Università degli studi di Sassari

Scuola di Dottorato di Ricerca

Scienze dei Sistemi Agrari e Forestali

e delle Produzioni Alimentari

Indirizzo Scienze e Tecnologie Zootecniche

Cycle XXIV

Approaches for the analysis of genetic diversity in cattle breeds farmed in Italy

Dr.: Elia Pintus

Direttore della Scuola: prof. Giuseppe Pulina

Referente di Indirizzo: prof. Nicolò P.P. Macciotta

Docente Guida: prof. Nicolò P.P. Macciotta

Tutor: Dott.ssa Silvia Sorbolini

Triennio Accademico 2009-2011

Alla mia famiglia e a Luana

Index

Chapter 1 - General Introduction

1. *The three levels of biodiversity* Pag. 1
2. *The resource biodiversity* Pag. 4
3. *The birth of agriculture* Pag. 5
4. *The phenomenon of cattle breeds* Pag. 7
5. *The principles of genetic selection of livestock species* Pag. 8
6. *The process of extinction* Pag. 10
7. *The breeds and molecular genetics* Pag. 11
8. *Molecular traceability* Pag. 14
9. *The identification of selection signatures* Pag. 17

Objectives of the thesis Pag. 19

References Pag. 20

Chapter 2 - Melanocortin 1 receptor (MC1R) gene polymorphisms in three Italian cattle breeds

1. *The case of the Sardo-Modicana breed* Pag. 27
2. *Principles of product traceaility* Pag. 30
3. *The pigmentation in mammals* Pag. 30

References Pag. 36

Abstract Pag. 42

Introduction Pag. 44

Materials and methods Pag. 47

Results Pag. 50

Discussion Pag. 53

Acknowledgements Pag. 56

References Pag. 57

Chapter 3 - Use of Locally Weighted Scatterplot Smoothing (LOWESS) regression to study genome signatures in Piedmontese and Italian Brown cattle breeds

1. Selection signatures Pag. 61

References Pag. 64

Summary Pag. 66

Introduction Pag. 67

Materials and methods Pag. 70

Results and discussion Pag. 74

Conclusion Pag. 84

Acknowledgments Pag. 84

References Pag. 85

Chapter 4 - General Conclusions Pag. 91

Acknowledgments Pag. 94

CHAPTER 1

General Introduction

General Introduction

The exact definition of the term *"biodiversity"*, was coined in 1988 by E.O. Wilson with the aim of replacing the term *"biological diversity"* which was considered less efficient in terms of communication. This definition opened, in the international scientific community, a debate that has not come yet to an end. Omitting definitions outdated or overly philosophical and eccentric, the most used definition is the one written in 1992, *"Biodiversity is the variety of ecosystems that include both communities of living organisms in their particular habitats, both the physical conditions under which they live"*.

Therefore biodiversity must be interpreted as diversity within species, between species and between ecosystems. Subsequent and different elaborations of the same concept have led to the definition adopted by the United Nations Convention on Biological Diversity of Rio de Janeiro: biodiversity is *"the variability among all living organisms including, the subsoil, of air, aquatic and terrestrial ecosystems, marine and ecological complexes of which they are part"* (UNEP, 1992).

1. *The three levels of biodiversity*

Biodiversity is, as already mentioned, the variety with which all the living parts of a place or territory occur, and it is the term commonly applied to different levels of biological organization (Harper and Hawksworth, 1995). Within species, individuals are all different from each other because of differences at DNA level and, therefore, genetic. It is now possible, by the use of

genetic techniques to quantify the diversity at this level and then talk about ***genetic diversity***.

Genetic diversity refers to the presence of different forms of genes in the genetic material of a single species (Templeton, 1995). In almost all multicellular organisms, the genetic information of an individual is not identical to the one of other individuals, because each of them represents a unique combination of genes within a species. This is a consequence of: sexual reproduction, genes recombination and spontaneous mutations induced in the structure of the genes.

The environment, with all its different aspects, acts on individuals determining death or survival. The final result is, therefore, a selection of various and possible combinations of genes (Falconer and Mackay, 1996). This is the reason why two isolated populations, even if they belong to the same species, may undergo a different selection due to the action of various environmental factors that, in the long run, can bring the two populations to have two distinct gene pools. This phenomenon can occur in relatively confined spaces and it is extremely important because it contributes to the creation of genetic diversity for determining the adaptability of the species during evolution. A population or a specie, which lost part of its gene pool, and then lack of genetic variability, is in danger of extinction because lose part of its potential adaptability to new and different environmental conditions (Colwell, 2009). Moreover, the loss of gene pool may lead to an increase of the frequency of unfavorable genes resulting in a further increase of the risk of extinction. All genes distributed in the totality of living beings in the world do not contribute equally to the global genetic

diversity. Genes that regulate fundamental biological processes are preserved unchanged under the different groups of species (taxa) and, generally, exhibit a lower degree of variation. More specialized genes show a larger range of variability.

It is possible to distinguish the different species that populate a certain environment. This level of diversity is ***species diversity*** and refers to the variety of species that live in close contact in a specific environment. Its aspects can be analyzed and studied in different ways. However the most popular types of measurement are:

- *species richness;*
- *abundance of species;*
- *phylogenetic or taxonomic diversity.*

The number of species is commonly defined as *species richness* and is one of the possible measures of the biodiversity of a specific environment. It can be used as a basis for comparison between different places. The *species richness* is considered the simplest measure of biodiversity, it is quite easy to evaluate (Christie et al., 2004). However, it is incomplete and it gives an approximation of the variability present among the living beings.

The estimate of *abundance of species,* evaluates the abundance of single species within the community. Changes of abundance of species is another aspect of diversity and it is measured with a standardized index on a scale ranging from values close to 0, indicating low uniformity or domination of a single species, to 1 that indicates the maximum homogeneity between species (Stirling and Wilsey,

2001). Another approach to measure species diversity is to consider the *phylogenetic and taxonomic diversity*. It is based on the study of genetic relationships between different groups of species (Faith, 1992). This type of measure leads to a hierarchical classification represented by a dendrogram whose branches represent the phylogenetic evolution of the taxa examined (Faith and Baker, 2006).

Measurements within the species are usually considered the most suitable to analyze the diversity between the organisms, because the species are the primary goal of the evolution and are relatively well defined.

Biodiversity is also defined as a measure of the complexity of an ecosystem and of the relationships between its components. The analysis of the availability of different ecosystems in a particular environment or in a distinct geographic area, is the analysis of ***diversity of ecosystem***. The assessment of the ecosystem diversity has critical points due to the complexity of finding the limits of the ecosystem (Christie et al., 2004). The classification of the immense variety of all ecosystems on Earth remains one of a major goal of science and it is important for the management and conservation of the biosphere.

The importance of protecting ecosystems to preserve nature and species, within Community rules, has been recognized with the Habitats Directive (92/43/EEC).

2. *The resource biodiversity*

Throughout its history man has gradually created a niche that, especially in urban areas, has excluded him from contact with the natural environment. Also

the continuous advancement of new technologies for the industrial exploitation of natural resources have made humanity able to radically change the appearance and the balance of the natural environment. Biodiversity is essential for humans because it yields the nutrients, oxygen for respiration, medicines, natural fibres for textiles, raw materials for the production of energy and even the processes of purification and recycling of waste products.

Therefore, the loss and the reduction of biodiversity not only changes the ecosystem functions essential for life but it also has negative economic impacts represented by reduced food, energy and genetic resources. Although the study of multiple forms of life on earth has a very far roots, it now represents a crucial tool to urgently address the problem of loss of biodiversity.

3. The birth of agriculture

The history of agriculture began about 13,000 years ago. In this period began the first attempts at domestication of the main species of livestock and crop plants. This process has inevitably led the man to have a high ability to control food productions. The main consequence has been the occurrence of major demographic and technological changes. The domestication of animals is still considered one of the most important moments of the history and, most likely, the spark that led to an initial growth of human civilizations (Diamond, 2002).

Thousands of years of evolution and selection have contributed to the growth of diversity (Groeneveld et al., 2010), creating the conditions to practice the farming of the species in different environment conditions. Diversity is

essential for all production systems, because it provides the raw material for the improvement of breeds and to adapt to changing conditions.

However, only a little part of the total species present on earth have been completely domesticated. In fact, the process of domestication has been extremely complex and gradual (FAO, 2007). Causes that first led man to domesticate animals remain a mystery and almost certainly they may be different from one geographical area to another or from one species to another. The tendency of the human to groped to tame wild animals is the basis of the domestication (Diamond, 2002). The great expansion of human populations, mainly due to climatic changes, probably represents the main cause that has led to the domestication of animals. Another cause is represented by the increased requirement of food. Finally, the same amount of calories of food energy could be produced by using less energy by means of agricultural practices rather than by hunting and gathering (Gupta, 2004).

Today some wild ancestor (i.e. auroc) and many breeds of farmed species are extinct or highly endangered with extinction (Taberlet et al., 2008). For these species, domestic animals are now a sort of biological bank that inherited diversity from their wild ancestors. Unfortunately most of this genetic diversity has been lost nowadays.

As already mentioned, only a small part of total animal species has been successfully domesticated. The explanation can be found in the characteristics or advantages required by the domestication itself (FAO, 2007). In fact, from the beginning of this phenomenon, some characteristics were more important:

adaptability; rapid growth rate; short intervals between births and offspring of large size (Diamond, 2002). Most of the ancestral species have been identified and through molecular studies was obtained a reconstruction of the history of breeds and ancestral populations (Groeneveld et al., 2010).

4. The phenomenon of cattle breeds

Since Neolithic age, cattle have spread all over the world following the migrations of human populations or because of trade. Once that new territories were reached, cattle gradually adapted to specific environmental conditions and was farmed in the new area. It was only 200 years ago that these differences between animals of the same species were defined and the concept of breed was introduced (Ajmone-Marsan and The GLOBALDIV Consortium, 2010).

After the industrial revolution, some of the traditional livestock productions lost their importance due to the availability of new industrial products. On the other hand, the demand of proteins of animal origin was continuously increasing. Therefore, an intense selection of breeds of livestock for food production started. Since then, specialized breeds and intensive production systems have spread around the world.

On the contrary, autochtonous populations not subjected to any selective pressure, have survived in areas where intensive farming had not been able to affirm due to economic, cultural or environmental conditions. Thus local or native breeds are now generally characterized by their limited geographical distribution (Hiemstra et al., 2010).

Breeds are today commonly divided into:

- present in a single country;
- cosmopolitan.

The former breeds are commonly referred to as “local”, whereas the latter as a “transboundary”. In agreement with FAO (2007) the transboundary breeds, can be subdivised in: breed present in more than one country but within a single region (regional transboundary), and breed present in all countries and more than one region (international transboundary).

5. The principles of genetic selection of livestock species

Genetic improvement is the process of modification of genetic heritage in order to improve the characteristics in the farmed species. This process has been often done, especially in the past, in unconscious and empirical way through the selection of phenotypes that were considered more favorable. Currently, thanks to modern techniques, this process is a combination of phenotypic observations with genotypic knowledge available from genome studies.

The results obtained in the field of genetic improvement in the millennia, since the domestication in the Neolithic period, are small compared to those obtained from the early years of the last century. In fact it is from the beginning of the twentieth century that the selection underwent to a revolution, largely due to the development of technical factors and scientific achievements that have made it a continuously evolving process. Today the main objective of livestock breeding is to be able to estimate with great accuracy the genetic merit of the individual. One of the first attempts to estimate the genetic value of selection candidates was the Selection Index (Hazel, 1943). According to this approach,

the breeding value for a quantitative trait is estimated using the phenotypes previously adjusted for some fixed effects. However, this method has some problems. It does not take into account genetic differences between generations or farms. For this reason, reliable results can be obtained only for animals farmed in the same environmental conditions.

The first applications of the BLUP (Best Linear Unbiased Prediction) method allowed to estimate simultaneously fixed effects and random additive genetic effects of the bulls (Henderson, 1975). This methodology has been used in genetic evaluation systems of many countries. However, early BLUP models considered only the male population (i.e. Sire and Maternal Grandsire models). Thus the estimated breeding value was only half of daughters additive genetic effect because only fathers were evaluated. With the Animal model, geneticists able to estimate the genetic merit of all animals within a breed. However, due to the large number of equations in the model, the routinely use of this approach had been feasible only when adequate computer resources were available.

The above mentioned methods take into account the total production per lactation of standardized length. The cumulated yields were obtained from Test Day (TD) data recorded on farm. The main limitation of the so called lactation models is that they are not able to take properly account of environmental effects (i.e. climate and feeding) that may affect specifically production in some lactation stages.

The Test Day Model (Stanton, 1992) provides the solution to this problem through the direct analysis of data obtained from daily production. Generally

these models require high computational resources. Moreover, they are very sensitive to the precision of the phenotypic data. Incorrect production, because obtained from imprecise controls, produce unreliable indices.

6. The process of extinction

Extinction is a natural process which is considerably accelerated by human activities (Martens et al., 2003). In general, the phenomenon involves all flora and fauna species. Recent studies reported that about 20% of all bovine breeds of the world resulted to be at risk. Actually they have a number of females less than or equal to 1,000 and about 9% of them are already extinct (FAO, 2007).

According to the information collected in European and worldwide databases (EFABIS (http://efabis.tzv.fal.de/) and DAD-IS (www.fao.org/dad-is/)), the local European cattle breeds present data even more alarming, with about 40-50% of them that be considered at risk and some other are actually extinct (www.fao.org/DAD-IS). For this reason, the majority of European cattle breeds can be classified as local breeds. Agriculture public organizations are increasingly oriented to understand the state of European local populations in order to develop best policies and strategies for the conservation and the maintenance of genetic diversity of cattle in Europe (Hiemstra et al., 2010).

The aim of conservation is to preserve breeds and agricultural production systems able to satisfy the maintenance of genetic variability (Negrini et al., 2006) and of cultural, social, economic and environmental values. From a genetic point of view, the importance of diversity safeguarding between and within the

breed is widely recognized. For all this reasons it is important to determine risk status or state of damage of a breed which is commonly estimated based on the number of animals. The level of risk is difficult to assess accurately. For this reason it is important to examine demographic and genetic factors, that have been defined as a probable indicators of a future extinction of a breed (Gandini et al., 2004).

7. The breeds and molecular genetics

Several investigations with molecular tools have been carried out on European local cattle breed to study their origin and genetic differentiation.

Archaeological findings indicates that the cattle entered in Europe through two main roads: the way of the Danube through the lowlands of Central Europe and the way along the Mediterranean coast (Pinhasi et al., 2005). Further molecular studies (Negrini et al, 2007) found that two main groups of cattle breeds can be distinguished in Europe:

- podolica, as many Italian and Hungarian breeds;
- other cattle breeds.

Molecular analysis are not only used for evolutionary studies but are now also used to measure the differences between or within breeds. The neutral markers reflect the overall genomic change and are able to highlight differences in breeds and the potential variation in traits not yet subjected to selection. The first research applications concerning genetic markers in livestock animals were made using biochemical and immunological markers. But it was with the use and

development of technologies for DNA markers that significant advancements in the knowledge of the structure of the genome of livestock species were achieved.

The first used DNA markers were Restriction Fragment Length Polymorphisms (RFLPs) (Kan and Dozy, 1978). These markers are very frequent in the genome and give the possibility to build genetic maps in species of zootechnical interest (Beckmann and Soller, 1983). Moreover, they allow for the identification of loci responsible for quantitative genetic variation (quantitative trait loci: QTL). However, RFLPs had a little application due to the identification method based on the technique of Southerm blotting which is long and laborious. Another reason is that, in general, RFLPs have only two alleles. Another type of markers, subsequently identified, were the minisatellites or Variable Number of Tandem Repeats (VNTRs) (Nakamura et al., 1987), which have the same problems of RFLP about laboratory analysis, but have the advantage to have high number of alleles.

RFLPs and VNTRs were used in the first phases of the construction of genetic maps. Currently they have been replaced by other markers such as microsatellites, which can be easily analyzed using the PCR technique. The development of microsatellites has allowed remarkable progress in the analysis of the genome. These markers are characterized by a variable number repetitions sequences of 1-5 nucleotides and are highly informative thanks to their high number of alleles (Litt and Luty, 1989; Weber and May, 1989). In general, microsatellites are found in anonymous DNA regions, i.e. regions without known function. The use of automated sequencers for their analysis and use of software

for data preservation and interpretation, has contributed to make microsatellite the most used markers for the genetic maps construction and QTL analysis. Other types of markers are:

- RAPDs (Random Amplified Polymorphic DNA) (Williams et al., 1990; Welsh and McClelland, 1991) that identify markers through short oligonucleotides as primers in PCR;
- The AFLPs (Amplified Fragment Length Polymorphisms) (Vos et al., 1995) that combine the restriction analysis of DNA with PCR and allow, using different combinations of enzymes and primers, for the simultaneous analysis of a large number of loci.

Other methods allow to identify more efficiently polymorphisms caused by point mutations, such as Single Nucleotide Polymorphisms (SNPs). These markers are the most widespread in the animal genome (one every 500-3000 nucleotides). Among these, the Single Strand Conformation Polymorphisms (SSCP) method (Orita et al., 1989) allows the identification of point mutations in amplified DNA fragments of 100-400 nucleotides. More recently, have been developed further methods of analysis commonly known as high-throughput, which allow high efficiency and speed in typing of SNPs. Among these we can mention:

- methods based on minisequencing primer extension (Syvänen, 1999);
- methods based on the chromatographic principles such as DHPLC (Denaturing High Performance Liquid Chromatography) (Huber et al., 1993);

– methods that use the mass spectrometry techniques, such as matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS).

