pathology, 49(2), 344-356. doi: 10.1177/0300985811402846.

Tuggle, C.K., Wang, Y.F., Couture, O., 2007. Advances in Swine Transcriptomics. Int J Biol Sci. 3, 132-152. doi: 10.7150/ijbs.3.132.

Wang, J., Zou, H., Chen, L., Long, X., Lan, J., Liu, W., Ma, L., Wang, C., Xu, X., Ren, L., Gu, Z., Li, N., Hu, X., Zhao, Y., Zhao, Y., 2017. Convergent and divergent genetic changes in the genome of Chinese and European pigs. Sci. Rep. 7, 8662. doi: 10.1038/s41598-017-09061.

Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D. M., Billis, K., Archibald, A. L., 2020. An improved pig reference genome sequence to enable pig genetics and genomics research. GigaScience. 9(6), giaa051. doi: 10.1093/gigascience/giaa051.

Wiedmann, S., Fischer, M., Koehler, M., Neureuther, K., Riegger, G., Doering, A., Baessler, A., 2008. Genetic variants within the LPIN1 gene, encoding lipin, are influencing phenotypes of the metabolic syndrome in humans. Diabetes. 57(1), 209-217. doi: 10.2337/db07-0083.

Chapter 4: Pigmentation

4.1 Molecular and development process affecting pigmentation in mammals

Melanin is the key pigment responsible for the diverse pigmentation found in animal and human skin, hair and eyes (Cappai et al., 2015; Yamaguchi et al., 2007). The synthesis, storage and transportation of melanin (melanogenesis) occurs in melanocytes through cascade of biochemical and enzymatic reactions (Ali et al., 2015, Ali and Naaz, 2015). Melanocytes are unique cells that produce melanosomes (Raposo and Marks, 2007). Melanosome biogenesis is categorized into four developmental stages (I-IV) according to their degree of maturation.

Stage I and II comprise immature, unmelanized premelanosomes called also early melanosomes: in this stages intraluminal fibrils begin to form in amorphous spherical stage I melanosomes and generate a network characteristic of stages II melanosomes. In these early stages structural proteins, as well as melanin synthesizing enzymes, such as tyrosinase (TYR), tyrosinase-related protein 1 (TYRP1), and tyrosinase-related protein 2 (TYRP2) are transported from other organelles to immature, non-pigmented stage II melanosomes. Melanin deposition begins at stage III where the melanins are deposited uniformly on the internal fibrils resulting in the production of the stage III melanosomes. The organelle is fully melanized by mature stage IV melanosomes (Wasmeier et al., 2008; Huanga et al., 2013; Marks and Seabra, 2001; Hearing, 2005; Chi et al., 2006). Melanosomes, with a exocytosis process, are transferred into the hair during their growing up. During embryo development, melanocytes, starting from the neural crest, migrate in the body, conferring pigmentation on the areas where they are present. White spots or white patterns of some breeds is given by melanocytes missing (Fontanesi and Russo, 2013).

Melanosomes are classified as lysosome-related organelles (LROs) and recent studies characterizing the proteomes of early melanosomes show that they are derived from the endoplasmic reticulum (ER), coated vesicles, lysosomes and endosomes (Chi et al., 2006; Kushimoto et al., 2001; Basrur et al., 2003; Dell'Angelica et al., 2000).

In mammals, melanogenesis is catalyzed by mainly three enzymatic components of melanosomes include i.e., tyrosinase (TYR), a critical copper-dependent enzyme required for melanin synthesis, tyrosinase-related protein 1 (TYRP1) or gp75 and TYRP2 or dopachrome tautomerase (DCT). Those three enzymes cooperate to synthesize two distinct types of melanins: black-brown eumelanins and yellow-reddish pheomelanins (Costin et al., 2005) .

During melanogenesis, mixtures of eumelanin as well as pheomelanin have been produced at different ratio. The ratio is decided by tyrosinase activity and the substrate concentration of tyrosine and sulfhydryl group (Simon et al., 2009).
Both eumelanin and pheomelanin are derived from the common precursor dopaquinone, which is formed following the oxidation of tyrosine by tyrosinase. Tyrosinase is considered as the rate-limiting enzyme of melanin synthesis that catalyzes the two important reactions in the biosynthetic pathway; the first rate-limiting step of hydroxylation of L-tyrosine to 3,4-dihydroxyphenylalanine (L-DOPA) and subsequent reaction of the oxidation to L-dopaquinone.
From dopaquinone, the eumelanin and pheomelanin pathways diverge: one pathway leading to eumelanogenesis while other leading to pheomelanogenesis.
Sulfhydryl groups such as L-cysteine or glutathione play crucial role in pheomelanogenesis. In the presence of sufficient concentration of sulfhydryl compound, L-dopaquinone immediately reacts with it to form 5-S-cysteinyldopa or 5-S-glutathionyldopa and quinones which are then further converted into benzothiazine afterward into benzothiazole. These products subsequently undergo oxidative polymerization resulted in the formation of pheomelanin. Instead, in the absence of thiol compounds (like cysteine or gluthathione) dopaquinone undergoes intramolecular cyclization, leading eventually to the formation of eumelanin (Marks and Seabra, 2001; Palumbo et al., 1994; Chedekel et al., 1992).

4.2 Major genes affecting pigmentation

The genetic structure of phenotypic diversity in domestic and wild animals, the characterisation of different breeds, adaptation and evolutionary processes associated with domestication, as well as studying the genetic architecture of complex traits, presenting the close relationship between phenotype and genotype, can all be studied using coat colour variation in domestic animals as a model (Cieslak et al., 2011; Protas and Patel, 2008).
In addition, coat colour associated genes have been shown to be the most important candidates for breeds traceability and authentication (Russo et al., 2007).
The study of a coat colour genes of mammals can be traced back to rediscovery of Mendel's law in 1900.
The pigmentation pattern is basically regulated by the two main melanins including eumelanin (black/brown) and phaeomelanin (yellow/red), and their related controlling genetic loci encompass the *Extension (E)* and *Agouti (A)*, respectively (Searle, 1968).

At least eight different genes and their related alleles have been detected to control the coat colour patterns in livestock species consisting of *MC1R* (extension - *E*), *ASIP* (agouti - *A*), *TYR* (albinism - *C*), *TYRP1* (brown - *B*), *KIT* (colour sideness, dominant white), *KITLG* (roan - *R*), *MITF* (white spotting) and *PMEL* (dilution - *D*).

In cattle, the *MC1R* gene has been the subject of several studies with the aim to elucidate the biology of coat colour and, in other species, like pig, is probably the better characterized coat colour gene.

The *Extension (E)* locus codes for the melanocyte stimulating hormone receptor (MSHR) or melanocortin 1 receptor (MC1R) that is expressed in melanocytes.

The *MC1R* gene is located on BTA18 and includes three main alleles (Klungland et al., 1995; Seo et al., 2007).

MC1R is a member of the superfamily of G-protein-coupled receptors consisting of seven transmembrane domains. The *E locus* encodes the melanocortin 1 receptor , which binds to melanocyte-stimulating hormone (*α-MSH*), which induces eumelanin synthesis (Chhajlaniet al., 1992; Mountjoyet al., 1992).Instead, the *A locus*, which encodes the protein *ASIP* (agouti-signalling protein), and is considered the antagonist of *MC1R* block α-*MSH* receptor interaction, resulting in the production of pheomelanin (Lu et al., 1994; Ollmann et al., 1998).

In cattle, the first molecular genetic studies identified three main alleles associated with coat colour at the *Extension* locus (Klungland et al., 1995): E^{+}, the wild type allele that produces a variety of colours depending on the *Agouti* locus (Adalsteinsson et al., 1995); E^{D}, the dominant allele, caused by a T>C missense mutation in the *MC1R* coding region determining an activation of the encoded receptor, which, in turn, gives black coat colour; *e*, the recessive allele caused by a single nucleotide deletion in the *MC1R* coding region that produces a non functional pre-maturely terminated receptor and that, in homozygous animals, yields red/yellow coat colour. Adalsteinsson et al. (1994) demonstarated the colours black, brown and red are determined by the occurrence of alleles from the *E* and *A* loci. The order of allele dominance is usually described as $E^{D} > E^{+} > e$.

Similarly, in a study by Russo et al. (2007), the E^{+} allele was identified among specimens of numerous breeds with a variety of coat colours: Modenese, Jersey, Simmental, Grigio Alpina, Piedmontese, Chianina, Romagnola, Marchigiana, Swedish Red and White, and Danish Red. In the study by Russo et al. (2007), Italian breeds analysed, with the imputation of Reggiana coat colour with individuals that had the e/e genotype. On the other hand, Brenig et al. (2013) found that individuals with the $E^{+}E^{+}$ genotype at the *MC1R* locus had a red coat colour.

Rouzaud et al. (2000) found that the coats of individuals with e/e genotypes were very light (Blonde d'Aquitaine), cream-and-white (Charolais), red (Limousin), and dark mahogany (Salers). According to the assumptions of Klungland et al. (1995), all individuals with e/e genotypes have a red coat.

In *Sus scrofa* The *MC1R* gene, localized on chromosome 6, is constituted by a single coding exon of about 950 bp that has been sequenced in many pig breeds with different coat colours (Mariani et al., 1996; Kijas et al. 1998; 2001; Fang et al., 2009). Five allelic groups have been reported so far: wild type alleles (E^{+}); dominant black alleles (E^{D1} and E^{D2}); the black spotted alleles (E^{P}); the recessive red *e* allele. The wild type E^{+} alleles have been identified in wild boar populations. The only European breed that carries the E^{+} allele is Mangalitza. Sequence data showed that the Dominant black *Extension* allele (E^{D}) identified by classical genetics studies is constituted by two different *MC1R* gene sequences identified as E^{D1} and E^{D2} (Fang et al., 2009). The recessive red allele (*e*), has been observed in the Duroc.

Agouti signaling protein (*ASIP*) is involved in the regulation of pigmentation in mammals by downregulating melanocortin 1 receptor (*MC1R*) activity. In wild type mice, *ASIP* is expressed in skin and testes. Widespread tissue expression of *ASIP* has been found in humans and cattle.

Although *ASIP* only codes for a small protein ranging from 131 to 135 amino acids in the mammals studied to date, the structure of the gene is far more complex than the *MC1R* locus. *ASIP* usually consists of 4 exons separated by large introns. Exons 2–4 are coding exons. Several different promoters determine which of several 5′-untranslated exons will be included in the transcript. Four different 5′ untranslated exons have been identified in mice; three were found in pigs.

Studies on *ASIP* in farm animals mainly focused on coat colour (Kim et al., 2004; Norris and Whan, 2008; Mao et al., 2010; Li et al., 2014; Han et al., 2015; Zhang et al., 2017).

ASIP is known as *A* or agouti locus on BTA13 which is responsible for agouti-signalling protein (ASP) and influences the expression of wild-type pattern. Identified alleles at this locus include A^{+} , *a*, A^{Br} , and a^{i} .

In *A* locus, A^{+} and *a* produce brown and recessive black (nonagouti) colours, they could express their effects only when accompanied with E^{+}/- genotypes and in the homozygous form (Adalsteinsson et al., 1995).

The mutant a^{i} at this locus has been considered to be the cause of the lighter belly observed in Limousin and Jersey cattle (Seo et al., 2007).

The *ASIP* locus in the pig is on chromosome 17 (SSC17q21) and also encodes a 131 amino acid paracrine signaling molecule.

Among the variety of coat colour phenotypes in the domestic pig, only two breeds, Duroc and Mangalitza, are characterized by general or regional pheomelanin expression.

Tang et al. (2008) suggested that T allele of this polymorphism might be related to black coat colour according to the results of the study on indigenous Chinese goat breeds. In our study Anatolian Black breed with black and white

coat colour has the highest T allele frequency (93.75%). T allele has also higher frequencies than G allele in Kilis (black coat) and Angora (white coat) breeds, 81.25% and 87.5%, respectively.

While porcine *MC1R* alleles have been previously characterized at the molecular level (Kijas et al., 1998, 2001), so far no functional mutation of the porcine *ASIP* gene has been reported. In pigs a recessively inherited *MC1R* loss-of-function mutation that is responsible for generalized pheomelanin production has been identified in the red-coloured Duroc breed (Kijas et al., 1998).

The “swallow-bellied” Mangalitza pig breed is characterized by black dorsal pigmentation and yellow or white ventral pigmentation and a sharp lateral line of demarcation. Similar phenotypes in black-and-tan or white-bellied agouti mice are associated with the a^{t} or A^{W}alleles at the *Asip* locus, which are recessive to the wild-type allele A.

4.2 White coat colour in livestock

After a long-term strong artificial selection, the phenotype of white coat colour has accumulated into a high frequency in many livestock species and shows a pattern of autosomal dominant inheritance (Adalsteinsson, 1970; Li et al., 2014).

Extended variability at the *KIT* gene is responsible for the allelic series of the Dominant White (I) locus (Giuffra et al., 1999; Pielberg et al., 2002; Fontanesi et al., 2010a; Rubin et al., 2012). The *KIT* gene encodes the mast/stem cell growth factor receptor (MGF) that is a large protein with an extracellular domain consisting of 5 Ig-like domains, a transmembrane region, and a tyrosine kinase domain (Ray et al. 2008). Normal expression of *KIT* and its ligand MGF is essential for migration and survival of neural-crest-derived melanocyte precursors. Mutations

in this gene cause pigmentation disorders in mice, called *Dominant white spotting/W* (Chabot et al., 1988), and in humans, called piebald trait (Fleischman et al. 1991; Giebel and Spritz 1991). Structural *KIT* mutations in mice are often lethal or sublethal in the homozygous form, exhibit pleiotropic effects on the development of melanocytes, hematopoietic cells, primordial germ cells, and interstitial cells in the small intestine, and may affect hearing.

A partial or complete duplication of the *KIT* gene causes different patterns of white coat colouration in pigs and in some of the cattle breeds such as Holsteins, Simmental, Hereford. *KIT* gene on BTA6 is responsible for colour-sided and white spotting coat colour phenotypes in cattle breeds (Grosz and MacNeil, 1999).

Alleles detected at the *spotting* locus (S) were S^{+} (non-spotted), S^{H} (Hereford pattern), S^{P} (Pinzgauer pattern or lineback) and *s* (spotted); this gene caused piebaldism in cattle (Fontanesi et al., 2010a).

Alleles S^{H} and S^{P} seem codominant to each-other and incompletely dominant over S^{+}. All these three alleles appear to be completely dominant over the s allele (Olson, 1981; 1999).

Reinsch et al. (1999) published that there was a strong correlation between the amount of white on Holsteins and markers near *KIT*.

Recently, Durkin et al. (2012) have shown that a duplication of a *KIT* gene segment from chromosome 6 and its aberrant insertion on chromosome 29 led to the “colour-sided” white coat colour phenotype in Belgian blue cattle.

The coat colour patterns in White Galloway and White Park cattle breeds encompassing homozygous animals (Cs_{29}/Cs_{29}) demonstrated a mismarked pattern and heterozygous animals (Cs_{29}/Wt_{29}) showed a range of strongly marked to fully black patterns which were found to be affected by the *KIT* gene on BTA29 through duplication and aberrant insertion (Brenig et al., 2013).

Fontanesi et al. (2010a) demonstrated that two selected synonymous single nucleotide polymorphisms (SNPs) identified in exon 2 (g.72779776C>T) and exon 3 (g.72783182A>G) of the *KIT* gene, could capture information of several *KIT* haplotypes in many autochthonous and commercial cattle breeds, as indicated in Holstein-Friesian, Angus and Hereford.

Therefore, Fontanesi et al. (2010b) conducted a study of haplotype analysis including the previously SNPs and results that only two haplotypes [T:G] and [C:A] were identified in Italian Holstein-Friesian and Italian Simmental, with one of them ([C:A]) having very high frequency in both breeds. The haplotype was the most frequent in Italian Brown and Rendena (62.5% and

53.9%, respectively). Reggiana and Modenese had 55.5% and 61.5% of the [C:A] haplotype. These results confirmed the high frequency of the [C:A] haplotype despite the coat colour pattern of the animals, as already reported by Fontanesi et al. (2010a), who analysed the haplotype structure of the *KIT* gene in Holstein-Friesian, Hereford and Angus based on 111 polymorphic sites.

In pigs white coat colour and white patterns are mainly influenced by the *Dominant white* which has a complex series of alleles determined by different copy number variations in *KIT* gene (Pielberg et al., 2002; Johansson et al., 2005; Rubin et al., 2012).

Four different Dominant White/KIT alleles with distinct phenotypic effects have been described so far (Giuffra et al., 1999; Johansson Moller et al., 1996;); the wild type (i), present in the Wild Boar and in coloured breeds; Patch (I^P), causing patches of white colour, found in Landrace and Large White pigs; Dominant White (I) causing a fully dominant white colour in Landrace and Large White pigs; and Belt (I^{Be}), causing a white belt across the shoulders and front legs in Hampshire pigs and most likely in other breeds with the Belt phenotype. I and I^P are associated with a duplication of the entire KIT coding sequence (Johansson Moller et al., 1996).

As says before, the Dominant White coat colour is determined by the duplication of the *KIT* gene (duplication of the whole gene) and by the presence of a splice mutation in intron 17 in one of the duplicated copies, that causes the skipping of exon 17 (allele I^1) in several important commercial breeds, like Large White and Landrace. The duplicated region is of about 450-kb (Fontanesi and Russo, 2013). These Dominant white alleles are indicated with I^2, I^3, etc., the number indicate the numbers of copies of the mutations (Pielberg et al., 2002). The presence of another Dominant white allele, I^L, has been hypothesised.

Another allele, with a single copy of the *KIT* gene and without the splice mutation, determining a spotted phenotype and that should be lethal if homozygous (Pielberg et al., 2002; Johansson et al., 2005). This allele has been named I^{Be*}, suggesting an accordance with allele that cause the belted phenotype. Earlier classical genetics studies on coat colour segregation between and within populations suggested the presence of an additional allele (I^d also indicated as I^{Rn}), giving a gray-roan phenotype and dominant over the i allele (Fontanesi et al., 2010b). Markers in the KIT gene were generally used to demonstrate that the I^d allele is determined by a single KIT copy gene (Fontanesi et al., 2010b).

The Dominant white colour, have been investigated also in other species like horses and donkeys.

In horses the trait is inherited as a monogenic autosomal dominant trait, also in this specie, the homozygous dominant genotype (W/W) was hypothesized to cause early embryonal lethality (Pulos and Hutt, 1969).

Furthermore, three mutations were detected to be involved in white phenotypes in donkeys. Haase et al. (2015) identified two variants in the *KIT* gene, as candidate mutations for Dominant White and white spotting phenotypes in donkeys. One missense mutation (c.662A > C) in exon 4 of the KIT gene was found in a Hungarian donkey with pink skin, white hair and dark eyes (Haase et al., 2015).

References

Adalsteinsson, S., 1970. Colour inheritance in Icelandic sheep and relation between colour, fertility and fertilisation. J. Agric. Res. Iceland. 2, 3–135.

Seo, H., Lee, D., 2007. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J. Neurosci. 27(31), 8366-8377. doi: 10.1523/JNEUROSCI.2369-07.2007.

Ali, S.A., Choudhary, R.K., Naaz, I., Ali, A.S., 2015. Understanding the challenges of melanogenesis, key role of bioactive compounds in the treatment of hyperpigmentary disorders. J. Pigment. Dis. 2,1-9. doi: 10.4172/2376-0427.1000223.

Ali, S.A., Naaz, I., 2015. Current challenges in understanding the story of skin pigmentation: Bridging the morpho-anatomical and functional aspects of mammalian melanocytes. In: Sakuma K, editor. Muscle Cell and Tissue. Europe, USA: In Tech Open House, 262-85. doi: 10.5772/60714.

Klungland, H., Vage, D. I., Gomez-Raya, L., Adalsteinsson, S., Lien, S., 1995. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat colour determination. Mamm. Genome. 6(9), 636-639.

Basrur, V., Yang, F., Kushimoto, T., Higashimoto, Y., Yasumoto, K., Valencia, J., Muller, J., Vieira, W.D., Watabe, H., Shabanowitz, J., Hearing, V.J., Hunt, D.F., Appella, E., 2003. Proteomic analysis of early melanosomes: identification of novel melanosomal proteins. J. Proteome Res. 2, 69–79. doi: doi.org/10.1021/pr025562r.

Brenig, B., Beck, J., Floren, C., Bornemann-Kolatzki, K., Wiedemann, I., Hennecke, S., Swalve, H. and Schütz, E., 2013. Molecular genetics of coat colour variations in White Galloway and White Park cattle. Anim. Genet. 44(4450-453. doi: 10.1111/age.12029.

Cappai, M.G., Picciau, M., Nieddu, G., Sogos, I., Cherchi, R., Pinna, W., 2015. Cutaneous metabolic pathway of tyrosine as a precursor to melanin in Asinara's white donkey, Equusa sinus L., 1758. Ital. J. Anim. Sci. 14, 3976.

Chabot, B., Stephensson, D.A., Chapman, V.M., Besmer, P., Berstein, A., 1988. The proto-oncogene *c-kit* encoding a transmembrane tyrosine kinase receptor maps to the mouse *W* locus. Nature. 335, 88–89.

Chedekel, M.R., Ahene, A.B., Zeise, L., 1992. Melanin standard method: Empirical formula 2. Pigment. Cell. Res. 5,240-6.

Chi, A., Valencia, J.C., Hu, Z.Z., Watabe, H., Yamaguchi, H., Mangini, N.J., Huang, H., Canfield, V.A., Cheng, K.C., Yang, F., Abe, R., Yamagishi, S., Shabanowitz, J., Hearing, V.J., Wu, C., Appella, E., Hunt, D.F., 2006. Proteomic and bioinformatic characterization of the biogenesis and function of melanosomes. J Proteome Res. 5, 3135–3144. doi: 10.1021/pr060363j.

Cieslak, M., Reissmann, M., Hofreiter, M., Ludwig, A., 2011. Colours of domestication. Biological Reviews. 86(4), 885-899. doi: 10.1111/j.1469-185X.2011.00177.

Costin, G.E., Valencia, J.C., Wakamatsu, K., Ito, S., Solano, F., Milac, A.L., Vieira, W.D., Yamaguchi, Y., Rouzaud, F., Petrescu, A.J., Lamoreux, M.L., Hearing, V.J., 2005. Mutations in dopachrome tautomerase (Dct) affect eumelanin/pheomelanin synthesis, but do not affect intracellular trafficking of the mutant protein. Biochem J. 391,249–259. doi: 10.1042/BJ20042070.

Dell'Angelica, E.C., Mullins, C., Caplan, S., Bonifacino, J.S., 2000. Lysosome-related organelles. Faseb J. 14,1265–1278. doi: 10.1016/j.ceb.2013.04.008.

Fang, M., Larson, G., Ribeiro, H.S., Li, N., Andersson, L., 2009. Contrasting mode of evolution at a coat colour locus in wild and domestic pigs. PLoS Genetics 5:e1000341. doi: 10.1371/journal.pgen.1000341

Ollmann, M. M., Lamoreux, M. L., Wilson, B. D., Barsh, G. S., 1998. Interaction of Agouti protein with the melanocortin 1 receptor in vitro and in vivo. Genes & development. 12(3), 316-330.

Lu, D., Willard, D., Patel, I. R., Kadwell, S., Overton, L., Kost, T., Cone, R. D., 1994. Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor. Nature, 371(6500), 799-802.

Brenig, B., Beck, J., Floren, C., Bornemann-Kolatzki, K., Wiedemann, I., Hennecke, S., Schütz, E., 2013. Molecular genetics of coat colour variations in White Galloway and White Park cattle. Anim. Genet. 44(4), 450-453. doi: 10.1111/age.12029.

