UNIVERSITA' DEGLI STUDI DI PADOVA

SEDE AMMINISTRATIVA: UNIVERSITA’ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE ANIMALI

SCUOLA DI DOTTORATO DI RICERCA IN: SCIENZE ANIMALI
INDIRIZZO: GENETICA, BIODIVERSITA’, BIOSTATISTICA E BIOTECNOLOGIE

CICLO XX

The use of molecular markers for analyzing genes and genomes of livestock

Direttore della Scuola : Ch.mo Prof. Luigi Gallo
Supervisore : Ch.mo Prof. Giovanni Bittante

Dottoranda : Marica Soattin

31 gennaio 2008

If we knew what it was we were doing,
it would not be called research....

-Albert Einstein-

Summary

The present thesis has been developed considering three different livestock species such as chicken, cattle and sheep. The aim of the study was the evaluation of the application of molecular makers in order to assay the genetic population structure on seven local breeds of chicken, to evaluate the applicability of candidate genes as support of conventional breeding on Piedmontese cattle breed and to detect new SNPs on a sheep population. The first two researchs were carried out at Department of Animal Science of University of Padova while the last one at Reprogen (Faculty of Veterinary Science, University of Sydney, NSW, AUS).

In recent years molecular markers have become available in both animal and plant systems for basic and applied studies. One of the most extensive use of molecular markers has been the development of detailed genetic and physical chromosome maps in a variety of organisms including animal species and humans and also plant systems. Another important application of molecular markers involves improvement in the efficiency of conventional breeding programs by carrying out indirect selection through molecular markers linked to the traits of interest both simple (Mendelian) and quantitative trait loci (QTL) because these markers are not influenced by the environment and can be scored at all stages of individual growth. In addition to these two major applications, DNA markers can also be used for germplasm characterization, genetic diagnostics, study of genome organization, phylogenetic analysis purposes, etc.

In Italy more than fifty different breeds of chicken (*Gallus gallus* L.) are known to have been locally present in the past. The overall situation is now critical since most of these breeds are becoming extinct or threaten and only a few are object of conservation plans. The adoption of molecular markers for the analysis of chicken populations could help characterizing their genetic variation and preserving them from genetic erosion. Both valuable and irreplaceable sources of chicken germplasm, such as indigenous populations of the Veneto region were analyzed by means of DNA fingerprinting with molecular markers designed on interspersed mini- and microsatellite repeats. The identification of either among-breed discriminant and breed-specific markers was based on the S-SAP and M-AFLP systems derived from the AFLP technology. Genomic DNA

fingerprints were generated in 84 individuals belonging to six local breeds (Ermellinata, Padovana, Pépoi, Polverara, Robusta lionata and Robusta maculate) and one commercial line used as reference standard. A number of variation statistics were computed for assessing the genetic variability within and relatedness among breeds: the effective number of alleles per locus (n_e=1.570), total and single-breed genetic diversity (H_T=0.365 and H_S=0.208, respectively) and the fixation index (G_{ST}=0.433). The mean genetic similarity coefficients within and between local breeds were 0.769 and 0.581, respectively. Analysis of the population structure along with individual reassignment tests successfully identified all breed clusters and subclusters. The vast majority of animals were correctly assigned to their breed of origin with a probability threshold of 90%. Markers exploitable for the genetic traceability of breeds revealed significant sequence similarities with either genic or intergenic regions of known chromosome position. Sequence tagged site primers were designed for the most discriminant markers in order to develop multiplex non-radioactive genomic PCR assays. The identification of single local breeds according to sequence-specific SNPs and haplotypes was preliminarily attempted and the polymorphism information content of genomic AFLP-derived markers was also reported and critically discussed.

A reliably tender meat product is one of the most important attributes for maintaining consumer satisfaction with beef steaks. Selection programs for genetic improvement in tenderness might be implemented in cattle if candidate genes and appropriate markers could be found. Calpain 1 (*CAPN1*), a gene encoding the enzyme μ-calpain which degrades myofibrillar proteins post-mortem, is thought to be one of the most promising candidates for beef tenderness. Three distinct single-nucleotide polymorphisms (SNPs) in the *CAPN1* gene were studied in a population of cattle to determine their frequency and the frequency of their haplotypes. A total of 170 animals were subjected to genotyping for the *CAPN1* gene-associated SNP markers 316 (G>C), 530 (G>A) and 4751 (T>C). The detection of nucleotides in the three sequence positions of the *CAPN1* gene was performed by using the ARMS-PCR and the PCR-RFLP methods. Genetic diversity statistics and inbreeding coefficients were computed, and population structure and animal ordination analyses were also performed. Moreover, Linkage Disequilibrium (LD) values were estimated using the standardized disequilibrium coefficient (D′) and the correlation coefficient (r^2). The relative proportions of the single nucleotides at each SNP site were equal to p(G)=0.9176 and q(C)=0.0824 for the marker 316, to p(G)=0.5710 and q(A)=0.4290 for the marker 530, and to p(T)=0.5178 and q(C)=0.4822 for the marker 4751. The observed proportions of single SNP markers were not significantly deviating from proportions expected in case of Hardy-Weinberg equilibrium. However, an excess of heterozygous animals was found for SNP markers 316 and 530, whereas an excess of homozygous animals was found for the SNP marker 4751. The pair-wise comparisons between SNP markers 316 and 530, and 316 and 4751 revealed linkage equilibrium in the population, whereas a highly significant (P<1‰) linkage disequilibrium was scored between SNP markers 530 and 4751. In particular, a marked excess was found for three haplotypes, *i.e.* GG/AA/TT, GG/GG/CC, and GG/GA/TC, while deficiency was found for most of the remaining haplotypes. The availability of preliminary data on SNP haplotypes can be useful to plan experiments aimed at testing the relationship between a candidate gene, like *CAPN1*, and a number of tenderness parameters, such as shear force and driploss.

Leptin is considered important in animal production and reproduction in livestock species. The gene encodes for a 16 kDa protein hormone that plays a key role in regulating energy intake and energy expenditure, including important physiological functions, such as the body weight regulation, reproduction, bone formation and growth as well as immunological functions. This study deals with the cloning and analysis of the sheep leptin gene in order to reconstruct its molecular structure and to discover its polymorphism content. Genomic DNA samples of four sires deriving from an extreme mating Awassi × Merino were used as template with the aim of amplifying the sheep leptin gene using primers designed on conserved regions of homologous genes. A sequence of the sheep leptin gene as long as 4,883 bp was recovered, including 714 bp of the promoter, 1,189 bp of exon 1 and intron 1 regions, 539 bp of exon II, 1,262 bp of intron II, 570 bp of exon III and 609 bp of the 3′-UTR. Multiple sequence analysis allowed to discover two SNPs in exon III at position 170 (G>A) and 332 (G>A). Both polymorphisms determined a missense mutation (Arg>Gly). The SNP G(170)A was adopted for the genotyping of animals by means of PCR-RFLP markers. A number of 255 animals belonging to the backcross and the double backcross Awassi × Merino progenies were analyzed in order to evaluate the frequency of the SNP in single families and in the population as a whole. The leptin gene-related SNP marker discovered in this study along with the PCR-RFLP protocol set for individual DNA genotyping will be useful for association mapping purposes and discovering quantitative trait nucleotides correlated to milk yield and quality components in sheep.

Riassunto

Il lavoro sviluppto nell'ambito della presente tesi di Dottorato è stato svolto su tre diverse specie animali: pollo, bovino e pecora. Gli scopi del lavoro sono stati: la valutazione dell'impiego di marcatori molecolari come strumenti a supporto dello studio della struttura genetica di sette razze locali di pollo, l'identificazione di geni candidati a sostegno del breeding convenzionale e la rilevazione di nuovi SNP a carico del gene della leptina in un campione di pecore. Le prime due ricerche sono state effettuate presso il Dipartimento di Scienze Animali dell'Università di Padova mentre l'ultima presso il centro Reprogen della Facoltà di Scienze Veterinarie dell'Università di Sydney (NSW, AUS)

Da alcuni anni i marcatori molecolari sono diventati disponibili sia in piante che in animali. Uno dei più diffusi impieghi dei marcatori molecolari è stato lo sviluppo di dettagliate mappe genetiche e fisiche in differenti organismi includenti animali, uomo e piante. Poiché i marcatori molecolari non sono influenzati dall'ambiente e posso essere individuati in qualsiasi stadio della crescita dell'individuo, essi possono essere vantaggiosamente impigati nel miglioramento del breeding convenzionale attraverso selezione indiretta sfruttando l'associazione dei marcatori molecolari con caratteri semplici (Mendeliani) e caratteri quantitativi (QTL). Tali strumenti possono essere usati per la caratterizzazione di germoplasma, per diagnosi genetica, per studi di organizzazione genetica e di filogenesi.