Additional methods are based on the use of solid supports (microarray) on which are fixed high-density oligonucleotides (Chee et al., 1996), which allow to simultaneously analyze hundreds of SNPs. All these technologies enable geneticists to analyze a large number of markers in a short time.

8. Molecular traceability

Today more than 40 animal species contribute to the production of food of animal origin. The combined selection pressure due to environmental factors and the controlled breeding imposed by humans, have led to the creation of large variety of genetically distinct breeds. The development of this diversity, which occurred over thousands of years, is a valuable resource for the breeding of livestock species. In fact, genetically different populations can positively deal problems such emerging threats, new human knowledge and nutritional requirements, fluctuating market conditions or, in general, changing societal needs (FAO, 2007).

It is clear the importance of biodiversity conservation and environmental protection, especially for biological areas particularly defined and limited. This phenomenon is of particular importance in the case of so-called minor livestock breeds, that are farmed in areas defined marginal and with which show a particular symbiosis. It is through the conservation, protection and rational

farming of these breeds that a sustainable economy can be created in areas that otherwise would face a gradual decline and abandonment (Davoli, 2011).

In Italy, 23 autochthonous cattle breeds have been recorded. They are distributed in highly fragmented and localized areas (Bigi and Zanon, 2009). To emphasize their great adaptability and strong ability to interact with the surrounding environment, these breeds are commonly defined as local breeds.

In general, they are characterized by:

- High capacity to adapt to the extreme environment conditions;
- ease of delivery, that is essential to ensure the survival of the calf in the wild farming;
- good maternal ability, remarkable ability to raise the calf in good food condition until weaning;
- high reproducibility, i.e. high sexual precocity, fertility and reproductive longevity;
- compatible with the farming environment, large size and bulk associated with strong skeletal framework.

Local breeds are farmed all over the world always in agronomically difficult areas that can not be used with specialized breeds or with an higher production performance.

A useful tool for the protection and enhancement of typical products that may lead to the development of marginal areas by encouraging the conservation of biodiversity and consequently the protection of local breeds is represented by molecular traceability (Crepaldi et al., 2008). It is defined as the ability to control

the origin of the products and the identity of the animals throughout the production chain through the use of technologies that allow direct analysis of DNA (McKean, 2001). Moreover, molecular traceability combined with a control system for food hygiene and safety, can protect consumers from fraud, help some categories as people suffering from food allergies or intolerances.

Also agriculture has suffered the effects of the markets globalization. Today most of the raw materials used for human nutrition are bought where they are cheaper, preventing the consumer to know the origin of food. Moreover, especially in Europe, in recent years, consumer confidence in food of animal origin declined significantly due to dioxin and BSE scandals.

All these facts has led to an ever-increasing attention of consumers to health and origin of food products. This represent a potential chance of development of marginal areas, typical productions and the consequent conservation of local breeds. The most actual example is represented by the growing interest in products marketed in areas very close to the place of production.

The traceability can be classified into (Crepaldi et al., 2008):

- individual traceability;
- traceability of breed;
- traceability of species.

The individual traceability allows to trace back a product to the individual it was obtained from. However, the implementation of this type of traceability is rather complex. A database of individual biological samples of farmed animals is

needed. In practice it is used only for those products that are really obtained from individual animals. The traceability of species can to attribute a product of animal origin to the species that produced it, and is useful in order to certify if a particular cheese was made using only milk from a particular species. The traceability of breed allows to assign a product, or the same animal, to a particular breed. This type of traceability has gained considerable importance due to the diffusion of so-called mono-breed products.

9. The identification of selection signatures

Molecular markers are used to study the genome at various levels and for different reasons. They are represented by locus-specific variations transmitted in Mendelian way from one generation to the next. Panels of high-density SNPs have made the use of markers a useful tool for identifying genome region affected by a selection (Colli et al., 2011).

Contrary to some evolutionary forces that act indiscriminately throughout the genome (Luikart et al., 2003), selection acts on specific points. It changes, for example, diversity within a breed or genetic distance between breeds that have been selected for different production attitudes.

Recent studies (Hayes et al., 2008; Prasad et al., 2008) investigated the difference in allele frequencies of breeds selected for different traits. Selection points in areas very close to genes that influence milk or meat production (i.e. STAT1, ABCG2, DGAT1 and TG) have been identified (Hayes et al., 2008).

Currently, different approaches and methods are used to identify signatures of selection (Biswas and Akey, 2006). Among these the fixation index

Fst, quantifies the level of differentiation between subpopulations (Weir and Cockerham, 1984) and one of his possible interpretations is the analysis of heterozygosity level between populations. Fst values higher than expected show a divergent selection, on the contrary lower values show an uniformity of selection. In general in the domestic breeds range of Fst values from 0,005 and 0.3, and values of 0.15 indicate significant differences between two populations (Frankham et al., 2002)

Additional approaches are: I) methods based on polymorphisms within species: Tajima's D (Tajima, 1989); Fu and Li's D and F (Fu and Li, 1993); Fay and Wu's H test (Fay and Wu, 2000); Long range haplotype (LRH) test (Sabeti et al., 2002); iHS (Voight et al., 2006); LD decay (LDD) (Wang et al., 2006), II) tests based on polymorphisms within species and the divergence between species: Hudson–Kreitman–aguade (HKA) test (Hudson et al., 1987); McDonald Kreitman (MK) test (McDonald and Kreitman, 1991), III) tests between species: dn/ds test (Suzuki and Gojobori, 1999).

Objectives of the thesis

General aim of the research developed during my PhD was the study of genetic differences between cattle breeds farmed in Italy. This purpose has been pursued by addressing two different issues. The first was the use of a specific gene as marker for the traceability of products obtained by local breeds farmed in low input systems. The second was the study of selection signatures in two Italian cattle breeds with different breeding goals, dairy and beef, using data generated by a high throughput SNP platform and a specifically adapted statistical procedure.

References

Ajmone-Marsan P. and The GLOBALDIV Consortium (2010) A global view of livestock biodiversity and conservation–GLOBALDIV. Animal Genetics, 41 (Suppl. 1), 1–5.

Beckmann J.S., Soller M. (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor. Appl. Genet. 67, 35-43.

Bigi D., Zanon A. (2009) Atlante delle razze autoctone bovine, equine, ovi-caprine e suine allevate in Italia. Edagricole Sole 24ore.

Biswas S. and Akey J.M. (2006) Genomic insights into positive selection. Trends in Genetics Vol. 22 No.8.

Chee M., Yang R., Hubbell E., Berno A., Huang X.C., Stern D., Winkler J., Lockhart D.J., Morris M.S., Fodor S.P.A. (1996) Accessing genetic information with high-density DNA arrays. Science 274, 610-614.

Christie M.,Warren J., Hanley N., Murphy K., Wright R., Hyde T. and Lyons N. (2004). Developing measures for valuing changes in biodiversity: Final Report. DEFRA London.

Colli L., Negrini R., Ajmone-Marsan P., GLOBALDIV CONSORTIUM (2011) Molecular markers, menome and animal genetic resources. Fondazioni iniziative zooprofilatiche e zootecniche. Brescia.

Colwell, R.K. (2009) "Biodiversity: Concepts, Patterns and Measurement". In Levin, S.A. (ed.) The Princeton Guide to Ecology. Princeton University Press, New Jersey.

Convention On Biological Diversity. United Nations Environment Programme, Nairobi, UNEP, 1992, p. 24.

Crepaldi P., Nicoloso L., Milanesi E., Negrini R. (2008) Supplemento Large Animal Review.14: 10-12.

Davoli R. (2011) Biodiversity: a saved heritage for the food quality. Fondazioni iniziative zooprofilatiche e zootecniche. Brescia.

Diamond J. (2002) Evolution, consequences and future of plant and animal domestication. Nature, 418: 700–707.

Faith D.P. (1992) Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1–10.

Faith D.P. and Baker A.M. (2006) Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. Evolutionary Bioinformatics Online 2006: 2.

Falconer D.S. and Mackay T.F.C. (1996) Introduction to Quantitative Genetics. Longman.

FAO (2007) The State of the World's Animal genetic Resources for Food and Agricolture. Edited by Barbara Rischkowsky & Dafydd Pilling. Rome.

Fay J.C. and Wu C.I. (2000) Hitchhiking under positive darwinian selection. Genetics 155, 1405–1413.

Frankham R., Ballou J.D., Briscoe D.A. (2002) Introduction to Conservation Genetics. Cambridge University Press, Cambridge, UK.

Fu Y.X. and Li W.H. (1993) Statistical test of neutrality of mutations.Genetics 133, 693–709.

Gandini G.C., Ollivier L., Danell B., Distl O., Georgoudis A, Groeneveld E., Martiniuk E., Van Arendonk J., and Woolliams J. (2004) Criteria to assess the degree of endangerment of livestock breeds in Europe. Livestock Production Science 91: 173-182.

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. and The GLOBALDIV Consortium (2010) Genetic diversity in farm animals–a review. Animal Genetics, 41 (Suppl. 1), 6–31.

Gupta A.K (2004) Origin of agriculture and domestication of plants and animals linked to early Holocene climate amelioration Current Science, Vol 87, No. 1, 10.

Harper J.L. and Hawksworth D.L. (1995) Preface. In: Hawksworth, D.L. (ed.) Biodiversity: measurement and estimation. Chapman and Hall, London, pp. 5-12.

Hayes BJ., Chamberlain AJ., MacEachern S., Savin K., McPartlan H., MacLeod I., Sethuraman L., Goddard ME. (2008). A genome map of divergent artificial selection between Bos taurus dairy cattle, Bos taurus beef cattle. Animal Genetics 40, 176–184.

Hazel L.N. (1943) The genetic basis for constructing selection indexes. Genetics 28, 476.

Henderson C.R. (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423-47.

Hiemstra S.J., de Haas Y., Mäki-Tanila A., Gandini G. (2010) Local cattle breeds in Europe Development of policies and strategies for self-sustaining breeds. Wageningen Academic Publishers The Netherlands.

Huber C.G., Oefner P.J., Preuss E., Bonn G.K. (1993) High-resolution liquid chromatography of DNA fragments on non-porous poly(styrene-divinylbenzene) particles. Nucleic Acids Res. 21, 1061-1066.

Hudson R.R., Kreitman M. and Aguadè M. (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116, 153–159.

Kan Y.S., Dozy A.M. (1978) Antenatal diagnosis of sickle-cell anemia by DNA analysis of amniotic fluid cells. Lancet 2, 910-912.

Litt M., Luty J.A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat whithin the cardiac muscle actin gene. Am. J. Hum. Genet. 44, 397-401.

Luikart G., England P., Tallmon D., Jordan S., Taberlet P. (2003) The power, promis of population genomics: from genotyping o typing. Nature Review Genetics, 981-984.

Martens P., Rotmans J., de Groot D. (2003) Biodiversity: luxury or necessity? Short communication. Global Environmental Change 13 75–81.

McDonald J.H. and Kreitman M. (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351, 652–654.

McKean J.D. (2001) The importance of traceability for public health and consumer protection. Revue scientifique et technique 20 (2): 363-378.

Nakamura Y., Leppert M., O'Connell P., Wolff R., Holm T., Culver M., Martin C., Fujumoto E., Hoff M., Kumlin E., White R. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235, 1616-1622.

Negrini R., Milanesi E., Bozzi R., Pellecchia M. And Ajmone-Marsan P. (2006) Tuscany autochthonous cattle breeds: an original genetic resource investigated by AFLP markers. J. Anim: Breed. Genetic. 123 10-16.

Negrini R., Nijman I.J., Milanesi E., Moazami-Goudarzi K., Williams J.L., Erhardt G., Dunner S., Rodellar C., Valentini A., Bradley D.G., Olsaker I., Kantanen J., Ajmone-Marsan P., Lenstra J.A. and the European Cattle Genetic Diversity Consortium (2007) Differentiation of European cattle by AFLP fingerprinting. Animal Genetics 38: 60-66.

Orita M., Iwahana H., Kanazawa H., Hayashi K., Sekiya T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl.Acad. Sci. USA 86, 2766-2770.

Pinhasi R., Fort J., Ammerman A.J. (2005) Tracing the origin and spread of agricultural in Europe. PloS Biol 3:e410.

Prasad A., Schnabel R., McKay S., Murdoch B., Stothard P., Kolbehdari D., Wang Z., Taylor J., Moore S. (2008). Linkage disequilibrium, signatures of selection on chromosomes 19, 29 in beef, dairy cattle. Animal Genetics 39, 597-605.

Sabeti P.C, Reich D.E., Higgins J.M., Levine H.Z.P., Richter D.J., Schaffner S.F., Gabriel S.B., Platko J.V, Patterson N.J., McDonald G.J., Ackerman

H.C., Campbell S.J., Altshuler D., Cooper R., Kwiatkowski D., Ward R. & Lander E.S. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–837.

Stanton T.L., Jones L.R., Everett R.W. & Kachman S.D. (1992) ESTIMATING MILK, FAT, AND PROTEIN LACTATION CURVES WITH A TEST DAY MODEL. Journal of Dairy Science 75, 1691-700.

Stirling G. and Wilsey B. (2001) Empirical Relationships between Species Richness, Evenness, and Proportional Diversity. American Naturalist vol. 158, no. 3.

Suzuki Y. and Gojobori T. (1999) A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. 16, 1315–1328.

Syvänen A.C. (1999) From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. Hum. Mutat. 13, 1-10.

Taberlet P., Valentini A., Rezaei H.R., Naderi S., Pompanon F., Negrini R. and Ajmone-Marsan P. (2008) Are cattle, sheep, and goats endangered species? Molecular Ecology (2008) 17, 275–284.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Templeton A.R. (1995) Biodiversity at the molecular genetic level: experiences from disparate macroorganisms. Phil. Trans. Roy. Soc. London B 345, 59–64.

Voight, B.F., Kudaravalli S., Wen X., Jonathan K., Pritchard J.K. (2006) A map of recent positive selection in the human genome. PLoS Biol. 4, e72.

Vos P., Hogers R., Bleeker M., Reijans M., Van De Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M., Zabeau M. (1995) AFLP – a new technique for DNA-fingerprinting. Nucleic Acids Res. 23, 4407-4414.

Wang E.T., Kodama G., Baldi P. and Moyzis R.K. (2006) Global landscape of recent inferred darwinian selection for Homo sapiens. Proc. Natl. Acad. Sci. U.S.A. 103, 135–140.

Weber J.L., May P.E. (1989) Abundant class of human DNA polymorphisms which can be typedusing polymerase chain reaction. Am. J. Hum. Genet. 44, 388-396.

Weir B.S. and Cockerham (1984) Estimating F-statistic for the analysis of population structure. Evolution Int. J. Org. Evolution 38, 1358-1370.

Welsh J. and Mcclelland M. (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18, 7213-7218.

Williams J.G.K., Kubelik A.R., Livak K.J., Rafalski A.J., Tingery S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18, 6531-6535.

Wilson E.O. (ed) 1988. Biodiversity, National Academy of Sciences/Smithsonian Institution, Washington.

CHAPTER 2

Melanocortin 1 receptor (MC1R) gene polymorphisms in three Italian cattle breeds

1. *The case of the Sardo-Modicana breed*

In the Island of Sardinia there are three local cattle breeds that are characterized by peculiar reproductive and productive traits (Brandano et al., 1984). They are Sarda, Sardo-Bruna and Sardo Modicana. The existence of the Sarda breed is documented since the prenuragic age. It almost certainly derived from the western Mediterranean cattle breeds (especially Iberian) with possible influences of North African and Middle Eastern breeds. During its evolution, the Sarda has been affected, sometimes very markedly, by other breeds. In particular, Brown Swiss bulls were imported from Switzerland in the northern areas to improve milk and meat production of the Sarda. In the south part of the Island, where agriculture was more developed, Modicana bulls from Sicily were imported to improve size and strength of the local cattle for work purposes (Brandano et al., 1984). Previous studies carried out on somatic measurements and on blood and milk genetic markers highlighted that the Sarda is actually a very heterogeneous population rather than a well-defined breed. Animals show marked differences in general conformation, coat color (which varies from black to red and from uniform to bi-color) and size. The total number of animals has been estimated in about 16,700 in 2011 (AIA, 2011) (Table 1).

The Sarda can be found in the most inaccessible areas of the Island (Barbagia, Iglesiente, Sarrabus and Gallura). The farming system is almost exclusively extensive. The breed is characterized by a poor attitude to meat yield: very low average daily gain and dressing percentage. However it is characterized by a relevant fertility, calving ease and maternal attitude.

Table 1: number of farms and animals raised for Sarda breed (From: thesis of Angelo Zedda)

Province	Herds number	Animals registered	Animals not registered	Undefined animals	Total
Ca	128	2,936	331	678	3,945
Nu	409	8,716	196	1,302	10,214
Or	25	168	6	101	275
SS	115	1,967	76	167	2,210
Tot.	677	13,787	609	2,248	16,644

For these reasons, the breed is mainly used to produce F1 crosses with specialized beef breeds in marginal areas that cannot be exploited by other animal farming systems (Brandano, 2008).