Tang, C. J., Zhou, R. Y., Li, X. L., Zhao, J. W., Li, L. H., Feng, F. J., Keng, J. F., 2008. Variation of 423G> T in the Agouti gene exon 4 in indigenous Chinese goat breeds. Biochemical genetics. 46(11-12), 770-780. doi: 10.1007/s10528-008-9192-4.

Giuffra, E., Evans, G., Törnsten, A., Wales, R., Day, A., Looft, H., Andersson, L. (1999). The Belt mutation in pigs is an allele at the Dominant white (I/KIT) locus. Mammalian Genome, 10(12), 1132-1136.

Rubin, C. J., Megens, H. J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., Andersson, L., 2012. Strong signatures of selection in the domestic pig genome. Proceedings of the National Academy of Sciences. 109(48), 19529-19536. doi: 10.1073/pnas.1217149109.

Ray, P., Krishnamoorthy, N., Ray, A., 2008. Emerging functions of c-kit and its ligand stem cell factor in dendritic cells. Cell. Cycle. 7(18), 2826-2832. doi: 10.4161/cc.7.18.6752.

Durkin, K., Coppieters, W., Drögemüller, C., Ahariz, N., Cambisano, N., Druet, T., Charlier, C. 2012. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. Nature. 482(7383), 81-84. doi.org/10.1038/nature10757.

Johansson Moller M., Chaudhary R., Hellmén E., Höyheim B., Chowdhary B. & Andersson L., 1996. Pigs with the dominant white coat colour phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. Mamm. Genome. 7, 822-30.

Johansson, A., Pielberg, G., Andersson, L., Edfors-Lilja, I., 2005. Polymorphism at the porcine Dominant white/KIT locus influence coat colour and peripheral blood cell measures. Anim. Genet. 36(4), 288-296. doi: 10.1111/j.1365-2052.2005.01320.

Haase B., Jagannathan V., Rieder S., Leeb T., 2015. A novel KIT variant in an Icelandic horse with white-spotted coat colour. Anim Genet. 46, 466. doi: 10.1111/age.12313.

Fleischman, R.A., Saltman, D.L., Stastny, V., Zneimer, S., 1991. Deletion of the c-kit prooncogene in the human developmental defect piebald trait. Proc. Natl. Acad. Sci. USA 88, 10885-10889.

Fontanesi, L., Russo, V., 2013. Molecular Genetics of coat Colour in Pigs. Acta Agric Slov. 4, 16.

Fontanesi, L., Scotti, E., Russo, V., 2010b. Analysis of SNPs in the KIT gene of cattle with different coat colour patterns and perspectives to use these markers for breed traceability and authentication of beef and dairy products. Ital. J. Anim. Sci. 2010. 9:e42.

Fontanesi, L., Tazzoli, M., Russo, V., Beever, J., 2010a. Genetic heterogeneity at the bovine *KIT* gene in cattle breeds carrying different putative alleles at the spotting locus. Anim Genet. 41, 295–303. doi: 10.1111/j.1365-2052.2009.02007.

Giebel, L.B., Spritz, R., 1991. Mutation of the KIT (mast/stem cell growth factor receptor) protooncogene in human piebaldism. Proc. Natl. Acad. Sci. USA 88, 8696-8699.

Grosz, M.D., MacNeil, M.D., 1999. The "spotted" locus maps to bovine chromosome 6 in a Hereford-cross population. J. Hered. 90:233-236.

Han, J.L., Yang, M., Yue, Y.J., Guo, T.T., Liu, J.B., Niu, C.E., Yang, B.H. 2015.Analysis of agouti signaling protein (ASIP) gene polymorphisms and association with coat colour in Tibetan sheep (Ovis aries), Genet. Mol. Res. GMR, 14, 1200–1209. doi: 10.4238/2015.February.6.22.

Hearing, V.J., 2015. Biogenesis of pigment granules: a sensitive way to regulate melanocyte function. J. Dermatol. Sci. 37, 3–14. doi: 10.1016/j.jdermsci.2004.08.014.

Huanga, Q., Lianga, W., Xub, D., Zhoua, Y., Wanga, T., Lianga, Y., 2013. Ultrastructural observations of human epidermal melanocytes cultured on polyethylene terephthalate film. Micron. 48, 49-53. doi: 10.1016/j.micron.2013.02.008.

Kijas, J.M.H., Moller, M., Plastow, G., Andersson, L., 2001. A frameshift mutation in *MC1R* and a high frequency of somatic reversions cause black spotting in pigs. Genetics. 158, 779–785. doi: 10.1093/genetics/158.2.779.

Kijas, J.M.H., Wales, R., Törnsten, A., Chardon, P., Moller, M., 1998. Melanocortin receptor 1 (*MC1R*) mutations and coat colour in pigs. Genetics. 150, 1177–1185. doi: 10.1093/genetics/150.3.1177.

Kim, T.H., Choi, B.H., Beever, J. E., 2004. Polymorphism in the porcine agouti signalling protein (ASIP) gene. Anim. Genet. 35, 418–420. doi: 10.1111/j.1365-2052.2004.01179.x

Klungland, H., Vage, D.I., Gomez-Raya, L., Adalsteinsson, S., Lien, S. 1995. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat colour determination. *Mamm. Genome*. *6*, 636–639. doi: 10.1007/BF00352371

Kushimoto, T., Basrur, V., Valencia, J., Matsunaga, J., Vieira, W.D., Ferrans, V.J., Muller, J., Appella, E., Hearing, V.J., 2001. A model for melanosome biogenesis based on the purification and analysis of early melanosomes. Proc. Natl. Acad. Sci. USA.. 98, 10698–703. doi: 10.1073/pnas.191184798.

Li, M.H., Tiirikka, T., Kantanen, J., 2014. A genome-wide scan study identifies a single nucleotide substitution in ASIP associated with white versus non-white coat-colour variation in sheep (Ovis aries). Heredity. 112, 122–131. doi: 10.1038/hdy.2013.83.

Mao, H., Ren, J., Ding, N., Xiao, S., and Huang, L.,2001. Genetic variation within coat colour genes of MC1R and ASIP in Chinese brownish red Tibetan pigs, Anim. Sci. J. Nihon Chikusan Gakkaiho. 81, 630–634, https://doi.org/10.1111/j.1740-0929.2010.00789.x

Mariani, P., Moller, M., Johansson, Hoyheim, B., Marklund, L., Davies, W., 1996. The extension coat colour locus and the loci for Blood Group O and Tyrosine Aminotransferase are on pig chromosome 6. J. Hered. 87, 272–276.

Marks, M.S., Seabra, M.C., 2001. The melanosome: membrane dynamics in black and white. Nature Rev. Mol. Cell Biol. 2, 738–748. doi: 10.1038/35096009.

Norris, B. J., Whan, V. A., 2008. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep, Genome Res. 18, 1282–1293. doi:10.1101/gr.072090.107

Olson, T.A., 1981. The genetic basis for piebald patterns in cattle. J. Hered. 72:113-116.

Olson, T.A., 1999. Genetics of colour variation.

Palumbo, A., Napolitano, A., De Martino, L., Vieira, W., Hearing, V.J., 1994. Specific incorporation of 2-thiouracil into biological melanins. Biochim Biophys Acta. 1200, 271-6.

Pielberg, G., Olsson, C., Syvänen, A.C., Andersson, L., 2002. Unexpectedly high allelic diversity at the *KIT* locus causing dominant white colour in the domestic pig. Genetics, 160, 305–311. doi: 10.1093/genetics/160.1.305.

Protas, M.E., Patel, N.H., 2008. Evolution of colouration patterns. Annual Review of Cell and Developmental Biology. 24, 425-446. doi: 10.1146/annurev.cellbio.24.110707.175302.

Pulos, W.L., Hutt, F.B., 1969. Lethal dominant white in horses. J Hered. 60, 59–63.

Raposo, G., Marks, M.S., 2007. Melanosomes-dark organelles enlighten endosomal membrane transport. Nat Rev Mol Cell Biol. 8, 786–97. doi: 10.1038/nrm2258.

Reinsch, N., Thomsen, H., Xu, N., Brink, M., Looft, C., Kalm, E., Brockmann, G.A., Grupe, S., Kühn, C., Schwerin, M. and Leyhe, B., 1999. A QTL for the degree of spotting in cattle shows synteny with the KIT locus on chromosome 6. Journal of Heredity, 90(6), pp.629-634. doi: 10.1093/jhered/90.6.629.

Russo, V.; Fontanesi, L.; Scotti, E.; Tazzoli, M.; Dall'Olio, S.; Davoli, R. 2007. Analysis of melanocortin 1 receptor (MC1R) gene polymorphisms in some cattle breeds: Their usefulness and application for breed traceability and authentication of Parmigiano Reggiano cheese. Ital. J. Anim. Sci. *6*, 257–272. doi: 10.4081/ijas.2007.257.

Searle, A.G., 1968. Comparative genetics of coat colour in mammals.

Simon, J.D., Peles, D., Wakamatsu, K., Ito, S., 2009. Current challenges in understanding melanogenesis: Bridging chemistry, biological control, morphology, and function. Pigment Cell Melanoma Res. 22, 563-79. doi: 10.1111/j.1755-148X.2009.00610.

Wasmeier, C., Hume, A.N., Bolasco, G., Seabra, M.C., Melanosomes at a glance. J Cell Sci. 121, 3995-9. doi: 10.1242/jcs.040667.

Yamaguchi, Y., Brenner, M., Hearing, V.J., 2007. The regulation of skin pigmentation. J Biol Chem. 282, 27557–61. doi: 10.1074/jbc.R700026200.

Zhang, X., Li, W., Liu, C., Peng, X., Lin, J., He, S., Li, X., Han, B., Zhang, N., Wu, Y., Chen, L., Wang, L., MaYila, N., Huang, J., and Liu, M., 2017. Alteration of sheep coat colour pattern by disruption of ASIP gene via CRISPR Cas9, Sci. Rep., 7, 8149. doi: 10.1038/s41598-017-08636-0.

Aim of the thesis

The phenotypic trait coat colour in livestock changed during the domestication process. Usually considered as monogenic, this exterior trait has been studied in many species (including also humans) to identify causative mutations. The genetic pathway involved in the pigmentation process affect also the pigmentation of the iris; assuming this, many studies were conducted in humans, cattle and horses to dissect the potential region causing different types of colours or abnormalities. Previous studies demonstrated that pigmentation processes are more complex than expected.

The aim of this thesis is to dissect, at the genome level, pigmentation related traits, in two livestock species. More specifically, we focus our researches on cattle and pig genomes in order to detect gene markers associated with coat colour and iris colour. We investigating coat colour in two Italian local cattle breeds (Reggiana and Modenese) using a population genomic approach, and, we investigated eye colour pigmentation in Italian Large White pigs using several GWAS approaches to find the genetic markers.

In the first study that we proposed, using the GeneSeek GGP Bovine 150k single nucleotide polymorphism (SNP) chip, we aimed to detect signatures of selection in the genome of the two autochthonous breeds using genome-wide information in comparative F_{ST} analyses.

The second study investigated the pigmentation process in the Italian Large White, because this is a white breed that whose coat colour is note due to albinism. We present, in the following chapters, our results obtain by the study of seven genome-wide association studies based on several comparison between different colour of iris (three groups of pigs with fully pigmented irides: pale brown, medium brown and dark brown; three groups of pigs with depigmented irides: both depigmented/pale grey eyes, heterochromia iridis i.e. depigmented iris sectors in pigmented irises and heterochromia iridum i.e. a whole eye iris of depigmented phenotype and the other eye with the iris completely pigmented).

Chapter 5: Research activity on cattle

5.1 Signatures of selection are present in the genome of two close autochthonous cattle breeds raised in the North of Italy and mainly distinguished for their coat colours

This work has been publish by the Journal of Animal Breeding and Genetics (2021), pp. 1-13; doi: 10.1111/jbg.12659. The current text is still not in the final format owned by the journal.

Introduction

Autochthonous cattle breeds constitute important genetic resources, in many cases unexploited or poorly characterized. Many local breeds have been developed by the combined action of several factors and events mainly driven by the recent interplay between economic, social and environmental conditions that contributed to define their genetic history (e.g. Felius et al., 2014). Directional selection, that finally shaped the current genetic pools, fixed or almost fixed inheritable exterior phenotypes that could be considered breed-specific traits useful to distinguish different breeds. One of the main exterior phenotypes that characterize different breeds is coat colour.

Two autochthonous cattle breeds, Reggiana and Modenese, are raised mainly in the production area of the well-known Protected Designation of Origin (PDO) Parmigiano-Reggiano cheese, in the North of Italy. Reggiana and Modenese are historically considered the ancestral cattle populations from which this cheese has been originated. Their names derive from the two geographically close provinces of Reggio Emilia and Modena, located in the Emilia-Romagna region where they have been constituted and where most of the farms raising Reggiana and Modenese cattle are now localized. The Herd Book of these two breeds were officially recognized in 1962 (Reggiana) and in 1957 (Modenese) even if the cattle populations that could be attributed to these breeds colonized extensively and much earlier a broader district that encompassed several areas of the Po valley (Pianura Padana), where these breeds accounted for the most numerous cattle over the first half of the last century. The progressing decline of the population size of the Reggiana and Modenese breeds reached a minimum of about 500 and 260 cows in the 1980' and in the first years of the 2000, respectively. Since then, the number of heads started to increase. The recovery of these two breeds can be mainly attributed to the development of two mono-breed branded Parmigiano Reggiano cheeses that can be produced from milk of only Reggiana cows or from milk of only Modenese cows. These niche products can obtain a higher market prize compared to the undifferentiated Parmigiano-Reggiano cheese, supporting economically the lower milk production of the cows of these two breeds compared

to the production of cows of cosmopolitan breeds (Gandini et al., 2007; Fontanesi, 2009; Petrera et al., 2016; Russo et al., 2007). In 2020, Reggiana and Modenese accounted a total of about 2800 cows (raised in about 100 farms) and 500 cows (raised in about 60 farms), respectively.

These two breeds can be distinguished according to their standard coat colour and muzzle colour (Figure 1): a classical red coat colour, indicated with the term "fromentino", over the whole body, with pink or pale muzzle colour are the main pigmentation features of the Reggiana cattle; white coat colour of the body with some pale grey shades and a black muzzle with a depigmented inverted "V" are the main characteristics of Modenese cattle, also known as Bianca Val Padana (Bianca = White) that is the second name of this breed that derives from its coat colour.

A few studies that investigated DNA markers in candidate genes and at the genome-wide level were carried out in Reggiana and Modenese breeds. Candidate gene markers were used in association studies with production traits in Reggiana sires (Fontanesi et al., 2015) and to compare the frequency of relevant alleles affecting milk production traits among breeds, including Reggiana and Modenese cattle (Fontanesi et al., 2007; Scotti et al., 2010). Polymorphisms in three coat colour affecting genes have been analysed in these breeds to identify markers useful to authenticate the breed-origin of the mono breed Parmigiano Reggiano cheeses (Russo et al., 2007; Fontanesi et al., 2010a, 2010b, 2012). The melanocortin 1 receptor (*MC1R*) allele causing a frameshift of the reading frame in this gene (allele e) has been indicated to determine the red coat colour in Reggiana cattle (Klungland et al., 1995; Russo et al., 2007). A genome-wide association study for exterior traits has been recently carried out in Reggiana (Bovo et al., 2021) and genome-wide analyses of signature of selection and population genomic parameters have been carried out in Reggiana against a few other cosmopolitan and local cattle breeds (Bertolini et al., 2015, 2018, 2020a, 2020b; Mastrangelo et al., 2016, 2018a). A study that applied single nucleotide polymorphism (SNP) chip information and that involved Modenese, Holstein and Maremmana cattle tentatively reported a few results on SNP allele frequencies that distinguished these breeds (Catillo et al., 2018). Genetic relationships among Italian cattle breeds that included SNP chip data from few Reggiana and Modenese cattle indicated that the two breeds are closely related compared to several other breeds (Mastrangelo et al., 2018b).

Considering the limited information that is however available in terms of genome differences between these two geographically close breeds, in this study we genotyped almost half of the extant cattle populations of Reggiana and Modenese breeds and used genome-wide information in comparative analyses to detect signatures of selection that might be derived by

the divergent directional selection and genetic drifts that have contributed to shape the genome structures of these two autochthonous cattle breeds.

Materials and methods

Ethic statement

Animal samples used in this study were collected following the recommendation of directive 2010/632.1.

Animals and genotyping data

A total of 1435 cattle included in this study (Reggiana, n. = 1109; Modenese, n. = 326) were genotyped with the GeneSeek GGP Bovine 150k SNP chip following the manufacturer's protocol. PLINK software v. 1.9 (Chang et al., 2015) was used to filter genotyping data. Only markers with minor allele frequency (MAF) <0.01 across the two breeds, with a call rate > 90% in each breed, mapped in unique positions in the autosomes of the ARS-UCD1.2 cattle genome version were retained. In addition, considering that the Herd Book of the Reggiana breed (ANABORARE, 2020) considers that the Reggiana cattle should have the homozygous recessive genotype *e/e* at the melanocortin 1 receptor (*MC1R*) gene, which causes the red coat colour of the breed (Klungland et al., 1995; Russo et al., 2007; Bovo et al., 2021), only Reggiana cattle with this genotype were included in this study. Genotyping data at the *MC1R* gene were retrieved from the GeneSeek GGP Bovine 150k SNP chip. At the end, the dataset accounted for a total of 1109 Reggiana cattle (98 samples were excluded because they were heterozygous *E/e* at the *MC1R* gene and 2 samples did not pass the quality criteria for genotyping) that had *e/e* genotype at the *MC1R* gene and 326 Modenese cattle and a total of 128574 SNPs.

Population genomic analyses

Observed and expected heterozygosity (H_O and H_E, respectively) were calculated with PLINK v. 1.9 (Chang et al., 2015). Inbreeding coefficient of an individual (I) relative to the subpopulation (S) (F_{IS}), fixation index (F_{ST}) and inbreeding coefficient of an individual (I) relative to the total (T) population (F_{IT}) were calculated with VCFtools software (Danecek et al., 2011).

Linkage disequilibrium (LD) was measured using r^2 for all SNP pairs of each chromosome using PLINK v 1.9 (Chang et al., 2015) and within breed LD decay was estimated using bins of 10 kb. Plots were generated in R v.3.5.1. (R Core Team, 2018) with the ggplot2 package (Wickam, 2016). Recent and historical effective population size (Ne) were estimated

using the SNeP software (Barbato et al., 2015), using the maximum distance between SNP to be analysed of 10 Mb and the binwidth of 100 kb for the calculation of LD.

To perform population structure analyses, pruning of SNPs in high LD was carried out of PLINK 1.9 (Chang et al., 2015) with the -- indep-pairwise command (options: window size of 50 kb, step size of 10 and r^2 threshold of 0.1). A total of 14131 SNPs was retained (487 ±166 SNPs for each chromosome). Multidimensional scaling (MDS) analysis was carried out by using a matrix of genome-wide identity-by-state (IBS) pairwise distances as implemented in PLINK 1.9 (Chang et al., 2015).

Population stratification was evaluated with the ADMIXTURE v3.1 software (Alexander et al., 2009). Analyses were preformed considering the number of subpopulations (K) that ranged from 1 to 39 and retaining the cross-validation error (CV) for each K.

F_{ST} analyses, gene annotation and haploblock analysis

Wright's F_{ST} for each SNP in the pairwise comparison between Reggiana and Modenese populations was calculated with PLINK 1.9 (Chang et al., 2015) using the method of Weir and Cockerham (1984). Overall averaged F_{ST} was calculated considering all SNPs in the pairwise comparisons.

Signatures of selection were determined using pairwise F_{ST} analyses using two approaches: (i) single marker pairwise F_{ST} analysis and(ii) averaged genome window F_{ST} comparative analysis. The SNP with the top 0.0005 % F_{ST} (99.95th percentile; top 64 SNPs) defined the threshold to detect signatures of selection. In the window-based approach, windows of 1 Mb with a step of 500 kb, were tested by computing an average F_{ST} based of SNPs overlapping the window. In total, 9953 windows harbouring 50 SNPs each were tested. Windows of 500 kb with a step of 250 kb were also tested. Analyses, based on the method of Weir and Cockerham (1984) method, were carried out with VCFtools (Danecek et al., 2011). All windows that contained at least four SNPs were then considered. The 99.8th percentile threshold was applied for the window-based analysis (top 50 windows).

Annotation of the genome regions including the SNPs and windows that trespassed the defined threshold was retrieved from the *Bos taurus* genome assembly version ARS-UCD1.2 and considering a region of ± 200 kb around the detected SNPs or considering the overlapping or partially overlapping windows. Genes were retrieved using Ensembl Biomart tool (http://www.ensembl.org/biomart/martview/) and then evaluated considering their functional roles according to an extensive literature search.

Functional gene enrichment analysis of genes closed to the SNPs detected in the single-marker F_{ST} analysis was carried out with Enrichr (Chen et al., 2013). Over-representation analysis run over the Gene Ontology v.2021 (http://geneontology.org), KEGG v.2021 (http://www.kegg.jp/) and GWAS catalog v.2019 (https://www.ebi.ac.uk/gwas/) human libraries. Terms with an adjusted p-value < 0.05 and at least two input genes were retained as statistically over-represented.

Haplotype block analysis of the *MC1R* and *ASIP* gene regions was performed using the software Haploview v. 4.2 (Barret et al., 2015) using default options to partition the region into segments of strong LD.

Results

Population genomic parameters and structures of the two cattle breeds

Table 1 summarises some basic population genomic parameters calculated in the two cattle breeds. The average within-breed MAF was higher in the Reggiana (mean = 0.271, s.d. = ± 0.147) breed than in the Modenese breed (mean = 0.257, s.d. = ± 0.151). The MAF distribution (Figure S1) confirmed the highest number of SNPs (n. = 8085) with the lowest MAF values (ranging from 0.01 to 0.05) that was detected in the Modenese breed. Within-breed H_O and H_E heterozygosity was lower in Modenese than in Reggiana (Table 1) reflecting the other SNP based information reported above.

Figure 2 reports the two-dimensional MDS plots obtained using genome information from the Reggiana and Modenese breeds. The two breeds are clearly separated by the first three coordinates into two compact and distinguishable clouds.

The results of the ADMIXTURE analysis are showed in Figure 3. Despite a high number of K was considered, the minimum value of CV error was not detected. However, the higher decrease in K is observed with K=2, and after that the K values remains quite constant (Figure 3a). The population stratification at K=2 (Figure 3b) is consistent with the clusters detected by the MDS analysis. If a higher K is considered (e.g. K=4; Figure 3c), a higher level of stratification of the Reggiana breed could be observed in contrast with the Modenese breed that tended to be more homogeneous.

Linkage disequilibrium was higher in the Modenese breed than in the Reggiana breed as it was evident from the LD decay (Figure 4a) and the average LD calculated for all autosomes in the two breeds (Figure 4b and Table S1). This information reflected the lower Ne value obtained in the Modenese breed than in the Reggiana breed (120 vs 215, respectively) as also evidenced from its progressive decline plot over the past generations (Figure 4c). The similar

trend in LD values estimated for each chromosome in the two breeds, which confirmed the general higher LD values in the Modenese than in the Reggiana breed, also evidenced that the SNPs present in the chip might largely affect the LD structure in the two cattle breeds.

F_{ST} derived signatures of selection between the two breeds

The global averaged F_{ST} value across all SNPs obtained comparing Reggiana and Modenese was 0.066. Figure 5 reports the Manhattan plots obtained in the single-marker (a) and window-based (b and c) F_{ST} analyses by contrasting genomic information of the Reggiana and Modenese breeds.