Negli ultimi anni, in Italia, sono state censite più di 50 differenti razze locali di pollo (*Gallus gallus* L.). Complessivamente però, gran parte di queste razze sono a rischio estinzione e solo poche sono oggetto di piani di conservazione. L'adozione dei marcatori molecolari per l'analisi di popolazioni di pollo potrebbe aiutare a caratterizzare la loro variazione a livello genomico e dunque preservarli da una potenziale erosione genetica. La loro estinzione determinerebbe una perdita irrecuperabile di geni utili per la loro rusticità, adattabilità e resistenza alle malattie. Alcune razze locali Venete di pollo sono state dunque studiate allo scopo di analizzare i profili genomici (fingerprints) generati attraverso marcatori molecolari sviluppati su regioni mini- e micro-satellite altamente ripetute. L'identificazione di marcatori razza-

specifici è stata attuata con l’ausilio di marcatori S-SAP e M-AFLP derivanti dai ben noti marcatori AFLP. I fingerprints dei DNA genomici sono stati generati per 84 individui appartenenti a sei razze locali (Ermellinata, Padovana, Pépoi, Polverara, Robusta lionata e Robusta maculata) e una linea commerciale utilizzata come riferimento. I dati molecolari raccolti sono stati usati per calcolare la variabilità genetica all’interno di ogni singola razza e tra le razze. Il numero effettivo di alleli per locus (n_e) è risultato pari a 1,570, la diversità genetica totale (H_T) e per singola razza (H_S) sono risultate pari a 0,365 e 0,208, rispettivamente e l’indice di fissazione (G_{ST}) pari a 0,433. I coefficienti di similarità genetica media entro e tra le razze sono risultati pari a 0,769 e 0,581, rispettivamente. L’analisi di popolazione, insieme con il test di assegnazione individuale, hanno identificato con successo tutte le razze ragruppandole in ben distinti gruppi e sottogruppi. La maggior parte degli animali è stata correttamente assegnata alla loro razza di origine con una probabilità del 90%. I marcatori razza-specifici e/o polimorfici tra le razze hanno inoltre, evidenziato significativa similarità per regioni geniche e intergeniche in posizioni cromosomiche note. Le sequenze dei marcatori più discriminanti sono state usate per disegnare primer specifici allo scopo di sviluppare marcatori di tipo SCAR in grado di identificare le razze attraverso semplici saggi di PCR. Il risultato più importante è stato ottenuto per il clone #38 che è risultato specifico per la razza Polverara. Infatti, quando i marcatori SCAR, sviluppati per questo clone sono stati impiegati per analizzare tutte le razze oggetto di studio, sono stati caratterizzati due distinti prodotti di amplificazione di 307 e 333 pd. Solo il più alto è risultato condiviso dalla maggior parte degli individui mentre il più basso è stato individuato soltanto negli individui della razza Polverara con una frequenza relativa del 52% (32 individui su 62). Il sequenziamento dei marcatori razza specifici ha consentito, inoltre, di caratterizzare SNP razza specifici e quindi aplotipi che potrebbero essere sfruttati per l’identificazione di razza e per lo sviluppo di saggi per la tracciabilità dei prodotti derivanti da queste razze.

La tenerezza della carne rappresenta una delle peculiarità maggiormente apprezzate dal consumatore. Programmi di selezione per il miglioramento genetico del carattere tenerezza possono essere implementati nella specie bovina se geni deputati al controllo di tale carattere fossero identificati. Il più promettente tra i geni deputati al controllo della tenerezza della carne risulta essere la Calpaina 1 (*CAPN1*), un gene che codifica l'enzima μ-calpaina che degrada le proteine miofibrillari durante la frollatura delle carcasse. Nel presente studio, tre differenti SNP a carico del gene *CAPN1* sono stati studiati in una popolazione bovina di razza Piemontese allo scopo di valutare la loro frequenza e quella dei loro aplotipi. In totale, 170 animali sono stati genotipizzati per i tre marcatori associati al gene *CAPN1* chiamati: 316 (G>C), 530 (G>A) e 4751 (T>C). La rilevazione dei nucleotidi nelle tre posizioni del gene è stata attuata attraverso i marcatori di tipo ARMS-PCR e PCR-RFLP. I dati raccolti da tali analisi sono stati usati per calcolare statistiche di diversità genetica, coefficienti di inbreeding e analisi della struttura di popolazione. Inoltre, sono stati calcolati i valori di *Linkage Disequilibrium* (LD) attraverso il coefficiente standardizzato di disequilibrio (D′) e il coefficiente di correlazione (r^2). I rapporti relativi dei singoli nucleotidi ad ogni SNP sono risultati uguali a p(G)=0.9176 e q(C)=0.0824 per il marcatore 316, a p(G)=0.5710 e q(A)=0.4290 per il marcatore 530 e a p(T)=0.5178 e q(C)=0.4822 per il marcatore 4751. Le proporzioni osservate per i singoli marcatori non sono risultate significativamente devianti da quelle attese nel caso di equilibrio di Hardy-Weinberg. Comunque, è stato rivelato un eccesso di animali eterozigoti per i marcatori 316 e 530 mentre un eccesso di individui eterozigoti è stato individuato per il marcatore 4751. Il confronto a coppie tra i marcatori 316 e 530 e i marcatori 316 e 4751 ha registrato una condizione di *linkage equilibrium* nella popolazione, mentre una situazione di LD è risultata altamente significativa (P<1‰) tra i marcatori 530 e 4751. In particolare, in elevato eccesso erano i tre aplotipi GG/AA/TT, GG/GG/CC e GG/GA/TC mentre i restanti aplotipi erano in netto difetto. La disponibilità di dati preliminari riguardanti aplotipi generati da diverse combinazioni di SNP può essere utile per pianificare esperimenti volti a saggiare le relazioni tra un gene candidato come *CAPN1* e numerosi parametri legati al carattere tenerezza come la forza di taglio e il driploss.

La leptina è considerata una molecola molto importante nella produzione e riproduzione negli animali di interesse zootecnico. Il gene codifica per un ormone di 16 kDa che gioca un ruolo chiave nella regolazione del bilancio energetico che include importanti funzioni fisiologiche come la regolazione del peso corporeo, la riproduzione, la formazione e sviluppo delle ossa come pure funzioni immunologiche. Il presente

studio riguarda il clonaggio e l’analisi del gene della leptina nelle pecore allo scopo di ricostruire la sua struttura molecolare e di identificare eventuali polimorfismi a caricoo della sequenza del gene. Il DNA genomico di quattro maschi F_1 derivanti da un incrocio Awassi × Merino è stato usato come templato allo scopo di amplificare il gene della leptina in pecora usando coppie di primer disegnate in regioni conservate di geni omologhi. In totale sono state sequenziate 4.883 pb del gene della leptina che includono: 714 pb del promotore, 1.189 pb delle regioni dell’esone 1 e dell’introne 1.539 pb dell’esone II, 1.262 pb dell’introne II, 570 pb dell’esone III e 609 pb del 3′-UTR. L’analisi multipla delle singole sequenze ha consentito di individuare due SNP nell’esone III alle posizioni 170 (G>A) and 332 (G>A). Entrambi i polimorfismi determinano una mutazione di tipo missense in cui l’amminoacido Arginina è sostituto dalla Glicina (Arg>Gly). Lo SNP G(170)A è stato, dunque, utilizzato per la genotipizzazione di pecore attraverso marcatori di tipo PCR-RFLP. Un numero di 255 animali (femmine) appartenenti ad una progenie derivante da incroci di tipo backcross e doppio-backcross Awassi × Merino è stato analizzato allo scopo di valutare la frequenza dello SNP nelle singole famiglie e nella popolazione nel suo complesso. Lo SNP G(170)A identificato nel gene della leptina insieme al protocollo PCR-RFLP, messo a punto per la genotipizzazione individuale, potranno essere utili per studi di mappaggio di associazione per caratteri legati alla resa e alla qualità del latte.

Table of contents

Chapter 1: Use of molecular markers in animal genomics: An overview on the state of the art

Chapter 2: Genomic DNA fingerprinting of chicken (*Gallus gallus* L.) indigenous breeds with M-AFLP and S-SAP markers designed on interspersed repeats

Chapter 4: Investigating SNPs and haplotypes for the calpain (CAPN1) gene in cattle (*Bos taurus* L.) populations: Preliminary analysis of linkage disequilibrium

Chapter 4: Discovering of novel leptin gene-based SNPs and sequence-tagged site genotyping of dairy sheep (*Ovis aries* L.) populations

Chapter 1

Use of molecular markers in animal genomics: An overview on the state of the art

1.1 Molecular markers: definition and classification

All living organisms are known to be made up of cells that are programmed by genetic material called DNA. Only a small fraction of the DNA sequence typically makes up genes coding for proteins, while the vast majority of the remaining DNA represents non-coding sequences. The genetic material is organized into sets of chromosomes, and the entire set is called the genome. In a diploid individual (*i.e.* where chromosomes are organized in pairs), there are two alleles of every gene – one from each parent (Ruan and Sonnino, 2007). Molecular markers can be considered as constant landmarks in the genome, as a matter of fact they cannot be thought as a gene because they do not have any known biological function. They are simply identifiable DNA sequences, found at specific locations of the genome, and transmitted by the standard laws of inheritance from one generation to the next. Moreover, molecular markers rely on DNA assays, in contrast to morphological markers that are based on visible traits and biochemical markers that are based on proteins produced by genes.

In recent years molecular markers have become available in both animal and plant systems for basic and applied studies. One of the most extensive use of molecular markers has been the development of detailed genetic and physical chromosome maps in a variety of organisms including animal species and humans and also plant systems (Gupta *et al.*, 1999). Another important application of molecular markers involves improvement in the efficiency of conventional breeding programs by carrying out indirect selection through molecular markers linked to the traits of interest both simple (Mendelian) and quantitative trait loci (QTL) because these markers are not influenced by the environment and can be scored at all stages of individual growth. In addition to these two major applications, molecular markers can also be used for germplasm characterization, genetic diagnostics, study of genome organization, phylogenetic analysis purposes, etc. (Rafalsky *et al.*, 1996; Barcaccia and Falcinelli, 2007)). Although each marker system is associated with some advantages and disadvantages, the choice of the marker system is dictated to a large extent by the intended application convenience and the cost involved. The information provided to the breeder by the markers varies depending on the type of marker system used. It is crucial the number of differences (polymorphisms) that the molecular marker system is able to discover between two of more individuals, breeds, populations at genome level. A powerful molecular marker is that skilled to identify a high number of polymorphisms.

The molecular markers can be classified in the following two groups:

- Southern-blot hybridization-based DNA markers such as Restriction Fragment Length Polymorphisms (RFLP) and Variable Number of Tandem Repeats (VNTR)

- PCR-based DNA markers such as Random Amplified Polymorphic DNA (RAPD), Arbitrarily Primed-Polymerase Chain Reaction (AP-PCR), Simple Sequence Repeats (SSR), Sequence-Tagged Sites (STS), Amplified Fragment Length Polymorphisms (AFLP), Inter-Simple Sequence Repeat (I-SSR), Sequence Characterized Amplified Regions (SCAR), Cleaved Amplified Polymorphic Sequences (CAPS), Selective Amplified Microsatellite Polymorphic Locus (SAMPL), Sequence-Specific Amplification Polymorphism (S-SAP), Single Nucleotide Polymorphism (SNP).