The Sardo-Modicana breed was obtained by cross of Sardinian hill-breed with bulls of Modicana breed from Sicily, imported from some local breeder around 1870 from the province of Ragusa. The aim of the crossbred was to improve the the size and strength of work of the animals. This was the main attitude of the breed until the spread of mechanization in agriculture. In the period of maximum diffusion (decade 1940-1950) the farming area covered the central (Montiferru, Planargia) and the southern part (Trexenta, Marmilla and Campidano) of the Island. After the massive introduction of mechanization in agriculture, the Sardo-Modicana breed lost its main productive function (Brandano et al., 1983) and a reduction of the number of animals started. Currently the Sardo-Modicana is farmed in the mountain areas of Montiferru and Planargia. The Sardo-Modicana is characterized by a robust skeleton, a red coat, medium size, high calving ease and good maternal ability. It is used, either pure-breed or in crossbreed with beef bulls, for meat production (Brandano, 2008). The milk that exceeds the amount suckled by the calf is used for the production

of the typical cheese "Casizolu". The present size of the Sardo–Modicana population is about 3,400 animals (AIA, 2011) (Table 2).

Table 2: number of farms and animals raised for Sardo-Modicana breed (From: thesis of Angelo Zedda)

Province	Herds number	Animals registered	Animals not registered	Undefined animals	Total
Ca	29	384	85	50	519
Nu	19	149	1	12	162
Or	79	1,473	723	351	2,547
SS	19	142	6	13	161
Tot.	146	2,148	815	426	3,389

Also this breed is farmed extensively. In spite of the quality of its production, that are highly appreciated by consumers, the farming of this breed experiences a deep crisis. Apart from the overall problems of agriculture, the breed suffers from the specific issue of local population, i.e. the markedly lower production levels compared to specialized breeds.

A strategy for the valorization of the Sardo-Modicana breed can be found in the genetic characterization and the development of methods for products identification and traceability. A successful example of genetic traceability for typical products in cattle breeds is represented by the MC1R gene polymorphism cattle breeds (Kantanen et al., 2000; Rouzard et al., 2000; Graphodatskaya et al., 2002; Maudet and Taberlet, 2002; Gan et al., 2007; Mohanty et al., 2008). Several authors have suggested that the MC1R gene alleles can be used as breed-specific markers for animal products traceability (Maudet and Taberlet, 2002; Crepaldi et al., 2003; Rolando and Di Stasio, 2006). In addition, the MC1R gene polymorphism has recently been analyzed in some Italian cattle breeds. In fact, in

Italy this gene has been used to distinguish the Parmigiano Reggiano cheese made exclusively with milk of Reggiana breed, from cheeses obtained from other breeds such as Holstein Friesian and Italian Brown (Russo et al., 2007). For this reason it can be used as a specific marker for the traceability of products obtained from local breeds such Sardo-Modicana.

2. *Principles of product traceability*

The assignment of a subject to a breed by using molecular methods can be carried out through two strategies:

1. the probabilistic approach;
2. the deterministic approach.

The first provides the creation, for each genotyped breed whit highly polymorphic markers, a database with information on the alleles present and their frequency. The individual to be assigned is analyzed with the markers mentioned above and the assignment is made probabilistically, starting from allele frequencies of each breed or from genetic distances between breeds.

The deterministic approach involves the search of specific molecular markers of a breed and/or of genes with specific allelic variants. The genotyping of these markers would allow to assign an animal directly to a specific breed without the need to carry out any probabilistic calculation (Mariani et al., 2005).

3. *The pigmentation in mammals*

The pigmentation in mammals is based on the presence or absence of the melanin in hair and skin. Melanins are formed by enzymatic oxidation of amino acid tyrosine. Two types of pigments are derived:

- Eumelanins;
- Pheomelanins.

The pigmentation is determined by the distribution of Eumelanins and Pheomelanins which are responsible of a black/brown and yellow/red colors, respectively (Prota, 1992; Nordlund et al., 1998).

The metabolic pathways that lead to the synthesis of these two types of melanin are largely unknown. The key enzyme is the tyrosinase, which catalyzes the metabolic steps that start from tyrosine idroxylation and leads to the synthesis of dopaquinone that is a common precursor of these two types of melanin. In absence of thiol compounds undergoes intermolecular cyclization leading to the production of eumelanin. In presence of thiols it gives rise thiol adducts of Dopa termed cysteinyldopas and leads to pheomelanin production (Figure 1) (Lamoreux et al., 2001).

The processes of synthesis and accumulation of melanin occur in melanosomes, which are specific cytoplasmatic organelles of specialized cells called melanocytes, which reside between dermis and epidermis. Subsequently, the melanosomes are transferred in the hairs during their growth through a exocytosis process. The migration of melanocytes occurs during embryo development. They start from the neural crest and move in different parts of the body conferring the pigmentation to the areas where they operate.

Moreover, in some parts of the body the same pigmentation can be changed depending on the level of activity of melanocytes (Seo et al., 2007). First studies on the genetics of coat color were made at the beginning of 1900

(Barrington and Pearson, 1906) just after the rediscovery of Mendel's law. These researches were followed by other studies on the pigmentation similarities between different mammals.

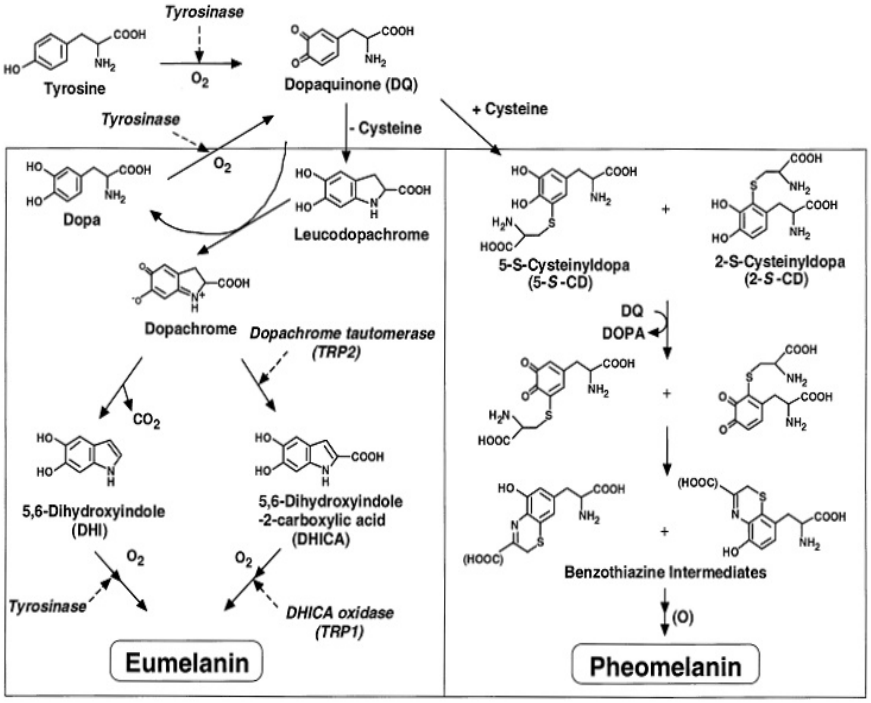

Figure 1: Metabolic pathways that lead to the synthesis of two types of melanin (From: Lamoreux et al., 2001)

The analyses of segregation of colors allowed for the identification of key genes that affect coat color in mammals (Searle, 1968; Olson, 1999). Thanks to the knowledge derived from embryology, biochemistry and molecular genetics has been possible to define the functions of these genes. According to Russo and Fontanesi (2004) they can be classified as follows:

Genes involved in the regulation of melanogenesis:

- the Extension locus (E) coding for melanocortin receptor 1 (MC1R);

- the Agouti locus (A) that encodes a protein (agouti signaling protein, ASIP), which acts as an antagonist of α-melanocyte-stimulanting hormone (α-MSH) in the MC1R receptor.

The E and A locus show epistatic effects. In various mammalian species, dominant alleles at the locus E produce a black coat color, whereas recessive alleles produce to a red/yellow color. Alleles at the locus A cause the recessive black color only when at the locus E the wild-type allele is present, but not dominant or recessive allele (Russo and Fontanesi, 2004).

Genes that influence the development of melanocytes and their migration during embryogenesis:

- the locus White Spotting (W), identified at molecular level in KIT gene;
- the locus Roan (R) coding for mast cell growth factor (MGF) that binds to the KIT gene.

Genes that encode enzymes for the biosynthesis of melanin:

- the Albino locus (C) coding for the enzyme tyrosinase (TYR);
- the Brown locus that encodes for the enzyme tyrosinase-related protein 1 (TYRP1);
- the Slaty locus that encodes for the enzyme tyrosinase-related protein 2 (TYRP2).

Genes that influence the morphology of melanocytes:

- the locus Dilute (D), which encodes for a type V myosin (MYO5A).

Genes that influence the structure and function of melanosomes:

- Locus Silver (PMEL17);

- pink eyed dilution locus (p) that encode for a melanosomes transmembrane proteins.

The Extension locus was initially characterized at the molecular level in mice (Robbins et al., 1993). This locus encodes for melanocortin receptor 1 (MC1R) also referred to as melanocyte stimulating hormone receptor. The MC1R is a transmembrane protein of Gprotein-coupled receptors family (Robbins et al., 1993). As well as in mice, in humans (Valverde et al., 1995), horse (Marklund et al., 1996), sheep (Vage et al., 1999), chicken (Takeuchi et al., 1997) and in pig (Kijas et al., 1998) different mutations in the MC1R gene have been associated with different coat colors. In cattle, the MC1R gene has been mapped to chromosome 18. It consists of a single exon of approximately 950 base pair (bp) and encodes for a protein of 45 kDa that belongs to the family of G protein-coupled receptor (Werth et al., 1996).

This protein, which contains seven transmembrane domains, is integrated in the cell membrane of melanocytes. It binds externally to the hormone MSH (melanocyte stimulating hormone) and to the product of the agouti gene (ASIP), to adjust the chain that leads to metabolic formation of eumelanin and pheomelanin (Mountjoy et al., 1992). Different alleles have been identified at MC1R locus in cattle. Three are the main ones (Klungland et al., 1995):

- allele "wild type" E^+ that produces different colors (Adalsteinsson et al., 1995);
- the dominant allele E^d (characterized by a point mutation that changes the amino acid in position 99 of protein sequence, Leu>Pro) that modifies the

receptor and makes it constitutively expressed and determines the black color (Crepaldi et al., 2003);

- the *e* allele, characterized by a deletion which causes a shift in the reading of codons, inserts a stop codon, and gives rise to a non-functional protein. In homozygous condition causes the coat red/yellow color (Russo et al., 2007).

Apart from the above described three main alleles, the MC1R locus exhibits other polymorphisms whose effect on coat pigmentation are still is not well clarified. Among these, the E^{1} allele, the allele E^{d1} and the e^{f} allels can be mentioned.

The E^{1} allele is characterized by an insertion of 12 bp, which creates a duplication of amino acids (Gly, Ile, Ala, Arg) in position 224 of the amino acid sequence (Rouzaud et al., 2000; Maudet and Taberlet, 2002). The allele E^{d1} is determined by a point mutation (C>T) in position 667 of the nucleotide sequence that causes an amino acid change (Arg>Trp) in position 223 of the amino acid sequence (Maudet and Taberlet, 2002; Graphodatskaya et al., 2002).

The e^{f} allele, found only in few subjects in the Simmental breed. It is determined by a point mutation in position 890 to the nucleotide sequence (C>T), which causes a change in an amino acid (Thr>Ile) in position 297 of the protein sequence (Graphodatskaya et al., 2002).

References

Adalsteinsson S., Bjarnadottir S., Vage D.I., Jonmundsson J.V. (1995) Brown coat color in Icelandic cattle produced by the loci Extension and Agouti. J. Hered. 86:395-398.

A.I.A. http://www.aia.it/aia-website/it/associate/apa-e-ara/sardegna

Barrington A., Pearson K. (1906) On the inheritance of coat color in cattle. I. Shorthorn crosses and pure Shorthorn. Biometrica 4, 427-437.

Brandano P., Asara P., Pulina G., Bolla P., Crivella C. (1983) La razza bovina Modicana della Sardegna. I - Le cararatteristiche morfologiche e biologiche. "Studi Sassaresi" sez.III - Annali della Facoltà di Agraria dell'Università di Sassari vol. XXX, 197-214.

Brandano P., Pulina G., Asara P. (1984) Le razze bovine rustiche della Sardegna. I - Le caratteristiche delle razze e del loro allevamento. Relazione presentata al corso di aggiornamento su tecniche di alimentazione per medici veterinari. Sassari 1984, 1-23.

Brandano P. (2008) L'allevamento dei Ruminanti. Facoltà di Agraria - Università degli studi di Sassari.

Crepaldi P., Marilli M., Meggiolaro D., Fornarelli F., Renieri C., Milanesi E., Ajmone-Marsan P. (2003) The MC1R gene polymorphism in some cattle breeds raised in Italy. Pigment Cell Research 16, 578.

Gan H.Y., Li J.B., Wang H.M., Gao Y.D., Liu W.H., Li J.P., Zhong J.F. (2007) Allele frequencies of TYR and MC1R in Chinese native cattle. Animal Science Journal 78, 484–488.

Graphodatskaya D., Joerg H., Stranzinger G. (2002) Molecular and pharmacological characterization of the MSH-R alleles in Swiss cattle breeds. J. Receptors Signal Transduction 22:421-430.

Lamoreux M.L., Wakamatsu K., and Ito S. (2001) Interaction of Major Coat Color Gene Functions in Mice as Studied by Chemical Analysis of Eumelanin and Pheomelanin. Pigment Cell Res 14: 23–31.

Kantanen J., Olsaker I., Brusgaard K., Eythorsdottir E., Holm L.K., Lien S., Danell B., Adalsteinsson S. (2000) Frequencies of genes for coat colour and horns in Nordic cattle breeds. Genetics, Selection, Evolution. 32, 561–576.

Kijas J.M.H., Wales R., Törnsten A., Chardon P., Moller M., Andersson L. (1998) Melanocortin Receptor 1 (MC1R) mutations and coat color in pigs. Genetics 150:1177-1785.

Klungland H., Vage D.I., Gomez-Raya L., Adalsteisson S., Lien S. (1995) The role of melanocytin stimulating hormone (MSH) receptor in bovine coat color determination. Mammalian Genome 6:636-639.

Mariani P., Panzitta F., Nardelli Costa J., Lazzari B., Crepaldi P., Marilli M., Fornarelli F., Fusi M., Milanesi E., Negrini R., Silveri R., Filippini F., Ajmone Marsan P. (2005) Molecular protocols for livestock product traceability. 4th World Italian beef Cattle Congress, Italy.

Marklund L., Johansson Moller M., Sandeberg K., Andersson L. (1996) A missense mutation in the gene for melanocyte-stimulating hormone

receptor (MC1R) is associated with the chestnut coat color in horses. Mamm. Genome 7, 895-899.

Maudet C., Taberlet P. (2002) Holstein's milk detection in cheeses inferred from melanocortin receptor 1 (MC1R) gene polymorphism. J. Dairy Sci. 85:707-715.

Mohanty T.R., Seo K.S., Park K.M., Choi T.J., Choe H.S., Baik D.H., Hwang I.H. (2008) Molecular variation in pigmentation genes contributing to coat colour in native Korean Hanwoo cattle. Animal Genetics 39, 550–553.

Mountjoy K.G., Robbins L.S., Mortrud M.T., Cone R.D. (1992) The cloning of a family of genes that encode the melanocortin receptors. Science 257: 1248-1251.

Nordlund J.J., Boissy R.E., Hearing V.J., King R.A., Ortonne J.P. (1998) Ito S. Advances in chemical analysis of melanins. In: The Pigmentary System: Physiology and Pathophysiology. New York: Oxford University Press; pp. 439–450.

Olson T.A. (1999) Genetics of colour variation. In: R. Fries, A. Ruvinsky (Eds); The genetics of the cattle. CABI Publishing, Wallingford, UK, pp. 33-53.

Prota G. (1992) Melanins and Melanogenesis. New York: Academic Press; pp 1–290.

Robbins L.S., Nadeau J.H., Johnson K.R., Kelly M.A., Roselli Rehfuss L., Baack E., Mountjoy K.G., Cone R.D. (1993) Pigmentation phenotypes of variant Extension locus alleles result from point mutations that alter MSH receptor function. Cell 72, 827-34.

Rolando A., Di Stasio L. (2006) MC1R gene analysis applied to breed traceability of beef. Italian Journal of Animal Science 5, 87–91.

Rouzaud F., Martin J., Gallet, P.F., Delourme D., Goulemont-Leger V., Amigues Y., Menissier F., Leveziel H., Julien R., Oulmouden A. (2000) A first genotyping assay of French cattle breeds based on a new allele of the extension gene encoding the melanocortin-1 receptor (MC1R). Genet. Sel. Evol. 32:511-520.

Russo V. and Fontanesi L. (2004) Coat colors gene analysis and breed traceability. 7th World Conference of the Brown Swiss Breeders.

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 the breed traceability and authentication of Parmigiano Reggiano cheese. Italian Journal of Animal Science 6, 257–272.