Table 2 reports the top 20 markers with the highest F_{ST} values which ranged from 0.977 to 0.637. The full set of markers (n. = 64) trespassing the 99.95^{th} percentile thresholds are reported in Table S2. Table 3 includes the top 0.5 Mb genome windows (n. = 50) and Table S3 contains information on the top 1 Mb genome windows identified using the two applied window-based analyses, respectively. Averaged F_{ST} values in these windows ranged from 0.344 to 0.222 in the top (99.98^{th} percentile) 0.5 Mb genome windows and from 0.255 to 0.163 in the top (99.98^{th} percentile) 1 Mb genome windows. The drastic drop of F_{ST} values from the single-marker to the window-based analyses might indicate that the two breeds could be distinguished by a few highly separated loci that experienced a rapid decay of LD apart from informative short genome region, diluting the F_{ST} values in the window-based approaches.

The 64 markers were distributed in 23 different autosomes (Table S2). Some of them were also captured in the window-based analyses: seven and eight genome windows in four and six autosomes, that encompassed regions that included several top markers, were detected in the 0.5 Mb and 1 Mb window analyses, respectively. The top marker (F_{ST} = 0.977) was the frameshift mutation in the *MC1R* gene that causes the *e* allele at the *Extension* locus and that determines the classical red coat colour of the Reggiana cattle (Klungland et al., 1995; Russo et al., 2007; Bovo et al., 2021). This result was due to the fact that Reggiana cattle selected for this study were fixed for this recessive *MC1R* allele whereas Modenese cattle were almost fixed for an alternative allele (only one Modenese animal carried the *e* allele). This polymorphic site, located on BTA18, was in the first top genome window of the 0.5 Mb analysis and in the third and fourth top sliding windows in the 1 Mb analysis. The haploblock structure of this region in Reggiana cattle indicated that a relatively low level of LD is present in the *MC1R* gene region in both Reggiana and Modenese breeds, with some LD blocks only upstream or downstream this gene (Figure S2).

The second top polymorphism in the single-marker analysis was localized on BTA7 about 9.6 kb from the *protein phosphatase 2 catalytic subunit alpha* (*PPP2CA*) gene (Table 2). Additional four markers in the same chromosome region (Table 2), that contributed to the second and fifth highest averaged F_{ST} values in the 0.5 Mb genome window analysis (Table 3), were within or close to the *transcription factor 7* (*TCF7*) and *voltage dependent anion channel 1* (*VDAC1*) genes.

Other three top 99.95th percentile SNPs, that were also contained in top genome windows considering both the 0.5 and 1 Mb size, were located on BTA13 within or very close to the *eukaryotic translation initiation factor 2 subunit beta* (*EIF2S2*) and *agouti signalling protein* (A*SIP*) genes (Table 2 and Table 3; Table S2). *ASIP* is the gene that determines the well known agouti locus which affects coat colour in many mammalian species (Searle, 1968). The LD structure of this region in Modenese cattle showed a haplotype block in the correspondence of the *RALY heterogeneous nuclear ribonucleoprotein* (*RALY*) and *eukaryotic translation initiation factor 2 subunit beta* (*EIF2S2*) genes, which are upstream the *ASIP* gene (Figure S3). The Reggiana breed had a main haploblock in the correspondence of the *itchy E3 ubiquitin protein ligase* (*ITCH*) gene (Figure S3).

Additional top SNPs on BTA5, BTA6, and BTA24 were also included in top genome windows detected with the 0.5 and/or 1 Mb window-approaches. Markers on BTA5 were within or close to the *myosin heavy chain 9* (*MYH9*) gene, SNPs on BTA6 were within the *glutamate ionotropic receptor delta type subunit 2* (*GRID2*) gene and the marker on BTA24 was in a desert region where the closest genes are *cadherin 2* (*CDH2*) and *cadherin related 23* (*CDH23*) (Table 3, Table S2 and Table S3).

Gene enrichment analysis returned significant results only for the Human GWAS catalog. Among the 12 over-represented phenotypes (Table S3), nine were related to pigmentation (e.g. skin colour, skin pigmentation, skin aging, freckles and tanning processes) and involved the *ASIP*, *ERBB4*, *EIF2S2* and *MC1R* genes. Melanogenesis was the most over-represented KEGG pathway (adjusted *p*-value = 0.14; *ASIP*, *MC1R* and *TCF7* genes) whereas the regulation of tyrosine phosphorylation of STAT protein (GO:0042509; adjusted *p*-value = 0.15; *ERBB4*, *GHR* and *PPP2CA* genes) was the most over-represented GO Biological Process.

Discussion

Reggiana and Modenese are considered two iconic breeds that are part of the history of the livestock production sector of the North of Italy from which the well-known PDO Parmigiano-Reggiano cheese was originated. Genomic population parameters calculated for the

two breeds are in agreement to those that in general describe the small population size of many autochthonous breeds.

These two local genetic resources have been originally shaped from the genetic pools of cattle populations that were present in the Po valley in the North of Italy over the last century and that subsequently experienced few additional (in many cases unregistered) introgressions from other populations that contributed to constitute the current breed genetic structures. The geographic closeness of the two breeds, as it could be expected, also resulted in a relatively high genetic closeness when Reggiana and Modenese were analysed together with many other Italian cattle breeds (Mastrangelo et al., 2018b). Despite this closeness, Reggiana and Modenese cattle were clearly distinguished using genomic information obtained with the GeneSeek GGP Bovine 150k SNP chip. Admixture patterns and MDS-plots were able to separate all animals belonging to the two breeds. That means that Reggiana and Modenese genomes contain many elements that tended to differentiate the animals of the two breeds. These genomic differences could be derived from the combined action of divergent artificial directional selection and genetic drift followed by genetic isolation due to the use of different male and female genetic stocks. A pairwise comparison between these two breeds highlighted the most relevant and peculiar differences as shown in Figure 5. Gene enrichment analysis confirmed that pigmentation and related traits explained the most relevant differences that emerged between these two breeds. These differences might be eventually masked or diluted if comparisons would have included more breeds in averaged F_{ST} analyses, which are the commonly used methodologies for these types of investigations (Munoz et al., 2019; Bovo et al., 2020).

Considering that both Reggiana and Modenese breeds originally derived from unspecialized triple purpose cattle (work-dairy-beef), one of the main drivers that contributed to separate them and that represents the main characterizing phenotype is their different coat colour. Solid red (Reggiana) and solid white with grey shades (Modenese) are the colours that define the standards of these two breeds. This phenotype is the most relevant descriptor used to admit animals in one or the other breed herd book. There is a story telling tradition that suggests that the selection for different coat colours in Reggiana and Modenese would derive from the ancient rivalry between the two close towns (i.e. Reggio Emilia and Modena) from which the two breeds took their names.

The classical red coat colour (fromentino) of the Reggiana cattle caused by the recessive *e* allele at the *MC1R* gene (Klungland et al., 1995; Russo et al., 2007; Bovo et al., 2021) was the source of the most relevant signature of selection that differentiated Reggiana from

Modenese cattle. The high F_{ST} value (almost equal to 1) reached by the causative mutation at the *MC1R* allele dropped in the window-based analyses, as it was averaged across all SNPs included in 0.5 Mb or 1 Mb. The relatively low LD that is present in the *MC1R* gene region of BTA18 in the Reggiana indicated an ancient origin of the fixed *e* allele in this breed and that more haplotypes or haplotype blocks containing this causative mutation were present in Reggiana cattle. The strong selection pressure that fixed (or almost fixed) this *Extension* allele, therefore, did not result in an extended fixation of several other close SNPs on BTA18.

It remains to explain the genetic determinism of the white coat colour with some pale grey shades of the Modenese cattle. This breed is almost fixed for the wild type allele at the *MC1R* gene, as also reported in a previous study (Russo et al, 2007). According to the classical epistatic interaction between the *Extension* and *Agouti* loci, wild type alleles at the *MC1R* gene would give the possibility to express mutated alleles at the *Agouti* locus (Searle, 1969). Therefore, it is quite remarkable that a strong F_{ST} signal between Reggiana and Modenese was detected in the *ASIP* gene region on BTA13. The signal was not as strong as it was observed for the *MC1R* gene region even if it was confirmed using single-marker and the two window-based approaches. The relatively lower F_{ST} value of this region if compared to that of the *MC1R* gene region could be due to the lack of the causative mutation in the SNP chip and/or to the masking effect of the mutated *MC1R* allele in Reggiana that would epistatically cover mutated alleles at the *ASIP* gene. Mutated *ASIP* alleles could be also present in the Reggiana breed but at lower frequency than in Modenese breed, reducing in this way allele frequency differences between the two breeds and, in turn the F_{ST} values in the *ASIP* region. The LD analysis in Reggiana and Modenese indicated high LD in different regions of the BTA13 that includes *ASIP*, suggesting the presence of different haplotype structures in the two breeds, with potential regulatory effects over *ASIP*.

Variability at the *ASIP* gene determined by copy number variations (CNVs), probably with regulatory effects on gene expression, has been already associated with the white coat colour in different sheep and goat breeds (Norris and Whan, 2008; Fontanesi et al., 2009, 2011). Therefore, it would be possible to speculate that, in Modenese breed, CNVs or other regulatory mutations affecting *ASIP* gene could determine a similar phenotypic effect on coat colour as already observed in the other two ruminant species (i.e. sheep and goat). Recently, Trigo et al. (2021) reported that in Nellore cattle (*Bos indicus*), which are selected for white coat colour, a structural variant affecting the *ASIP* gene expression is associated with darker coat pigmentation on specific parts of the body. Few studies have investigated variability in the *Bos taurus ASIP* gene. None of the *ASIP* polymorphisms reported in Korean cattle were associated

to any coat colours (Do et al., 2007). Girardot et al. (2006) reported in Normande cattle an insertion in a regulatory region of the *ASIP* gene that was suggested to be implicated in the brindle coat colour pattern of the breed. It will be important to characterize the *ASIP* gene in Modenese cattle to disentangle its expected effect on coat colour that it could be possible to predict from the results of this study.

Other signatures of selection were evident from the F_{ST} comparative analyses between the two breeds. Considering their history, these signatures might be mainly due to genetic drift that would be subsequently due to the constrains generated by the use of sires and dams that could assure the requested coat colour phenotype needed to register the animals to their herd books. These genomic differences could contribute to further differentiate these two breeds for some production performances or other phenotypic traits, but their effect should be demonstrated using other approaches. The use of other pair-wise methods to detect signature of selection (e.g. Bertolini et al., 2020) could also identify additional genomic regions that might be involved in differentiating these breeds or that could provide more complete genomic pictures of the results of the genetic drift, bottleneck and admixture with other breeds or populations that have probably contributed to shape the current genetic pool of these two cattle autochthonous breeds.

Conclusions

Population genomic analyses applied to compare the genome architecture of two closely related cattle genetic resources (Reggiana and Modenese) made it possible to capture some hints that could explain their main phenotypic differences. Signatures of selection were evidenced in two genome regions encompassing major coat colour affecting genes. One region on BTA18, including the *MC1R* gene, whose role in determining the red coat colour of Reggiana was already well established, could provide a proof of concept for the general interpretation of the results obtained in a region of BTA13 which includes the *ASIP* gene. We are sequencing the genome of Reggiana and Modenese cattle to characterise variability in the *ASIP* gene and investigate their association with the white coat colour in Modenese breed.

This study demonstrates how population genomic approaches designed to take advantage from the diversity between local genetic resources could provide interesting information to explain exterior traits not yet completely investigated in cattle.

Acknowledgements

This work is dedicated to the memory of Prof. Vincenzo Russo, Emeritus of the University of Bologna. We thank ANABORARE personnel, all Reggiana farmers, veterinary and technician staff that collaborated in this study. This study was funded by the PSRN (Programma di Sviluppo Rurale Nazionale) 2014-2020 and 2020-2023, measure 10.2, Dual Breeding and Dual Breeding 2 projects (co-funded by the European Agricultural Fund for Rural Development of the European Union and by the Italian Ministry of Agriculture, Food and Forestry - MiPAAF) and by the University of Bologna RFO2019-2020 programs.

Conflict of interest statement

The authors declare they do not have any competing interests. Data reported in this work can be shared after signature of an agreement on their use with University of Bologna.

Author contribution

L.F. designed the study and obtained funding. L.F., F.B. and G.M. wrote the paper. A.R. and G.M. performed the wet lab work. G.M., F.B., G.S., S.B. and M. Ballan conducted bioinformatic analyses. M. Bonacini. and M.P. provided samples and data. S.B., G.S., S.D. and M. Ballan contributed to data interpretation. All authors read and approved the submitted version.

References

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

ANABORARE (2020). Retieved from https://www.razzareggiana.it/.

Barbato, M., Orozco-terWengel, P., Tapio, M., & Bruford, M. W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics*, *6,* 109.doi:10.3389/fgene.2015.00109.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*, 263–265. https://doi.org/10.1093/bioinformatics/bth457

Bertolini, F., Galimberti, G., Calò, D. G., Schiavo, G., Matassino, D., & Fontanesi, L. (2015). Combined use of principal component analysis and random forests identify population-informative single nucleotide polymorphisms: application in cattle breeds. *Journal of Animal Breeding and Genetics*, *132*, 346–356. https://doi.org/10.1111/jbg.12155

Bertolini, F., Galimberti, G., Schiavo, G., Mastrangelo, S., Di Gerlando, R., Strillacci, M. G., Bagnato, A., Portolano, B., & Fontanesi, L. (2018). Preselection statistics and Random Forest classification identify population informative single nucleotide polymorphisms in cosmopolitan and autochthonous cattle breeds. *Animal*, *12*, 12–19. https://doi.org/10.1017/S1751731117001355

Bertolini, F., Schiavo, G., Bovo, S., Sardina, M. T., Mastrangelo, S., Dall'Olio S., Portolano B., & Fontanesi L. (2020). Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano-Reggiano cheese production system. *Animal*, *14*, 921–932. https://doi.org/10.1017/S1751731119003318

Bovo, S., Ribani, A., Muñoz,M., Alves, E., Araujo, J. P., Bozzi, R., Čandek-Potokar, M., Charneca, R., Di Palma, F., Etherington, G., Fernandez, A. I., García, F., García-Casco, J., Karolyi,D., Gallo, M., Margeta,M., Martins, M. J., Mercat, M. J., Moscatelli, G., Núñez, Y., Quintanilla, R., Radović, Č., Razmaite, V., Riquet, J., Savić, R., Schiavo, G., Usai, G., Utzeri, V. U., Zimmer, C., Ovilo, C., & Fontanesi, L. (2020).Whole-genome sequencing of European autochthonous and commercial pig breeds allows the detection of signatures of selection for adaptation of genetic resources to different breeding and production systems. Genet*.* *52*, 33. https://doi.org/10.1186/s12711-020-00553-7

Bovo, S., Schiavo, G., Kazemi, H., Moscatelli, G., Ribani, A., Ballan, M., Bonacini, M., Prandi, M., Dall'Olio, S., & Fontanesi, L. (2021). Exploiting within-breed variability in the autochthonous Reggiana breed identified several candidate genes affecting pigmentation-related traits, stature and udder defects in cattle. Anim. Genet. https://doi.org/10.1111/age.13109

Catillo, G., Moioli, B., Napolitano, F., & Steri, R. (2018). Identification of genomic regions harboring diversity between Holstein and two local endangered breeds, Modenese and Maremmana. *Livest. Sci.*, *216*, 75–83. https://doi.org/10.1016/j.livsci.2018.07.011

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,* s13742–015–0047–8. https://doi.org/10.1186/s13742-015-0047-8

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*, 128. https://doi.org/10.1186/1471-2105-14-128

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). 1000

Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, *27*, 2156-2158. https://doi.org/10.1093/bioinformatics/btr330

Do, K. T., Shin, H. Y., Lee, J. H., Kim, N. S., Park, E. W., Yoon, D. H., & Kim, K. S. (2007). Investigation of coat colour candidate genes in Korean cattle (Hanwoo). *Journal of Animal Science and Technology*, *49*, 711-718. https://doi.org/10.3390/d6040705

Felius, M., Beerling, M. L., Buchanan, D. S., Theunissen, B., Koolmees, P. A., & Lenstra, J. A. (2014). On the history of cattle genetic resources. *Diversity*, *6*, 705–750. https://doi.org/10.3390/d6040705

Fontanesi, L. (2009). Genetic authentication and traceability of food products of animal origin: new developments and perspectives. *Ital. J. Anim. Sci.*, *8*, 9–18. https://doi.org/10.4081/ijas.2009.s2.9

Fontanesi, L., Beretti, F., Riggio, V., González, E. G., Dall'Olio, S., Davoli, R., Vincenzo, R. & Portolano, B. (2009). Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colours. *Cytogenetic and Genome Research*, *126*, 333–347. https://doi.org/10.1159/000268089

Fontanesi, L., Dall'Olio, S., Beretti, F., Portolano, B., & Russo, V. (2011). Coat colours in the Massese sheep breed are associated with mutations in the agouti signalling protein (ASIP) and melanocortin 1 receptor (MC1R) genes. *Animal*, *5*, 8–17. https://doi.org/10.1017/S1751731110001382

Fontanesi, L., Scotti, E., & Russo V. (2010a). Analysis of SNPs in the KIT gene of cattle with different coat colour patterns and perspectives for using these markers for breed traceability and authentication of beef and dairy products. *Ital. J. Anim. Sci.*, *9*, e42. https://doi.org/10.4081/ijas.2010.e42

Fontanesi, L., Scotti, E., & Russo, V. (2012). Haplotype variability in the bovine MITF gene and association with piebaldism in Holstein and Simmental cattle breeds. *Animal Genetics*, *43*, 250–256. https://doi.org/10.1111/j.1365-2052.2011.02242.x

Fontanesi, L., Scotti, E., Samorè, A. B., Bagnato, A., & Russo, V. (2015). Association of 20 candidate gene markers with milk production and composition traits in sires of Reggiana breed, a local dairy cattle population. *Livest. Sci.*, *176*, 14–21. https://doi.org/10.3168/jds.2017-12666

Fontanesi, L., Scotti, E., Tazzoli, M., Beretti, F., Dall'Olio, S., Davoli, R., & Russo, V. (2007). Investigation of allele frequencies of the growth hormone receptor (GHR) F279Y mutation in dairy and dual purpose cattle breeds. *Ital. J. Anim. Sci.*, *6*, 415–420. https://doi.org/10.4081/ijas.2007.415

Fontanesi, L., Tazzoli, M., Russo, V., & Beever, J. (2010b). Genetic heterogeneity at the bovine KIT gene in cattle breeds carrying different putative alleles at the spotting locus. *Animal Genetics*, *41*, 295–303. https://doi.org/10.1111/j.1365-2052.2009.02007.x

Gandini, G., Maltecca, C., Pizzi, F., Bagnato, A., & Rizzi, R. (2007). Comparing local and commercial breeds on functional traits and profitability: the case of Reggiana dairy cattle. *Int. J. Dairy Sci. 90*, 2004–2011. https://doi.org/10.3168/jds.2006-204

Girardot, M., Guibert, S., Laforet, M. P., Gallard, Y., Larroque, H., & Oulmouden, A. (2006). The insertion of a full-length Bos taurus LINE element is responsible for a transcriptional deregulation of the Normande Agouti gene. *Pigment Cell Research*, *19*, 346–355. https://doi.org/10.1111/j.1600-0749.2006.00312.x.

Klungland, H., Vage, D. I., Gomez-Raya, L., Adalsteinsson, S., & Lien, S. (1995). The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat colour determination. *Mammalian Genome, 6*, 636–639. https://doi.org/10.1007/BF00352371

Mastrangelo, S., Ciani, E., Ajmone-Marsan, P., Bagnato, A., Battaglini, L., Bozzi, R., Carta, A., Catillo, G., Cassandro, M., Casu, S., Ciampolini, R., Crepaldi, P., D'Andrea, M., Di Gerlando, R., Fontanesi, L., Longeri, M., Macciotta, N. P. P., Mantovani, R., Marletta, D., Matassino. D., Mele, M., Pagnacco, G., Pieramati, C., Portolano, B., Sarti, M. F., Tolone, M., & Pilla, F. (2018a). Conservation status and historical relatedness of Italian cattle breeds. *Genet.*, *50*, 35. https://doi.org/10.1186/s12711-018-0406-x

Mastrangelo, S., Sardina, M. T., Tolone, M., Di Gerlando, R., Sutera, A. M., Fontanesi, L., & Portolano, B. (2018b). Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. *Animal*, *12*, 2480–2488. https://doi.org/10.1017/S1751731118000629

Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M. T., & Portolano, B. (2016). Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal*, *10*, 746–754. https://doi.org/10.1017/S1751731115002943

Muñoz, M., Bozzi, R., García-Casco, J., Núñez, Y., Ribani, A., Franci, O., García, F., Skrlep, M., Schiavo, G., Bovo, S., Utzeri, V. J., Charneca, R., Martins, J. M., Quintanilla, R., Tibau, J., Margeta, V., Djurkin-Kusec, I., Mercat, M. J., Riquet, J., Estellè J., Zimmer, C., Razmaite, V., Araujo, J. P., Radovic, C., Savic, R., Karolyi, D., Gallo, M., Candek-Potokar, M., Fernandez, A. I., Fontanesi, L., & Ovilo, C. (2019). Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed

with a high density SNP chip. *Scientific Reports*, *9*, 13464. https://doi.org/10.1038/s41598-019-49830-6

Norris, B. J., & Whan, V. A. (2008). A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Research*, *18*, 1282–1293. https://doi.org/10.1101/gr.072090.107

Petrera, F., Catillo, G., Napolitano, F., Malacarne, M., Franceschi, P., Summer, A., & Abeni, F. (2016). New insights into the quality characteristics of milk from Modenese breed compared with Italian Friesian. *Ital. J. Anim. Sci.*, *15*, 559–567. https://doi.org/10.1080/1828051X.2016.12228899

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/org/.

Russo, V., Fontanesi, L., Scotti, E., Tazzoli, M., Dall'Olio, S., & Davoli, R. (2007). Analysis of melanocortin 1 receptor (MC1R) gene polymorphisms in some cattle breeds their usefulness and application for breed traceability and authentication of Parmigiano Reggiano cheese. *Ital. J. Anim. Sci.*, *6*, 257–272. https://doi.org/10.4081/ijas.2007.257

Scotti, E., Fontanesi, L., Schiavini, F., La Mattina, V., Bagnato, A., & Russo, V. (2010). DGAT1 p. K232A polymorphism in dairy and dual purpose Italian cattle breeds. *Ital. J. Anim. Sci.*, *9*, e16. https://doi.org/10.4081/ijas.2010.e16

Searle, A. G. (1968). Comparative Genetics of Coat Colour in Mammals. Logos Press: London, UK.

Trigo, B. B., Utsunomiya, A. T., Fortunato, A. A., Milanesi, M., Torrecilha, R. B., Lamb, H., Nguyen, L., M. Ross, E., Hayes, B., Padula, R. C. M., Sussai, T. S., Zavarez, L. B., Cipriano, R. S., Caminhas, M. M. T, Lopes, F. L., Pelle, C., Leeb, T., Bannasch, D., Bickart, D., Smith, T. P. L., Sonstegard, T. S., Garcia, J. F., & Utsunomiya, Y. T. (2021). Variants at the ASIP locus contribute to coat colour darkening in Nellore cattle. *Genet.*, *53*, 40. https://doi.org/10.1186/s12711-021-00633-2

Weir, B. S. & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution, 38*, 1358–1370. https://doi.org/10.2307/2408641

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org

Tables

Table 1. Population genomic parameters calculated in the Reggiana and Modenese cattle breeds.

Breed	No. of animals	Average MAF[1]	H_O[2]	H_E[3]	F_{IS}	F_{IT}
Reggiana	1109	0.271 (± 0.147)	0.359	0.360	0.003	0.068
Modenese	326	0.257 (± 0.151)	0.354	0.349	-0.017	0.050

[1] Minor allele frequency and standard deviation in brackets.

[2] Observed heterozygosity.

[3] Expected heterozygosity.