These two main groups of molecular marker can be further subdivided depending on their ability to detect variations at single locus or at multiple locus level. The SCAR, CAPS, SSR, SNP, RFLP and VNTR markers belong to the former subgroup, whereas RAPD, AP-PCR, AFLP, SAMPL, S-SAP and I-SSR markers belong to the latter subgroup. For some kind of molecular markers included in the second subgroup, such as RAPD, AP-PCR, AFLP, no preliminary DNA sequence information is necessary. By contrast, older molecular marker systems, such RFLP and VNTR, or SNP markers need preliminary DNA sequence information to be investigated. In the future, due to the increasing development of sequence information for different species and organisms, new DNA marker systems will be set up to meet the new research area needs. An overview of the nowadays most important molecular markers is reported below.

RFLP markers: they have been the first genomic DNA-based molecular markers developed from a specific application of Southern blot analysis (Southern, 1975) in which a single stranded DNA used as probe hybridize to complementary DNA sequences. RFLP markers detect variation in DNA sequences at the same loci in different individuals or accessions. Technically, RFLP technology involves the hybridization of cloned DNA to restriction fragments of differing molecular weights from restriction enzyme-cleaved genomic DNA. The digested DNA fragments are separated by agarose gel electrophoresis and transferred as single stranded fragments to filters through capillary action. Filters are then incubated with specific labelled genes or anonymous fragments of single stranded DNA, washed and exposed to X-ray film. The identification of polymorphisms is possible between individuals: the genomic DNA extracted from each individual is digested with various restriction enzymes to find those that produce fragments (bands) that differ in molecular weight between individuals and can be distinguished by hybridization with a given probe. To ensure that probes hybridize to single fragments on a gel, the DNA used as a probe should be from a single or low copy (non-repetitive) region of the genome. Initially, probes were usually produced by shearing or digesting DNA and cloning the fragments into a plasmid vector that allowed the replication of the inserted fragment, several times when it was put in an appropriate host (bacteria). This method is still employed but it is time consuming compared to that in which the probes are obtained by polymerase chain reaction commonly named PCR. The DNA differences detected by RFLP markers result from single base changes causing a loss of restriction sites or a gain of new restriction sites, or from insertions and/or deletions (indels) between restriction sites (McCouch *et al.*, 1988; Edwards *et al*., 2004).

PCR-based markers: since when Mullis *et al.* (1986) have introduced, in the research world, the Polymerase Chain Reaction (PCR), many new and advanced molecular marker technologies have come through. The PCR exploits a thermo-stable DNA polymerase enzyme that makes copies of a specific sequence (target sequence) included between two short primers that are complementary to regions at the borders of the target sequence. The whole PCR process comes through consecutive temperature changes in which the double stranded DNA is separated at 94-95°C, the primers anneal to the complementary regions at ranging temperature of 35-70°C and the target sequence is synthesized at 72°C. These three steps are repeated 25-40 times to allow the exponential amplification of the target sequence flanked by primer binding sites. In general, PCR-based DNA markers require much less DNA per assay than RFLP markers and are more compatible with automated high-throughput genotyping that is the ability to process large numbers of samples quickly and efficiently (Edwards and McCouch, 2007).

One of the first employment of the PCR has been the development of RAPD markers in which sequences of DNA are amplified using simple ten-base long primers. The small size of primer and the low annealing temperature allow the amplification of many loci simultaneously that are assayed in one lane of an agarose gel. This technology has the advantage that it can be applied without any prior sequence knowledge because primer are arbitrary and it is useful when there is a need to assay several loci across the entire genome. The RAPD marker system belongs to the category of dominant markers in which the polymorphisms are detected only as the presence versus absence of a band of a particular molecular weight and it is not possible to differentiate between homozygous and heterozygous genotypes. Overall, RAPD markers are not always reliable because the success of amplification of any DNA fragment may be sensitive to many factors, including DNA template quality, PCR conditions, reagents and equipment (Edwards and McCouch, 2007).

More sophisticated than RAPD markers but also more powerful are AFLP markers (**Fig. 1.1**). They have been firstly proposed by Vos *et al*. (1995) and provide the amplification of restriction fragments ligated to specific adaptors. The isolated total genomic DNA is digested with arbitrary two restriction enzymes [usually one is a four-base cutter (for example *MseI* or *TaqI*) and the other one is a six-base cutter (for example *EcoRI* or *PstI*)] and then ligated to the ends of each fragments with oligonucleotide adaptors. The new resulting fragments are amplified using primers that anneal to the adaptor sequences and extend into the restriction fragment. Only a portion of restriction fragments within the range of 50 and 350 bp in size can be amplified by PCR and visualized on polyacrylamide gels. Large genomes usually require additional selective bases to the primers to reduce the number of co-amplified bands. The reproducibility of AFLP fingerprints is very high compared to the RAPD marker system. Moreover, this technology has great potentials for wide genome screenings since many marker loci can be assayed simultaneously without any prior information. By contrast, AFLP technology needs particular technical skills because the AFLP fragments run on polyacrylamide gels instead of agarose and they also require a larger investment in equipment than RAPD markers.

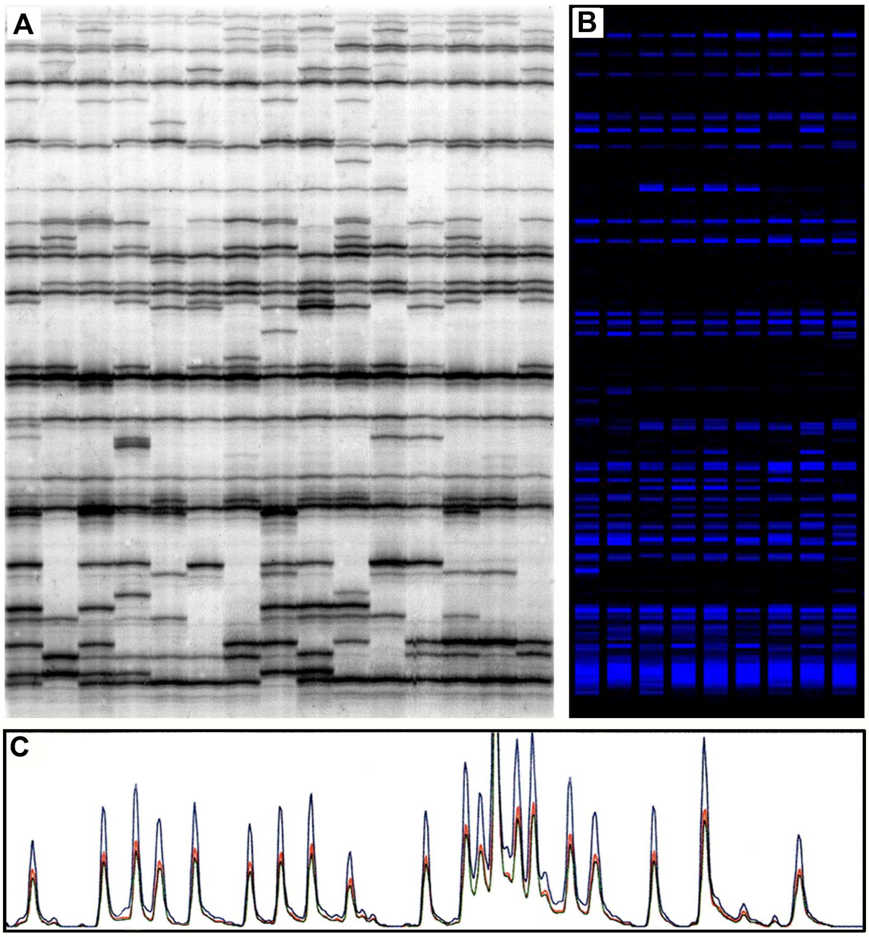


Figure 1.1 Examples of AFLP fingerprints generated using primers with radioactive and fluorescent labels by conventional PAGE analysis (A and B, respectively) or capillary DNA sequencer (C).

At the beginning of their diffusion, AFLP markers were detectable only using silver staining or labelling of the primers with a radioactive isotope (Edwards and McCouch, 2007). Also AFLP markers are included in the class of dominant molecular marker systems. Nowadays, it is possible to detect AFLP markers using an automated DNA sequencer with fluorescently labelled primers. This allows to obtain an higher number of analyzed individuals in few steps, saving time and reducing cost when the size of sample is quite big. A modification of the AFLP procedure is the Diversity Array Technology (DArT) that uses a microarray platform that greatly increases throughput (Jaccoud *et al*., 2001). Exploiting this technology, DNA fragments from one sample are distributed on an array and used to detect polymorphisms for the fragments in other samples by differential hybridization (Wenzl *et al*.., 2004).

Thanks to the increasing knowledge of the genome of different organisms (animal and plants) it is possible to exploit the sequence information to design specific primers for amplification of specific loci throughout the entire genome. An example of this application is represented by Simple Sequence Repeat (SSR) markers, also known as microsatellites. They consist of di-, tri- or tetra-nucleotide motifs repeated several times [for instance, $(CA)_n$, $(GAT)_n$ or $(AGCT)_n$] which are a common feature of most eukaryotic genomes. The number of repeats is highly variable because slipped strand mis-pairing causes frequent gain or loss of repeat units. As this particular characteristic, SSR markers are able to detect a high level of allelic diversity, and so they are valuable as molecular markers, particularly for studies of closely related individuals. For analyzing the polymorphisms at these regions, primers complementary to unique sequences surrounding the highly repeated regions need to be designed (Weber and May, 1989). By measuring the molecular weight of the resulting PCR fragments, differences in the number of tandem repeats are so assayed. The differences between two alleles may be as small as two base pairs and the DNA fragments to be detected are separated by PAGE systems or using capillary DNA sequencers. The SSR can be employed having prior knowledge about the flanking-repeated regions: these regions can be discovered by screening libraries of clones containing the repeat motif that must be sequenced to find unique sites for primer design flanking the repeats. Nowadays, the nucleotide sequence information acquired on many genomes, including human and animal species, is collected in some public databases such as NCBI (available at web site http://www.ncbi.nlm.nih.gov/) in which genome- and gene-specific features can be recovered. Exploiting the public nucleotide databases, the microsatellite marker development from pre-existing sequences is easier and faster (McCouch *et al*. 1997; Zane *et al.*, 2002). Microsatellites discovered in non-coding sequences often have a higher rate of polymorphism than microsatellites discovered in genes.