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

Seo K., Mohanty T.R., Choi T., and Hwang I. (2007) Biology of epidermal and hair pigmentation in cattle: a mini-review. 18; 392-400.

Takeuchi S., Suzuki H., Yabuuchi M., Takahashi S. (1997) A possible involvement of melanocortin 1-receptor in regulating feather color pigmentation in the chicken. Biochimica et Biophysica Acta 1308, 164-8.

Våge D.I., Klungland H., Lu D., Cone R.D. (1999) Molecular and pharmacological characterization of dominant black coat color in sheep. Mamm. Genome 10:39-43.

Valverde P., Healy E., Jackson I., Rees J. L., Thody A.J. (1995) Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nature Genet. 11: 328-330.

Werth L.A., Hawkins G.A., Eggen A., Petit E., Elduque C., Kreigesmann B., Bishop M.D. (1996) Rapid communication: melanocyte stimulating hormone receptor (MC1R) maps to bovine chromosome 18. Journal of Animal Science 74, 262.

Zedda A. (2011) Indagine conoscitiva sulla biodiversità dei bovini rustici allevati in Sardegna. Thesis, Università degli Studi di Sassari, Sassari, (SS), Italy.

Melanocortin 1 receptor (MC1R) gene polymorphisms in three Italian cattle breeds

Anna Maria Guastella[A], Silvia Sorbolini[B], Antonio Zuccaro[A], Elia Pintus[B], Salvatore Bordonaro[A], Donata Marletta[A,C] and Nicolò Pietro Paolo Macciotta[B]

[A]DISPA Sezione di Scienze delle Produzioni Animali, Università degli Studi di Catania, via Valdisavoia, 5. 95123 Catania, Italy.

[B]Dipartimento di Scienze Zootecniche, Università degli Studi di Sassari, via De Nicola, 9. 07100 Sassari, Italy.

[C]Corresponding author. Email: d.marletta@unict.it

This research has been pubished as:

A.M. Guastella, S. Sorbolini, A. Zuccaro, E. Pintus, S. Bordonaro, D. Marletta, and N.P.P. Macciotta. 2011. Melanocortin 1 receptor (MC1R) gene polymorphisms in three Italian cattle breeds. Animal Production Science, 2011, 51, 1039–1043.

Animal Production Science, 2011, 51, 1039–1043. Received 3 May 2011, accepted 8 September 2011, published online 21 October 2011.

Abstract

The Melanocortin 1 receptor (MC1R) is one of the main genes implicated in the determination of the coat colour in mammals. This locus showed a relevant genetic variation between breeds that can be exploited for breed traceability of the animal productions. Modicana, Cinisara and Sardo-Modicana are three Italian endangered cattle breeds. Genetic characterization by molecular markers is a fundamental prerequisite for managing genetic resources and for developing potential genetic traceability protocols. In order to improve the knowledge on Modicana, Cinisara and Sardo-Modicana breeds and to evaluate the possibility to develop DNA-based protocols for their mono-breeds products traceability, the genetic structure of MC1R gene was analysed. Four main alleles were observed in a representative sample of 162 animals. In the black coated Cinisara breed (n=42), the E^D and E^+ alleles segregated with a frequency of 0.93 for E^D allele. In the red coated Modicana (n=60) and Sardo-Modicana (n=60) breeds the E^+ and E^1 alleles segregated with frequencies of 0.42, 0.57 and 0.52, 0.47, respectively. The recessive allele e showed a low frequency (0.01) in both breeds. Sequencing a subsample of 34 animals the rare E^2 allele was found only in Modicana and Sardo-Modicana at a good frequency (0.50). A new PCR-RFLP test, based on BstOI restriction endonuclease, was devised to assay for this allele. Results of the work indicate that red coat in Modicana and Sardo-Modicana cattle is genetically determined by the E^+ and E^1 alleles instead of the e allele at homozygote status, as occurs in other red European breeds. In these three Italian breeds of local importance, MC1R polymorphisms can be used to discriminate Cinisara from

Modicana and Sardo-Modicana, but it was not able to distinguish between the two red coat populations.

Additional keywords: breed traceability, Cinisara, coat colour gene, genetic diversity, Modicana, Sardo-Modicana.

Introduction

Coat colour in mammals is determined by the distribution and relative amount of two pigments, eumelanin (black or brown pigment) and pheomelanin (red or yellow pigment) (Klungland and Våge, 2000). Melanin production is mainly regulated by two loci, namely Extension and Agouti (Seo et al., 2007). The Extension locus encodes for Melanocortin 1 Receptor (MC1R), a seven trans-membrane domain receptor. In cattle, the MC1R gene is located on chromosome 18 and consists of a single exon 954 bp long (Werth et al., 1996). This gene shows a polymorphism related to the coat colour (Olson, 1999). More recently, it has been proposed as breed-specific DNA marker for the genetic traceability of the animal productions (Chung et al., 2000; Maudet and Taberlet, 2002; Crepaldi et al., 2003; Rolando and Di Stasio, 2006).

Four main alleles responsible for coat colour determination have been identified at the MCR1 locus in cattle (Klungland et al., 1995; Joerg et al., 1996; Rouzaud et al., 2000; Kriegesmann et al., 2001; Maudet and Taberlet, 2002): (1) the wild-type E^{+}, which may produce a wide range of colours, depending on genotype at the Agouti locus; (2) the dominant E^{D}, that results in black coat; (3) the recessive e, which is associated with red/yellow coat colour in homozygotes; and (4) the E^{1} with an unclear role in colour determination (Crepaldi et al., 2005; Russo et al., 2007).

Furthermore a rare allele, now named E^{2}, was previously observed in some Italian breeds (Maudet and Taberlet, 2002). The genetic polymorphism at MC1R gene has been investigated in several cattle breeds (Kantanen et al., 2000;

Rouzaud et al., 2000; Graphodatskaya et al., 2002; Maudet and Taberlet, 2002;Gan et al., 2007; Mohanty et al., 2008). This locus is a potential candidate marker for a genetic traceability test that could be used to certify typical livestock production. In Italy, for example, MC1R was found to be effective in distinguishing Parmigiano Reggiano cheese made from milk of the local breed Reggiana from other breeds as Holstein Friesian or Italian Brown (Russo et al., 2007). It may, therefore, also be used in other breeds of local importance as a population-specific marker.

An interesting situation is represented by three local cattle breeds farmed in extensive traditional systems in the two main Italian Islands. The Modicana (MO), characterised by a solid red coat, and the Cinisara (CI), characterised by a uniform black coat, are farmed in Sicily and their economic importance lies on the production of two typical "pasta filata" cheeses: Ragusano P.D.O. (Protected Designation of Origin) and "Caciocavallo Palermitano" cheese (Marletta et al., 1998; Guastella et al., 2006). The Sardo-Modicana (SM), derived by the cross of local Sarda cows with MO bulls (Dattilo and Brandano, 1969) is characterised by a wine red coat colour more intense in males. It is farmed extensively in Sardinia and the milk is used to produce the typical "Casizolu" cheese. More information about the breeds is available at the following link (http://eng.agraria.org/cattle.htm, verified 22 September 2011).

In the last 50 years these local breeds have experienced a progressive reduction in size, mainly due to the mechanization of agriculture and to the introduction of cosmopolitan breeds, more specialized and productive. European

Union policy supports their conservation, however they could definitely benefit from the creation of P.D.O. labels for their mono-breed products. An essential prerequisite for such an application is the knowledge of the genetic polymorphism of some candidate genes. In this paper, the genetic polymorphism of MC1R locus in MO, CI and SM cattle breeds was investigated to asses the feasibility of DNA-based traceability protocols for the identification of their mono-breed products.

Materials and methods

Sampling

Blood samples were obtained from a representative sample of 162 cattle of the three breeds: 60 MO, 42 CI and 60 SM. Modicana was collected all over Sicily, CI mainly in the West of Sicily and SM in the Monti Ferru area of Middle-West Sardinia. Unrelated or minimally related individuals were chosen. Genomic DNA was extracted using the commercial GenElute Blood Genomic DNA kit (Sigma-Aldrich, St Louis, MO, USA).

Polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP) and polymerase chain reaction–amplified product length polymorphism (PCR-APLP) methods

The four main alleles (E^D, E^+, E^1 and e) at the MC1R locus were determined by different protocols. A PCR-RFLP method, using MspI and MspaI1 restriction enzymes (New England BioLabs Inc., Milano, Italy), was used to identify E^+, E^D and e alleles (Rolando and Di Stasio, 2006). A PCR-APLP method was able to detect the 12-bp duplication that characterises the E^1 allele (Russo et al., 2007). Amplifications were performed using a GenAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA) thermal cycler. To resolve the presence of nucleotidic duplication, the PCR products were run on 5% polyacrylamide gel in a vertical apparatus (Sequi-Gen Sequencing Cell, BIO-RAD, Laboratories, Hercules, CA, USA).

DNA sequencing and PCR-RFLP method for detection of the E^2 allele

For sequence analysis, in order to confirm the insertion of 12 bp starting from position 669, a fragment was amplified in a subsample of 34 cows (9 MO, 4 CI and 21 SM) by using the following primers (forward 5'-TCG TGG AGA ACG TGC TGG TAG-3'; reverse 5'-TCC ACA ATG GCG TTG CAA ATG ATG-3') designed from the *MC1R* gene sequence (GenBank accession number Y19103). The PCR reaction was performed in a 25-µl mixture, containing 7–100 ng of genomic DNA, 1X PCR buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 10 pmol of each primer, 2 U of Ampli Taq DNA Gold Polymerase (Applied Biosystems). After 5 min of denaturation at 95°C, the PCR conditions were for 35 cycles at 95°C for 30 s, 62°C for 30 s, 72°C for 30 s and a final extension at 72°C for 10 min using a 2720 Thermal Cycler (Applied Biosystems). The amplified region ranged from positions 158 to 882 and contained all the known mutation sites. PCR products were resolved in 1.5% agarose gel, purified by Wizard Vs Gel and PCR Cleaning-up System (Promega Corporation, Madison, WI, USA) and sequenced using the BigDye Terminator Kit, on an ABI PRISM 3130 Genetic Analyser equipped with Sequencing Analysis software (Applied Biosystems).

The transition C667T that characterises the E^2 allele creates an additional restriction site for BstOI (CC^vTGG). A PCR-RFLP procedure for detection of the E^2 allele was applied using the abovementioned primers and conditions. The amplicons were digested for 4 h at 60°C with 5 units of BstOI restriction enzyme (Promega, Carlsbad, CA, USA). Restriction fragments were separated on 4% MS-12 (Molecular Screen) agarose gels (PRONADISA, Torrejon de Ardoz,

Madrid, Spain) with GeneRule 50-bp DNA Ladder, stained with ethidium bromide and visualized under UV light.

Results

The Extension locus showed a genetic polymorphism related to coat colour in all the breeds considered. All the four alleles investigated by PCR-RFLP and PCR-APLP methods were found (Table 1 and Table 2). As expected, the E^D allele was detected only in black coat CI with allele frequency of 0.86. In this breed, the wild allele E^+ was also observed but only at low frequency and in heterozygous state. Basically, only two alleles were detected in the two red coat breeds: the most frequent allele was E^1 and E^+ in MO and SM, respectively. The recessive allele e was observed at very low frequency (0.01) in both the breeds and only in a few heterozygote animals.

A large coding region was sequenced in a subsample of 34 animals in order to discover and describe genetic polymorphism: a total of three SNP and an insertion were found. The missense substitution T296C that characterized allele E^D was confirmed in CI, as well as the G310 del in e allele in red coated breeds (MO and SM). The insertion of 12 bp GGCATTGCCCGG starting from nucleotide 669 was established in all the carriers of E^1 allele previously identified by PCR-APLP. Moreover sequence analysis revealed the presence in MO and SM of a non-synonymous substitution (C667T, accession number GU982927) already described by Maudet and Taberlet (2002) in some Italian breeds and now named E^2.

This mutation results in an amino acid substitution r >W at the 223 position, in the third intracellular protein domain. No specific effect on coat colour was associated with this allele, so far, but this SNP could represent a

promising marker for further studies. A modified PCR-RFPL method was applied to analyse C667T polymorphism. The procedure proved to be very effective in amplifying a long fragment that resulted in a large size of the digested bands, allowing a good electrophoretic resolution in agarose gels.

Table 1. Genotype distribution and genotypic (in brackets) frequencies at MC1R locus in Modicana (MO), Sardo-Modicana (SM) and Cinisara (CI) cattle breeds

Breed	**N**	**Origin**	**Genotypes**			**Alleles**		
			E^DE^D	E^DE^+	E^+E^+	E^D	E^+	e
MO	60	East-Sicily	-	-	7 (0.12)	-	0.12	-
SM	60	West-Sardinia	-	-	17 (0.28)	-	0.28	-
CI	42	West-Sicily	36 (0.86)	6 (0.14)	-	0.93	0.07	-
Breed	**N**	**Origin**	**Genotypes**			**Alleles**		
			E^+E^1	E^+e	E^1E^1	E^+	E^1	e
MO	60	East-Sicily	34 (0.57)	2 (0.03)	17 (0.28)	0.3	0.57	0.01
SM	60	West-Sardinia	28 (0.47)	1 (0.02)	14 (0.23)	0.24	0.47	0.01
CI	42	West-Sicily	-	-	-	-	-	-

Table 2. Allelic frequencies at MC1R locus in Modicana (MO), Sardo-Modicana (SM) and Cinisara (CI) cattle breeds

Breed	**N**	**Tot. Alleles**			
		E^D	E^+	E^1	e
MO	60	-	0.42	0.57	0.01
SM	60	-	0.52	0.47	0.01
CI	42	0.93	0.07	-	-

The wild type sequence (E^+ allele) presents five restriction sites for the enzyme BstOI that produces six fragments ranging from 9 to 447 bp of length: E^2 allele is characterised by an additional restriction site that divides the 447-bp fragment in two bands of 263 and 184 bp easily distinguishable on gel (Figure 1). Since this SNP is very close to the insertion site of the 12 nucleotide sequence (position 669) characteristic of the E^1 allele, haplotypes were reconstructed. About 75% of the 34 sequences were informative: all four CI, eight MO and 13 SM. The E^1 allele was always associated with the nucleotide C^{667} whereas, when

the insertion was not present (allele E^+), the most common nucleotide was T^{667} (only two MO and one SM had the C^{667} nucleotide in heterozygous state). This evidence, once again, confirmed that E^1 derived from the mutation of the wild type at 669 position, whereas a different mutational event seems to be responsible for the occurrence of the allele E^2 characterised by the transition C667T.

Discussion

Local endangered breeds are worthy of investigation and need to be preserved because they represent an important reservoir of rare alleles and often possess allelic combinations that have disappeared in highly selected breeds (Gandini and Villa, 2003). Molecular biology can provide useful tools for the conservation of breeds. Modicana, CI and SM cattle breeds are recognised as "endangered" according to FAO and their conservation is supported by the European Union.

Notwithstanding the risk of extinction, these breeds still have an economic function due to the production of typical cheeses obtained in low-impact farming systems. The genetic authentication and traceability of their mono-breed products (meat and cheese) could contribute to their economic sustainability. MC1R locus is a candidate gene for breed genetic traceability in cattle. This study reveals that in MO and SM the red coat pigmentation is determined by E^+ and E^1 alleles, as observed in Tarentaise (Maudet and Taberlet, 2002), whereas in other European cattle breeds (Limousin, Salers, Red Holstein, Reggiana, Pezzata Rossa d'Oropa) the red coat is caused by the recessive allele e (Rouzaud et al., 2000; Maudet and Taberlet, 2002; Russo et al., 2007). In general, allelic frequencies of E^+ and E^1 in MO and SM are similar, probably because of their genetic relationships, whereas in most Italian cattle breeds the wild allele is largely predominant or fixed, as in Marchigiana cattle (Russo et al., 2007) and Piedmontese (Rolando and Di Stasio, 2006). Actually, MC1R allele frequencies show a great variation between breeds expressing different coat colours. For example, in multi-colored breeds like

Icelandic cattle or in Dolafe the allele E^+ is one of the most represented (Klungland et al., 2000) whereas E^1 allele was observed at high frequency in Aubrac, Gasconne (Rouzaud et al., 2000), Rendena and, especially, in Italian Brown (Russo et al., 2007). In Brown Swiss, this allele was found with a moderate frequency (Kriegesmann et al., 2001).