Table 2. Top 20 (99.95th percentile) single nucleotide variants identified in the single-marker F_{ST} analysis between the two breeds. The markers are ranked according to the F_{ST} value. All 99.95th percentile markers are reported in Table S2.

Markers[1]	BTA[2]	Position[3]	F_{ST}	Closest gene (bp)[4]
MC1R	18	14705645	0.977	*MC1R* (0)*
BovineHD0700013748	7	45833400	0.783	*PPP2CA* (9618)*
ARS-BFGL-NGS-114140	21	21791054	0.773	*FURIN* (0)
ARS-BFGL-NGS-28154	18	26500840	0.752	*GOT2* (53100)
BovineHD0600033381	6	112511216	0.736	*LDB2* (357854)
BovineHD1300018297	13	63480254	0.717	*EIF2S2* (0), *ASIP* (182542)*
BTA-78954-no-rs	7	45800275	0.705	*TCF7* (0)*
ARS-BFGL-NGS-55059	4	5545419	0.702	*IKZF1* (0)
ARS-BFGL-NGS-5595	7	45766695	0.694	*TCF7* (2218)*
ARS-BFGL-NGS-73679	7	45729837	0.692	*TCF7* (39076)*
BTA-86548-no-rs	11	16591322	0.671	*RASPGRP3* (476598)
BovineHD2500007120	25	24908014	0.665	*IL4R* (0)
BovineHD2400015179	24	53014583	0.661	*DCC* (0)
BovineHD0600009128	6	31158986	0.652	*GRID2* (0)*
BovineHD0600009122	6	31135482	0.647	*GRID2* (0)*
ARS-BFGL-NGS-35081	14	46102133	0.647	*SAMD12* (36424), *EXT1* (62323)
BovineHD2100006752	21	22531247	0.645	*SLC28A1* (0)
ARS-BFGL-NGS-20141	7	45691037	0.639	*VDAC1* (0)*
BovineHD0500003920	5	12981358	0.638	*TMTC2* (405765)
ARS-BFGL-NGS-16203	3	99840480	0.637	*RAD54L* (0)

[1] Marker name in the GeneSeek GGP Bovine 150k SNP chip.

[2] *Bos taurus* chromosome.

[3] Position of the marker in the ARS-UCD1.2 cattle genome version.

[4] Distance in bp of the marker with the indicated gene is reported within the bracket. When the marker overlap the gene, a value equal to 0 bp is indicated. The star symbol indicates those genes that are also included in the top 0.5 and/or 1 Mb windows in the window-based F_{ST} analyses (see also Figure 5).

Table 3. Top 0.5 Mb genome windows identified in the F_{ST} analysis between the two breeds. The windows are ranked according to the average F_{ST} value.

BTA[1]	Bin start[2]	Bin end[3]	No. of SNPs[4]	Average F_{ST}[5]	Genes[6]
18	14500001	15000000	18	0.344	*CPNE7, DPEP1, CHMP1A, CDK10, SPATA2L, VPS9D1, ZNF276, FANCA, SPIRE2, TCF25, MC1R, TUBB3, DEF8, DBNDD1, GAS8, U1, SHCBP1, VPS35*
7	45750001	46250000	15	0.328	*TCF7, SKP1, PPP2CA, CDKL3, UBE2B, CDKN2AIPNL, JADE2, SAR1B, U6, SEC24A, CAMLG*
22	13250001	13750000	15	0.310	*ENTPD3, RPL14, ZNF619, ZNF621, 7SK, U6*
6	1	500000	19	0.287	*U6, APELA*
7	45500001	46000000	19	0.286	*C7H5orf15, VDAC1, TCF7, SKP1, PPP2CA, CDKL3, UBE2B, CDKN2AIPNL*
18	14250001	14750000	17	0.281	*CDH15, SLC22A31, ANKRD11, SPG7, RPL13, CPNE7, DPEP1, CHMP1A, CDK10, SPATA2L, VPS9D1, ZNF276, FANCA, SPIRE2, TCF25, MC1R, TUBB3, DEF8*
16	42250001	42750000	10	0.276	*UBIAD1, MTOR, ANGPTL7, EXOSC10, SRM, MASP2, TARDBP*
8	93250001	93750000	21	0.259	*SMC2*
8	93000001	93500000	23	0.257	-
13	63250001	63750000	21	0.257	*CHMP4B, PXMP4, E2F1, ZNF341, NECAB3, RALY, EIF2S2, ASIP, AHCY*
16	44250001	44750000	6	0.256	*U1, GPR157, CA6, ENO1, U6*
11	2750001	3250000	28	0.249	*CNNM4, CNNM3, ANKRD23, ANKRD39, SEMA4C, COX5B, ACTR1B*
15	1	500000	22	0.243	-
6	30000001	30500000	22	0.238	*PDLIM5, 7SK, HPGDS, SMARCAD1*

6	35000001	35500000	28	0.234	*SNCA*
18	15500001	16000000	22	0.227	*NETO2, ITFG1, U6, PHKB*
15	250001	750000	23	0.225	-
6	94250001	94750000	15	0.224	*U6, ANTXR2*
6	65250001	65750000	30	0.224	*COX7B2, H4C14, GABRA4, GABRB1*
16	44000001	44500000	13	0.222	*SPSB1, H6PD, U1, GPR157, CA6*

[1] *Bos taurus* chromosome.

[2] Start position of the genome window in the ARS-UCD1.2 cattle genome version.

[3] End position of the genome window in the ARS-UCD1.2 cattle genome version.

[4] Number of SNPs included in the 0.5 Mb genome window.

[5] Average F_{ST} value based on SNPs included in the genome window.

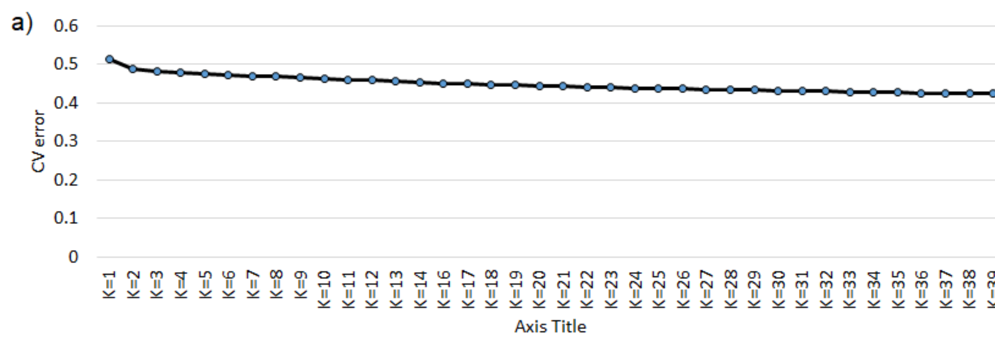
[6] Genes annotated in the reported genome window (ARS-UCD1.2 cattle genome version).

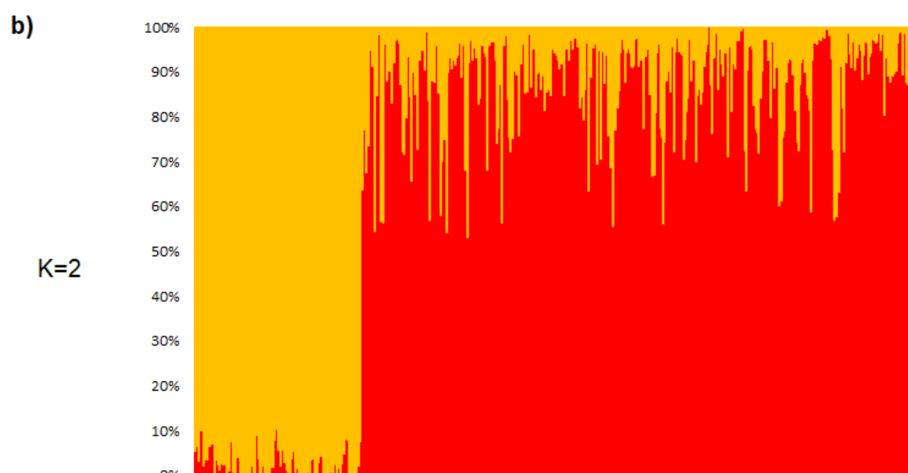
Figures

Figure 1. Pictures of Reggiana (a) and Modenese (b) sires.

Figure 2. Multidimentional scaling plots produced using genotyping information from each investigated cattle of the Reggiana (red dots) and Modenese (blue dots) breeds. Different components (C) are considered in the plots.

Figure 3. Results of the ADMIXTURE analysis. **a**) Cross validation (CV) error with K from 1 to 39. **b**) Plot distribution with K=2. **c**) Plot distribution with K=4.





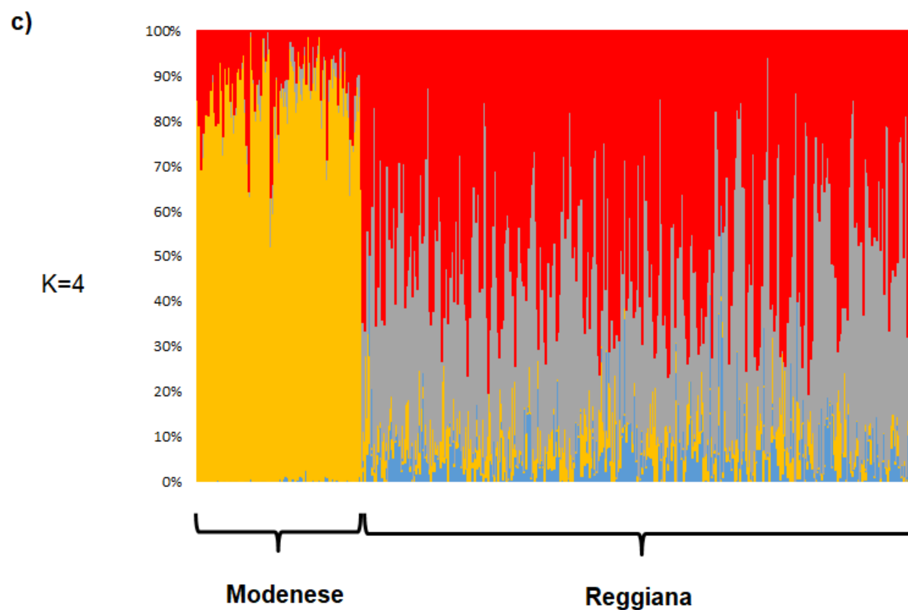


Figure 4. Population genomic parameters represented in the Reggiana (red points and lines) and in Modenese (blue points and lines) breeds: (**a**) linkage disequilibrium (LD) decay over distance; (**b**) average LD calculated for all autosomes; (**c**) effective population size (Ne) over the past generations.

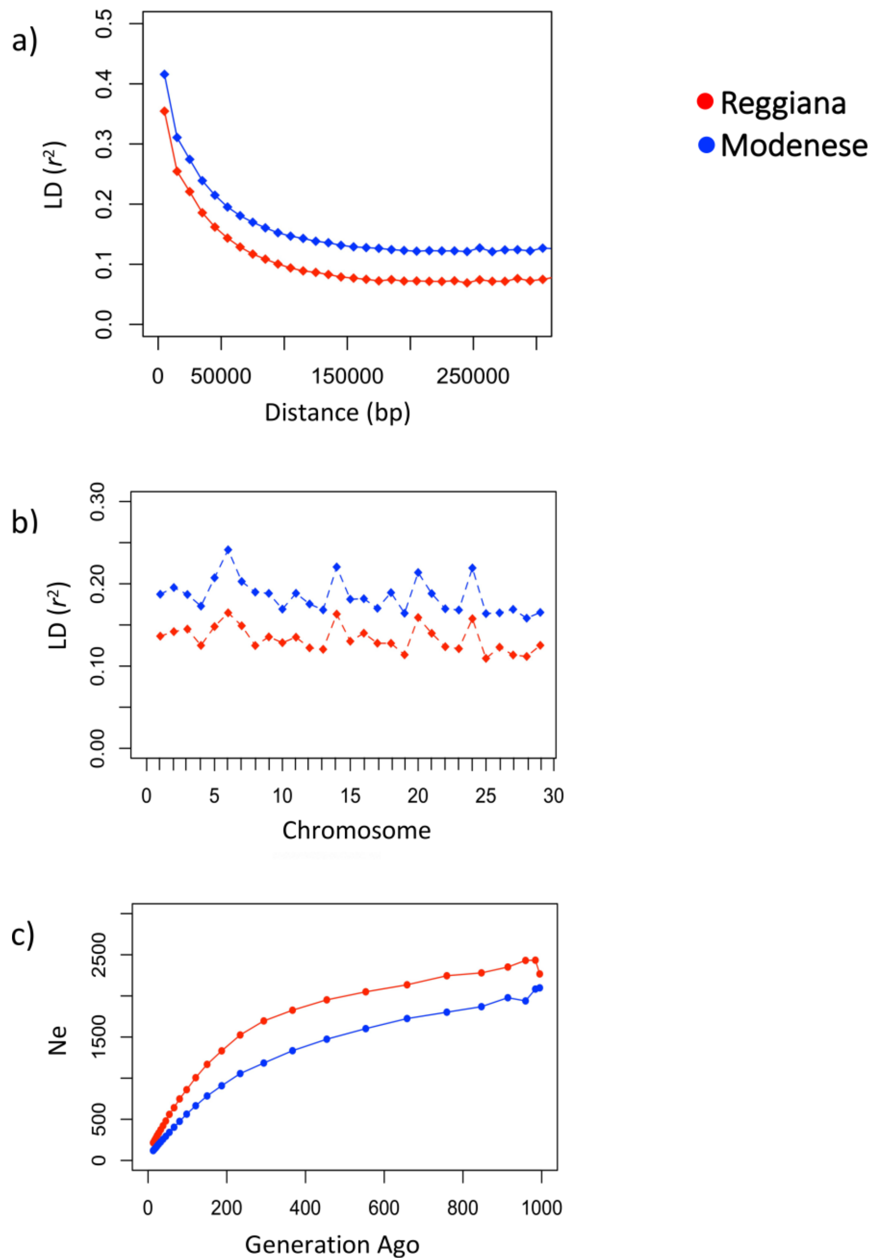
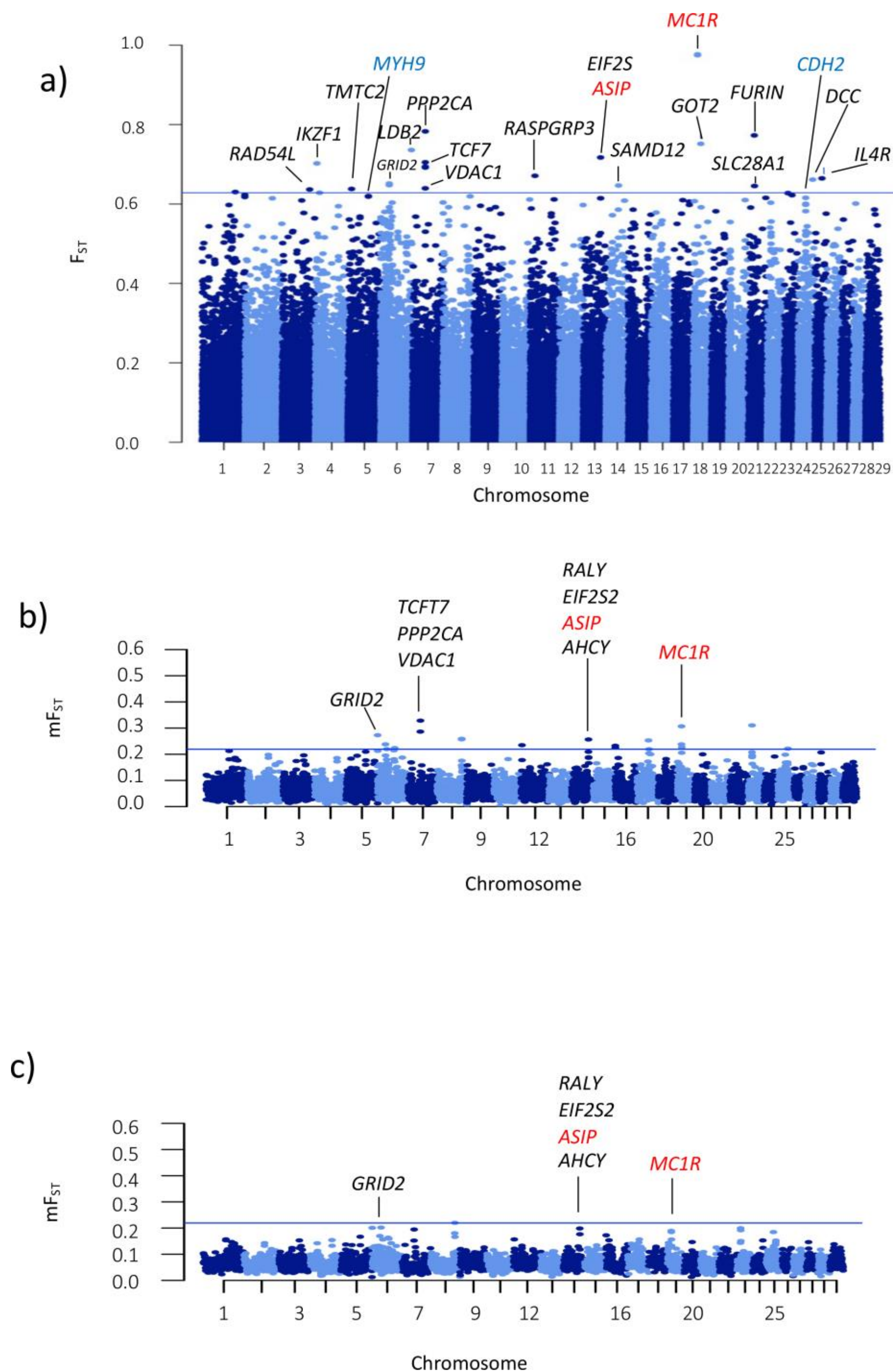


Figure 5. Manhattan plots obtained in the single-marker (**a**) and window-based F_{ST} analyses using windows of 0.5 Mb (**b**) or windows of 1 Mb (**c**), in which the y axis reports the mean F_{ST} values (mF_{ST}). In the single-marker analysis, the top 20 markers have been annotated, including two markers within the top 70 list which are close to genes (in blue) that have been also contained in windows detected with the window-based approaches. The regions detected with the window-based approaches (b and c) are annotated with the genes close to SNPs reported in the single-marker analysis. The two main coat colour genes are indicated in red. A few genes in the *ASIP* region on BTA13 identified in the window-based analyses are annotated.



Supplementary material

Figure S1. Distribution of the different classes of minor allele frequencies (MAF) for the markers genotyped in the Reggiana and Modenese breeds.

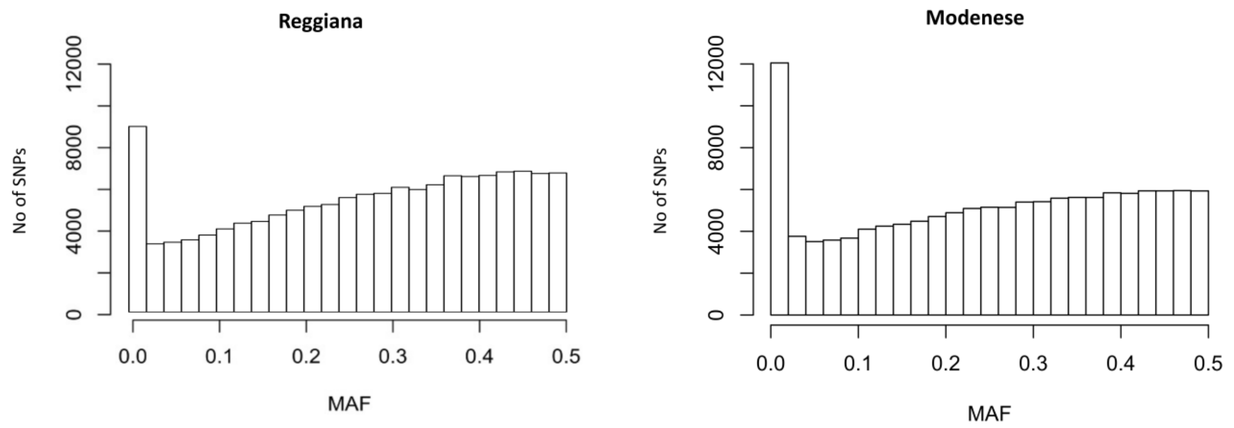


Figure S2. Linkage disequilibrium (r^2) plot of the chromosome 18 region encompassing the *MC1R* gene, obtained for the Reggiana and Modenese breeds. The genes annotated in this region are reported at the top of the two haploblocks.

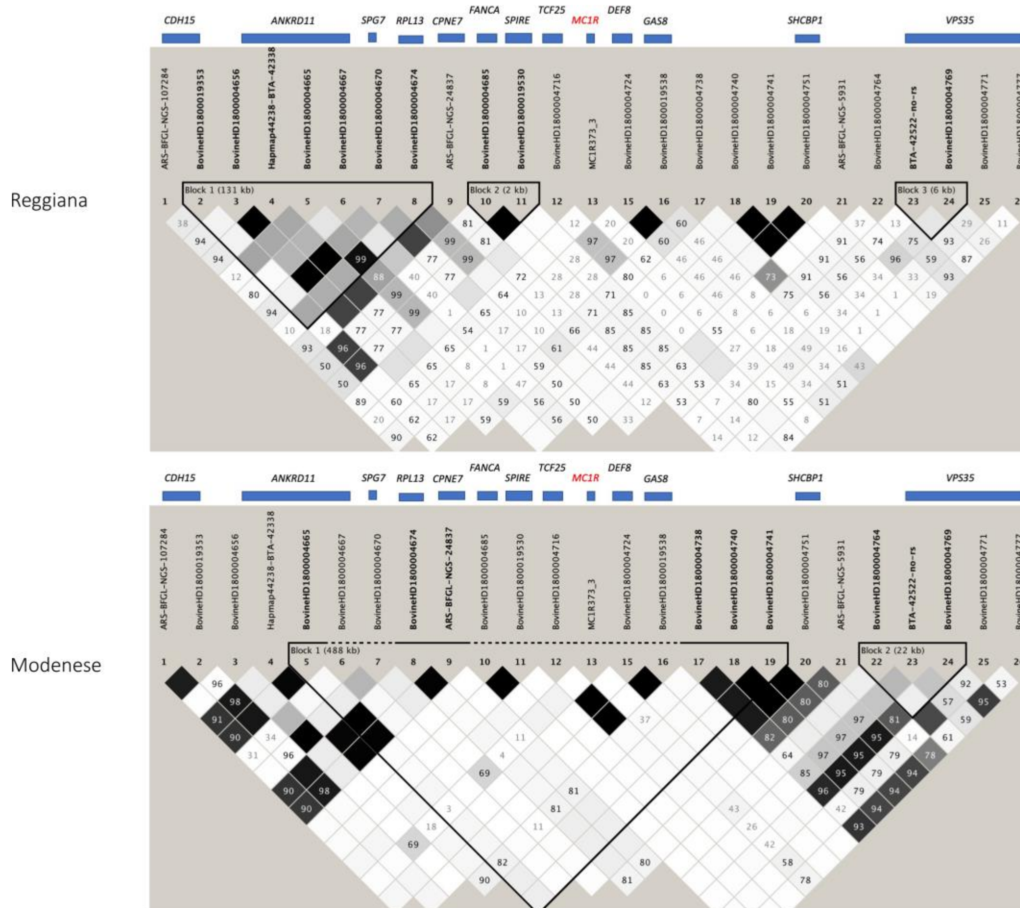


Figure S3. Linkage disequilibrium (r^2) plot of the chromosome 13 region encompassing the *ASIP* gene, obtained for the Reggiana and Modenese breeds. The genes annotated in this region are reported at the top of the two haploblocks.