Microsatellite markers have several advantages over other molecular marker systems: i) they represent a co-dominant assay (*i.e.*, the heterozygous state is easily discerned from the homozygous state); ii) are simply detectable using florescent primers on an DNA sequencer allowing to multiplex several markers with non-overlapping size ranges on a single electrophoresis run; iii) supply highly reproducible polymorphisms; and iv) they are easily exchangeable among different laboratories only by distributing primer sequences. Although SSRs are abundant in most eukaryotic genomes, their genomic distribution may vary and basic information is not always available about repeated motifs and their flanking regions. Therefore, in many cases it is more practical to exploit heterologous primers from species closely related to that under study (for instance: chicken genome-derived SSR primers used for studying turkey microsatellite polymorphisms). Microsatellite polymorphisms are shown in **Figure 1.2**.

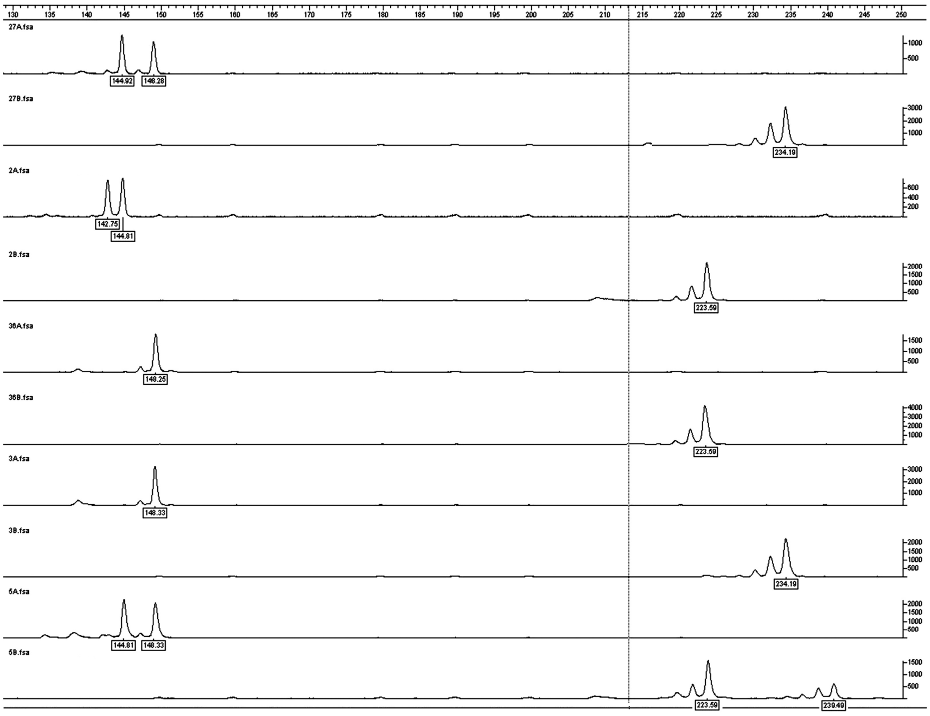


Fig. 1.2. Examples of SSR polymorphisms.

Another type of molecular marker system is represented by Inter-SSR markers that utilize a repeated motif and a region upstream it as anchor point of primers (Zietkietkiewicz *et al*., 1994). The amplification can occur using a single I-SSR primer or a combination of primers anchored to different microsatellite regions or combining a I-SSR primer with an universal AFLP primer (performing the Microsatellite-AFLP or M-AFLP markers). The I-SSRs are a multi-locus molecular markers and they can be visualized on agarose or polyacrylamide gels if they are silver-stained or on automated sequencer if they are fluorescently labelled.

Similarly to M-AFLP, a transposon-anchored primer or a minisatellite-specific primer can be used in combination with a conventional AFLP primer to detect what are often called S-SAP markers. The success of amplification depends on the presence of transposable elements or minisatellites within a restriction fragment. This approach can also be used to assay and visualize polymorphisms, simultaneously at many loci throughout the genome, related to multigene families using primers designed on gene domains. The same approach can be exploited to assay for a TE insertion at a specific locus designing specific primers located in unique target sequences. These markers are usually dominant and biologically informative because they provide evidence of both complete or incomplete insertion or excision events (Edwards and McCouch, 2007).

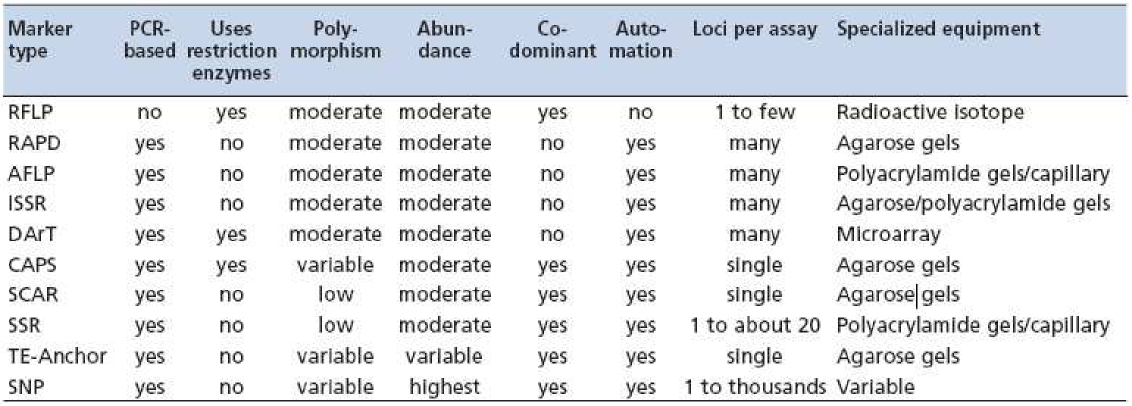
Single nucleotide polymorphism (SNP) is a molecular marker technology that directly assays single base changes in single genes throughout the genome. SNP markers are an abundant source of sequence variants and they provide the highest gene polymorphism information content and the greatest genome marker density. SNPs are often the only option for finding markers very near to or within a gene of interest, and can even be used to detect a known functional nucleotide polymorphism (FNP). SNP investigation needs an initial DNA sequencing in a reference individual followed by some form of re-sequencing in other varieties to find variable base pairs. Many authors (Kwok, 2001; Comai *et al.*, 2004) suggest alternative *in vitro* systems for SNP discovering and they consist in ecotilling with the *CelI* enzyme or by denaturing high pressure liquid chromatography (DHPLC) to measure small conformational differences when PCR-amplified sequences are hybridized to a reference sequence. Moreover, the SNPs can be detected by *in silico* analysis of polymorphisms among EST (Expression

Sequence Tag) clones than can be find in public databases. Once SNPs have been detected the most common assay method is the micro-sequencing in which a primer, having its 3′-end on the nucleotide just before the mutation, is labelled with a fluorescent molecule and the sample is then resolved in a automatic sequencer. Many SNPs can be detected in the same reaction only by changing the primer size linking oligonucleotide tails to the 5′-primer end. In this way the run products will have a different molecular weight and they will be easily detectable in the final chromatogram. This assay method turns to be cheap if the number of sample is quite high. In many other cases, it is worth to verify if the discovered mutation allows the loss or the gain of some restriction enzyme sites. If it happens, it will be feasible to assay the allele variations at one genome position only by a simple PCR and a restriction enzyme reaction followed by agarose electrophoresis of restriction products (CAPS or PCR-RFLP markers).

Another type of SNPs assay is represented by the ARMS-PCR (tetraprimer amplification refractory mutation system-PCR) in which two allele-specific amplicons are generated using two pairs of primers (outer and inner primers), one pair producing an amplicon representing, for example, the A allele and the other the G allele. Allele specificity is conferred by a mismatch between the 3′-terminal base of an inner primer and the template. In order to enhance this specificity, a second deliberate mismatch is usually introduce at position -2 from 3′-terminus of inner primer. After amplification, the PCR products are easily discriminated by gel electrophoresis (Ye *et al.*, 2001).

Overall, it is not a easy task to say whether the best molecular marker system for all situations does exist. Factors influencing the decision may include the goals of the study, availability of organism specific sequences, equipment and technical resources, and biological features of the species. **Table 1.1** summarizes the main features for each type of molecular markers (Powell *et al.*, 1996). The main characteristics of molecular markers are their high reliability and capacity to be easily shared among researchers. In particular for MAS purposes, co-dominant markers are preferred to avoid the need for progeny testing. Dominant markers such as RAPD, I-SSR and AFLP markers are useful for finding markers linked to the desired allele.

Table 1.1: Main features of common molecular marker technologies.



Marker type	PCR-based	Uses restriction enzymes	Poly-morphism	Abun-dance	Co-dominant	Auto-mation	Loci per assay	Specialized equipment
RFLP	no	yes	moderate	moderate	yes	no	1 to few	Radioactive isotope
RAPD	yes	no	moderate	moderate	no	yes	many	Agarose gels
AFLP	yes	no	moderate	moderate	no	yes	many	Polyacrylamide gels/capillary
ISSR	yes	no	moderate	moderate	no	yes	many	Agarose/polyacrylamide gels
DArT	yes	yes	moderate	moderate	no	yes	many	Microarray
CAPS	yes	yes	variable	moderate	yes	yes	single	Agarose gels
SCAR	yes	no	low	moderate	yes	yes	single	Agarose gels
SSR	yes	no	low	moderate	yes	yes	1 to about 20	Polyacrylamide gels/capillary
TE-Anchor	yes	no	variable	variable	yes	yes	single	Agarose gels
SNP	yes	no	variable	highest	yes	yes	1 to thousands	Variable

Once a specific marker is found, it is possible to elute and sequence the corresponding band. This sequence can be used to develop co-dominant markers such as cleaved amplified polymorphic sequence (CAPS) markers (Konieczny and Ausubel, 1993) or to sequence characterized polymorphic region (SCAR) markers (Paran and Michelmore, 1993). These two marker systems can be used to screen easily a large number of individuals and also be employed for molecular marker positioning on a specific genetic linkage map.