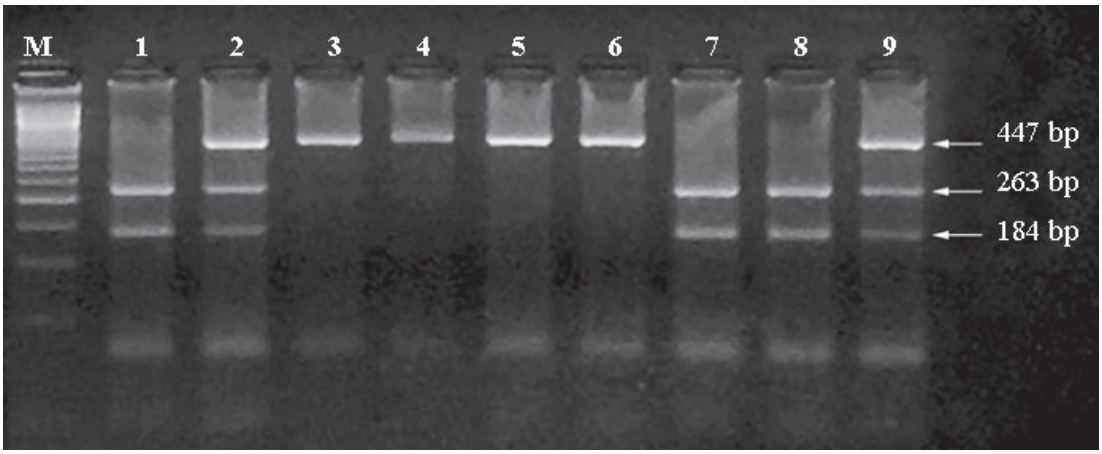

Fig. 1. Polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP) analysis of the bovine MC1R gene by BstOI. In comparison with E^+ allele, E^2 is characterised by the transition C667T. The enzyme BstOI cuts the wild type sequence (E^+ allele) at five restriction sites (CC^vTGG) producing six fragments of 9, 32, 78 (two fragments), 80 and 447 bp long. E^2 allele is characterized by an additional restriction site that divides the 447-bp fragment in two bands of 263 and 184 bp. The smaller fragments (10 and 32 bp) are not visible, short fragments (78 and 80 bp) are co-electrophoresed as a unique band. M: molecular weight marker. The genotypes at nucleotide 667 are: TT (lanes 1, 7, 8), CT (lanes 2, 9) and CC (lanes 3, 4, 5, 6).

Finally in MO and SM the rare E^2 allele that has been observed only in some Italian breeds was also detected. This allele E^2, now easily detectable by PCR-RFLP, represents a potential marker for biodiversity studies and could be included in further investigation at this locus. This study provides new knowledge on MO, CI and SM breeds. The results can be partially used to develop protocols for genetic traceability aimed to distinguish mono-breed

productions. On the basis of detected alleles in MO and SM, that can be ascribed to the crosses between the two breeds that occurred for long times, it can be concluded that the MC1R polymorphism is not able to distinguish between their products and also from Italian Brown, which is a breed that is diffusely farmed in Sicily and Sardinia. On the contrary, the occurrence of the E^D allele exclusively in the CI allows to easily trace back its products in this set of breeds. Nonetheless the same allele is present in Holstein Friesian and in Valdostana Black Pied, but they are characterised by a different coat colour pattern.

The polymorphism at another locus could be helpful to distinguish between uniform and pied coat: in fact classical genetic studies indicated that in cattle spotted coats are genetically determined by homozygote recessive genotype at Spotted or KIT locus (Olson, 1999). Even if a more complex scenario has been recently described in three cosmopolitan and in seven Italian breeds (Fontanesi et al., 2010a, 2010b), it can be supposed that a genetic test combining MC1R and KIT gene analysis should be able to distinguish between this spotted black and white breeds and the uniform black coat breed such as CI.

In conclusion the study investigated the genetic polymorphism at MC1R locus in three endangered Italian cattle breed revealing a notable genetic variation associated to coat colour. However, the alleles observed in MO, CI and SM are not unique/exclusive of these breeds and a cunning fraud cannot be excluded “a priori”. All these findings lead to the consideration that, in spite of the first promising results obtained in some sets of breeds, in our case, MC1R gene can distinguish CI from MO and SM, but it is unable to discriminate

between red coated MO and SM breeds. In perspective, the polymorphism at this locus could be usefully exploited including some of the observed SNP in more complex panels of markers for genetic breed traceability, together with other candidate loci, as Myostatin (MSTN), KIT, MLPH, and SILV genes, as recently suggested (D'Andrea et al., 2009; Nicoloso et al., 2009).

Acknowledgements

This research was funded by the "Fondazione Banco di Sardegna" and "PRA Università degli studi di Catania". Authors thank the Associazione Italiana Allevatori della Sardegna, B. Licheri, F. Deiala and A. M. Cubadda for their assistance.

References

Chung E.R., Kim W.T., Kim Y.S., Han S.K. (2000) Identification of Hanwoo meat using PCR-RFLP marker of MC1R gene associated with bovine coat color. Korean Journal of Animal Science 42, 379–390.

Crepaldi P., Marilli M., Meggiolaro D., Fornarelli F., Renieri C., Milanesi E., Ajmone-Marsan P. (2003) The MC1R gene polymorphism in some cattle breeds raised in Italy. Pigment Cell Research 16, 578.

Crepaldi P., Fornarelli F., Marilli M. (2005) MC1R gene: comparison between different farm animal species. Italian Journal of Animal Science 4, 43–45.

D'Andrea M., Merigioli A., Scarano M.T., Pilla F. (2009) Cattle breed traceability in meat and cheese using DNA polymorphism detected by the SnaPShot method. Italian Journal of Food Science 21, 365–373.

Dattilo M., Brandano P. (1969) La razza Modicano-Sarda. I. I caratteri etnici. Annali della Facoltà di Agraria dell'Università di Sassari, XV, 1–24.

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. Animal Genetics 41, 295–303.

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. Italian Journal of Animal Science 9, 217–221.

Gan H.Y., Li J.B., Wang H.M., Gao Y.D., Liu W.H., Li J.P., Zhong J.F. (2007) Allele frequencies of TYR and MC1R in Chinese native cattle. Animal Science Journal 78, 484–488.

Gandini G., Villa E. (2003) Analysis of the cultural value of local livestock breeds: a methodology. Journal of Animal Breeding and Genetics 120, 1–11.

Graphodatskaya D., Joerg H., Stranzinger G. (2002) Molecular and pharmacological characterization of the MSH-R alleles in Swiss cattle breeds. Journal of Receptor and Signal Transduction Research 22, 421–430.

Guastella A.M., Marletta D., Bordonaro S., D'Urso G. (2006) Sicilian cattle breeds and typical cheeses. Italus Hortus 13(2), 836–838.

Joerg H., Fries H.R., Meijerink E., Stranzinger G.F. (1996) Red coat color in Holstein cattle is associated with a deletion in the MSHR gene. Mammalian Genome 7, 317–318.

Kantanen J., Olsaker I., Brusgaard K., Eythorsdottir E., Holm L-K., Lien S., Danell B., Adalsteinsson S. (2000) Frequencies of genes for coat colour and horns in Nordic cattle breeds. Genetics Selection Evolution 32, 561–576.

Klungland H., Våge D.I. (2000) Molecular genetics of pigmentation in domestic animals. Current Genomics 1, 223–242.

Klungland H., Våge D.I., Gomez-Raya L., Adalsteinsson S., Lien E. (1995) The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. Mammalian Genome 6, 636–639.

Klungland H., Olsen H.G., Hassanane M.S., Mahrous K., Våge D.I. (2000) Coat colour genes in diversity studies. Journal of Animal Breeding and Genetics 117, 217–224.

Kriegesmann B., Dierkes B., Leeb T., Jansen S., Brenig B. (2001) Two breed specific bovine MC1-R alleles in Brown Swiss and Saler breeds. Journal of Dairy Science 84, 1768–1771.

Marletta D., Summer A., Bordonaro S., Mariani P., D'Urso G. (1998) Chemical composition, percentage distribution of caseins and rennet-coagulation properties of herd milks from Modicana and Italian Friesian reared in the Ragusa province. Zootecnica e Nutrizione Animale 24, 185–192.

Maudet C., Taberlet P. (2002) Holstein's milk detection in cheeses inferred from Melanocortin 1 receptor (MC1R) gene polymorphism. Journal of Dairy Science 85, 707–715.

Mohanty T.R., Seo K.S., Park K.M., Choi T.J., Choe H.S., Baik D.H., Hwang I.H. (2008) Molecular variation in pigmentation genes contributing to coat colour in native Korean Hanwoo cattle. Animal Genetics 39, 550–553.

Nicoloso L., Milanesi E., Passerotti S., Malavolta M., Gilmozzi I., Crepaldi P. (2009) Assessing SNPs in coat colour genes for cattle breed traceability. Italian Journal of Animal Science 8, 120–122.

Olson T.A. (1999) Genetics of colour variation. In 'The genetics of the cattle'. (Eds R Fries, A Ruvinsky) pp. 33–53. (CABI Publishing: Wallingford, UK).

Rolando A., Di Stasio L. (2006) MC1R gene analysis applied to breed traceability of beef. Italian Journal of Animal Science 5, 87–91.

Rouzaud F., Martin J., Gallet P.F., Delourme D., Goulemot-Leger V., Amigues Y., Ménissier F., Levéziel H., Julien R., Oulmouden A. (2000) A first genotyping assay of French cattle breeds based on a new allele of the extension gene encoding the melanocortin-1 receptor (MC1R). Genetics Selection Evolution 32, 511–520.

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 the breed traceability and authentication of Parmigiano Reggiano cheese. Italian Journal of Animal Science 6, 257–272.

Seo K.S., Mohanty T.R., Choi T., Hwang I.H. (2007) Biology of epidermal and hair pigmentation in cattle: a mini-review. Veterinary Dermatology 18, 392–400.

Werth L.A., Hawkins G.A., Eggen A., Petit E., Elduque C., Kreigesmann B., Bishop M.D. (1996) Rapid communication: melanocyte stimulating hormone receptor (MC1R) maps to bovine chromosome 18. Journal of Animal Science 74, 262.

CHAPTER 3

Use of Locally Weighted Scatterplot Smoothing (LOWESS) regression to study genome signatures in Piedmontese and Italian Brown cattle breeds

1. Selection signatures

In population genetics the identification of selection signatures is based on a decrease in nucleotide diversity, or an increase in linkage disequilibrium (LD) or a change in frequency of certain alleles in the population under examination (Doebley et al., 2006). Currently several statistical methods are used for the selection signatures detection, and many of these are based on a comparison of allele frequencies within and between breeds (Biswas and Akey, 2006).

The recent availability of platforms capable of genotyping thousands of SNPs simultaneously provides a powerful tool for the assessment of genetic diversity throughout the genome (Andersson and Georges, 2004; Maki-Tanila et al., 2010; Winding and Engelsma, 2010). Genome wide analyses (GWA) studies have been performed to clarify the role of selection in evolutionary processes (Biswas and Akey, 2006). The assumption is that the replacement of a favorable allele at a particular site will reduce the variability in sites closely linked and lead to fixation of alleles in a population (Przeworki et al., 2005; Charlesworth, 2007).

The detection of selection signatures through the study of population genetics is an approach for identifying regions of the genome that contain any genes of interest (Black et al., 2001; Luikart et al., 2003). The identification of these areas involves various aspects such as: 1) the analysis and the study of phenomena such as genetic drift and evolutionary history of populations, that can also affect neutral loci scattered throughout the genome; 2) analysis and the study of patterns of change, loss of diversity and linkage disequilibrium which can present the loci under selection pressure; 3) the study of hitchhiking

phenomenon, through which the effect of selection is also reflected on markers associated with interesting genes allowing the detection of selection signatures.

The identification of loci under to strong selective pressure or that can be fixed in a particular breed involves also genes of great importance for livestock, such as those governing characters involved in adaptation to extreme environments or resistance to disease, which are difficult to study by classical QTL mapping and association studies (FAO, 2007). The identification of genetic variations that control important traits for livestock represents a fundamental point for developing future breeding programs and it is essential for effective management of genetic resources of farm animals (Groeneveld et al., 2010).

The analysis and comparison of the allele frequencies distribution can be executed directly or through various statistics such as the fixation index Fst. The fixation index Fst is widely used and it is considered a simple and robust method (Cavalli-Sforza, 1966; Weir et al., 2005; Barendse et al., 2009). This index uses the differences in allele frequencies to detect any differences between sub-populations or breeds. However, these differences could be due to reasons other than selection as, for example, genetic drift caused by the reduced size of a population/breed or consanguinity phenomena. However, deviations of the Fst values caused by phenomena of inbreeding or genetic drift can be excluded, because the inbreeding affects the entire genome in a similar way while genetic drift affects all loci in a totally random manner (MacEachern et al., 2009). The deviation of the Fst value in a small area of the genome can be observed only as result of selection that affects only small areas of the genome, involving the

selected locus and markers closely associated. The size of the area under selective pressure depend to variables such as, the number of generations under selection or the recombination rate of the specific genomic region.

References

Andersson L., and Georges M. (2004) Domestic-animal Genomics: Deciphering the genetics of complex traits. Nature Reviews Genetics, Vol 5: 202-212.

Barendse W., Harrison B.E., Bunch R.J., Thomas M.B. and Turner L.B. (2009) Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. BMC Genomics 10:178.

Biswas S. and Akey J.M. (2006) Genomic insights into positive selection. TRENDS in Genetics, Vol. 22: 437-446.

Black W.C., Baer C.F., Antolin M.F. & DuTeau N.M. (2001) Population genomics: genome-wide sampling of insect populations. Annual Review of Entomology 46: 441–469.

Cavalli-Sforza L.L. (1966) Population structure and human evolution. Proc. R. Soc. Lond. B. Biol. Sc., 164:362-379.

Charlesworth B. (2007) A hitch-hiking guide to the genome: a commentary on "The hitch-hiking effect of a favourable gene by John Maynard Smith and John haig". Genet. Res. Camb., Vol. 89: 389-390.

Doebley J.F., Gaut B.S. and Smith B.D. (2006) The Molecular Genetics of Crop Domestication. Cell 127 29 1309-1321.

FAO (2007) The State of the World's Animal genetic Resources for Food and Agricolture. Edited by Barbara Rischkowsky & Dafydd Pilling. Rome.

Groeneveld L.F., Lenstra J.A., Eding H., Toro M.A., Scherf B., Pilling D., Negrini R., Finlay E.K., Jianlin H., Groenveld E., Weigend S., and The

GLOBALDIV Consortium (2010) Genetic diversity in farm animals- a review. Animal Genetics, Vol. 41, Suppl. 1 6-31.

Luikart G., England P.R., Tallmon D., Jordan S. & Taberlet P. (2003) The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics, 4: 981–994.

MacEachern S., Hayes B., McEwan J. and Goddard M. (2009) An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (Bos taurus) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. BMC genomics 2009, 10:181.

Mäki-Tanila A., Fernandez J., Toro M. and Meuwissen T. (2010) Local Cattle Breeds in Europe. Development of policies and strategies for self-sustaining breeds. Wageningen Academic Publishers, Chap. 6, 98-117.

Przeworki M., Graham C., and Wall J.D. (2005) The Signature of Positive Selection on Standing Genetic Variation. Evolution, Vol. 59, Issue 11, 2312-2323.

Weir B.S., Cardon L.R., Anderson A.D., Nielsen D.M. and Hill W.G. (2005) Measures of human population structure show heterogeneity among genomic regions. Genome Res., 15:1468-1476.

Winding J.J. and Engelsma K.A. (2010) Perspective of genomics for genetic conservation livestock. Conserv. Genet., Vol. 11 635-641.

Summary

Selection is the major force affecting local levels of genetic variation in a species. Genomics and the use of dense marker maps offer new opportunities for a detailed understanding of genetic diversity across the animal genome. Over the last fifty years, cattle breeds have been subjected to intense artificial selection. Consequently, regions controlling traits of economic importance are expected to exhibit selection signatures. The fixation index (Fst) is an estimate of population differentiation, based on genetic polymorphism data. It is calculated using the relationship between inbreeding and heterozygosity. The aim of this study is to develop a new statistical approach to detect signature selection of the genome among cattle breeds with different production types. Fst and observed heterozygosity (Het) were calculated for 43,766 SNPs marker loci randomly distributed across the genome in 749 Italian Brown and 364 Piedmontese bulls. Then, both Fst and Het were fitted with a Locally Weighted Scatterplot Smoothing (LOWESS) regression to yield chromosomal smoothed patterns. LOWESS predicted trends of both Fst and Het were compared. The statistical significance of Fst values was assessed by using a control chart. The LOWESS technique was efficient in removing noise from the raw data and was able to highlight selection signatures in the whole bovine genome using Het and Fst values. Examples are the peaks detected for BTA6 in the regions harbouring ABCG2 and casein clusters and for BTA2 in the region were the myostatin locus is located.

Key words: SNPs, Fst, Lowess, cattle breeds

Introduction

The study of the genetic basis of differences between animal populations is a key topic of modern biology. Aims and implications of biodiversity researches are different depending on the diffusion and on the economic importance of the considered ethnic group. The inventory of the number of alleles and the evaluation of their association with phenotypes represents a way for a safeguard and potential economic valorisation of local breeds, usually characterized by a small number of individuals and farmed in low input farming systems for typical productions. In highly specialized breeds, on the other hand, the identification of genomic regions involved in differences between individuals for phenotypes of economic importance may allow for the genetic dissection of quantitative traits providing useful knowledge for the enhancement of breeding programme efficiency.

Specialised breeds have been subjected to intense selection that has resulted in a progressive erosion of local levels of genetic variation that may have compromised the ability to challenge environmental factors (Mäki-Tanila et al., 2010; Palaisa et al., 2003). Thus a genetic comparison with autochthonous populations may allow for the identification of genomic regions involved in the control of fitness traits. On the other hand, studies involving highly selected breeds with different production aptitudes, as the case of dairy and beef cattle breeds for example, provide an exciting opportunity for studying signatures of selective breeding (Hayes et al., 2008; Burt, 2009; Qanbari et al., 2010). Actually, little is known about the effects of intensive, directional and prolonged

selection on genome sub-structure of domestics species (Sonstegard et al., 2009; Karlsson and Moen, 2010).