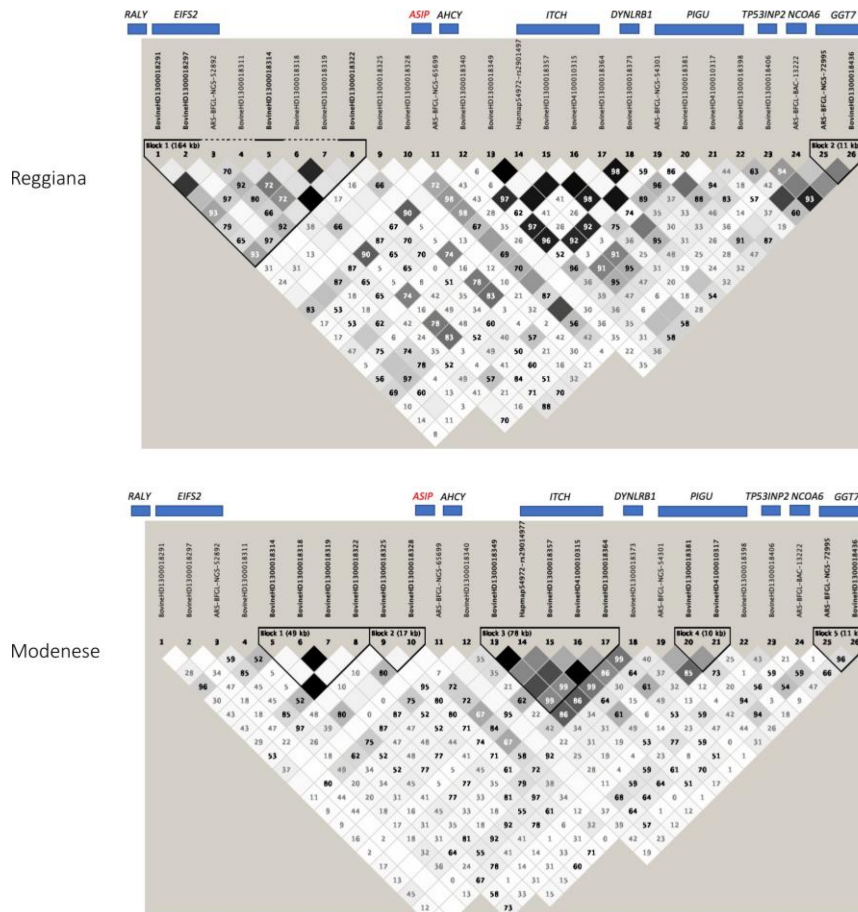


Table S1. The average linkage disequilibrium value (r^2) for the markers in all autosomes reported for the Reggiana and Modenese breeds, the standard deviation is reported in brackelets.

BTA[1]	r^2 Modenese (s.d.)	r^2 Reggiana (s.d.)
1	0.187 (±0.243)	0.136 (±0.205)
2	0.195 (±0.250)	0.142 (±0.210)
3	0.187 (±0.247)	0.145 (±0.222)
4	0.173 (±0.228)	0.125 (±0.199)
5	0.207 (±0.267)	0.148 (±0.228)
6	0.241 (±0.296)	0.165 (±0.248)
7	0.203 (±0.261)	0.149 (±0.227)
8	0.190 (±0.244)	0.125 (±0.195)
9	0.188 (±0.244)	0.135 (±0.210)
10	0.169 (±0.222)	0.128 (±0.193)
11	0.188 (±0.246)	0.135 (±0.203)
12	0.175 (±0.231)	0.122 (±0.192)
13	0.168 (±0.225)	0.120 (±0.191)
14	0.220 (±0.287)	0.163 (±0.250)

15	0.181 (±0.236)	0.130 (±0.204)
16	0.181 (±0.237)	0.140 (±0.212)
17	0.170 (±0.225)	0.128 (±0.196)
18	0.189 (±0.241)	0.128 (±0.199)
19	0.164 (0.217)	0.114 (±0.180)
20	0.213 (±0.273)	0.159 (±0.244)
21	0.188 (±0.245)	0.140 (±0.210)
22	0.170 (±0.226)	0.123 (±0.189)
23	0.169 (±0.231)	0.121 (±0.197)
24	0.219 (±0.302)	0.157 (±0.240)
25	0.164 (±0.218)	0.109 (±0.178)
26	0.165 (±0.219)	0.123 (±0.187)
27	0.169 (±0.223)	0.113 (±0.181)
28	0.158 (±0.209)	0.111 (±0.175)
29	0.165 (±0.221)	0.125 (±0.189)

[1] *Bos taurus* chromosome.

Table S2. All markers of the 99.8th percentile identified in the single-marker F_{ST} analysis between the two breeds.

Markers[1]	BTA[2]	Position[3]	F_{ST}	Closest gene (bp)[4]
MC1R	18	14705645	0.977	*MC1R* (0)
BovineHD0700013748	7	45833400	0.783	*PPP2CA* (9618)
ARS-BFGL-NGS-114140	21	21791054	0.773	*FURIN* (0)
ARS-BFGL-NGS-28154	18	26500840	0.752	*GOT2* (53100)
BovineHD0600033381	6	112511216	0.736	*LDB2* (357854)
BovineHD1300018297	13	63480254	0.717	*EIF2S2* (0), *ASIP* (182542)
BTA-78954-no-rs	7	45800275	0.705	*TCF7* (0)
ARS-BFGL-NGS-55059	4	5545419	0.702	*IKZF1* (0)
ARS-BFGL-NGS-5595	7	45766695	0.694	*TCF7* (2218)
ARS-BFGL-NGS-73679	7	45729837	0.692	*TCF7* (39076)
BTA-86548-no-rs	11	16591322	0.671	*RASPGRP3* (476598)
BovineHD2500007120	25	24908014	0.665	*IL4R* (0)
BovineHD2400015179	24	53014583	0.661	*DCC* (0)
BovineHD0600009128	6	31158986	0.652	*GRID2* (0)
BovineHD0600009122	6	31135482	0.647	*GRID2* (0)
ARS-BFGL-NGS-35081	14	46102133	0.647	*SAMD12* (36424), *EXT1* (62323)
BovineHD2100006752	21	22531247	0.645	*SLC28A1* (0)
ARS-BFGL-NGS-20141	7	45691037	0.639	*VDAC1*(0)
BovineHD0500003920	5	12981358	0.638	*TMTC2* (405765)
ARS-BFGL-NGS-16203	3	99840480	0.637	*RAD54L* (0)
Hapmap55173-rs29024142	1	121487166	0.630	*PLSCR5* (365630)
Hapmap50265-BTA-13206	4	15203507	0.628	*ASNS* (0)
ARS-BFGL-NGS-25325	23	14974356	0.627	*APOBEC2* (0)
BovineHD0100046358	1	156520533	0.622	*KCNH8* (0)
BTA-94041-no-rs	23	29663663	0.622	*ORP2* (0)
BovineHD0500021352	5	74767531	0.620	*MYH9* (0)
BovineHD0500021349	5	74753566	0.620	*MYH9* (0)

BovineHD0500021332	5	74715894	0.620	*MYH9* (0)
BTB-00370549	8	99625705	0.619	*PALM2* (96246)
ARS-BFGL-NGS-40570	5	74717288	0.618	*MYH9* (0)
BovineHD0100046348	1	156507162	0.617	*KCNH8* (0)
BovineHD2400007601	24	27560729	0.616	-
Hapmap24484-BTA-136154	21	63972179	0.616	*BCL11B* (221357)
BovineHD1300018322	13	63629244	0.615	*ASIP* (33552)
BovineHD0200028587	2	98950479	0.614	*ERBB4* (273296)
BovineHD1300018314	13	63579877	0.614	*ASIP* (82919)
ARS-BFGL-NGS-43715	11	86912838	0.612	*ATP6V1C2* (0)
BovineHD1000029475	10	100564643	0.611	*TTC8* (2400)
BovineHD2000016914	20	60190001	0.610	-
BTB-01307961	17	33562176	0.609	*ANKRD50* (393929)
BovineHD0300020926	3	70918424	0.609	*LRRIQ* (200093)
ARS-BFGL-BAC-20448	14	67928364	0.608	*PTDSS1* (0)
BovineHD1800000963	18	3286986	0.607	*CNTNAP4* (133529)
BovineHD1700016954	17	57518753	0.606	*KSR2* (0)
BovineHD0800000070	8	503066	0.604	*ANXA10* (0)
BovineHD0600011794	6	41996260	0.604	*ADGRA3* (0)
BovineHD2200003649	22	12417613	0.603	*WDR48* (0)
Hapmap41322-BTA-64648	28	9215236	0.601	*HEATR1* (0)
BovineHD2400007580	24	27449561	0.600	-
BovineHD4100016540	24	27570575	0.598	*CDH23* (133529)
ARS-BFGL-NGS-112404	17	57495561	0.598	*KSR2* (0)
BovineHD0100028276	1	98305974	0.597	*MECOM* (0)
BovineHD1200026502	12	86367927	0.597	*ATP11A* (0)
BTB-01073083	9	24493304	0.595	*RASPO3* (306182)
BovineHD0400023583	4	84535781	0.594	*KCND2* (132580)
BovineHD4100004382	6	31590806	0.592	*GRID2* (50650)
BovineHD0900022313	9	79439745	0.592	*NMBR* (138468)
BovineHD0800006437	8	20952293	0.592	-

BovineHD1800008609	18	27586155	0.592	-
BovineHD2100001969	21	8498932	0.591	*ARRDC4* (235849)
BovineHD0800026512	8	87886263	0.591	-
BovineHD1100001176	11	3287227	0.589	*TMEM131* (0)
BovineHD0900016418	9	58935377	0.586	-
BovineHD2900009743	29	32109387	0.586	*GHR* (0)

[1] Marker name in the GeneSeek GGP Bovine 150k SNP chip.

[2] *Bos taurus* chromosome.

[3] Position, in base pairs, of the marker in the ARS-UCD1.2 cattle genome version.

[4] Distance in base pairs of the marker with the indicated gene is reported within the brackets. When the marker overlaps the gene, a distance equal to 0 bp is indicated.

Table S3. Top 1 Mb genome windows identified in the F_{ST} analysis between the two breeds. The windows are ranked according to the average F_{ST} value.

BTA[1]	Bin start[2]	Bin end[3]	No. of SNPs[4]	Average F_{ST}[5]	Genes[6]
8	93000001	94000000	42	0.255	*SMC2*
6	34500001	35500000	51	0.213	*CCSER1, MMRN1, SNCA*
18	14000001	15000000	39	0.209	*CDT1, APRT, GALNS, TRAPPC2L, CBFA2T3, ACSF3, CDH15, SLC22A31, ANKRD11*
18	14500001	15500000	37	0.206	*CPNE7, DPEP1, CHMP1A, CDK10, SPATA2L, VPS9D1, ZNF276, FANCA, SPIRE2, TCF25, MC1R, TUBB3, DEF8, DBNDD1, GAS8, U1, SHCBP1, VPS35, ORC6, MYLK3, C18H16orf87, GPT2, DNAJA2, NETO2*
6	1	1000000	50	0.205	*APELA*
22	13000001	14000000	41	0.199	*EIF1B, ENTPD3, ZNF619, RPL14, ZNF621, CTNNB1, ULK4*
7	45500001	46500000	41	0.199	*C7H5orf15, VDAC1, TCFT7, SKP1, PPP2CA, CDKL3, UBE2B, CDKN2AIPNL, JADE2, SAR18, SEC24A, CAMLG, DDX46, PCBD2, TXNDC15, C7H5orf24, CATSPER3, PITX1*
13	63000001	64000000	38	0.198	*C13H20orf144, CHMP4B, PXMP4, E2F1, ZNF341, NECAB3, RALY, EIF2S2, ASIP, AHCY, ITCH, DUNLRB1, PIGU, MP1LC3A, NCOA6, TP53INP2*
22	13500001	14500000	46	0.192	*CTNNB1, ULK4, TRAK1, CCK*
24	27500001	28500000	77	0.189	*CDH2*
8	92500001	93500000	49	0.180	*CYLC2*

16	44000001	45000000	18	0.178	*SPSB1, H6PD, GPR157, CA6, ENO1, RERE, SLC45A1*
13	63500001	64500000	35	0.176	*ASIP, AHCY, ITCH, DYNLRB1, MAP1LC3A, PIGU, TP53INP2, NCOA6, GGT7, ACSS2, GSS, MYH7B, bta-mir-499, TRPC4AP, EDEM2, PROCR*
15	1	1000000	41	0.175	-
6	30000001	31000000	49	0.173	*PDLIM5, 7SK, Metazoa_SRP, HPGDS, SMARCAD1, ATOH1, GRID2*
8	93500001	94500000	31	0.171	*SMC2, OR13C3, OR13C8*
5	74000001	75000000	52	0.165	*RBFOX2, APOL3, MYH9, U6, TXN2, FOXRED2, EIF3D, CACNG2*
23	38500001	39500000	64	0.163	*5S_rRNA, RNF144B, DEK, KDM1B, 5S_rRNA, TPMT, NHLRC1, U6*
6	62000001	63000000	45	0.163	*KCTD8*
11	2500001	3500000	65	0.163	*NEURL3, KANSL3, ARID5A, U6, FER1L5, LMAN2L, CNNM4, CNNM3, ANKRD23, ANKRD39, SEMA4C, COX5B, ACTR1B, ZAP70, TMEM131, VWA3B*

[1] *Bos taurus* chromosome.

[2] Start position, in base pairs,of the genome window in the ARS-UCD1.2 cattle genome version.

[3] End position, in base pairs, of the genome window in the ARS-UCD1.2 cattle genome version.

[4] Number of single nucleotide polymorphisms (SNPs) included in the 1 Mb genome window.

[5] Average F_{ST} value based on the SNPs included in the genome window.

[6] Genes annotated in the reported genome window (ARS-UCD1.2 cattle genome version).

Table S4. Results of the gene enrichment analysis. Significant results were obtained only with the Human GWAS catalog. The most significant KEGG pathways and GO Biological Processes are also
reported.

Library	**Term**	**Overlap**[1]	***p*-value**[2]	**Genes**[3]
GWAS catalog	Skin sensitivity to sun	2/6	0.016	*MC1R*, *ASIP*
	Facial pigmentation	2/7	0.016	*MC1R*, *ASIP*
	Skin colour saturation	2/9	0.017	*MC1R*, *ASIP*
	Tanning	2/11	0.017	*MC1R*, *ASIP*
	Freckles	2/12	0.017	*MC1R*, *ASIP*
	Skin aging (microtopography measurement)	2/13	0.017	*MC1R*, *ERBB4*
	Skin pigmentation	2/13	0.017	*ASIP*, *EIF2S2*
	Non-melanoma skin cancer	2/23	0.040	*MC1R*, *ASIP*
	Post bronchodilator FEV1/FVC ratio	4/206	0.040	*GRID2*, *PALM2*, *HEATR1*, *KSR2*
	Low tan response	2/24	0.040	*MC1R*, *ASIP*
	Feeling fed-up	2/29	0.049	*ERBB4*, *DCC*
	Monoclonal gammopathy of undetermined significance	2/29	0.049	*ERBB4*, *KSR2*
KEGG pathways*	Melanogenesis	3/101	0.137	*MC1R*, *TCF7*, *ASIP*
GO Biological Process*	Regulation of tyrosine phosphorylation of STAT protein (GO:0042509)	3/68	0.147	*PPP2CA*, *GHR*, *ERBB4*

[1] Number of genes of the input set over the number of genes annotated with the term.

[2] Adjusted *p*-value.

[3] Genes of the input set annotated with the term.

* Not statistically valid (adjusted *p*-value > 0.05).

Chapter 6: research activity on pig

6.1 Genome-wide association studies for iris pigmentation and heterochromia in Large White pigs identified several genomic regions affecting eye colours

This work has been publish by the Journal Animal Genetics (2020),51(3), pp. 409-419; doi: 10.1111/age.12930. The current text is still not in the final format owned by the journal.

Introduction

Eye colour genetics has been extensively studied in humans and animals since the rediscovery of the Mendel's laws, at the beginning of the last century. After the first report of Davenport and Davenport (1907), who defined a simplistic monogenic dominant-recessive model of inheritance for brown-blue eye transmission in humans, it became clear that eye colours can be better explained following quantitative genetic rules, as several genes are involved in determining iris pigmentation (Sturm & Frudakis 2004; Sturm & Larsson 2009).

The iris is a small connective and smooth muscle structure that controls the amount of light captured by the pupil, through contraction or dilation in bright or dark conditions. It is anatomically composed by two layers of different embryological origin (Rennie 2012). The physical basis of iris colour variation is mainly determined by variable amounts and qualities of the melanosome particles where the melanin pigments (eumelanin and pheomelanin) are packaged in different ratios and quantity within the iridial stromal melanocytes (of the anterior layer), which are of the same neural crest embryological origin as the dermal melanocytes. The iridal posterior layer (known as the iris pigment epithelium) is neuroectodermal in origin, deriving from the anterior extremity of the optic cup. This layer is not pigmented only in case of albinism (Rennie 2012).

Heterochromia is an alteration of iris colour and structure that can have congenital origin. The colour change may involve one eye alone or both eyes and may be partial, segmental or complete (Gladstone 1969). Heterochromia iridis (or sectorial heterochromia) arises when distinct areas of the same iris have different colours. This condition may be unilateral or bilateral when one eye or both eyes are interested, respectively. Heterochromia iridum (binocular heterochromia) is considered when the whole iris of the two eyes have different colours (Rennie 2012). In humans, heterochromia iridis and iridum have been described in a few complex defects, including Waardenburg syndrome type I, II, III and IV variants, congenital Horner syndrome and piebaldism (Read & Newton 1997; Rennie 2012). Iris heterochromia has been also described in livestock species, namely cattle, water buffalo and

horse where it seems sporadic (Huston *et al.* 1968; Kaswan *et al.* 1987; Misk *et al.* 1998; Harland et al. 2006). This eye alteration was estimated to occur in about 5-16% of pigs and determined by a supposed recessive allele at the *Het* locus (Yoshikawa 1935; Dürr 1937; Searle 1968; Gelatt *et al.* 1973). Heterochromia is not however the only source of eye colour variability in pigs. Therefore, following what emerged from more recent studies in humans that applied genome wide association analyses on eye colours (Sturm and Frudakis 2004; Sturm and Larsson 2009), it could be expected that more complex genetic mechanisms might be also involved in explaining iris pigmentation diversity also in *Sus scrofa*.

Among livestock species, a few genomic studies have been recently carried out in several cattle breeds and horses to identify genes determining coat colour anomalies which also affect iris pigmentation (Philipp *et al.* 2011; Haase *et al.* 2015; Rothammer *et al.* 2017). To our knowledge, there is no similar investigation in pigs focused on iris colours.

In this study we analysed eye colour diversity in a white pig breed (i.e. Italian Large White) and report the results of genome-wide association studies based on several comparisons which demonstrated that the within breed variability for this exterior phenotype is controlled by several loci.

Materials and methods

Animals and eye colour phenotypes

All animals used in this study were kept according to the Italian and European legislations for pig production. All procedures described were in compliance with Italian and European Union regulations for animal care and slaughter. Pigs were not raised or treated in any way for the purpose of this study. Eye colours were registered just after slaughtering of the animals in a commercial abattoir that followed standard procedures for slaughtering heavy pigs at the commercial weight and age.

Eye colours were recorded in 897 nine months-old performance tested Italian Large White pigs (Fontanesi *et al.* 2014; Bovo *et al.* 2019), adapting a classical score system for iris colours developed in humans, based on different colour categories (Sturm & Frudakis 2004). Four main colour categories were considered in this pig population, according to the observed iris colour variability. Three of which included pigmented eyes with different grades of brown: a) pale brown; b) medium brown; c) dark brown. Another category included both eyes completely depigmented (pale grey, having a marked marginal area. In addition, heterochromia patterns were scored considering d) heterochromia iridis, i.e. depigmented/pale grey iris sectors in brown irises of one of the three grades (present in only one eye: unilateral; or present in both

eyes: bilateral) and e) heterochromia iridum, i.e. a whole eye iris of depigmented/pale grey colour and the other eye with the iris completely of one of the three grades of brown. Figure 1 shows examples of eye colours described above.

Genotyping data

DNA was extracted from blood (collected at the slaughterhouse during jugulation), using the Wizard Genomic DNA Purification kit (Promega Corporation, Madison, WI, USA). Animals were then genotyped with the Illumina PorcineSNP60 BeadChip v.2 (Illumina Inc., San Diego, CA, USA), which interrogates a total of 61565 single nucleotide polymorphisms (SNPs), using standard procedures. Genotype calls were conducted by using the Genotyping Module in GenomeStudio software 1.0.2.20706 (Illumina Inc.). Genotypes with an Illumina GenCall score (GC; GenCall Version 6.3.0) below 0.15 were assigned as missing. PLINK 1.9 software (Chang *et al.* 2015) was used for quality check. Samples with a genotype missing rate > 0.9 were discarded while SNPs were discarded if they presented a call rate < 0.9, a Hardy-Weinberg equilibrium (HWE) p-value < 0.0001 and a minor allele frequency (MAF) < 0.01. Only SNPs located on autosomal chromosomes were retained. After filtering, the dataset comprised a total of 857 animals (Table 1).

Genome wide association analyses

To dissect the potential different genetic mechanisms affecting the recorded eye colour variability, seven genome-wide association analyses (GWAS1-7) were carried out including distinct groups of pigs as detailed below:

1) GWAS1: Pigs having all three categories of brown grades, excluding pigs with bi-ocular depigmented eyes;
2) GWAS2: Pigs having completely depigmented eyes *vs* pigs having completely dark brown eyes;
3) GWAS3: Pigs having completely depigmented eyes *vs* all pigs having completely brown eyes (pale, medium or dark brown);
4) GWAS4: Pigs having heterochromia iridis *vs* all pigs having completely brown eyes (pale, medium or dark brown);
5) GWAS5: Pigs having heterochromia iridum *vs* all pigs having completely brown eyes (pale, medium or dark brown);
6) GWAS6: Pigs having heterochromia patterns (iridum and iridis) *vs* all pigs having completely brown eyes (pale, medium or dark brown);

7) GWAS7: Pigs having completely depigmented eyes + pigs having heterochromia patterns (iridum + iridis) *vs* all pigs having completely brown eyes (pale, medium or dark brown).

The number of pigs included in these comparisons are reported in Table S1.

Genome-wide association analyses were carried out following a linear mixed effect model:

$$\boldsymbol{y} = \boldsymbol{W\alpha} + \boldsymbol{x}\beta + \boldsymbol{u} + \boldsymbol{e} \qquad (1)$$

where **y** ($n \times 1$) is a vector containing parameter for n^{th} animal, $\boldsymbol{W}$ ($n \times k$) is a matrix of a covariates with $k = 2$ (a column of 1s and a dummy variable coding the two operators involved in the animal phenotyping), $\boldsymbol{\alpha}$ is the k-dimensional vector of covariates effects, $\boldsymbol{x}$ ($n \times 1$) is the vector containing genotypes for the ith SNP (coded as 0, 1, 2, according to the number of copies of the minor allele), β is the additive fixed effect of the ith SNP on the trait, $\mathbf{u} \sim N(\mathbf{0}, \sigma^2_u \mathbf{K})$ is a multivariate Gaussian polygenic effect, with covariance matrix proportional to the relatedness matrix **K** ($n \times n$) and $\mathbf{e} \sim N(\mathbf{0}, \sigma^2_e \mathbf{I})$ is a multivariate Gaussian vector of uncorrelated residuals. The assessment of the association between each SNP and trait was obtained by testing the null hypothesis $H_0{:}\beta = 0$. Significance was tested by using the Wald test. All analyses were performed using GEMMA v. 0.96 (Zhou *et al.* 2012) after computing the relatedness matrix **G** as a centred genomic matrix controlling the population structure. A Bonferroni's corrected threshold equal to a nominal value of 0.05 was used to define significant markers. Details are reported in Table S1. For each trait GEMMA estimated from the whole set of available genotypes the chip heritability (or SNP heritability; h^2_{SNP}).