AFLP, DArT, ISSR and RAPD markers are usually employed for a first analysis of unknown genomes in which sequence information or plentiful investment are not required. Molecular marker technology that assay the presence or absence of PCR products are often subject to changes in PCR conditions and the quality of sample DNA, and the data from separate experiments may diverge. Moreover, the data can be difficult to score and reproducibility requires a lot of technical skill. Also methods depending on accurate measurement of molecular weight (*e.g.* SSRs) have some limitations: the exact molecular weights assigned to each marker allele may be different in each analysis because of differences in labelling of PCR products. Furthermore, the collected data need to be accurately checked and it can also be difficult to merge separate sets of data. Although divergences in the exact data derived from molecular markers, the results and conclusions should be reliable within independent experiments. SNP markers are preferred for reliability in making inferences across independent datasets. They produce precise datasets that may be simply integrated based on sequences and the low mutation rate of SNPs are particularly valuable for evolutionary inference too (Nielsen, 2000).

Microsatellite and SNP markers are the best choice for genome studies although they still remain the two most expensive technologies. As they are costly, the adoption of SSR markers can be facilitated, by preparing kits of selected microsatellites for certain species to provide a trustworthy set of markers with good amplification, reasonable polymorphism and good genome coverage (Edwards and McCouch, 2007).

Similarly, for SNP markers there is a need to develop useful sets of markers that are widely available and can be mass-produced (at reduced cost) for distribution to the international community. For many secondary species, SNP and SSR markers are not always available or numerous so to overcome this trouble it is possible to transfer molecular markers from closely related species (Gupta *et al.*, 2003; La Rota *et al.*, 2005; Zhang *et al.*, 2005). According to financial resources, species, application and technical skill requirements the choice of the best molecular marker technology may be summarized as follow: RAPD, AFLP and I-SSR systems can give large numbers of markers with a limited basic investment, AFLP, SSR and I-SSR markers can provide high throughput using an automated sequencer, while RAPD and I-SSR markers can be run on agarose gels with minimal investment in equipments. It is therefore comprehensible that the best marker approach, in particular at the beginning of the research, is to try to use more than one method and than try to focus on the approach that provide a good rate in terms of cost and given data (Edwards and McCouch, 2007).

Concluding, the molecular markers may be used for a wide range of different tasks, such as to estimate the genetic diversity and relationships within and between livestock populations and breeds, to investigate biological processes such as mating systems and gene flow levels or to identify specific genotypes (Ruane and Sonnino, 2007).

1.2 Genetic linkage maps and marker assisted selection (MAS)

Phenotypic selection has created a wide diversity of breeds of domestic animals that are adapted to different climatic conditions and purposes. The phenotypic variation that is observed within and among breeds of domestic animals is irresistible compared with that observed in natural populations. Genomics now provides more and more powerful tools for sorting out the molecular basis of phenotypic diversity in domestic animals (Andersson, 2001).

The genome research in farm animals of simple monogenic disease loci is not much important, because animals with inherited disorders as well as their parents tend to be eliminated from breeding. Major part of the interest traits, such as growth, milk production and meat quality, have a multifactorial background and are controlled by an unknown number of quantitative trait loci (QTL), that are regions of the genome containing one or more genes affecting a trait with continuous phenotypic variation. Mutations that modify gene function or gene expression dominate over mutations with pathological consequences because the latter tend to be eliminated by natural selection. Selective breeding has been going on for thousands of years and with increasing intensity during recent centuries. Such selection, over many generations and in large populations, has driven the accumulation of new mutations with favourable phenotypic effects, as well as the development of alleles (reviewed by Andersson *et al.*, 2001).

All strategies for MAS are based on the use of a molecular score, although the composition of this score differs from application to application. The applications of molecular data in genetic programs include their use for genotype building and introgression programs, recurrent selection programs, parentage verification or identification studies, crossbred or hybrid performance tests, and in genetic conservation programs to identify unique genetic resources and quantify genetic diversity (Dekkers and Hospital, 2002).

The main vehicle for genetic improvement in livestock is the recurrent selection program, that has the aim of improving a breed or line as a source of superior germplasm for commercial production through within-breed or within-line selection. This involves recording the phenotypes of numerous individuals and the use of these phenotypes to estimate the breeding value of selection candidates. An example of performance testing is the progeny test, in which the breeding values are estimated on the basis of the phenotype of progeny that have been created through test mating (Dekkers and Hospital, 2002).

Introgression programs are commonly used in plants and mouse and less in livestock. The aim of an introgression program is to introduce a target gene, which can be a single gene, a quantitative trait locus or a transgenic construct, from a low-productivity line or breed, used as donor, into a productive line that lacks that particular gene, used as recipient. Introgression starts by crossing the donor and recipient lines, followed by repeated backcrosses to the recipient line to recover the recipient-line genome. The effectiveness of introgression schemes is limited by the ability to identify backcross or intercross individuals with the target gene and by the ability to identify backcross individuals that have a high proportion of the recipient genome, in particular in regions around the target gene (Tanksley *et al*., 1989; Dekkers and Hospital, 2002).

Another genetic approach is represented by genotype building programs. If many QTL are known, and favourable alleles are present in different lines or breeds, genotype building strategies can be used to design new genotypes that combine favourable alleles at all loci. Selection is then based on the molecular score alone, which is determined by the genotype at those loci possibly estimated through indirect markers, along with information on linkage and linkage phase between those loci, if available. Starting from a cross between two parental lines, the simplest genotype building strategy involves screening a population for individuals that are homozygous at the relevant loci (Van Berloo *et al*., 1998). More than one generation of mating and selection might be needed to produce individuals that are homozygous for a larger number of loci (Charmet *et al*., 1999; Hospital *et al.,* 2000). When more than two parental lines are involved, gene pyramiding can be used to create individuals that are homozygous at all loci. Gene pyramiding involves multiple initial crosses between several parents. Because the above strategies involve several generations of specific matings and the production of numerous offspring, they are more applicable to plants than animals.

Crossbred or hybrid performance is the other main genetic program by which the genetic improvement can be developed. In theory, crosses between lines that are genetically distant are expected to show greater hybrid vigour or heterotic effects than those between more closely related lines, because differences in allele frequencies between genetically distant lines are expected to be greater. Genetic distance can be measured from differences in allele frequencies at anonymous markers spread throughout the genome. Evaluation of this concept for many species (Melchinger *et al.*, 1999) shows that marker-based prediction of hybrid performance can be efficient if hybrids include crosses between parental lines that are related by pedigree or which trace back to common ancestral populations. By contrast, prediction is not efficient for crosses between parental lines that are unrelated or that originated from different populations, because the associations through LD between marker loci and QTL that are involved in heterosis are not the same in the different populations (Charcosset *et al.*, 1994).

The basic prerogative for MAS development is the definition of genetic linkage maps that consists in the relative locations of specific DNA markers along the chromosomes. Precisely, a linkage genetic map is a graphic representation of marker linkage groups that shows the order and relative position of the genes depending on their recombination frequencies. Once the linkage map of the interest species is available it will be possible of identifying the chromosome traits carrying QTL or Mendelian genes responsible for the variation or expression of a given trait. The genome of the most important animal farm, such as cow, chicken, pig, and sheep, has been mapped by different types of molecular markers (mainly SSRs and SNPs) and they are freely available at http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome web site. High-density genetic maps need to be developed in order to restrict the area in which molecular markers can be found associated to the polygene or gene controlling the trait.

A high-density map is represented by several markers (dominant or co-dominant) of known location, interspersed at relatively short intervals throughout the genome. Nowadays, such kind of linkage maps have been constructed for a range of economically important agricultural and animal species. High-density map is the first step for the eventual application of MAS and it can be used to test the association between marker variants and any trait of interest. These traits might be genetically simple because controlled by one or a few genes. However, most of the economically important traits, such as milk yield and meat quality, are affected by many genes and environmental effects. Once markers, physically located beside or even within genes of interest have been detected, in the next step it will possible to select this identifiable marker variants (*i.e.*, alleles) in order to select for non-identifiable favourable variants of the genes of interest (Ruane and Sonnino, 2007). This is the basic principle of MAS in which the selection of individuals is depending on the presence or absence of a combination of markers tightly linked with the trait. An example could be the following: a molecular marker M, with two variants M1 and M2, discovered by DNA assay is proved to be linked to a gene controlling the expression of an economically important trait. Let is say that the two variants of gene are G1, which determines an increasing of milk yield and G2, which decreases the milk yield. If the marker M1 and the unknown gene G1 are quite close to each other, they will be inherited together. This implies that the selection for the marker M1 allows the selection of those animals carrying the favourable gene G1 for the milk yield.

The mapping and identification of trait loci, in particular QTL, require powerful genome resources. After human genome many other animal farms genome projects started. As reported at the US Livestock Genome Mapping Projects and Roslin Institute web sites, dense microsatellite maps comprising greater than 1,000 markers have been developed for all the main farm animals (Andersson 2001).

Libraries made up of large-insert yeast artificial chromosome (YAC) and bacterial artificial chromosome (BAC) are also extensively used for cloning and characterizing trait loci for all the main farm animals. High-resolution maps and comparative maps (Band *et al.*, 2000) have been developed by employment of radiation hybrid (RH) panels that are available for pig (Hawken *et al.*, 1999), cattle (Yang and Womack, 1998) and chicken (Kwok *et al.*, 1998). Following the example of human, preliminary transcript maps using ESTs have been reported in farm animal species. Even though they are still small compared with that in humans the size of these maps is constantly increasing mainly due to the large scale initiative at the US Meat Animal Research Center (MARC) (Andersson, 2001).

In order to increase the knowledge about the gene functions, the researchers exploited genetic maps of other known species as human and mouse to carry out a comparative mapping approaches. Chromosome painting and linkage mapping were used for comparative mapping of human/cattle genomes (Chowdhary *et al.*, 1998) and human/chicken genomes (Burt *et al.*, 1999), respectively. The analysis of comparative mapping data shows that the organization of the human genome is closer to that of the chicken than to the mouse. This shows that the rate of chromosomal evolution varies considerably between species (Andersson, 2001).