The most common method in animal breeding for comparing individuals on genetic basis is the prediction of their genetic merit using jointly phenotype and pedigree data in a mixed model framework. The primary goal is to rank animal for their breeding potential thus no specific interest in the definition of the genetic architecture of the trait is present. In population genetics, the identification of a locus target of selection is based on the existence of a decrease in nucleotidic diversity, or on an increase linkage disequilibrium (LD) and/or a changed population frequency of alleles (Doebley et al., 2006). Currently, different statistical methods are used for the detection of selection signatures. Many of them are based on the comparison of allele frequencies both within and across species (for a review see Biswas and Akey, 2006). The most commonly used measures are Linkage Disequilibrium (LD) and the Fst. However, for many of these methods it is difficult to develop a proper statistical test. This is particularly true when searching for selective signatures within a single population.

The recent availability of platforms able to simultaneously genotype many thousands of SNP offers a powerful tool for the assessment of the genetic diversity across the genome (Andersson and Georges, 2004; Mäki-Tanila et al., 2010; Winding and Engelsma, 2010). Genome wide analysis (GWA) have been performed to clarify the role of selection and drift in the evolutionary processes (Biswas and Akey, 2006). Several recent studies have proposed the hitchicking

mapping approach for identification of target of positive selection. The basic assumption is that the substitution of favorable allele at one site results in a reduction of variability at closely linked sites and lead to the allele fixation in a population (Przeworki et al, 2005; Charlesworth, 2007). Actually, the abundance of SNP throughout the genome made them particularly useful in the detection of such selection sweeps (Andersson and Georges, 2004).

However, such a huge amount of information has become rather problematic to handle. A major issue is represented by the great variability of the signal pattern (for example heterozygosity or other related statistics as Fst) along the chromosome. A common empirical practice to smooth data is to work on average values of sliding windows of predetermined size (Weir et al., 2005, Hayes et al., 2008, Barendse et al., 2009, Flori et al., 2009). A further problem is represented by the development of a suitable statistical test able to assess when the measure of the genetic difference between two population can be considered significant. Stella et al. (2010) have recently proposed a permutation test based on a simple binomial distribution of the SNP allelic frequencies. This approach is more robust because is based upon the specific distribution of allelic frequencies observed in the data rather than on a theoretical distribution.

The present study tests two statistical methodologies for studying selection signatures in two Italian cattle breeds, Italian Brown and Piedmontese. In particular, a local regression is used to smooth raw Fst data and a control chart is applied to predicted data for identifying values that are different from the mean.

Materials and methods

A total of 749 Italian Brown and 364 Piedmontese bulls were genotyped at 54,001 SNP loci with the Illumina Bovine SNP50TM bead-chip (http://www.illumina.com). Data were generated within the SELMOL research project. The choice of the two breeds was based on their different breeding goals: dairy traits for Brown, beef for Piedmontese. Only those SNP that had been assigned to the 29 autosomes (X chromosome was not considered) in the Btau-4.0 build of the Bovine Genome assembly were considered. SNP were discarded if: monomorphic, with more than 2.5% missing data, with a MAF lower than 1%. After edits, 43,766 markers were retained for the study. Missing data were replaced with the most frequent allele at that specific locus. Allele frequencies, and observed and expected heterozygosities were calculated for each breed. Total allelic frequencies for each locus, f_A and f_B, considering all animals as a single population were calculated as:

$$f_A = f_{pop.1} * (2 * pop.1) + f_{pop.2} * (2 * pop.2) / pop.1 + pop.2;$$

Where pop.1=number of individuals in population1 and pop.2=number of individuals in population2.

$$f_B = 1 - f_A$$

Then, expected heterozygosities in populations (Hs) and overall (Ht) were calculated. Finally, Fst was calculated according to Weir and Cockerham (1984):

$$Fst = Ht - Hs / Ht$$

In order to smooth the pattern of the obtained Fst, data were fitted with a Locally Weighted Scatterplot Smoothing (LOWESS) regression (Cleveland,

1979). In this non parametric local regression, the space of the independent variable is fragmented into different intervals for which separate regressions are fitted. The aim is to remove noise from raw data and to clarify the graphical presentations. A critical point in fitting LOWESS is the identification of a suitable dimension of the data interval to be included in the analysis. In other words, if q is the number of adjacent points to be used in estimation procedure of a set of n data, each region contains a fraction of points given by q/n. This ratio is defined as the smoothing parameter *f* of the LOWESS regression. As *f* increases, the fitted line will be smoother until *f* = 1 that corresponds to a single line (is the standard linear regression). Consequently, the goodness of fit depends strongly on the smoothing parameter used (Cohen, 1999).

The number of markers differs between chromosomes, being directly associated with their length. Therefore the use of the same parameter in all chromosomes could not be feasible, because different sizes of intervals (i.e. number of markers) will be considered. A preliminary analysis revealed that a smoothing parameter corresponding to an interval of 20 SNPs for each separated regression gave the best results. The different smoothing parameters for each chromosome are reported in table 1.

Fst values smoothed by the LOWESS were then subjected to a control chart analysis. The aim was to identify SNP with values significantly different from the average pattern and, therefore, that could may be considered indicators of possible selection signatures. The control chart approach is based on the partitioning and quantification of natural variability present in every process. The

rationale is that some variation may be due to causes which are not normally present in the process, whereas the remaining part can be ascribed to numerous, ever-present differences.

Table 1: different smoothing parameter and length in base pair (bp) for each chromosome

BTA	Smoothing parameter	Length in bp
1	0.007057163	161,021,444
2	0.008594757	140,672,838
3	0.009017133	127,908,629
4	0.009363296	124,125,394
5	0.011049724	125,804,605
6	0.009149131	122,543,360
7	0.01048218	112,064,213
8	0.009813543	116,938,581
9	0.011554015	107,962,209
10	0.010775862	119,596,824
11	0.010487677	110,120,689
12	0.014224751	85,277,438
13	0.013236267	84,344,187
14	0.013689254	81,323,942
15	0.013831259	84,598,267
16	0.014869888	77,895,388
17	0.014534884	76,454,249
18	0.017421603	66,116,595
19	0.017035775	65,213,966
20	0.014673514	75,705,448
21	0.017021277	69,171,298
22	0.018298262	61,825,382
23	0.021574973	53,329,482
24	0.018281536	64,945,342
25	0.024009604	44,021,516
26	0.021881838	51,726,098
27	0.023781213	48,726,297
28	0.024691358	46,020,951
29	0.022271715	51,979,343

The first source is named special cause of variation, that in the present study can be ascribed to the effect of selection that causes a drop of genetic diversity, revealed by an increase of Fst. The second source is named common

cause variation and in this study may represent the random variation of Fst values along the chromosome. Control charts are graphically displayed as stream of data falling within control limits based on plus or minus 3 standard deviations of the centerline. When variation exceeds these limits the control chart highlight outliers signals (Shewart, 1931). In this study were represented by smoothed Fst values plotted against their position along the chromosome.

Results and discussion

Genetic diversity can be considered divisible in two aspects: genetic differences between breeds and genetic differences between individuals within a breed (Kantanen et al., 2000). Understand genetic variation in traits of interest, using molecular data, is the basis for future breeding programmes and it is essential for effective management of farm animal genetic resources (Groeneveld et al., 2010). The continuous use and study of genetic markers for traits of economic interest is important for establishing marker-assisted selection as a tool in cattle industry (Allan et al., 2007).

In this work the comparison of average heterozigosity per chromosome between the two breeds (Figure 1) highlights lower values for the Italian Brown (average difference of 0.04).

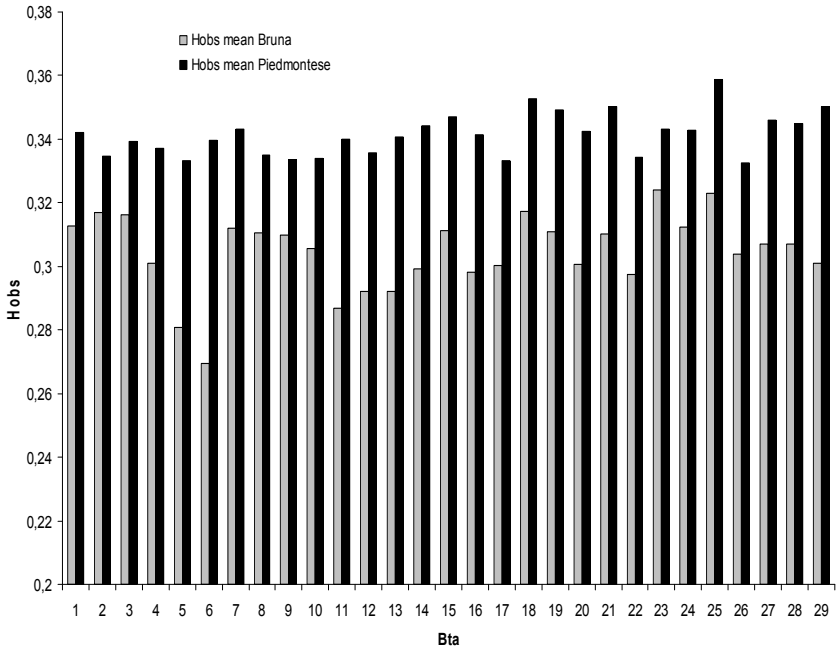

Figure 1: comparison of average heterozygosity per chromosome between the two breeds (grey = Bruna, Black = Piedmontese)

Difference in eterozygosity level, for different cattle breeds, was found also by other authors (Peelman et al., 1998; Cañón et al., 2001; Kantanen et al., 2000 and Ciampolini et al., 1995). The largest difference of heterozygosity was found for BTA6 (0.07). These figures agree with the selection history of the two breeds. The Italian brown has been strongly selected for milk production traits and the BTA6 is known to harbour important genes affecting milk yield and, mainly, composition (Cohen-Zinder et al., 2005). Results of the LOWESS were reported for BTA6, BTA2, BTA20 and Bta14. Such a choice was motivated both by the observed large difference in heterozygosity between breeds and by the known presence of genes known to affect milk and beef traits on these chromosomes. Figure 2 shows the pattern of observed heterozygosity (Hobs) calculated for SNP located along chromosome 2.

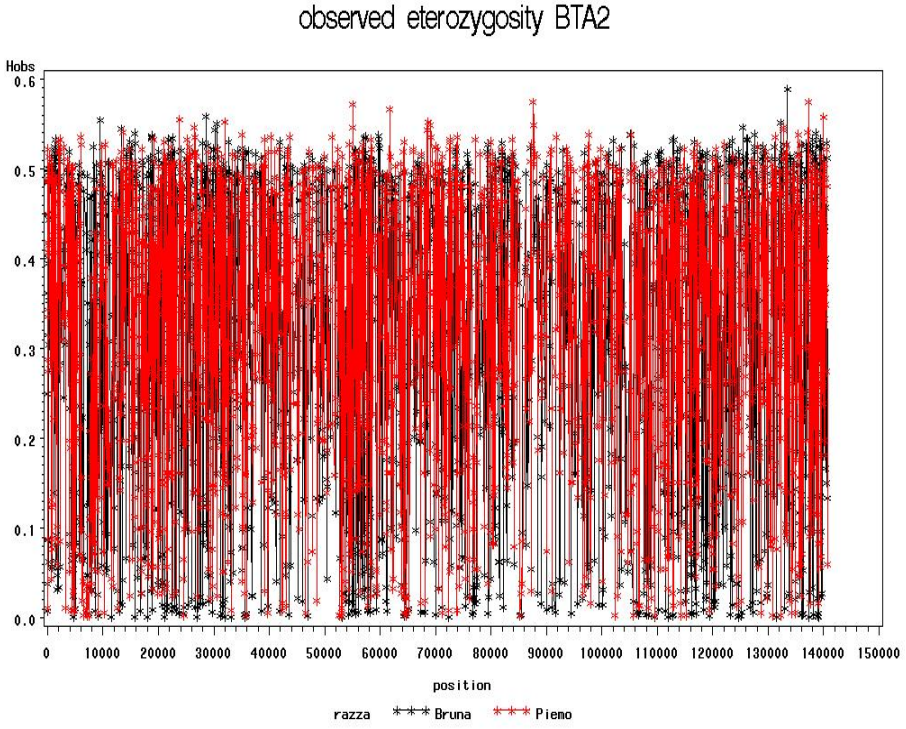

Figure 2: pattern Hobs calculated for SNP located along chromosome 2, in black are represented the brown breed and in red the Piedmontese breed

In red are represented the Piedmontese and in black the Brown data, respectively. It can be clearly seen the great variability of the signal and it is rather difficult to distinguish any defined pattern.

Fst values calculated for the same SNPs are reported on figure 3. A pattern characterized by some sharp and high peaks and some background noise can be noticed.

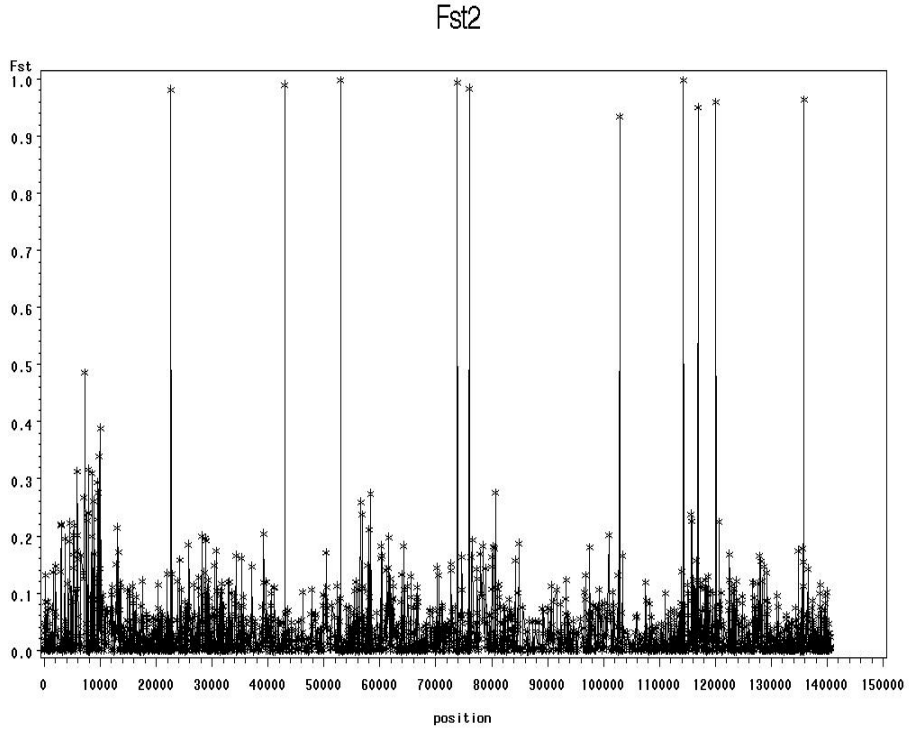

Figure 3: pattern of raw Fst calculated for SNP located along chromosome 2

The effect of the LOWESS correction on data can be observed in figure 4, where data predicted with a smoothing parameter of 0.008 are reported. A part from a reduction in scale due to the regression, a smaller number of high peaks can be observed compared to the raw data (Figure 3).

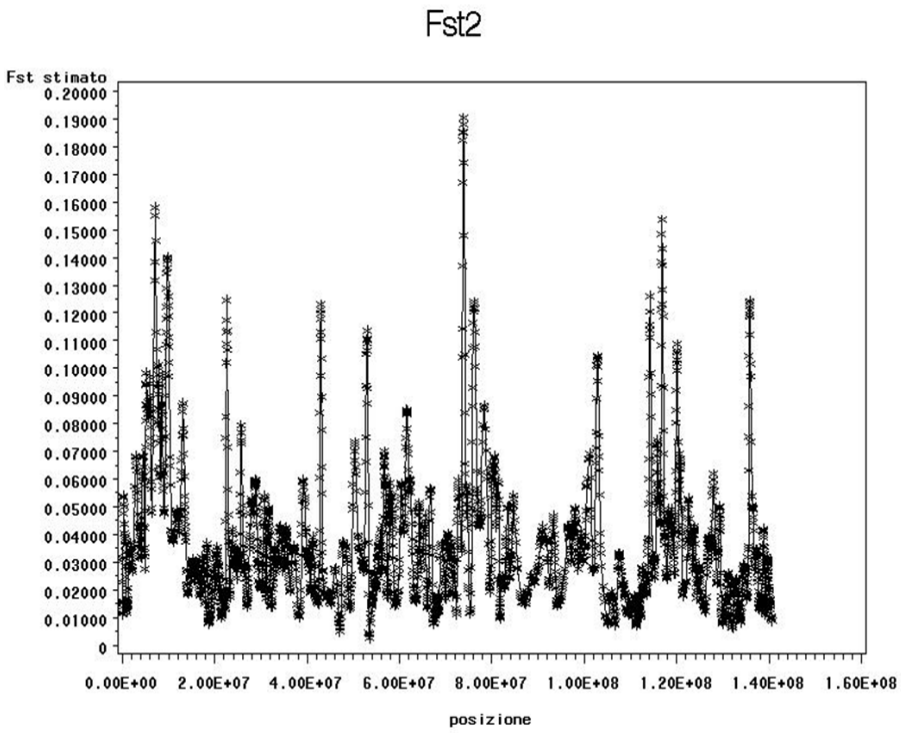

Figure 4: the effect of the LOWESS correction on pattern Fst calculated for SNP located along chromosome 2 with a smoothing parameter of 0.008

Finally, the application of a control chart to the model predicted Fst values are reported in Figure 5. Among the values that exceed the threshold of three standard deviations from the central value, the largest can be observed between 75,861,230-76,399,248 bp. Among annotated genes of potential interest that have been mapped in this region there is Transilin (TSN) a protein expressed on the mammary gland in position 77,018,913-77,026,632. Gene content information was derived from the UCSC Genome Broswer Gateway (http://genome.ucsc.edu/) consultation. A clear peak can be also observed between 6,502,337-7,520,464 bp. It is well known that in position 2q14-q15 between bp 6,532,697 and 6,539,265 is located the MSTN locus that control double muscling phenotype in cattle.