QQplots and Manhattan plots were generated in R v. 3.5.1 by using the "qqman" package while the genomic inflation factors (λ) were computed with the function "estlambda" function within the "GenABEL" package and reported in Table S1. Figure S1 reports the corresponding QQplots.

Genes annotated in the Sscrofa11.1 genome version spanning a region of ± 500 kb around all significant and suggestively significant SNPs were retrieved using Ensembl Biomart tool (http://www.ensembl.org/biomart/martview/) and then considered relevant in affecting the associated phenotypes according to a detailed analysis of the literature.

Results

Table 1 reports the number of pigs that were scored for the described eye colours and heterochromatic patterns. About 17.9%, 14.8% and 54.3% of the phenotyped pigs had both

eyes classified as pail brown, medium brown and dark brown, respectively. Both completely depigmented eyes were identified in about 3.8% of the pigs. The proportion of pigs with heterochromia iridis and heterochromia iridum was equal to 3.2% and 5.9%, accounting for a total of 9.1% of animals having heterochromia patterns. Bilateral heterochromia iridis was less frequent than the mono-lateral defect (Table 1). No effect of the sex could be evidenced as the proportion of females and males belonging to the same eye colour classes or heterochromatic pattern did not statistically differ from the general proportion of the two sexes in the whole recorded population, i.e. about ⅔ of females and ⅓ of males (P >0.05; Chi-Square test).

Considering the observed heterogeneity on these iris phenotypes, it is possible that different biological mechanisms and therefore different genes could be involved in determining the variability recorded in the Italian Large White population. Therefore, to dissect the potential underlying genetic mechanisms affecting this iris colour diversity, seven genome-wide association studies (GWAS1-7) were carried out based on groups of pigs showing different scored eye colour classes, depigmented or heterochromatic patterns (Table S1).

The genomic inflation factors (λ) in these studies ranged from 1.013 to 1.067 indicating that stratification was corrected for all analyses (Table S1). The h^2_{SNP} ranged from 0.095 (in the genome-wide association study that compared heterochromia iridis against brown eyes: GWAS5) to 0.504 (in the genome-wide association study that included the three brown colour types: GWAS1; Table S1). Table 2 list the most significant markers identified in the seven genome-wide association studies that were analysed. Table S2 reports all significant markers identified in the all genome-wide association studies that were carried out.

The genome-wide association study (GWAS1) including pigs categorized according to the eye colour based on the three brown classes evidenced a significant SNP (ALGA0110382, rs81338630: $P = 3.71\times10^{-08}$) on SSC16 at position 19997882 within the *solute carrier family 45 member 2* (*SLC45A2*) gene (Table 2 and Figure 2). SLC45A2 (also known as MATP) is a transporter protein that mediates melanin synthesis by playing a role in the distribution and processing of tyrosinase and other enzymes of the pigmentation machinery (Costin *et al.* 2003; Cook *et al.* 2009). Mutations in the human *SLC45A2* gene determine variations in skin, hair and eye colour (e.g. Wilde *et al.* 2014).

To dissect the different depigmented eye phenotypes, several other genome-wide association studies were defined based on different groups of pigs with depigmented and pigmented iris patterns. The study that included pigs having extreme iris colours, i.e. pigs having both eyes with depigmented irises *vs* pigs with dark brown irises (GWAS2; Table S1), evidenced a main significant SNP peak on porcine chromosome (SSC) 11 (Table 2 and Figure

2). The most significant SNP (INRA0036477 or rs334119989; P = 4.40×10^{-09}) was an intergenic marker at position 50,141,902 close to the *endothelin receptor type B* (*EDNRB*) gene, annotated from positions 50073161 to 50102814 bp. EDNRB is well known to be involved in the regulation processes that control the development of several cell types from the neural crest, which in turn affect pigmentation at different levels, together with several other physiological effects (e.g. Hosoda *et al.* 1994). Another significant marker was localized on SSC6 (SIRI0000289 or rs334333448) at position 13446699, in a region without any obvious candidate gene.

The third genome-wide association study (GWAS3) was similar to the previous one and compared pigs having complete bi-ocular depigmented irises *vs* all pigs with both pigmented eyes (with pale brown, medium brown and dark brown irises). The results confirmed the same significant SNP peak on SSC11 of the previous analysis (Table 2 and Figure 2), with the same most significant marker (INRA0036477; P = 1.06×10^{-08}). Another significant region was identified on SSC4 with a SNP (ALGA0027376 or rs81382510; position 101085844) located in an intron of the *notch receptor 2* (*NOTCH2*) gene (Table 2 and Figure 2). Notch signaling regulates a vast spectrum of fundamental developmental processes that also affect eye components (Mašek & Andersson 2017).

The genome-wide association study that investigated pigs with heterochromia iridis *vs* pigs with completely brown eyes (i.e. from all three brown classes; GWAS4; Table S1) identified four significant SNP peaks, one on SSC5, two on SSC6 (separated by about 5 Mbp) and one on SSC14 (Table 2 and Figure 2). The significant region on SSC5 (with ASGA0026910 at position 94284630, being the most significant marker: P = 1.01×10^{-11}) encompasses the *KIT ligand* (*KITLG*) gene (positions: 94017384-94110216), which encodes the ligand of the tyrosinase-kinase receptor produced by the *KIT* gene and known to have a key role in cell development and migration with demonstrated effects on pigmentation (Wehrle-Haller 2003; Amyere *et al.* 2011; Picardo & Cardinali 2011). The first peak on SSC6 (at about 25.0 Mbp) was evidenced in a region with few annotated genes of unknown functions. The most significant markers of the first SSC6 region (ALGA0035018 and ASGA0027932; both with P = 1.27×10^{-07}) were within or just a few hundred bp upstream the *solute carrier family 6 member 2* (*SLC6A2*) gene, whose known functions might not be directly related to the described phenotype. No other obvious candidates could emerge in this region according to the literature survey on the function of the mapped genes. The same could be for the second significant region on SSC6, at about 24.7 Mbp, for which no obvious candidate could be identified. Another significant marker was located on SSC14 (ASGA0066255, at position 114987510) within the

collagen type XVII alpha 1 chain (*COL17A1*) gene. Mutations in this gene are associated with epidermolysis bullosa (characterized by mucocutaneous blistering and chronic epithelial fragility) also associated with pigmentation defects (Pasmooij *et al.* 2012).

The genome-wide association study that included pigs with heterochromia iridum *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown; GWAS5) identified a peak on SSC8, within the *electron transfer flavoprotein dehydrogenase* (*ETFDH*) gene which encodes for a component of the mitochondrial electron-transfer system. No obvious function related to the investigated ocular defect could be retrieved by searching the literature for this gene or other genes in this chromosome region.

To further mine the potential genetic mechanisms affecting the heterochromia patterns in *Sus scrofa* and supposing that the two heterochromia defects could share, at least in part, similar developmental patterns and mechanisms (as also hypothesized by the old literature in this field: Yoshikawa 1935; Dürr 1937; Gelatt *et al.* 1973), pigs having heterochromia iridis and heterochromia iridum were grouped together and their SNP data were contrasted with those of pigs with completely pigmented eyes (i.e. pale brown, medium brown and dark brown; GWAS6 in Table S1). Results of this combined study confirmed the significant peak of the *KITLG* gene region on SSC5, already observed in GWAS4 for heterochromia iridis, in addition to significant markers on SSC4 at position ~67.0 Mbp, within or just upstream the *phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2* (*PREX2*) gene (Table 2 and Figure 2). Frequent mutations in *PREX2* are amongst the most common causes of human melanoma (Berger *et al.* 2012).

In addition to this combined analysis for the heterochromia defects, another genome-wide association analysis was conducted grouping all pigs with eye depigmented patterns (i.e. having both depigmented/pale grey eyes and the two heterochromia types) in comparison with pigs having pigmented (pale, medium or dark brown) eyes (GWAS7 in Table S1). This study hypothesized that similar mechanisms might determine all depigmentation defects (Yoshikawa 1935). In this analysis, significant markers were only on SSC11 (*EDNRB* region) and SSC5 (*KITLG* region), the most relevant regions already observed in the separated studies for both depigmented eyes (GWAS2 and GWAS3) and heterochromia iridis (GWAS4), respectively. These two different traits dominated the result in this combined analysis, further confirming that completely depigmented eyes and partially depigmented eyes are mainly affected by different genetic mechanisms, involving developmental processes regulated by EDNRB and KITLG, respectively.

Discussion

Several studies in humans and other species already indicated that eye colours and heterochromia patterns are complex traits that are determined by variants at several loci explaining different biological mechanisms of pigmentation and developments in mammals (e.g. Sulem *et al.* 2007; Kayser *et al.* 2008; Liu *et al.* 2010; Deane-Coe *et al.* 2018). Variants in several genes affecting hair and skin pigmentation have been also associated with eye colour in humans (Han et al. 2008; Walsh et al. 2017; Wollstein et al. 2017). Moreover, eye heterochromia defects have been frequently reported in several multi-trait syndromes caused by mutations showing pleiotropic effects and, in most cases, incomplete penetrance (e.g. Pingault *et al.* 2010).

Iris colour variability is also present in pig populations, particularly in white pigs in which a few previous studies reported frequent eye heterochromatic patterns (Yoshikawa 1935; Dürr 1937; Searle 1968; Gelatt *et al.* 1973). The percentage of pigs with depigmented eye sections (i.e. both completely depigmented eyes, heterochromia iridis and heterochromia iridum) reported in our study (12.9%) was similar to what was reported in descriptive investigations involving other white pig populations not purposely developed to study eye pigmentation defects and in which these three classes were not well separated (Yoshikawa 1935; Gelatt *et al.* 1973). The frequency of white pigs with depigmented eyes, resulting by crossing animals with heterochromia iridis, increased reaching about 38% in a Miniature pig line in which an autosomal recessive locus was suggested to be the main cause of this defect (Gelatt *et al.* 1973). Results from these studies could be better explained including incomplete penetrance, suggesting that other modifier loci might be involved in determining the heterogeneity of the different heterochromatic patterns in pigs.

In our genome-wide association studies we categorized the Italian Large White pigs into six different groups based on their eye pigmentation phenotypes (three groups of pigs with fully pigmented irides: pale brown, medium brown and dark brown; three groups of pigs with depigmented irides; both depigmented/pale grey eyes, heterochromia iridis and heterochromia iridum) and analysed these pig cohorts to dissect the heterogeneity of eye pigmentation.

The results indicated that *SLC45A2*, *EDNRB* and *KITLG* are the main candidates to affect the different grades of brown pigmentation of the eyes, the bilateral eye depigmentation defect and the heterochromia iridis defect recorded in these white pigs, respectively.

Studies in humans have already reported that variants in the *SLC45A2* gene affect iris pigmentation, in addition to effects of hair and skin colouration (Branicki *et al.* 2008; Walsh *et al.* 2017; Wollstein *et al.* 2017). Several alleles and haplotypes have been indicated to influence

different levels of darkness of the observed colouration (Han *et al.* 2008; Fracasso *et al.* 2017), similarly to what we might expect from the observed different grades of brown colouration intensity in which we classified the Italian Large White pigs. It is worth mentioning that the way in which we classified the level of brown pigmentation in the eyes is an approximation that followed a putative continuous distribution, from pale to dark (almost black) brown. Other improvements of this study could be derived by a more precise exploitation of the full spectrum of eye colouration using more precise recording methods including digital photographic systems, as already proposed in humans (Wollstein *et al.* 2017), but of difficult application in field experiments in pigs. In other domestic species, mutations in the *SLC45A2* gene determine dilution of coat colour and oculocutaneous albinism (Gunnarsson *et al.* 2007; Caduff *et al.* 2017; Rothammer *et al.* 2017; Holl *et al.* 2019; Sevane *et al.* 2019).

Mutations in the *EDNRB* gene in heterozygous or homozygous condition have been frequently reported to determine several human syndromes with eye colour defects and heterochromia (Syrris *et al.* 1999; Pingault *et al.* 2001; Verheij *et al.* 2003; Issa *et al.* 2017; Morimoto *et al.* 2018). A few studies in pigs have already indicated that variants in the *EDNRB* affect different coat colour spotted patterns in Chinese breeds and in traditional European breeds (Ai *et al.* 2013; Wilkinson *et al.* 2013; Lü *et al.* 2016; Wang *et al.* 2015, 2018; Zhang *et al.* 2018) but no information was reported on eye colour phenotypes. A missense mutation in the *EDNRB* causes the Lethal White Foal Syndrome in horses (Metallinos *et al.* 1998; Santschi *et al.* 1998; Yang *et al.* 1998) but in no other livestock species mutations in this gene have been associated with eye pigmentation defects.

In humans, thus far only a study identified a few mutations in the *KITLG* gene determining eye heterochromia as a symptom of a complex form of Waardenburg Syndrome Type 2 (Zazo Seco *et al.* 2015). A selection signature in the *KITLG* gene region was reported in Berkshire and Meishan pig breeds compared to other European breeds pointing out a role of this gene in determining coat colour patterns specific for the first two breeds (Wilkinson *et al.* 2013). A missense mutation in the *KITLG* gene causes the roan coat colour in Belgian Blue and Shorthorn cattle breeds (Seitz *et al.* 1999) and variants in this gene may determine the same colour phenotype in goats (Talenti *et al.* 2018) but no reports have identified mutations in this gene affecting eye colour in any other animal species.

The results we obtained indicated that the eye depigmented patterns are however not only influenced by the *EDNRB* and *KITLG* genes. Significant chromosome regions on SSC4, SSC6, SSC8 and SSC14 included several other genes. For some of them, the involvement in affecting the studied iris depigmentation traits could be derived by what is currently known about their

biological functions. Particularly, significant markers on SSC4 reported in two different regions indicated the *NOTCH2* and the *PREX2* genes as additional candidates affecting the bilateral complete depigmented iris and the heterochromia defects, respectively. Notch regulatory pathways are involved in several developmental processes including the formation of major organs (Mašek & Andersson 2017). NOTCH2 plays important roles in a variety of these mechanisms, including the development of eye components and eye functions (Ma *et al.* 2007; Penton *et al.* 2012; Grisanti *et al.* 2016; Liu *et al.* 2017). Our study is however the first report that indicates *NOTCH2* as a candidate gene for iris depigmentation. *PREX2* is highly expressed in corneal endothelium (Frausto *et al.* 2014) and is frequently mutated in melanomas that derive from transformed melanocytes (Berger *et al.* 2012). Its function and its role in melanoma development could be interesting links with our genome-wide association study results that indicated *PREX2* as a candidate gene for eye heterochromia in white pigs. On SSC14, *COL17A1* was suggested as another candidate associated to the heterochromia defects. This gene encodes the alpha chain of type XVII collagen that is a transmembrane structural protein of hemidesmosomes mediating the adhesion of keratinocytes to the underlying membrane zone. This structure is altered in epidermolysis bullosa that is associated frequently with altered pigmentation (Pasmooij *et al.* 2012; Gostyński *et al.* 2014).

Summarizing, the results of the six genome-wide association studies that we carried out using distinct groups of pigs with depigmented eyes tended to confirm that the complete and bilateral depigmentation of the eyes is functionally separated from the partial eye depigmentation phenotypes (i.e. heterochromia iridis and heterochromia iridum): the first depigmented trait is not the extreme result of the partial eye depigmented patterns as demonstrated by the different loci involved in determining these eye colour alterations. This is, to some extent, in contrast to what was suggested by the classical studies that described in simple Mendelian fashion the segregation of the depigmented eye phenotypes in pigs (Yoshikawa 1935; Dürr 1937; Searle 1968; Gelatt *et al.* 1973). Other studies, including a larger number of individuals (considering the relatively low frequency of the animals having eye decolouration phenotypes) are needed to better clarify the interactions of variants in the candidate genes identified for these two main eye depigmented traits. Other improvements of this study could be derived by histological analyses that might be also needed to identify not only the eye structures involved in determining different grades of pigmentation patterns but also the anatomical features that are modified in depigmented regions.

The effect of the identified loci were evidenced at the iris pigmentation level only, affecting structures having different embryological origin from that of the dermal lineage. No

similar effect could be observed on coat colours in the Italian Large White pigs, whose completely white colour pattern is due to the dominant and epistatic effects of the *KIT* gene (derived by copy number variants; Johansson Moller *et al.* 1996; Fontanesi *et al.* 2010). This study, provided useful information to understand eye pigmentation mechanisms without any other expression of coat colour patterns of the animals (that were completely white, as determined at the *Dominant white* locus) that might have simplified the interpretation of complex genetic mechanisms affecting the development of iris pigmentation. Results obtained in this study provide additional information further valuing the pig as animal model to study complex phenotypes in humans.

Acknowledgements

The authors thank the Italian Pig Breeders Association (ANAS) for the collaboration in this study. This work was supported by fund from the MiPAAF Innovagen project, the MIUR PRIN2017 PigPhenomics project and from the University of Bologna RFO 2018-2019 programmes.

Competing interests

The authors declare they do not have any competing interests.

References

Ai H., Huang L. & Ren J. (2013) Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. *PLoS ONE* **8**, e56001.

Amyere M., Vogt T., Hoo J., Brandrup F., Bygum A., Boon L. & Vikkul, M. (2011) KITLG mutations cause familial progressive hyper-and hypopigmentation. *Journal of Investigative Dermatology* **131**, 1234-1239.

Anderson M.G., Norman L.H., Colleen M.T., Bo C. & Simon W.M.J. (2008) Iris phenotypes and pigment dispersion caused by genes influencing pigmentation. *Pigment Cell Melanoma Research* **21**, 565-78.

Berger M.F., Hodis E., Heffernan T.P., Deribe Y.L., Lawrence M.S., Protopopov A., Ivanova E., Watson I.R., Nickerson E., Ghosh P., Zhang H., Zeid R., Ren X., Cibulskis K. Sivachenko A.Y., Wagle N., Sucker A., Sougnez C., Onofrio R., Ambrogio L., Auclair D., Fennell T., Carter S.L., Drier Y., Stojanov P., Singer M.A., Voet D., Jing R., Saksena G., Barretina J., Ramos A.H., Pugh T.J., Stransky N., Parkin M., Winckler W., Mahan S.,

Ardlie K., Baldwin J., Wargo J., Schadendorf D., Meyerson M., Gabriel S.B., Golub T.R., Wagner S.N., Lander E.S., Getz G., Chin L. & Garraway L.A. (2012) Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature* **485**, 502-6.

Bovo S., Mazzoni G., Bertolini F., Schiavo G., Galimberti G., Gallo M., Dall'Olio S., Fontanesi L. (2019) Genome-wide association studies for 30 hematological and blood clinical-biochemical traits in Large White pigs reveal genomic regions affecting intermediate phenotypes. *Sci Rep* **9**, 7003.

Branicki W., Brudnik U., Draus-Barini J., Kupiec T. & Wojas-Pelc A. (2008) Association of the SLC45A2 gene with physiological human hair colour variation. *Journal of Human Genetics* **53**, 966-71.

Caduff M., Bauer A., Jagannathan V. & Leeb T. (2017) A single base deletion in the SLC45A2 gene in a Bullmastiff with oculocutaneous albinism. *Animal Genetics* **48**, 619-21.

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.

Cook A.L., Chen W., Thurber A.E., Smit D.J., Smith A.G., Bladen T.G., Brown D.L., Duffy D.L., Pastorino L., Bianchi-Scarra G., Leonard J.H., Stow J.L. & Sturm R.A. (2009) Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. *Journal of Investigative Dermatology* **129**, 392-405.

Costin G.E., Valencia J.C., Vieira W.D., Lamoreux M.L. & Hearing V.J. (2003) Tyrosinase processing and intracellular trafficking is disrupted in mouse primary melanocytes carrying the underwhite (uw) mutation. A model for oculocutaneous albinism (OCA) type 4. *Journal of Cell Science* **116**, 3203-12.

Davenport G.C. & Davenport C.B. (1907) Heredity of eye-colour in man. *Science* **26,** 589-92.

Deane-Coe P.E., Chu E.T., Slavney A., Boyko A.R., Sams A.J. (2018) Direct-to-consumer DNA testing of 6,000 dogs reveals 98.6-kb duplication associated with blue eyes and heterochromia in Siberian Huskies. *PLoS Genetics* **14**, e1007648. doi: 10.1371/journal.pgen.1007648.

Dürr M. (1937) Die Vererbung der Glasäugigkeit beim Schwein 1. *Zeitschrift für Züchtung. Reihe B, Tierzüchtung und Züchtungsbiologie einschließlich Tierernährung* **37**, 129-58.

Fontanesi L., D'Alessandro E., Scotti E., Liotta L., Crovetti A., Chiofalo V., Russo V. (2010) Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. *Animal Genetics* **41**, 478-92.

Fontanesi L., Schiavo G., Galimberti G., Calò D. G., Russo V. (2014) A genomewide association study for average daily gain in Italian Large White pigs. *J Anim Sci* **92**, 1385–1394.

Fracasso N.C.A., de Andrade E.S., Wiezel C.E.V., Andrade C.C.F., Zanão L.R., da Silva M.S., Marano L.A., Donadi E.A., C Castelli E., Simões A.L. & Mendes-Junior C.T. (2017) Haplotypes from the SLC45A2 gene are associated with the presence of freckles and eye, hair and skin pigmentation in Brazil. *Legal Medicine (Tokyo)* **25**, 43-51.

Frausto R.F., Wang C., & Aldave A.J. (2014) Transcriptome analysis of the human corneal endothelium. *Investigative Ophthalmology & Visual Science* **55**, 7821-30.

Gelatt K.N., Rempel W.E., Makambera T.P.E. & Anderson J.F. (1973) Heterochromia irides in miniature swine. *Journal of Heredity* **64**, 343-347.

Gladstone R.M. (1969) Development and significance of heterochromia of the iris. *Archives of Neurology* **21**, 184-92.

Gostyński A., Pasmooij A.M., Del Rio M., Diercks G.F., Pas H.H. & Jonkman M.F. (2014) Pigmentation and melanocyte supply to the epidermis depend on type XVII collagen. *Experimental Dermatology* **23**,130-2.

Grisanti L., Revenkova E., Gordon R.E. & Iomini C. (2016) Primary cilia maintain corneal epithelial homeostasis by regulation of the Notch signaling pathway. *Development* **143**, 2160-71.

Gunnarsson U., Hellström A.R., Tixier-Boichard M., Minvielle F., Bed'Hom B., Ito S.I., Jensen P., Rattink A., Vereijken A. & Andersson L. (2007) Mutations in SLC45A2 cause plumage colour variation in chicken and Japanese quail. *Genetics* **175**, 867-77.

Haase B., Jagannathan V., Rieder S. & Leeb T. (2015) A novel KIT variant in an Icelandic horse with white-spotted coat colour. *Anim Genetics* **46**, 466.

Han J., Kraft P., Nan H., Guo Q., Chen C., Qureshi A., Hankinson S.E., Hu F.B., Duffy D.L., Zhao Z.Z., Martin N.G., Montgomery G.W., Hayward N.K., Thomas G., Hoover R.N., Chanock S., Hunter D.J. (2008) A genome-wide association study identifies novel alleles associated with hair colour and skin pigmentation. *PLoS Genetics* **4**,e1000074.

Harland M.M., Stewart A.J., Marshall A.E. & Belknap E.B. (2006) Diagnosis of deafness in a horse by brainstem auditory evoked potential. *The Canadian Veterinary Journal* **47**, 151.