In conclusion, the main goal of genome research in farm animals is to map and characterize trait loci that control various phenotypic characters: it is obviously comprehensible that the MAS success is dependent on the relationships among the markers and the gene of interest. The application of molecular genetics in the identification of loci and chromosomal regions affecting traits of importance in livestock production might enhance and help the genetic improvement programs by direct selection of genes or genomic regions through marker-assisted selection (Andersson, 2001; Dekkers and Hospital, 2002).

1.3 Genome scan and Linkage disequilibrium approaches for MAS

For MAS purposes, it is crucial to understand the relationships between the traits and the discovered markers. Exist two main types of relationships among molecular markers and genes or chromosome regions of interest that easily contribute to the success of MAS, as reported by Ruane and Sonnino (2007): i) the molecular marker is located within the gene of interest (*i.e.*, within the gene G, using the example above). In this situation, one can refer to gene-assisted selection (GAS). This is the most favourable situation for MAS since, this is a direct marker that codes for a functional mutation easily detectable by following inheritance of the M alleles. In contrast, these kinds of markers are the most uncommon and are thus the most difficult to find. When the selection uses these markers it is normally called GAS (Gene Associated Selection); ii) the marker is in linkage disequilibrium (LD) with the functional mutation (coded by G) throughout the whole population. LD is the predisposition of certain alleles combinations (*e.g.*, M1 and G1) to be inherited together. Population wide LD can be found when markers and genes of interest are physically very close to each other and/or when lines or breeds have been crossed in recent generations. Selection using these markers are usually called LD-MAS.

An easy example will help to understand the LD bases: consider a marker locus with alleles *M* and *m* and a QTL with alleles *Q* and *q* that is on the same chromosome as the marker, *i.e*. the marker and the QTL are linked. An individual that is heterozygous for both loci would have genotype *MmQq*. Alleles at the two loci are arranged in haplotypes on the two chromosomes of a homologous pair that each individual carries. An individual with genotype *MmQq* could have the following two haplotypes *MQ*/*mq* or the second two haplotypes *Mq*/*mQ*. This alternative arrangement of linked alleles on homologous chromosomes is referred to as the marker-QTL linkage phase. The arrangement of alleles in haplotypes is important because progeny inherit one of the two haplotypes that a parent carries, barring recombination (Dekkers and van der Werf, 2007). The presence of linkage equilibrium or disequilibrium refers to the relative frequencies of alternative haplotypes in the population. When a population is in a phase of linkage equilibrium, alleles at two loci are randomly assorted into haplotypes. This means that the frequency of the *MQ* haplotypes is equal to the product of the population allele frequency of *M* and the frequency of *Q*. Thus, if a marker and QTL are in linkage equilibrium, there is no value in knowing an individual marker genotype because it provides no information on QTL genotype. If the marker and QTL are in linkage disequilibrium, however, there will be a difference in the probability of carrying *Q* between chromosomes that carry *M* and *m* marker alleles and, therefore, a difference in mean phenotype between marker genotypes would also be expected (Dekkers and van der Werf, 2007).

Figure 1.3 reports a schematic representation of selection base on linkage disequilibrium. As it is shown, markers that are tightly linked to a quantitative trait locus (QTL) can be in complete or partial population-wide LD with the QTL, such that some marker-QTL haplotypes are more frequent than expected by chance (for example, *MQ* and *mq* versus *Mq* and *mQ*). In this case, selection can be directly on marker genotype. The probability of population-wide LD is higher for closely linked markers and in selected populations of small effective size. Population-wide LD can also be created by crossing lines or breeds and will then exist between loosely linked markers for several generations. When a marker and a QTL are in linkage equilibrium, all marker-QTL haplotypes are present and at random-mating frequencies, and marker genotype gives no information about QTL genotype. This will be the case for most linked markers in an outbreeding population. However, the marker and QTL will be in partial disequilibrium within a family. The extent of within-family disequilibrium depends on the recombination rate (*r*), but will occur even with loose linkage (for example, $r = 0.2$). This disequilibrium can be used to detect QTL and for selection on a within-family basis (Dekkers and Hospital, 2002).

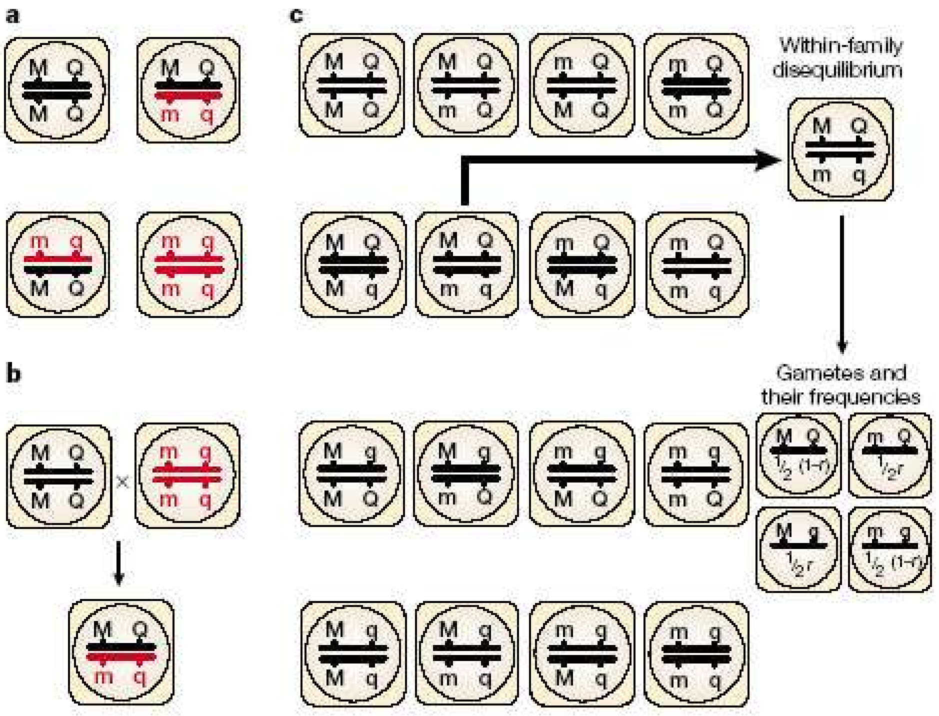


Figure 1.3. Selection programs based on linkage disequilibrium estimated in natural and experimental populations.

The LD in a population is mainly due to mutation, selection, inbreeding, and migration or crossing. Recombination is the main factor that breaks down LD because it rearranges haplotypes that exist within a parent in every generation. The **Figure 1.4** shows the rate of the LD erosion that depends on the rate of recombination between the loci. For tightly linked loci, any LD that has been created will persist over many generations but, for loosely linked loci ($r>0.1$), LD will decline rapidly over generations (Dekkers and van der Werf, 2007).

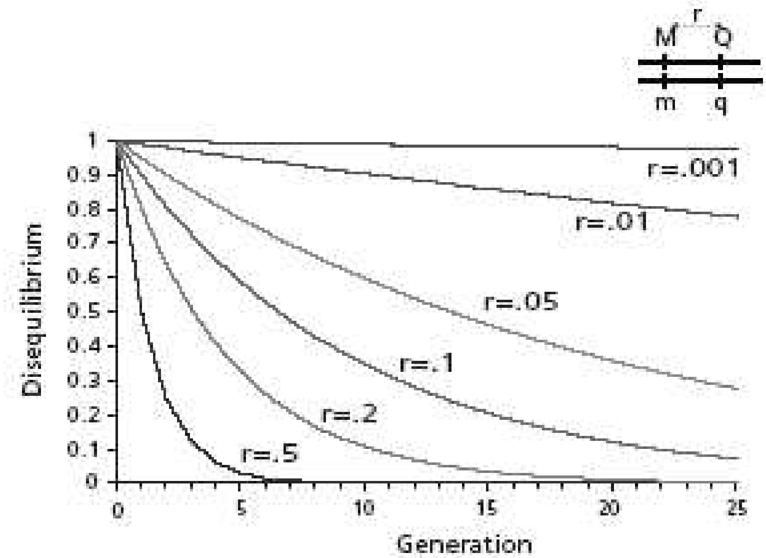


Figure 1.4: Break-up of LD over generations. LD is continuously eroded by recombination. Tighter are two loci and less is the probability to gain recombination and lose disequilibrium phase (Dekkers and van der Werf, 2007).

Since the extent of LD is due to the number of genomic recombination events, Rafalsky and Morgante (2004) discussed that gene-dense regions recombination will be high (also known as "*hot spots*") and LD will be low (**Figure 1.5**). Such gene-dense regions occur frequently close to the telomeres. By contrast, in gene-poor segments of the genome (around centromeres) recombination will be low, resulting in high LD. Genome segments devoid of genes will show high LD, and will constitute linkage LD blocks. LD could therefore decay rapidly within genes but be maintained from one end of a gene to the beginning of the next one. The identification of genes responsible for a trait of interest is possible by screening a limited number of candidate genes. Individual SNPs or SNP haplotypes within a candidate gene are systematically tested for association with the phenotype of interest. The interest in the study of LD has increased dramatically in recent years because of genomic technology enables rapid identification of haplotypes at many genetic loci, either by DNA sequencing or by high-throughput SNP analysis. Moreover, in the presence of significant LD, of the order of tens of kilobases or more, it can be possible to identify genetic regions that are associated with a particular trait of interest (e.g. disease susceptibility) by a systematic, high-density genome scan of individuals from an existing population (Weiss *et al*., 2002; Rafalsky and Morgante 2004)