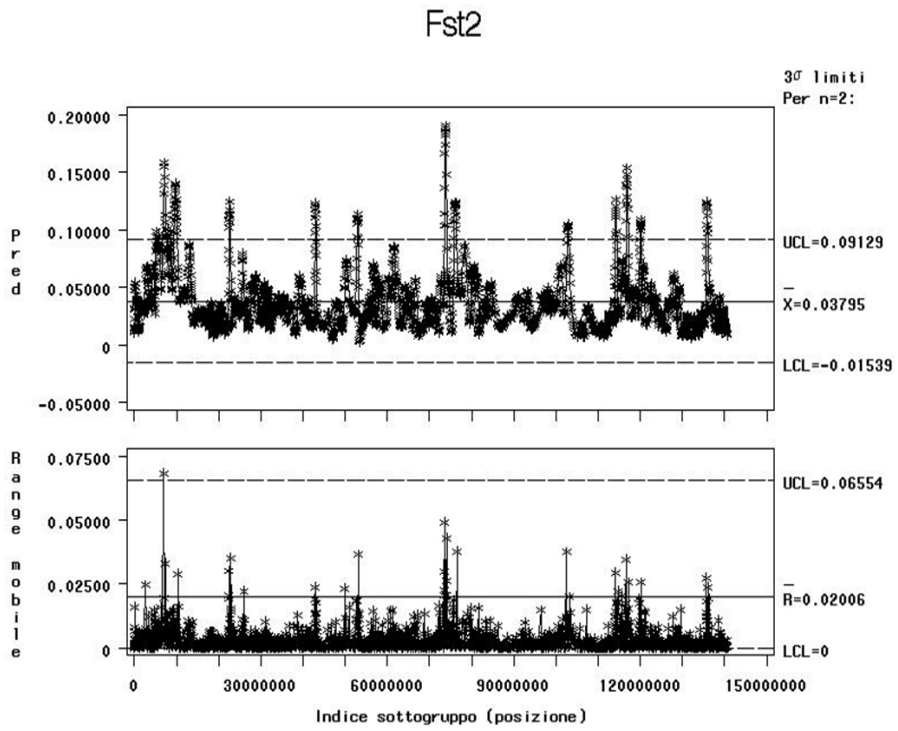

Figure 5: the application of a control chart along chromosome 2

Actually this gene is reported to be fixed for the favourable allele in the Piedmontese breed (Casas et al., 1999). A statistically significant difference in allelic frequencies between different cattle breeds has been reported for this region by Stella et al. (2010). Moreover, another significant peak can be observed between 7,066,570-7,148,685. In this region has been mapped the SLC40A1locus (solute carrier family 40) a gene codifying for a protein called ferroportin 1 (FPN1) that plays an essential role in the regulation of iron levels on the body. Even though the causative mutation of this gene is not present in the SNP chip, the signal has been detected in the adjacent markers. This result agrees with other reports obtained using either allelic frequencies (Stella et al., 2010) or Fst values (Hayes et al., 2008). Similar results can be observed for BTA6.

Compared to BTA2, the pattern of raw Fst (Figure 6) is characterized by a larger number of high peaks.

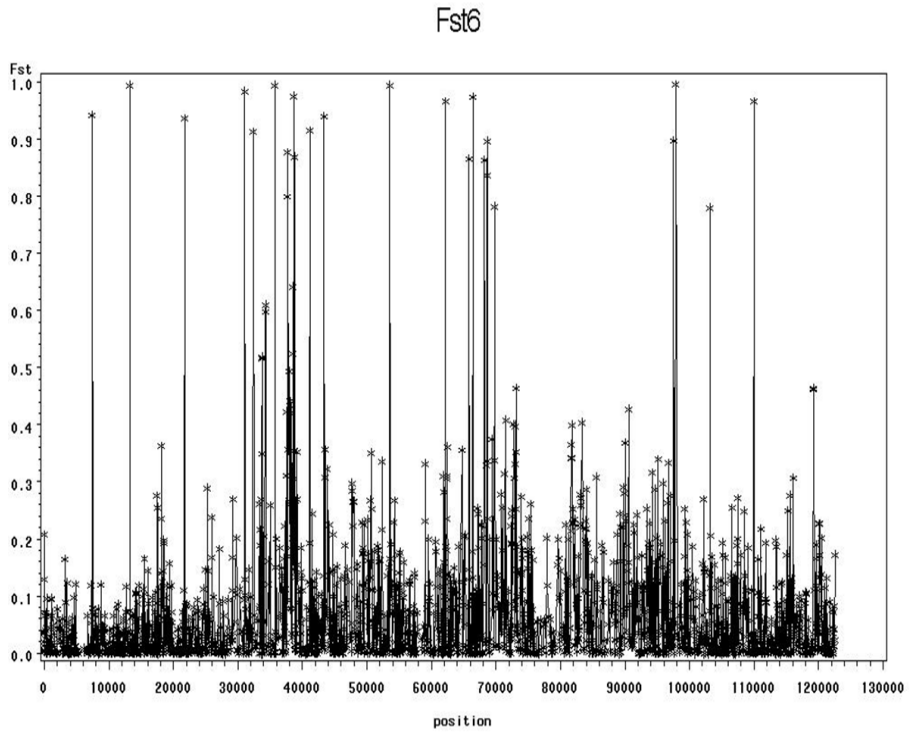

Figure 6: pattern of raw Fst calculated for SNP located along chromosome 6

Actually this is an expected result being this chromosome characterized by a linked multi-QTL architecture (Weikard et al., 2011) and highly subjected to selection in dairy cattle breeds.

Moving from raw Fst (Figure 6) to LOWESS predicted (Figure 7) data, a reduction in the number of relevant peaks can be clearly observed. Among values that exceed limits of the control chart (Figure 8) the highest peak occurs at about 40,000 K bp.

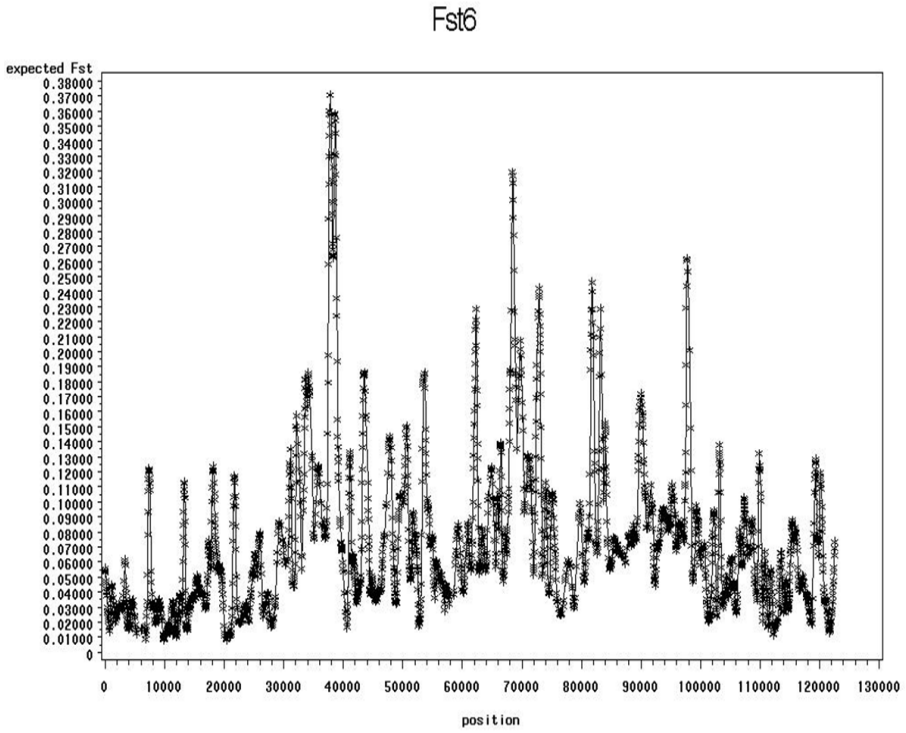

Figure 7: the effect of the LOWESS correction on pattern Fst calculated for SNP located along chromosome 6 with a smoothing parameter of 0.009

In this region are localized some genes known to affect milk production traits like FAM13A1 (36,740,247-36,843,133 bp) (Cohen et al., 2004), ABCG2 (37,342,201-37,433,870 bp), OPN (37,511,672-37,511,830 bp) (de Koning, 2006; Ron and Weller, 2007; Sheehy et al., 2009) and PPARGC1A (44,797,216-44,935,623 bp) (Cohen-Zinder et al., 2005).

A further peak region occurs at around 70,000,000 bp, near the DPP10 *locus* located in position between 70,240,970 and 71,099,937 bp. The DPP10 gene encodes for a protein called dipeptidyl peptidase. In human, mutations in this gene have been associated with asthma (Gao et al., 2010).

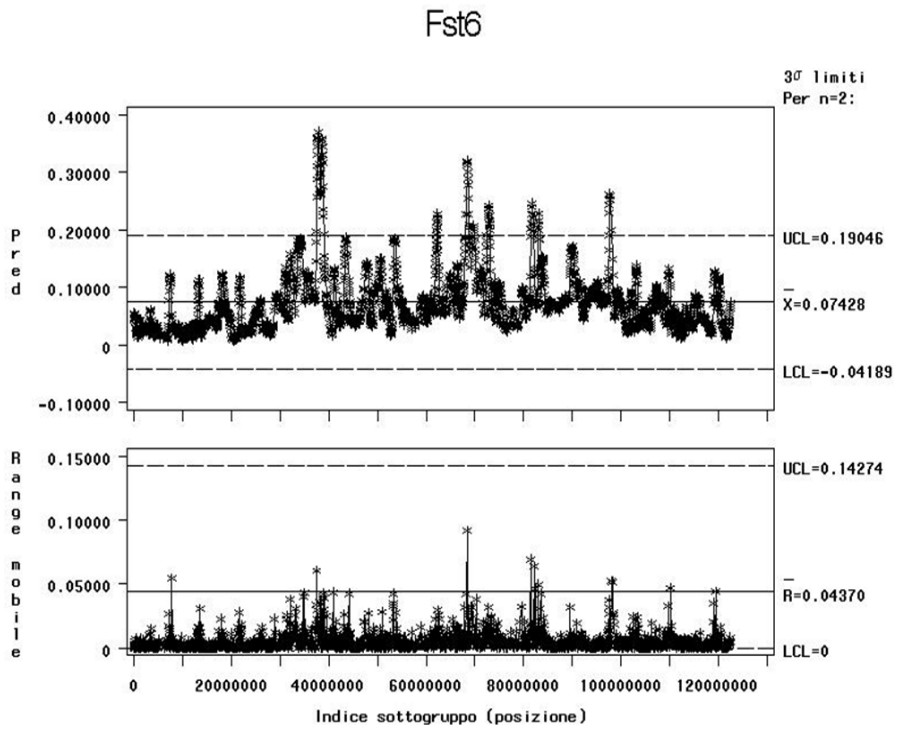

Figure 8: the application of a control chart along chromosome 6

Results obtained on BTA14 (Figure 9) and 20 (Figure 10), are in agreement with previous reports on cattle selection signatures (Hayes et al., 2008; Stella et al., 2010).

In particular, a peak was observed on chromosome 14 (Figure 11) at around 10,000,000 bp, near the thyroglobulin (TG) gene. This TG presents a polymorphism that has been found to be associated with fatness and marbling traits in beef cattle (Barendse 1999; Barendse et al., 2004). On BTA20 a relevant peak around to the 10,000,000 bp and different peaks between 20,000,000-30,000,000 bp were observed (Figure 12). In this case the results obtained by Hayes et al. (2008) showed a more clear peak in close proximity to the GHR gene, that contains a mutation with large effects on protein percentage in milk from dairy cattle (Blott et al., 2003).

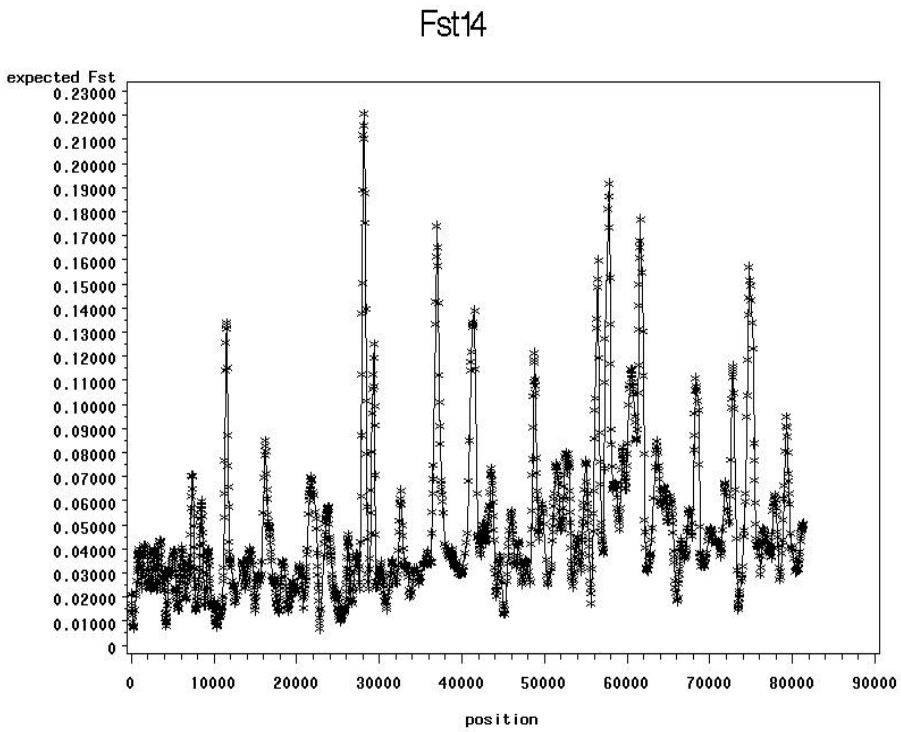

Figure 9: the effect of the LOWESS correction on pattern Fst calculated for SNP located along chromosome 14 with a smoothing parameter of 0.013

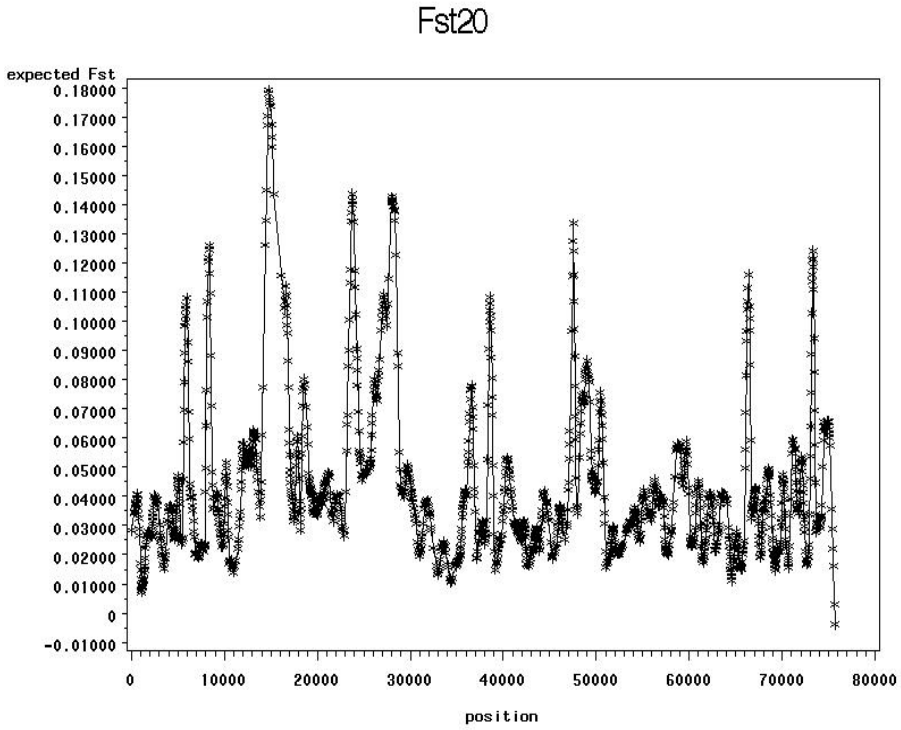

Figure 10: the effect of the LOWESS correction on pattern Fst calculated for SNP located along chromosome 20 with a smoothing parameter of 0.014