Holl H.M., Pflug K.M., Yates K.M., Hoefs-Martin K., Shepard C., Cook D.G., Lafayette C. & Brooks S.A. (2019) A candidate gene approach identifies variants in SLC45A2 that

explain dilute phenotypes, pearl and sunshine, in compound heterozygote horses. *Animal Genetics* **50**, 271-4.

Huston K., Leipold H. W. & Freeman A. E. (1968) Heterochromia irides in dairy cattle. *Int. J. Dairy Sci.* **51**, 1101-2.

Issa S., Bondurand N., Faubert E., Poisson S., Lecerf L., Nitschke P., Deggouj N., Loundon N., Jonard L., David A., Sznajer Y., Blanchet P., Marlin S. & Pingault V. (2017) EDNRB mutations cause Waardenburg syndrome type II in the heterozygous state. *Human Mutations* **38**, 581-93.

Johansson Moller M., Chaudhary R., Hellmén E., Höyheim B., Chowdhary B. & Andersson L. (1996) Pigs with the dominant white coat colour phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. *Mammalian Genome* **7**, 822-30.

Kaswan R.L., Collins L.G., Blue J.L. & Martin C.L. (1987) Multiple hereditary ocular anomalies in a herd of cattle. *Journal of the American Veterinary Medical Association* **191**, 97-9.

Kayser M., Liu F., Janssens A.C., Rivadeneira F., Lao O., van Duijn K., Vermeulen M., Arp P., Jhamai M.M., van Ijcken W.F., den Dunnen J.T., Heath S., Zelenika D., Despriet D.D., Klaver C.C., Vingerling J.R., de Jong P.T., Hofman A., Aulchenko Y.S., Uitterlinden A.G., Oostra B.A. & van Duijn C.M. (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris colour gene. *American Journal of Human Genetics* **82**, 411-23.

Liu F., Wollstein A., Hysi P.G., Ankra-Badu G.A., Spector T.D., Park D., Zhu G., Larsson M., Duffy D.L., Montgomery G.W., Mackey D.A., Walsh S., Lao O., Hofman A., Rivadeneira F., Vingerling J.R., Uitterlinden A.G., Martin N.G., Hammond C.J. & Kayser M. (2010) Digital quantification of human eye colour highlights genetic association of three new loci. *PLoS Genetics* **6**, e1000934.

Liu L., Zhou X., Kuang X., Long C., Liu W., Tang Y., Liu H., He J., Huang Z., Fan Y., Zhang Q., Shen H. (2017) The inhibition of NOTCH2 reduces UVB-induced damage in retinal pigment epithelium cells. *Molecular Medicine Reports* **16**, 730-6.

Lü M.D., Han X.M., Ma Y.F., Irwin D.M., Gao Y., Deng J.K., Adeola A.C., Xie H.B. & Zhang Y.P. (2016) Genetic variations associated with six-white-point coat pigmentation in Diannan small-ear pigs. *Scientific Reports* **6**, 27534.

Ma A., Boulton M., Zhao B., Connon C., Cai J. & Albon J. (2007) A role for notch signaling in human corneal epithelial cell differentiation and proliferation. *Investigative Ophthalmology & Visual Science* **48**, 3576-85.

Mašek J. & Andersson E.R. (2017) The developmental biology of genetic Notch disorders. *Development* **144**, 1743-63.

Metallinos D.L., Bowling A.T. & Rine J. (1998) A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung disease. *Mammalian Genome* **9**, 426-31.

Misk N.A., Semieka M.A. & Fathy A. (1998) Heterochromia iridis in water buffaloes (*Bubalus bubalis*). *Veterinary Ophthalmology* **1**, 195-201.

Morimoto N., Mutai H., Namba K., Kaneko H., Kosaki R. & Matsunaga T. (2018) Homozygous EDNRB mutation in a patient with Waardenburg syndrome type 1. *Auris Nasus Larynx* **45**, 222-6.

Pasmooij A.M., Nijenhuis M., Brander R. & Jonkman M.F. (2012) Natural gene therapy may occur in all patients with generalized non-Herlitz junctional epidermolysis bullosa with COL17A1 mutations. *Journal of Investigative Dermatology* **132**, 1374-83.

Penton A.L., Leonard L.D. & Spinner N.B. (2012) Notch signaling in human development and disease. *Seminars in Cell and Developmental Biology* **23**, 450-7.

Philipp U., Lupp B., Mömke S., Stein V., Tipold A., Eule J. C., Rehage J. & Distl O. (2011) A MITF mutation associated with a dominant white phenotype and bilateral deafness in German Fleckvieh cattle. *PLoS ONE* **6**, e28857.

Picardo M. & Cardinali G. (2011) The genetic determination of skin pigmentation: KITLG and the KITLG/c-Kit pathway as key players in the onset of human familial pigmentary diseases. *Journal of Investigative Dermatology* **131**, 1182-5.

Pingault V., Bondurand N., Lemort N., Sancandi M., Ceccherini I., Hugot J.P., Jouk P.S., Goossens M. (2001) A heterozygous endothelin 3 mutation in Waardenburg-Hirschsprung disease: is there a dosage effect of EDN3/EDNRB gene mutations on neurocristopathy phenotypes? *Journal of Medical Genetics* **38**, 205-9.

Pingault V., Ente D., Dastot-Le Moal F., Goossens M., Marlin S., Bondurand N. (2010) Review and update of mutations causing Waardenburg syndrome. *Human Mutations* **31**, 391-406.

Read A.P. & Newton V.E. (1997) Waardenburg syndrome. *Journal of Medical Genetics* **34**, 656-65.

Rennie I.G. (2012) Don't it make my blue eyes brown: heterochromia and other abnormalities of the iris. *Eye (Lond)* **26**, 29-50.

Rothammer S., Kunz E., Seichter D., Krebs S., Wassertheurer M., Fries R., Brem G. & Medugorac I. (2017) Detection of two non-synonymous SNPs in SLC45A2 on BTA20

as candidate causal mutations for oculocutaneous albinism in Braunvieh cattle. *Genet.* **49**, 73.

Santschi E.M., Purdy A.K., Valberg S.J., Vrotsos P.D., Kaese H. & Mickelson J.R. (1998) Endothelin receptor B polymorphism associated with lethal white foal syndrome in horses. *Mammalian Genome* **9**, 306-9.

Searle A.G. (1968) *Comparative Genetics of Coat Colour in Mammals*. Logos Press, London.

Seitz J.J., Schmutz S.M., Thue T.D. & Buchanan F.C. (1999) A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle. *Mammalian Genome* **10**, 710-2.

Sevane N., Sanz C.R. & Dunner S. (2019) Explicit evidence for a missense mutation in exon 4 of SLC45A2 gene causing the pearl coat dilution in horses. *Animal Genetics* **50**,275-8.

Sturm R.A. & Frudakis T.N. (2004) Eye colour: portals into pigmentation genes and ancestry. *Trends in Genetics* **20**, 327-32.

Sturm R.A. & Larsson M. (2009) Genetics of human iris colour and patterns. *Pigment Cell & Melanoma Research* **22**, 544-62.

Sulem P., Gudbjartsson D.F., Stacey S.N., Helgason A., Rafnar T., Magnusson K.P., Manolescu A., Karason A., Palsson A., Thorleifsson G., Jakobsdottir M., Steinberg S., Pálsson S., Jonasson F., Sigurgeirsson B., Thorisdottir K., Ragnarsson R., Benediktsdottir K.R., Aben K.K., Kiemeney L.A., Olafsson J.H., Gulcher J., Kong A., Thorsteinsdottir U. & Stefansson K. (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39**, 1443-52.

Syrris P., Carter N.D. & Patton M.A. (1999) Novel nonsense mutation of the endothelin-B receptor gene in a family with Waardenburg-Hirschsprung disease. *American Journal of Medical Genetics* **87**, 69-71.

Talenti A., Bertolini F., Williams J., Moaeen-Ud-Din M., Frattini S., Coizet B., Pagnacco G., Reecy J., Rothschild M.F., Crepaldi P.; Italian Goat Consortium (2018) Genomic Analysis Suggests KITLG is Responsible for a Roan Pattern in two Pakistani Goat Breeds. *Journal of Heredity* **109**, 315-9.

Verheij J.B., Kunze J., Osinga J., van Essen A.J. & Hofstra R.M. (2002) ABCD syndrome is caused by a homozygous mutation in the EDNRB gene. *American Journal of Medical Genetics* **108**, 223-5.

Walsh S., Chaitanya L., Breslin K., Muralidharan C., Bronikowska A., Pospiech E., Koller J., Kovatsi L., Wollstein A., Branicki W., Liu F. & Kayser M. (2017) Global skin colour prediction from DNA. *Human Genetics* **136**, 847-63.

Wang C., Wang H., Zhang Y., Tang Z., Li K. & Liu B. (2015) Genome-wide analysis reveals artificial selection on coat colour and reproductive traits in Chinese domestic pigs. *Molecular Ecology Resources* **15**, 414-24.

Wang C., Wang X., Tang J., Chen H., Zhang J., Li Y., Lei S., Ji H., Yang B., Ren J. & Ding N. (2018) Genome-wide association studies for two exterior traits in Chinese Dongxiang spotted pigs. *Animal Science Journal* **89**, 868-75.

Wehrle-Haller B. (2003) The role of Kit-ligand in melanocyte development and epidermal homeostasis. *Pigment Cell Research* **16**, 287-96.

Wilde S., Timpson A., Kirsanow K., Kaiser E., Kayser M., Unterländer M., Hollfelder N., Potekhina I.D., Schier W., Thomas M.G. & Burger J. (2014) Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proceedings of the National Academy of Sciences of the USA* **111**, 4832-7.

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

Wollstein A., Walsh S., Liu F., Chakravarthy U., Rahu M., Seland J.H., Soubrane G., Tomazzoli L., Topouzis F., Vingerling J.R., Vioque J., Böhringer S., Fletcher A.E. & Kayser M. (2017) Novel quantitative pigmentation phenotyping enhances genetic association, epistasis, and prediction of human eye colour. *Scientific Reports* **7**, 43359.

Yang G.C., Croaker D., Zhang A.L., Manglick P., Cartmill T. & Cass D. (1998) A dinucleotide mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome (LWFS); a horse variant of Hirschsprung disease. *Human Molecular Genetics* **7**, 1047-52.

Yoshikawa T. (1935) On Heterochromia Iridis in Swine. *Proceedings of the Imperial Academy* **11**, 125-8.

Zazo Seco C., Serrão de Castro L., van Nierop J.W., Morín M., Jhangiani S., Verver E.J., Schraders M., Maiwald N., Wesdorp M., Venselaar H., Spruijt L., Oostrik J., Schoots J.; Baylor-Hopkins Center for Mendelian Genomics, van Reeuwijk J., Lelieveld S.H., Huygen P.L., Insenser M., Admiraal R.J., Pennings R.J., Hoefsloot L.H., Arias-Vásquez A., de Ligt J., Yntema H.G., Jansen J.H., Muzny D.M., Huls G., van Rossum M.M., Lupski J.R., Moreno-Pelayo M.A., Kunst H.P. & Kremer H. (2015) Allelic mutations of KITLG, encoding KIT Ligand, cause Asymmetric and Unilateral Hearing Loss and Waardenburg Syndrome Type 2. *American Journal of Human Genetics* **97**, 647-60.

Zhang Z., Xiao Q., Zhang Q.Q., Sun H., Chen J.C., Li Z.C., Xue M., Ma P.P., Yang H.J., Xu N.Y., Wang Q.S. & Pan Y.C. (2018) Genomic analysis reveals genes affecting distinct phenotypes among different Chinese and Western pig breeds. *Scientific Reports* **8**, 13352.

Zhou X. & Stephens M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821-4.

Table 1. Number of pigs with different iris pigmentation classes and heterochromia patterns.

EYE COLOUR CLASSES	**TOTAL NO. OF PIGS**	**NO. OF FEMALES**	**NO. OF MALES**
PAIL BROWN[1]	161	110	51
MEDIUM BROWN[1]	133	98	35
DARK BROWN[1]	487	320	167
DEPIGMENTED/PAIL GREY[1]	34	19	15
HETEROCHROMIA IRIDIS[2]	29	22	7
HETEROCHROMIA IRIDUM	53	37	16
TOTAL	897	607	290

[1] Both eyes had the same colour.

[2] Heterochromia iridis was defined as unilateral (the first number in parenthesis) or bilateral (the second number in parenthesis).

Table 2. List of the most significant single nucleotide polymorphisms (SNPs) that identified different loci in the seven genome-wide association studies (GWAS) that were carried out and candidate genes located in the corresponding genome regions.

GWAS[1]	SSC[2]	SNP	Position (nt)[3]	m/M allele[4]	MAF[5]	P[6]	Candidate gene[7]
1	16	ALGA0110382	19997882	A/G	0.431	3.71×10^{-08}	*SLC45A2*
2	11	INRA0036477	50141902	G/A	0.064	4.40×10^{-09}	*EDNRB*
	6	SIRI0000289	13446699	A/G	0.059	6.55×10^{-07}	-
3	11	INRA0036477	50141902	G/A	0.065	1.06×10^{-08}	*EDNRB*
	4	ALGA0027376	101085844	A/G	0.018	9.65×10^{-07}	*NOTCH2*
4	5	ASGA0026910	94284630	A/G	0.093	1.01×10^{-11}	*KITLG*
	6	ALGA0035018	29957350	A/G	0.059	1.27×10^{-07}	-
	6	ASGA0027932	29935029	A/G	0.059	1.27×10^{-07}	-
	14	ASGA0066255	114987510	A/G	0.017	7.13×10^{-07}	*COL17A1*
	6	ALGA0111634	24735903	A/C	0.075	9.84×10^{-07}	-
5	8	ASGA0038797	47624697	C/A	0.202	4.00×10^{-07}	-
6	5	ASGA0026806	93644796	A/G	0.113	2.28×10^{-08}	*KITLG*
	4	ASGA0019987	66926527	G/A	0.015	5.07×10^{-07}	*PREX2*
	4	MARC0005122	66696704	C/A	0.015	5.07×10^{-07}	*PREX2*
7	11	INRA0036466	49283058	A/G	0.068	3.23×10^{-09}	*EDNRB*
	5	ASGA0026806	93644796	A/G	0.115	3.98×10^{-08}	*KITLG*

[1] Genome-wide association studies (GWAS) were the following: 1) Pigs with pale brown eyes, medium brown and dark brown eyes; 2) Pigs with completely depigmented/pale grey eyes *vs* pigs with completely dark brown eyes; 3) Pigs with completely depigmented/pale grey eyes *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); 4) Pigs with heterochromia iridis *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); 5) Pigs with heterochromia iridum *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); 6) Pigs with heterochromia patterns (iridis and iridum) *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); 7) Pigs with completely pale grey/depigmented eyes + pigs with heterochromia patterns (iridis + iridum) *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown).

[2] Porcine chromosome.

[3] Position of the SNP on the corresponding chromosome in the Sscrofa11.1 genome version.

[4] Minor (m) and Major (M) alleles.

[5] Minor allele frequency.

[6] *P* from GEMMA (Wald test). Only SNPs significantly associated after Bonferroni correction (Table 1) are reported.

[7] Candidate genes identified in the significant SNP region (± 500 kbp) according to their function and potential role in the analysed phenotypes. Several other SNPs were significant in most of these regions (see Table S2). Two SNPs in the *PREX2* gene region had the same *P*. For some regions, no obvious candidate gene could be identified.

Figure 1. Examples of eye colour phenotypes recorded in pigs: a) pale brown eye; b) dark brown eye; c) depigmented/pale grey eye; d) heterochromia iridis; e) Heterochromia iridum.

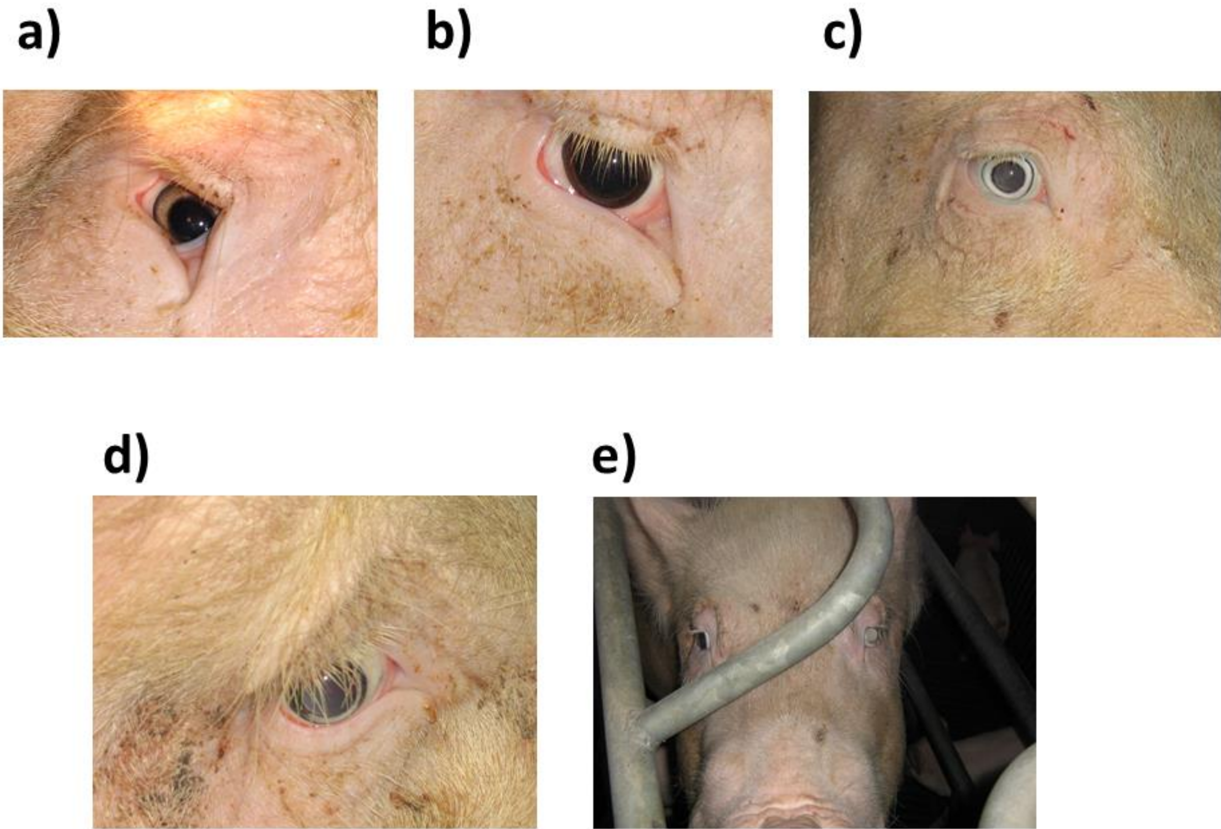
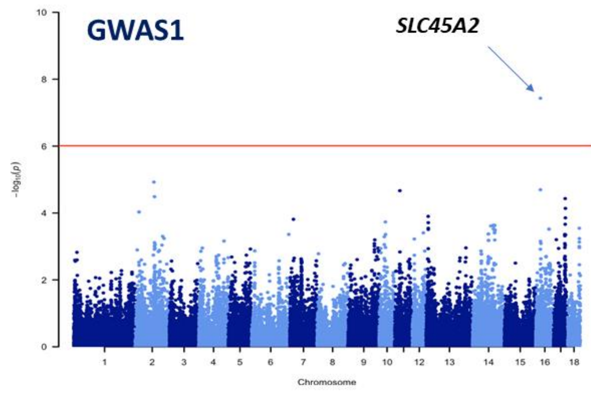
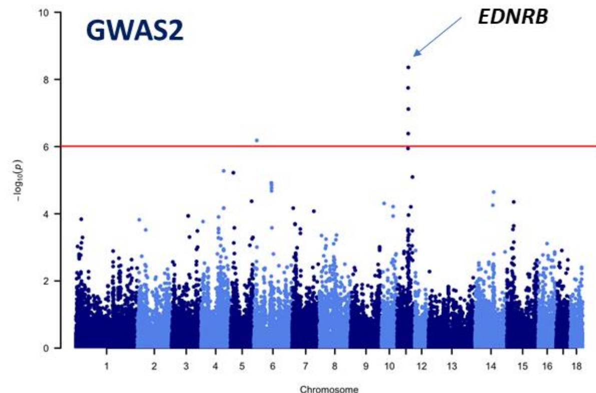
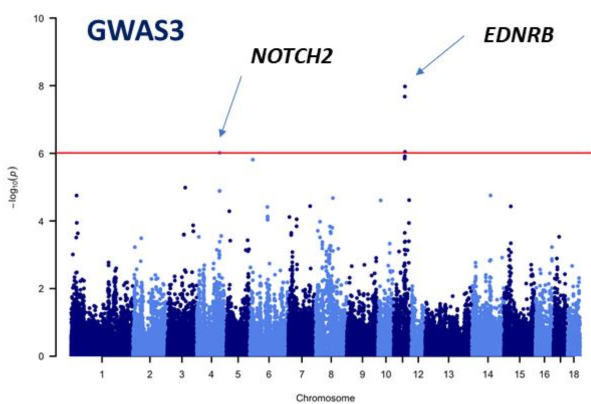
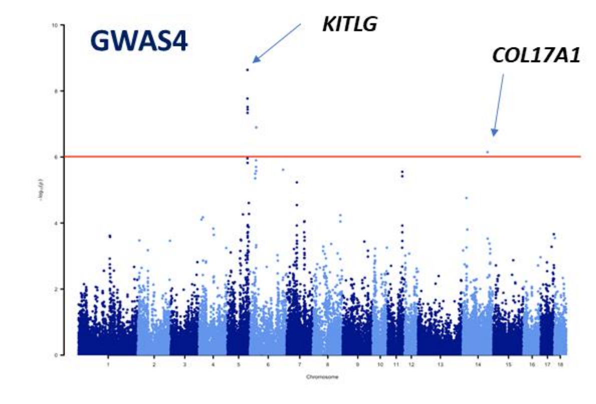


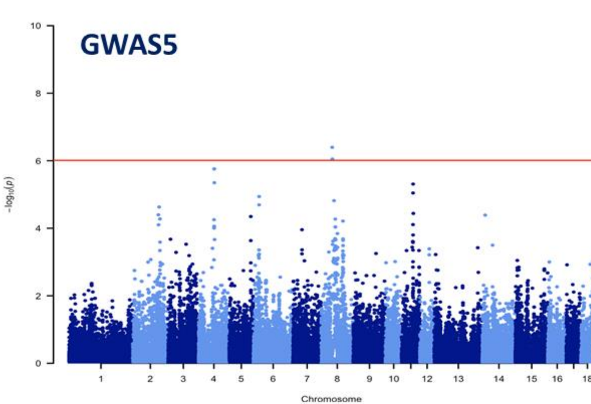
Figure 2. Manhattan plots obtained for the seven genome-wide association studies (GWAS1-7) that analysed different eye colours and heterochromia patterns in Italian Large White pigs. Each point represents a SNP. The redline marks the threshold for statistical significance (Table S1). GWAS1: Pigs with pale brown eyes, medium brown and dark brown eyes; GWAS2: Pigs with completely depigmented/pale grey eyes *vs* pigs with completely dark brown eyes; GWAS3: Pigs with completely depigmented/pale grey eyes *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); GWAS4: Pigs with heterochromia iridis *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); GWAS5: Pigs with heterochromia iridum *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); GWAS6: Pigs with heterochromia patterns (iridis and iridum) *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); GWAS7: Pigs with completely pale grey/depigmented eyes + pigs with heterochromia patterns (iridis + iridum) *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown).