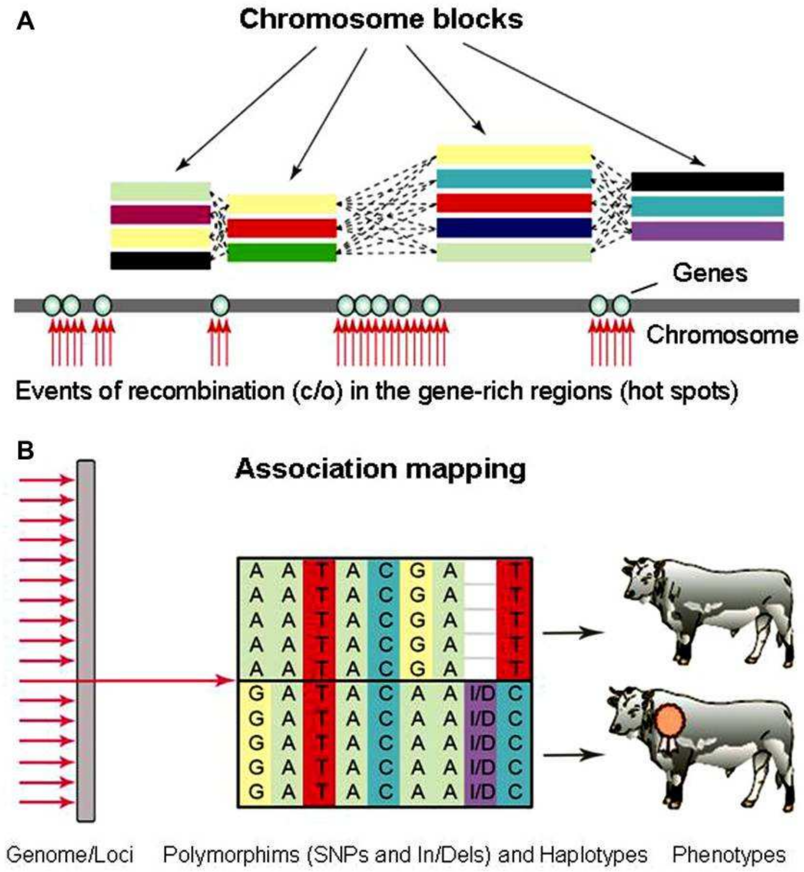


Figure 1.5. Haplotype block arrangement resulting from interspersed hot spots, indicated by red arrows, and cold spots for recombination (A); Schematic representation of genetic association mapping (B).

Thus, the crucial issue is how closely a QTL must be mapped with molecular markers to be such information useful for MAS. Several simulation studies have shown that for MAS based on within-family LD, markers that flank a QTL within a 5 cM distance seem adequate (Moreau *et al*., 1998). Given that markers are not fully informative in practice, this can be achieved by using haplotypes of several markers within a 10 cM region surrounding the QTL. For example, Spelman and Bovenhuis (1998) found that a flanking marker interval of 5 cM around the QTL achieved 85–90%

of the extra response over selection without markers, relative to a flanking marker interval of 2 cM. Although further fine mapping of QTL might provide limited benefits for MAS based on within-family LD, the occurrence of population-wide LD will increase substantially if the markers are more tightly linked to the QTL. Selection on markers that are in population-wide LD with QTL is much preferred because QTL effects and linkage phase can be estimated from population-wide data instead of the limited data that would be available using a within-family LD approach (Smith and Smith, 1993). For individual QTL, individual markers or marker haplotypes within 1 or 2 cM of the causative locus might be required for substantial population-wide LD to be present, depending on population size and selection history.

LD can be exploited at a genome-wide level when molecular marker data are available from a high-density genetic map (*e.g.*, with one marker per cM). The potential of using such data was illustrated by Meuwissen *et al.* (2001), who simulated genome-wide data for a breeding population based on the historical accumulation of mutations which gives rise to QTL at locations throughout the genome in the context of a high-density marker map. They then computed molecular scores based on statistical associations of phenotype with marker haplotypes to capture population wide LD. For populations that are representative of livestock with an effective population size of 100, they showed that sufficient LD was available and that the molecular score had an accuracy of 85% as a predictor of the total genetic value of an individual, when marker spacing was 1 cM. Accuracy dropped to 81 and 74% for marker spacings of 2 and 4 cM, respectively.

1.4 Strategies for mapping trait loci

The last advances in molecular genetics allowed the identification of molecular markers associated with genes that affect traits of interest in livestock, including single-gene traits and genomic regions that affect quantitative traits. These findings provided opportunities to enhance response to selection, in particular for traits that are difficult to improve by conventional selection such as low heritability or traits for which measurement of phenotype is difficult, expensive, only possible late in life, or not possible on selection candidates (Dekkers, 2004).

Andersson (2001) and Dekkers (2004) have given a detailed classification and explanation of the strategies employed for finding trait loci in the main animal species. Application of molecular genetics for genetic improvement relies on the ability to genotype individuals for specific genetic loci.

The most common strategy for finding trait loci is the use of existing pedigrees deriving from many years of collecting and analysing data on phenotypic traits for breeding purposes in farm animals. This approach exploits large family sizes mainly deriving from artificial insemination in which half-sib families comprise usually more than 1,000 progeny from a single male. A useful strategy is to increase the statistical power in QTL mapping by using breeding values based on phenotypic data from progeny. The pedigree shown in **Figure 1.6** illustrates the grand-daughter design for mapping QTL using half-sib families proposed by Weller *et al*. (1990).

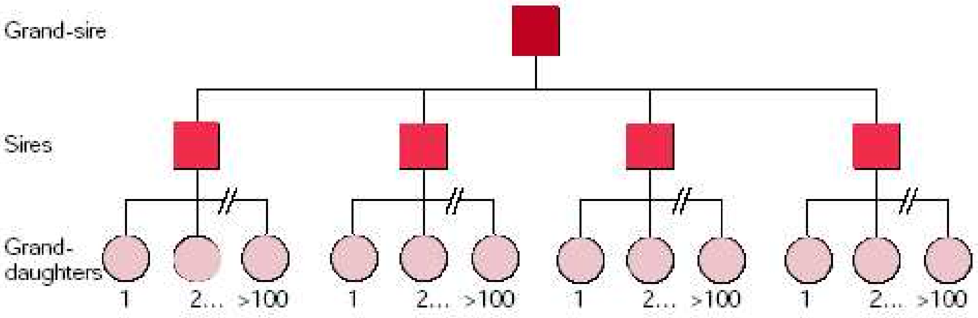


Figure 1.6 Half-sib families and breeding value (adapted from Andersson, 2001).

Each family consists of one grand-sire whose sons, 20 or more, have been selected as sires. Each son has 100 or more daughters with phenotypic data, and they are employed to estimate accurate breeding values for the sires, even for traits with a considerable environmental influence. Marker genotypes are only collected for the grand-sire and his sons, and QTL mapping is carried out by analysing the segregation of markers from the grand-sire to the sires in relation to differences in breeding values. If the grand-sire is heterozygous for a QTL with a major effect, the sons that have received the favourable allele will present a higher breeding values than those that received the unfavourable allele. This strategy is used extensively to identify QTL for milk production traits, as well as for other traits of interest in dairy cattle and sheep (Georges *et al*., 1995; Zhang *et al*., 1998).

The three main strategies for trait loci identification are based on association tests using candidate genes, genome scans based on linkage mapping and intercrosses between divergent populations with anonymous DNA markers.

The candidate gene approach can be very powerful and can detect loci even with small effects, provided that the candidate gene represents a true causative gene. Applying this approach it is necessary to take into account that often many candidate genes interact to determine the trait of interest. The evaluation of all potential candidate genes that affect the trait is normally more time-consuming if compared to a genome scan approach. In addition, the candidate gene approach might be unsuccessful for the detection of a major trait locus simply because there is no enough knowledge about gene function. Candidate gene tests must also be interpreted with prudence because false results can occur because of linkage disequilibrium to linked or non-linked causative genes. (Andersson, 2001). Instead of scanning a whole genome, one or more candidate genes might be identified on the basis of biochemical knowledge, gene expression analysis, genetic mapping or other clues. SNPs or haplotypes are catalogued in a diverse population at the candidate gene locus or loci and the phenotypic values of individuals carrying each haplotype are compared, to identify statistically significant differences. Haplotype sharing refers to association analysis using haplotypes, instead of individual polymorphisms (Rafalsky and Morgante, 2004).

The candidate gene-QTL association approaches rely on combining multiple lines of evidence to restrict the numbers of genes that are evaluated. Comparative genomics, genome sequencing, transcriptome profiling, low-resolution QTL analysis and large-scale knockouts all provide opportunities to develop and refine candidate gene lists. These approaches are powerful at identifying candidate genes, but not at evaluating allelic affects on the quantitative traits. The candidate gene approach can substantially reduce the amount of genotyping required, but most importantly, it can reduce the multiple issues created by testing thousands of sites across the genome. The statistical issues in combining these disparate types of evidence have not been resolved.

Despite of candidate gene approach, the genome scan will always find the map location of a trait locus with a major effect, provided that an accurate genetic model has been postulated, a reasonable sample size has been used and that the marker set provides full genome coverage. Conversely, all these advantages are overcome by the inability of genome scan to detect trait loci with smaller effects. It occurs when they do not reach the stringent significance thresholds that must be applied when doing a large number of tests in a full genome scan (Andersson, 2001).

The other powerful approach for mapping trait loci is represented by the use of intercrosses between divergent populations. The F_1 animals show a high heterozygosity at marker loci and, in particular, at those loci that account for phenotypic differences between the two populations. In literature many example of intercrosses have been reported for animal farm such as between the European Wild Boar and Large White domestic pigs (Andersson *et al*., 1994), Asian and European breeds of pig (Rohrer and Keele 1998; de Koning *et al*., 1999) *Bos taurus* and *Bos indicus* cattle (Brenneman *et al*., 1996) and Awassi and Merino sheep breeds (Raadsma *et al*., 1999). All these populations allows to researchers to find out several QTLs controlling body composition, growth and fatness in pig (Walling *et al*., 2000).

Animals deriving from backcross or advanced intercross lines (AIL) can be used for the obtaining of high-resolution mapping of QTL (Darvasi and Soller, 1995; Darvasi, 1998). The employment of these two animal types is costly for the larger farm animals and it is usually used for particularly important resource populations (Andersson, 2001). For example, QTL affecting protein yield and lactation persistency in dairy sheep have been detected in extreme breed back-cross and inter-cross between Awassi fat-tail sheep and Merino superfine and medium wool sheep (Singh *et al*., 2007; Raadsma *et al*., 1999). Once the trait locus has been localized, the next step involves the identification of the genes that cause the phenotype variations (*i.e*., causative genes) through the positional candidate cloning strategy that keeps being the main approach for this purpose. High-resolution mapping is a first step towards restricting the region of interest and thereby the number of potential candidate genes (Singh *et al*., 2007).