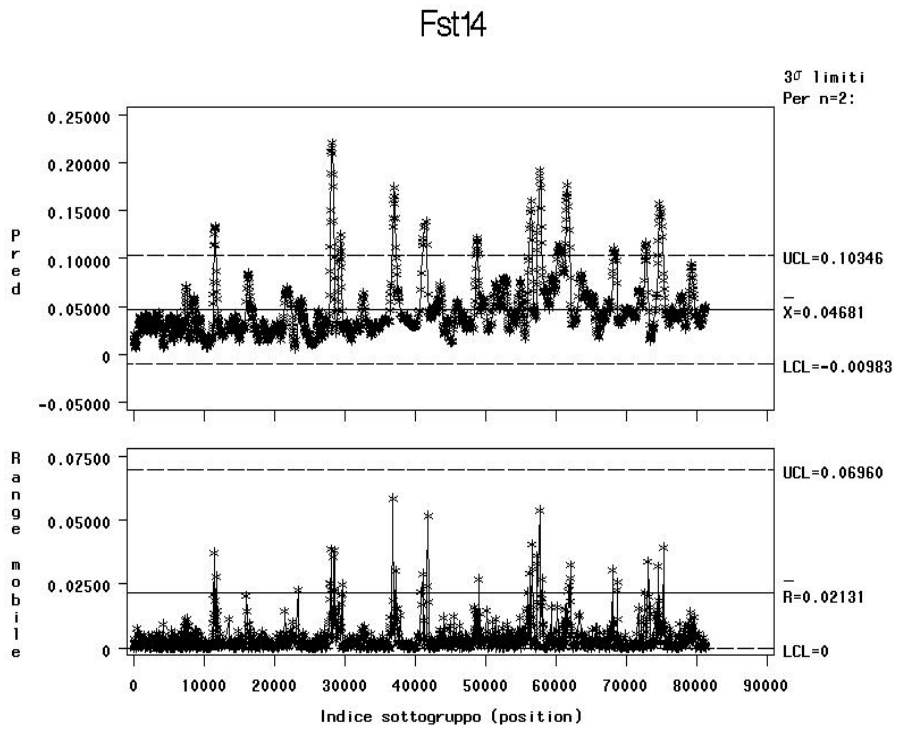

Figure 11: the application of a control chart along chromosome 14

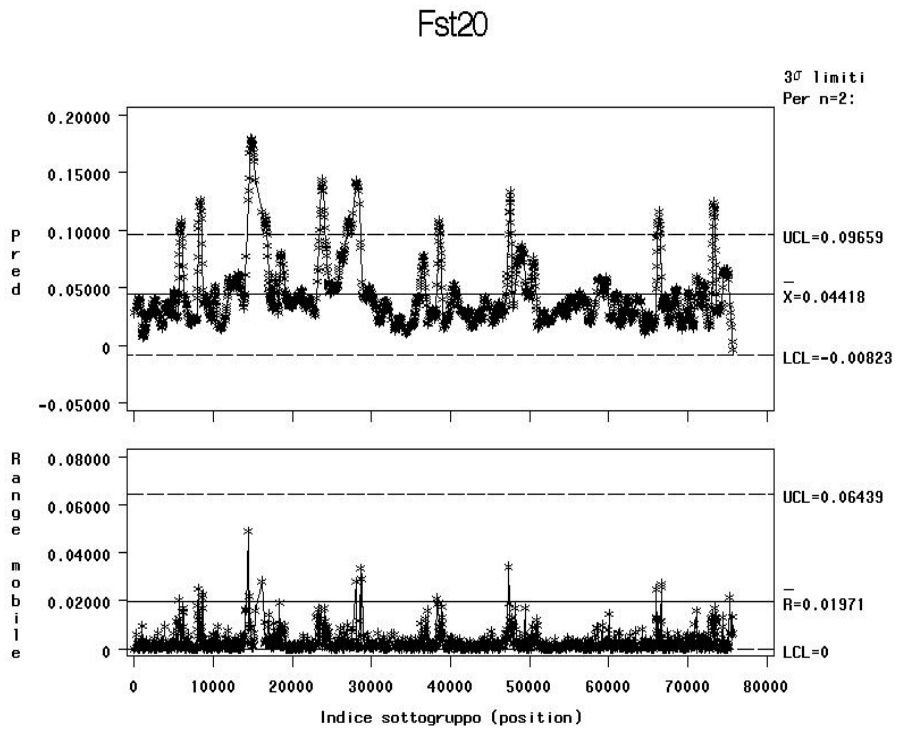

Figure 12: the application of a control chart along chromosome 20

Conclusions

The combined use of a LOWESS regression and a control chart approach was effective in studying the genetic differences between the Piedmontese and the Italian Brown cattle breeds. In particular, the local regression was able to yield a smooth Fst pattern, easy to interpret compared to raw data. The control chart allowed for a quite simple detection of significant Fst values that may indicate selection signatures. The reliability of the method was assessed by results obtained on BTA2, BTA6, BTA14 and BTA20, that are in agreement with previous reports in cattle. Moreover, some regions harbouring genes not yet associated to traits of economic importance for livestock have been detected. The methodology could be proposed as an easy approach for performing a whole genome scan in studies aimed at identifying selection signatures.

Acknowledgments

This work was carried out using data obtained in studies made in SELMOL project.

References

Allan M.F., Thallman R.M., Cushman R.A., Echternkamp S.E., White S.N., Kuehn L.A., Casas E., and Smith T.P.L. (2007) Association of a single nucleotide polymorphism in SPP1 with growth traits and twinning in a cattle population selected for twinning rate. J ANIM SCI, 85:341-347.

Andersson L., and Georges M. (2004) Domestic-animal Genomics: Deciphering the genetics of complex traits. Nature Reviews Genetics, Vol 5: 202-212.

Barendse W.J. (1999) Assessing Lipid Metabolism. Patent number WO 99/23248.

Barendse W.J., Bunch R., Thomas M., Armitage S., Baud S. & Donaldson N. (2004) The TG5 thyroglobulin gene test for a marbling quantitative trait loci evaluated in feedlot cattle. Australian Journal of Experimental Agriculture 44, 669-74.

Barendse W., Harrison B.E., Bunch R.J., Thomas M.B. and Turner L.B. (2009) Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. BMC Genomics 10:178.

Biswas S. and Akey J.M. (2006) Genomic insights into positive selection. TRENDS in Genetics, Vol. 22: 437-446.

Blott S., Kim J.J., Moisio S., Schmidt-Kuntzel A., Cornet A., Berzi P., Cambisano N., Ford C., Grisart B., Johnson D., Karim L., Simon P., Snell R., Spelman R., Wong J., Vilkki J., Georges M., Farnir F. and Coppieters W. (2003) Molecular dissection of a quantitative trait locus: a

phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yeld and composition. Genetics 163, 253-66.

Burt D.W. (2009) The cattle genome reveals its secrets. Journal of Biology, Vol. 8: 36.

Cañón J., Alexandrino P., Bessa I., Carleos C., Carretero Y., Dunner S., Ferran N., Garcia D., Jordana J., Laloë D., Pereira A., Sanchez A., Moazami-Goudarzi K. (2001) Genetic diversity measures of local European beef cattle breeds for conservation purposes. Genet. Sel. Evol., 33: 311-332.

Casas E., Keele J.W., Fahrenkrug S.C., Smith T.P., Cundiff L.V. and Stone R.T. (1999) Quantitative analysis of birth, weaning, and yearling weights and calving difficulty in Piedmontese crossbreds segregating an Inactive Myostatin Allele. J ANIM SCI., 77: 1686-1692.

Charlesworth B. (2007) A hitch-hiking guide to the genome: a commentary on "The hitch-hiking effect of a favourable gene by John Maynard Smith and John haig". Genet. Res. Camb., Vol. 89: 389-390.

Ciampolini R., Moazami-Goudarzi K., Vaiman D., Dillmann C., Mazzanti E., Foulley J.L., Leveziel H. and Cianci D. (1995) Individual multilocus genotypes using microsatellite polymorphisms to permit the analysis of the genetic variability within and between Italian beef cattle breeds. J. Animal Sci., 73:3259-3268.

Cleveland W.S. (1979) Robust Locally Weighted Fitting and Smoothing Scatterplots. Journal of the American Statistic Association, Vol. 74: 829-836.

Cohen M., Reichenstein M., Everts-van der Wind A., Heon-Lee J., Shani M., Lewin H.A., Weller J.I., Ron M. and Seroussi E. (2004) Cloning and characterization of FAM13A1 a gene near a milk protein QTL on BTA6: evidence for population-wide linkage disequilibrium in Israeli Holstein. Genomics, Vol. 84: 374-383.

Cohen-Zinder M., Seroussi E., Larkin D.M., Loor J.J., Everts-van der Wind A., Lee J.H., Drackley J.K., Band M.R., Hernandez A.G., Shani M., Lewin H.A., Weller J.I., Ron M. (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res., Jul. 15(7): 936-44.

Cohen R.A. (1999) An introduction to PROC LOESS for local regression. SAS Institute Inc. Cary, North Carolina, USA.; paper 273.

de Koning D-j. (2006) Conflicting candidates for cattle QTLs. Trends in Genetics, Vol. 22 Issue 6: 301-305.

Doebley J.F., Gaut B.S. and Smith B.D. (2006) The Molecular Genetics of Crop Domestication. Cell 127 29 1309-1321.

Flori L., Fritz S., Jaffrézic F., Boussaha M., Gut I., Heath S., Foulley J-L., and Gautier M. (2009) The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. PloS One, Vol. 4 Issue 8: e6595.

Gao J., Li W., Willis-Owen S.A., Jiang L., Ma Y., Tian X., Moffat M., Cookson W., Lin Y., and Zhang Y. (2010) Polymorphisms of DHF11 and DPP10 are associated with asthma and related traits in a Chinese polulation. Respiration, Vol.79 Issue 1: 17-24.

Groeneveld L.F., Lenstra J.A., Eding H., Toro M.A., Scherf B., Pilling D., Negrini R., Finlay E.K., Jianlin H., Groenveld E., Weigend S., and The GLOBALDIV Consortium (2010) Genetic diversity in farm animals- a review. Animal Genetics, Vol. 41, Suppl. 1 6-31.

Hayes B.J., Chamberlain A.J., Maceachern S., Savin K., McPartlan H., MacLeod I. (2008) A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle. Animal Genetics, Vol. 40, Issue 2, 176–184.

Kantanen J., Olsaker I., Holm L.-E., Lien S., Vilkki J., Brusgaard K., Eythorsdottir E., Danell B., and Adalsteisson S. (2000) Genetic Diversity and Population Structure of 20 North European Cattle Breeds. The American Genetic Association, Vol. 91: 446-457.

Karlsson S. and Moen T. (2010) The power to detect artificial selection acting on single loci in recently domesticated species. BMC Research Notes, 3 232.

Mäki-Tanila A., Fernandez J., Toro M. and Meuwissen T. (2010) Local Cattle Breeds in Europe. Development of policies and strategies for self-sustaining breeds. Wageningen Academic Publishers, Chap. 6, 98-117.

Palaisa K.A., Morgante M., Williams M. and Rafalski A. (2003) Contrasting Effects of Selection on Sequence Diversity and Linkage Disequilibrium at Two Phytoene Synthase Loci. The Plant Cell, Vol. 15: 1795-1806.

Peelman L.J., Mortiaux F., Van Zeveren A., Dansercoer A., Mommens G., Coopman F., Bouquet Y., Burny A., Renaville R., Portetelle D. (1998) Evaluation of the genetic variability of 23 bovine microsatellite markers in four Belgian cattle breeds. Animal Genetics, 29, 161-167.

Przeworki M., Graham C., and Wall J.D. (2005) The Signature of Positive Selection on Standing Genetic Variation. Evolution, Vol. 59, Issue 11, 2312-2323.

Qanbari S., Pimentel E.C.G., Tetens J., Thaller G., Lichtner P., Sharifi A.R., Simianer H. (2010) A genome-wide scan for signatures of recent selection in Holstein cattle. Animal Genetics, Vol. 41, Issue 4, 377–389.

Ron M. and Weller J.I. (2007) From QTL to QTN identification in livestock-winning points rather than knock-out: a review. Animal Genetics, Vol. 38, 429-439.

Sheehy P.A., Riley L.G., Raadsma H.W., Williamson P., and Wynn P.C. (2009) A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. Animal Genetics, Vol. 40, Issue 4, 492-498.

Shewart W.A. (1931) Economic Control of Quality of Manifactured Products ASQC Milwaukee.

Sonstegard T., Ma L., Van Tassell C., Wiggans G., Crooker B., Garbe J., Fahrenkrug S., Mariani B., Liu G. and Da Y. (2009) Genome Signatures of artificial Selection in Holstein Cattle. Cold Spring Harbor Meeting.

Stella A., Ajmone-Marsan P., Lazzari B., Boettcher P. (2010) Identification of Selection Signatures in Cattle Breeds Selected for Dairy Production. Genetics, Vol. 185, 1451-1461.

Weikard R., Widmann P., Buitkamp J., Emmerling R. and Kuehn C. (2011) Revisiting the quantitative trait loci for milk production traits on BTA6. Animal Genetics.

Weir B.S. and Cockerham C.C. (1984) Estimating F-Statistics for the Analysis of Population Structure. Evilution, Vol. 38, No. 6, 1358-1370.

Weir B.S., Cardon L.R., Anderson A.D., Nielsen D.M. and Hill W.G. (2005) Measures of human population structure show heterogeneity among genomic regions. Genome Res., 15:1468-1476.

Winding J.J. and Engelsma K.A. (2010) Perspective of genomics for genetic conservation livestock. Conserv. Genet., Vol. 11 635-641.

CHAPTER 4

General Conclusions

General Conclusions

The availability of high resolution molecular tools has represented a landmark for the developement of researches on the genome structure in several animal species. Particularly, studies on the assessment of genetic diversity between breeds or populations largely benefit from the implementation of genomic methods. The concept of biodiversity is usually associated with the safeguard of wild species or populations and, in general, with the conservation of the existing richenss of alleles. Processes of domestication and artificial selection from which livestock species and breeds have originated, are considered somewhat unfavourable events for the conservation of genetic diversity. Therefore biodiversity issues could be considered not relevant for the general purposes of animal breeding. On the contrary, the study of the genetic basis of differences between ethnic groups represents a key topic also for farm animals. Identification of genomic regions that harbour genes affecting traits of economic interest, the development of traceability protocols based on molecular markers, and the assessment of genetic relationships between breeds are examples of specific requirements of animal production industry that have been fulfilled by researches on genetic diversity.

A wide range of molecular markers (single genes, microsatellites, SNPs) and statistical approaches are available for assessing genetic diversity. Features such as degree of polymorphism, location across the genome, map density, linkage with possible causal mutation, markedly differ between marker categories. The choice of the most suitable molecular methodology should be

strictly consistent with the aim of the research. Also the statistical approach should be evaluated according to this criterion even though most of the algorithims, also implemented in popular softwares, basically rely on the comparison of allelic frequencies between groups of individuals.

The two experimental contributions reported in this PhD thesis are an example of studies on genetic diversity that are developed for addressing specific issues of animal breeding and production. Two different aims, breed traceability and detection of selection signatures, have been pursued using different approaches.

The development of protocols for detecting breed-specific typical products represents a widely accepted strategy for the valorization of local breeds. In the first experimental contribution, the three breeds considered were of limited population size, not subjected to strong artificial selection and genetically related. Thus the comparison to be perfomed was not easy, giving the potential genetic similarity. In any case, the use of a single gene (MC1R) was effective in distinguishing between the Cinisara and the other two populations. So a relatively simple approach, based on a single locus with a moderate polymorphism (in this specific case), may allow for the development of a traceability protocol. The techniques used, PCR-RFLP and DNA sequencing, can now be considered as "traditional" in the field of molecular genetics. A further interesting result has been the detection of the rare allele E^2 in two of the three breeds considered. The detection of this variation still underlines the role of local breeds of livestock as sources of genetic variation that has been probably lost in

specialized breeds. Meaning and possible phenotypic effects of the E^2 allele needs further investigations. A more comprehensive approach, including also other genes that affect coat colour is probably needed.

The second study was developed in a completely different scenario. Two highly selected breeds, with different breeding goals (dairy and beef), and the use of an high density (50K) biallelic marker map. In this case the major issue was represented by the processing of such a huge amount of data and by their interpretation. This is a rather common problem in genomics and it will be more relevant in the very next future when sequences of the whoe genome will be stored and processed. The combined use of a smoothing technique based on local regression and of a control chart allowed for the reduction of the background variation that usually characterizes measures of diversity in dense marker maps and the identification of points of possible divergent selection.

Acknowledgments

Un sentito ringraziamento va al Prof. Nicolò P.P. Macciotta per essere stata una ottima guida in questi tre anni. Ringrazio tutto il Dipartimento di Scienze Zootecniche ed in particolare Roberto, Alberto, lo "stimato" Giustino Ziu Gaspa, Maura, M.Grazia, M.Annunziata, Gianfranco, Claudio, Matteo, Oscar Natalia e il Dott. Gianni Battacone. Un grazie speciale va inoltre a Silvia per avermi insegnato e per la sua preziosa collaborazione.