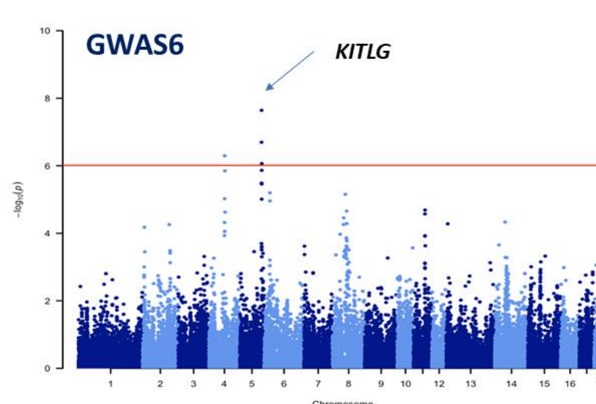

GWAS1
SLC45A2
Chromosome

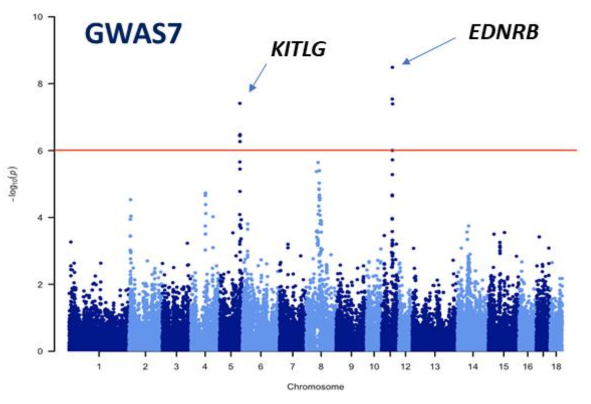

GWAS2
EDNRB
Chromosome


GWAS3
NOTCH2
EDNRB
Chromosome


GWAS4
KITLG
COL17A1


GWAS5
Chromosome


GWAS6
KITLG
Chromosome


GWAS7
KITLG
EDNRB
Chromosome

Supplemental material

Table S1. Information on the seven genome-wide association studies carried out including different groups of pigs defined according to their eye colour phenotypes.

GWAS[1]	GROUPS OF PIGS INCLUDED IN THE GWAS (N. OF PIGS) [2]	N. OF INTERROGATED SNPS[3]	P^4	LAMBDA[5]	h^2_{SNP} (S.E.)[6]
1	PIGS WITH PALE BROWN EYES (N. 157+ 24*); PIGS WITH MEDIUM BROWN EYES (132); PIGS WITH DARK BROWN EYES (N. 456+47*)	50840	9.81×10^{-07}	1.009	0.504 (0.062)
2	PIGS WITH PALE GREY/DEPIGMENTED EYES (N. 32); PIGS WITH COMPLETELY DARK BROWN EYES (N. 456)	51136	9.78×10^{-07}	1.035	0.331 (0.101)
3	PIGS WITH PALE GREY/DEPIGMENTED EYES (N. 32); PIGS WITH COMPLETELY BROWN EYES (PALE BROWN AND DARK BROWN)- (N. 744)	51200	9.77×10^{-07}	1.027	0.116 (0.055)
4	PIGS WITH HETEROCHROMIA IRIDIS (N. 29); PIGS WITH COMPLETELY BROWN EYES (PALE BROWN AND DARK BROWN)- (N. 744)	51247	9.76×10^{-07}	1.066	0.135 (0.062)

5	PIGS WITH HETEROCHROMIA IRIDUM (N. 52); PIGS WITH COMPLETELY BROWN EYES (PALE BROWN AND DARK BROWN) (N. 744)	51245	9.76×10^{-07}	1.039	0.095 (0.054)
6	PIGS WITH HETEROCHROMIA PATTERNS (IRIDIS AND IRIDUM) (N. 81); PIGS WITH COMPLETELY BROWN EYES (PALE BROWN AND DARK BROWN) (N. 744)	51217	9.76×10^{-07}	1.047	0.214 (0.064)
7	PIGS WITH COMPLETELY PALE GREY/DEPIGMENTED EYES + PIGS WITH HETEROCHROMIA PATTERNS (IRIDIS AND IRIDUM) (N. 113); PIGS WITH COMPLETELY BROWN EYES (PALE BROWN AND DARK BROWN) (N. 741)	51157	9.77×10^{-07}	1.061	0.285 (0.069)

[1] Progressive number of the genome-wide association studies that were carried out.

[2] Tested groups of pigs. A few individuals were not included in the GWAS as these animals did not pass the quality control in terms of number of genotyped SNPs

[3] Autosomal single nucleotide polymorphisms (SNPs) filtered according to the criteria described in Materials and methods and used in the corresponding genome-wide association studies.

[4] *P* threshold for statistical significance. It considers Bonferroni correction, based on a nominal $P = 0.05$[5] Genomic inflation factors (λ) value in the seven genome-wide association studies.

[6] Chip heritability (standard error) estimated by GEMMA from the whole set of available genotypes.

* In addition to pigs already included in the categories, there's also been added pigs that presented heterochromia iridis and iridum.

Table S2. List of all significant single nucleotide polymorphisms identified in the seven genome-wide association studies.

GWAS[1]	SSC[2]	SNP	Position (nt)[3]	m/M allele[4]	MAF[5]	P[6]	Closest gene (distance in bp)[7]
1	16	ALGA0110382	19997882	A/G	0.431	3.71×10^{-08}	*SLC45A2 (0)*
2	11	INRA0036477	50141902	G/A	0.064	4.40×10^{-09}	*EDNRB (39018)*
	11	INRA0036466	49283058	A/G	0.059	1.79×10^{-08}	*MYCBP2 (0)*
	11	INRA0036487	50245271	A/G	0.029	7.62×10^{-08}	*EDNRB (142387)*
	11	INRA0036473	49483919	G/A	0.085	4.10×10^{-07}	*MYCBP2 (51061)*
	6	SIRI0000289	13446699	A/G	0.059	6.55×10^{-07}	-
	11	ALGA0062287	48749391	A/G	0.089	1.13×10^{-06}	*ENSSSCG00000045771 (61008)*
	4	ALGA0027376	101085844	A/G	0.019	5.28×10^{-06}	*NOTCH2 (0)*
	5	MARC0114123	13421753	A/G	0.046	6.01×10^{-06}	-
	11	H3GA0032382	68644922	A/G	0.015	8.03×10^{-06}	*ENSSSCG00000039610 (15052)*
3	11	INRA0036477	50141902	G/A	0.065	1.06×10^{-08}	*EDNRB (39018)*
	11	INRA0036466	49283058	A/G	0.060	2.10×10^{-08}	*MYCBP2 (0)*
	11	INRA0036487	50245271	A/G	0.032	9.05×10^{-07}	*EDNRB (142457)*
	4	ALGA0027376	101085844	A/G	0.018	9.65×10^{-07}	*NOTCH2(0)*
	11	INRA0036473	49483919	G/A	0.090	1.23×10^{-06}	*MYCBP2 (87524)*
	11	ALGA0062287	48749391	A/G	0.087	1.43×10^{-06}	*KCTD12 (197061)*
	6	SIRI0000289	13446699	A/G	0.068	1.54×10^{-06}	*ST3GAL2 (0)*
4	5	ASGA0026910	94284630	A/G	0.093	$1{,}01\times10^{-11}$	*TMTC3 (0)*
	5	ASGA0026904	94238901	C/A	0.094	$1{,}99\times10^{-11}$	*TMTC3 (37868)*

	5	ASGA0100714	94149478	A/G	0.095	$2,65\times10^{-11}$	*KITLG (37807)*
	5	ASGA0026806	93644796	A/G	0.106	$2,30\times10^{-09}$	*ENSSSCG00000042946 (106067)*
	5	ASGA0026812	93575560	G/A	0.115	$1,70\times10^{-08}$	*ENSSSCG00000042946 (36831)*
	5	ASGA0097186	94040632	A/G	0.108	$3,07\times10^{-08}$	*KITLG (0)*
	5	ALGA0033764	94457579	G/A	0.125	$3,65\times10^{-08}$	*CEP290 (0)*
	5	ASGA0096350	94038358	G/A	0.120	$4,61\times10^{-08}$	*KITLG (0)*
	6	ALGA0035018	29957350	A/G	0.059	$1,27\times10^{-07}$	*SLC6A2 (0)*
	6	ASGA0027932	29935029	A/G	0.059	$1,27\times10^{-07}$	*SLC6A2 (0)*
	14	ASGA0066255	114987510	A/G	0.017	$7,13\times10^{-07}$	*COL17A1 (0)*
	6	ALGA0111634	24735903	A/C	0.075	$9,84\times10^{-07}$	-
	5	ALGA0033641	93546479	A/G	0.103	$1,10\times10^{-06}$	*ENSSSCG00000042946 (7750)*
	6	ALGA0034962	29065531	A/G	0.069	$1,26\times10^{-06}$	*SMPD3 (40174)*
	5	ASGA0092476	93956402	A/G	0.197	$1,51\times10^{-06}$	*KITLG (61167)*
	6	ALGA0034982	28964755	A/C	0.073	$2,00\times10^{-06}$	*SMPD3 (0)*
	6	ASGA0096677	154283999	A/C	0.021	$2,43\times10^{-06}$	*ENSSSCG00000041035 (0)*
	6	ALGA0034995	29486603	A/G	0.075	$2,65\times10^{-06}$	*GNAO1 (0)*
	11	ASGA0051613	67683899	G/A	0.165	$2,81\times10^{-06}$	*SLC15A1 (0)*
	6	MARC0036639	25103864	G/A	0.078	$3,21\times10^{-06}$	*ENSSSCG00000043247 (1827)*
	11	ASGA0051621	67714628	A/G	0.168	$3,83\times10^{-06}$	*SLC15A1 (3415)*
	6	MARC0049743	25014694	G/A	0.079	$4,42\times10^{-06}$	*ENSSSCG00000047893 (9602)*
	7	ASGA0033639	47475050	A/C	0.490	$5,89\times10^{-06}$	*HYKK (0)*

5	8	ASGA0038797	47624697	C/A	0.202	4.00×10^{-07}	*ETFDH (0)*
	8	CASI0007301	48319713	A/C	0.292	8.92×10^{-07}	*ENSSSCG00000042238 (4035)*
	4	ASGA0019987	66926527	G/A	0.013	1.74×10^{-06}	*PREX2 (0)*
	4	MARC0005122	66696704	C/A	0.013	1.74×10^{-06}	*PREX2 (18549)*
	4	INRA0014612	67435002	C/A	0.014	4.48×10^{-06}	*CPA6 (23389)*
	11	INRA0036466	49283058	A/G	0.061	4.90×10^{-06}	*MYCBP2 (0)*
	3	MARC0010425	30875848	A/G	0.020	5.20×10^{-06}	*SNX29 (0)*
	11	ALGA0062299	49001427	C/A	0.038	9.02×10^{-05}	*KCTD12 (0)*
6	5	ASGA0026806	93644796	A/G	0.113	2.28×10^{-08}	*ENSSSCG00000042946 (106067)*
	5	ASGA0026812	93575560	G/A	0.121	2.01×10^{-07}	*ENSSSCG00000042946 (36831)*
	4	ASGA0019987	66926527	G/A	0.015	5.07×10^{-07}	*PREX2 (0)*
	4	MARC0005122	66696704	C/A	0.015	5.07×10^{-07}	*PREX2 (18530)*
	5	ASGA0026910	94284630	A/G	0.096	8.60×10^{-07}	*TMTC3 (0)*
	5	ASGA0026904	94238901	C/A	0.097	1.37×10^{-06}	*TMTC3 (37868)*
	4	INRA0014612	67435002	C/A	0.015	1.42×10^{-06}	*CPA6 (23409)*
	5	ASGA0100714	94149478	A/G	0.098	1.53×10^{-06}	*KITLG (37807)*
	5	ALGA0033641	93546479	A/G	0.109	3.26×10^{-06}	*ENSSSCG00000042946 (7750)*
	5	ASGA0097186	94040632	A/G	0.112	3.48×10^{-06}	*TMTC3 (236137)*
	6	MARC0036639	25103864	G/A	0.083	6.35×10^{-06}	*ENSSSCG00000043247 (1827)*
	8	ALGA0047923	55526190	G/A	0.370	7.02×10^{-06}	*C8orf34 (23951)*
	4	MARC0082820	66094932	A/C	0.049	9.46×10^{-06}	*ENSSSCG00000050048 (199680)*

	5	ASGA0096350	94038358	G/A	0.125	9.76×10^{-06}	*KITLG (0)*
7	11	INRA0036466	49283058	A/G	0.068	3.23×10^{-09}	*MYCBP2 (0)*
	11	ALGA0062287	48749391	A/G	0.095	2.88×10^{-08}	*ENSSSCG00000045773 (61008)*
	5	ASGA0026806	93644796	A/G	0.115	3.98×10^{-08}	*ENSSSCG00000042946 (166774)*
	11	INRA0036477	50141902	G/A	0.073	4.04×10^{-08}	*EDNRB (39018)*
	5	ASGA0026910	94284630	A/G	0.098	3.33×10^{-07}	*TMTC3 (0)*
	5	ASGA0026812	93575560	G/A	0.123	3.60×10^{-07}	*ENSSSCG00000042946 (36910)*
	5	ASGA0026904	94238901	C/A	0.099	5.38×10^{-07}	*TMTC3 (37868)*
	5	ASGA0100714	94149478	A/G	0.100	5.38×10^{-07}	*KITLG (37747)*
	11	ALGA0062299	49001427	C/A	0.041	9.88×10^{-06}	*KTCD12 (0)*
	11	INRA0036473	49483919	G/A	0.096	1.90×10^{-06}	*MYCBP2 (51061)*
	5	ALGA0033641	93546479	A/G	0.111	2.19×10^{-06}	*ENSSSCG00000042946 (7750)*
	8	ALGA0047923	55526190	G/A	0.377	2.28×10^{-06}	*KIAA1211 (0)*
	5	ASGA0097186	94040632	A/G	0.113	3.57×10^{-06}	*KITLG (0)*
	8	ALGA0047986	61064094	A/G	0.336	4.00×10^{-06}	*ADGRL3 (433977)*
	8	ALGA0047989	61296982	A/G	0.336	4.00×10^{-06}	-
	8	CASI0007301	48319713	A/C	0.305	4.00×10^{-06}	*ENSSSCG00000042238 (4035)*
	11	ALGA0062302	49088705	G/A	0.047	5.22×10^{-06}	*ACOD1 (360)*
	8	ALGA0047983	60870775	A/G	0.331	9.37×10^{-06}	*ENSSSCG00000031193 (76633)*

[1] Genome-wide association studies (GWAS) were the following: 1) Pigs with pale brown eyes, medium brown and dark brown eyes; 2) Pigs with completely depigmented/pale grey eyes *vs* pigs with completely dark brown eyes; 3) Pigs with completely depigmented/pale grey eyes *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); 4) Pigs with heterochromia iridis *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); 5) Pigs with heterochromia iridum *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); 6) Pigs with heterochromia patterns (iridis and iridum) *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown); 7) Pigs with completely pale grey/depigmented eyes + pigs with heterochromia patterns (iridis + iridum) *vs* pigs with completely brown eyes (pale brown, medium brown and dark brown).

[2] Porcine chromosome.

[3] Position of the SNP on the corresponding chromosome in the Sscrofa11.1 genome version.

[4] Minor (m) and Major (M) alleles.

[5] Minor allele frequency.

[6] Candidate genes identified in the significant SNP region (± 500 kbp) according to their function and potential role in the analysed phenotypes. The distance between the reported SNP and the corresponding gene is obtained using the annotation available on Sscrofa11.1 genome version.

[3] Position of the SNP on the corresponding chromosome in the Sscrofa11.1 genome version.

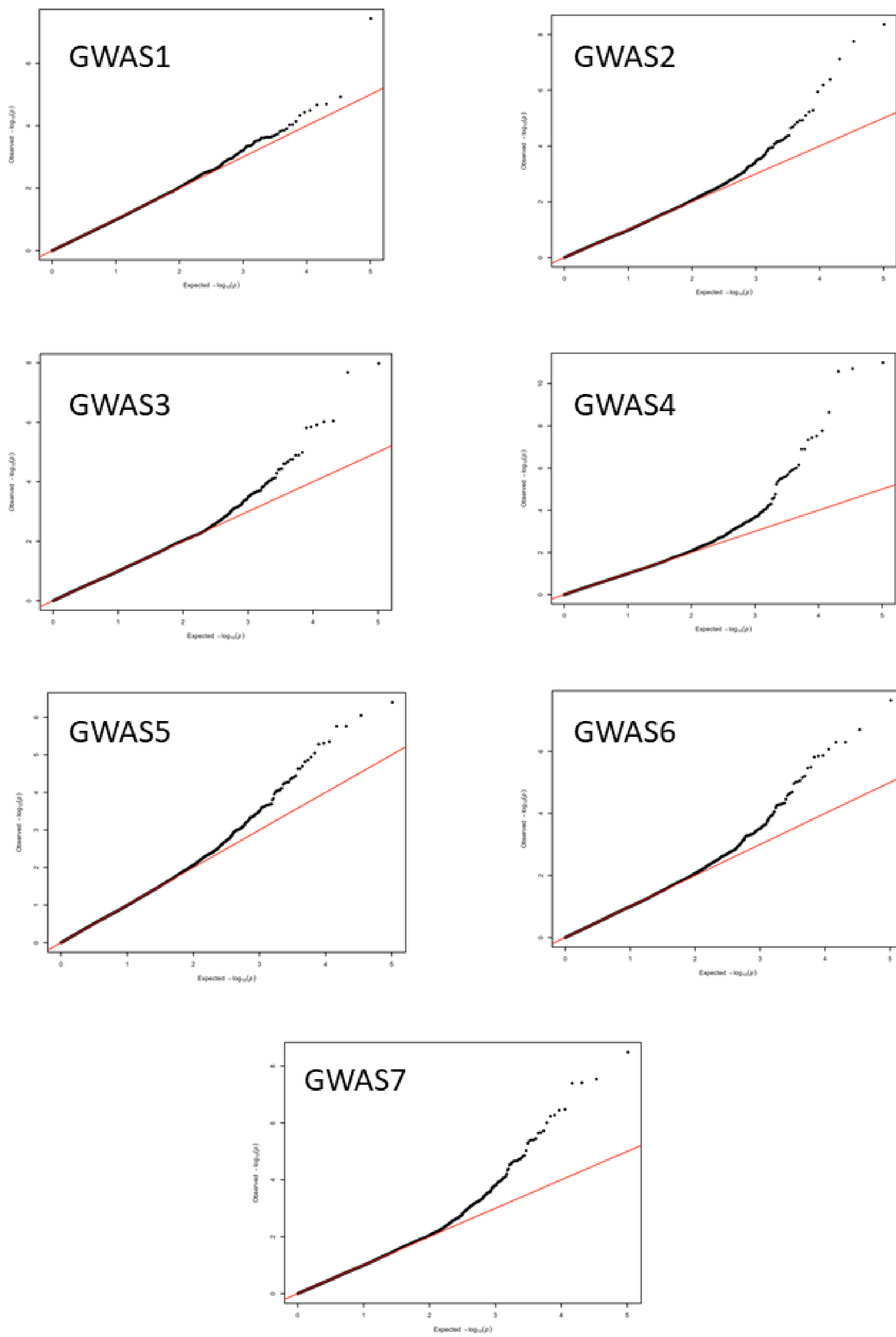
[4] Minor (m) and Major (M) alleles

[5] Minor allele frequency

[6] P from GEMMA (Wald test). Only SNPs significantly associated after Bonferroni correction (Table 1) are reported.

[7] Closest annotated gene and the distance from the corresponding SNP in bp. When no gene was within ± 500 kbp, no information was reported.

Figure S1. QQ plots for the seven genome-wide association studies (GWAS1-7) described in Table S1.



Conclusions

This thesis provides some insights into the genetic architecture of pigmentation related traits in autochthonous and cosmopolitan breeds.

Local breeds could be considered as important animal genetic resources. They are the results of the combination of many different genetic events driven by the agricultural production systems where they have been developed.

In addition to their cultural value, local breeds have unrealized economic potential due to their production qualities, which should be studied further in order to build better management and breeding programs that will benefit their conservation.

We have dissected the Reggiana and Modenese genomes with several population genomic analyses.

This study provides a primary investigation in the genetic pathway of pigmentation traits.

The identification of polymorphisms in genes affecting coat colour could be important to assure the origin of animal products. Traceability, obtained by molecular analysis is a reliable approach for the authentication and valorisation of animal products and an important issue to prevent and detect frauds. In the case of dairy cattle, DNA could be extracted also from milk, as it contains somatic cells of the cow released during milking. Recently, DNA extracted from Parmigiano Reggiano cheese was studied using PCR-RFLP approaches, that made it possible to establish a method to authenticate the breed of origin of the milk using markers in coat colour genes in different cattle breeds.

With the investigation of new coat colour related genes, it could be possible to establish novel assays to authenticate the breed of origin of the milk and thus of the derived cheese.

Analyses of pigmentation genetic marker could be also used to authenticate meat product.

This study could be considered as a starting point to use commercial GeneSeek GGP Bovine 150k Array to study, in a very efficient way, the genomes of local cattle breeds.

Further investigations are needed to design informative SNP-chip panels able to analyse polymorphic markers in genes responsible for other exterior traits that would be useful to easily discriminate other breeds.

Moreover, the use of denser SNP panels could be more informative for the dissection of genetic factors affecting many traits in local cattle breeds. The ascertain bias that all pre-designed SNP chips have can be reduced when more dense SNP chip are used or whole genome resequencing data are produced in local breeds.

The application of NGS in local breeds could facilitate the identification of breeds-specific polymorphisms, which could be then directly used for breeds traceability and authentication of dairy or beef products.

The second research presented in this thesis was the first one that investigated using a genomic approach the genetic factors affecting iris pigmentation in pigs. Different coat colour could arise from different mechanisms governing pigmentation, which could in turn affect in different ways iris pigmentation. It could be also interesting to evaluate if iris colour can have the characteristics to become a breed specific trait that can be useful to discriminate animals for their registration in the breed herd book, similarly to what is considered for other exterior traits (e.g. coat colour, tail and ear shape and position).

Furthermore, eye pigmentation could be also studies in accordance with productive traits and fitness. Studies related with feed intake and rate of growth could give us information about defects in vision associated with different eye colours. We assume that pigs with depigmented eye colour have more incidence to became myopic before pigs with brown eye colour: their fitness in nature will be reduced compared with the other pigs; in farm they will be stressed by the competition for food. So, in accordance with other parameters, like number of receptors of smell or taste, we could study also the impact of eye colour on animal welfare and performance traits. The pig is also an important animal model for many physiological and biologically determined traits. Iris coat colour in pigs could lead to better understand defects in eye pigmentation in humans. This genetically determined trait in pigs can be very useful in this context, according to the broad variability and high frequency of potential iris defects that are present in Large White pigs.
In general, this thesis could be considered an interesting contribution to the genetics of pigmentation in mammals. Additional studies are needed to characterize the causative

mutations underlying pigmentation related patterns (in the coat and in the eyes) and better understand the biological mechanisms that lead to these interesting phenotypes.

Acknowledgments

A mia mamma,
a cui dedicherò sempre tutto

Vorrei esprimere la mia più sincera gratitudine al Prof. Luca Fontanesi. La ringrazio per la sua infinita pazienza, per il supporto e per l'incoraggiamento datomi in questi tre anni, per la sua umanità e sensibilità.
Ringrazio la Prof.ssa Stefania Dall'Olio, mi è stata vicina in questi anni, e in Lei ho sempre trovato sostegno e incoraggiamento.
Senza la vostra guida, oggi non avrei raggiunto questo traguardo.
Ringrazio la mia famiglia, sempre presente.
Grazie ai miei colleghi (e non), che sono diventati i miei amici, con voi ho vissuto momenti indelebili.
Ringrazio ANABORARE, e in particolar modo, il Dott. Massimo Bonacini, per avermi dato la possibilità di praticare il tirocinio presso l'associazione, implementando le mie competenze.