Information on map location and gene function is then combined to identify positional candidate genes, which are subsequently evaluated by mutation screening and functional analysis. The increasing knowledge about gene function, expression patterns and the completion of different species map will help the power of positional candidate cloning (Andersson, 2001).

The whole collected molecular information can finally be used for powerful marker assisted selection purposes even though it has to be seen as a support tool of the conventional breeding programs.

1.5 Statistical issues and economic aspects of MAS in livestock

Genetic information acquired by using molecular markers can be used to enhance several breeding strategies in livestock species through what is broadly referred to as marker-assisted selection (MAS). It is a matter of fact that individual-genome fingerprinting and wide-genome scanning based on mapped molecular markers allows genetic characterization and dissection of phenotypic variability in farm animals. Unless genetic markers capture most of the genetic variation for a given quantitative trait, selection must be based on a combination of molecular marker and conventional phenotypic data. Although several useful genes, primarily molecular markers linked to them, have been discovered as well as candidate genes identified in livestock species, their application and success have been limited because the genes were not identified in breeding populations, or because they interact with other genes or environmental factors (Andersson, 2001).

A key role in the association of molecular markers to quantitative traits is played by statistical procedures. The vast majority of applications of molecular markers in genetic selection programs are preceded by an analysis aimed at identifying and mapping chromosome regions carrying either major or minor QTL. Only QTL that are shown to have a significant effect on the phenotype under study will be subsequently exploited for selection assisted by molecular markers. This application raises two important statistical issues: i) the setting of statistical thresholds for deciding which QTL to use; and ii) dealing with the inherent overestimation of QTL effects. For QTL detection, very stringent methods are used to control the false-positive error rate, as suggested by Lander and Kruglyak (1995). Several studies have, however, shown that greater gains from MAS can be obtained by allowing a higher rate of false-positives, to increase the power of detecting QTL effects and reduce the number of false-negative results (Moreau *et al*., 1998; Dekkers and Hospital, 2002). For instance, Spelman and Garrick (1998) quantified the genetic and economic potentials for pre-selection of bulls for progeny testing in dairy cattle. Alternative strategies needed to more adequately balance the cost of false-positive against false-negative results for MAS were identified by Weller *et al*. (1998). This balance might differ depending on the particular applications and animal species. Thresholds could be lowered even further if proper statistical methods would be used to account for the degree of uncertainty about estimates of QTL effects. For example, Meuwissen *et al.* (2001) obtained a molecular score with high predictive ability on the basis of high-density marker genotyping data by using all estimated marker effects, regardless of their statistical significance. In particular, they demonstrated the potentials to capitalize on population-wide linkage disequilibrium (LD) between markers and QTL in breeding populations. Overestimation of QTL effects has been shown to occur both by theoretical studies (Beavis, 1994; Bost *et al*., 2001) and by experimental programs (Melchinger *et al*., 1998; Barton and Keightley, 2002). Overestimation of QTL effects leads to too much emphasis on molecular scores in selection relative to phenotypic data, and results in a less than optimal response to selection. In part, biases can be caused by the use of only significant QTL effects, and they can be reduced, although not entirely removed, by re-estimation of significant QTL effects in an independent sample (Bost *et al*., 2001). Eshed and Zamir (1995) described for the first time a powerful, though cumbersome, method to accurately estimate QTL effects and QTL interactions, whereas Fernando and Grossman (1989) described in their keystone paper the application of mixed statistical methods based on best linear unbiased prediction to MAS. However, alternative statistical methods for the analysis of QTL data that avoid overestimation or reduce their impact on selection response are needed.

A more general point about the statistical aspects of MAS is that the existing models and theory do not adequately accommodate the more complex genetics that underlies quantitative traits. Furthermore, although existing quantitative genetic theory provides a satisfactory basis to derive selection strategies that maximize response to selection in the short term (one or two generations), the theory has been much less developed for selection over several generations (Dekkers and Hospital, 2002). This was most clearly seen in several simulation studies that showed that combined selection on an index of molecular score and phenotype results in greater genetic gain in the short term. However, in the long term, selection on phenotype alone resulted in a greater response to selection (Gibson, 1994; Larzul *et al*. 1997), because selection is better distributed over all loci (Hospital *et al*., 1997). A theory to optimize selection on molecular score, in combination with phenotype, has been developed (Dekkers *et al*., 1998; Manfredi *et al*., 1998), but for genetic models and selection strategies of limited complexity (Dekkers and Hospital, 2002). Further theoretical work is needed to accommodate multilocus Mendelian inheritance and phenomena such as epistasis, genetic background effects and interactions between genetic and environmental factors.

Most applications of molecular genetics to breeding programs have attempted to incorporate molecular data into the existing programs (Dekkers and Hospital, 2002). The effective use of molecular data might, however, require a complete redesign of breeding programs. In animals, strategies are required that integrate the collection and evaluation of phenotypic data for QTL detection with the use of this information for MAS (Spelman and Garrick, 1998). Furthermore, breeding strategies must be developed that take better advantage of the unique features of molecular marker data.

Economics is the key determinant for the application of molecular markers in genetic improvement programs. The use of markers in the selection of animals incurs the costs that are inherent to molecular techniques. Apart from the initial costs related to QTL detection and mapping, costs for MAS include the costs of DNA isolation, genotyping, and data analysis. The economic assessment of MAS is straightforward in some cases, but complex in others, and has been addressed in few studies (Brascamp *et al*., 1993; Davis *et al*., 1998; reviewed by Dekkers and Hospital, 2002). Several studies have focused primarily on genetic and economic modelling because the results are extremely difficult to verify using replicated experiments. Cases in which the economic merit of MAS is clear include situations in which molecular costs are more than offset by the savings in phenotypic evaluation. In other cases, the ability for early selection offsets the extra costs that are associated with the use of DNA markers for genetically characterizing animals. The benefits of being able to release new genetic material more quickly can be substantial, particularly in competitive markets.

The economic merit of MAS becomes questionable and more difficult to evaluate in cases in which MAS is expected to provide greater genetic gain at increased costs. This is particularly the case for selection schemes that rely on a combination of phenotype traits and molecular markers, because animal genotyping costs are in addition to, not in place of, animal phenotyping costs. In such cases, MAS might not be economically more advantageous than quantitative genetic selection, although the economic merit of MAS could be restored by reducing the frequency of re-evaluation of marker effects, as proposed by Hospital *et al*. (1997). Another consideration is that the resources allocated to MAS could also be allocated to enhance phenotypic selection programs. For example, improvement by conventional selection could also be enhanced by increasing the number of individuals that are tested for phenotypic evaluation (Moreau *et al*., 1998). Further work on the economic evaluation and optimization of strategies for the use of molecular genetics in breeding programs is required.

Even though the good advance in molecular makers, today the MAS is still seen as a future perspective. Overall, there are still few reports of successful MAS experiments or applications. Most refer to the use of molecular markers in genotype building programs, at various levels of complexity. Successful reports include marker-assisted background selection with introgression of genes for which the functional variant is known or which have clearly identifiable phenotypic effects. Marker-assisted introgression of such known genes is now widely used in plants, in particular by private plant-breeding companies, where in animals its use is still scanty. However, even in this case, more work is needed to optimize the information provided by markers and reduce costs (Visscher *et al.*, 1999; Servin and Hospital 2002).

For instance, traits that are controlled by several QTL of moderate or low effect, or that are subject to high environmental variation, genotype-environment interactions, epistasis between QTL or epistasis between QTL and the genetic background, it is risky to carry out selection solely on the basis of marker effects, without confirming the estimated effects by phenotypic evaluation. This is true in particular if QTL were initially detected in a different population or genetic background. Although no documented reports are available, industrial applications of molecular data in livestock are limited and have mainly been in the context of recurrent selection programmes, which are the principal vehicles for genetic improvement in animals. A mixture of causal and indirect markers is used. In swine, the indirect markers used were primarily identified by using candidate-gene approaches or positional cloning, whereas in dairy cattle, indirect markers identified using genome scans are also used (reviewed by Dekkers and Hospital, 2002; see also references listed in Dekkers and Hospital, 2002) This species difference is partially explained by the different strategies that are used for QTL detection. In swine, genome scans are primarily based on crosses between divergent lines. These identify QTL that differ between breeds but have limited direct application for within-breed selection. Direct access to closed breeding populations has, however, made candidate-gene approaches relatively successful. In dairy cattle, QTL detection capitalizes on the large half-sib family sizes that result from extensive use of artificial insemination (Andersson 2001). This allows genome scans to detect QTL that segregate within rather than between breeds.

Recent advances in molecular marker technology for genome and gene analysis will soon create a wealth of information that can be exploited for the genetic improvement of animals. High-throughput genotyping, for example, will allow direct selection on marker information based on population-wide LD. Methods to effectively analyse and use this information in selection are still to be developed. The eventual application of these technologies in practical breeding programs will be on the basis of economic grounds and will require further evidence of predictable and sustainable genetic advances using MAS. Until complex traits can be fully dissected, the application of MAS will be limited to Mendelian genes and to QTL-related genes of moderate-to large effect as well as to applications that do not endanger the response to conventional selection. Until then, observable phenotype will remain an important component of genetic improvement programs, because it takes account of the cumulative effect of all genes.

It is however important to note that the identification of the causative gene for a trait locus is not a prerequisite for practical applications. As a matter of fact, several cattle and pig breeding companies are now using marker-assisted selection with markers flanking QTL as a complement to phenotypic selection of breeding animals. It is likely that large-scale marker analysis will be used routinely, as soon as the cost for genotyping has been consistently reduced.

An increasing number of trait loci in farm animals have been characterized at the molecular level in recent years. The future prospects for cloning trait loci are bright, even for major QTL, provided that the QTL is due to one or more mutations in a single gene and not a haplotype effect. The reason for this optimism is the continuous development of better tools and new resources for genomics. Current initiatives to develop physical maps of farm animal genomes will provide researchers with a large-insert contig covering the region of interest as soon as a trait locus has been mapped. Such large-insert contigs can then be used to build a preliminary transcript map of the region by high-resolution comparisons with the corresponding region in humans or mice. It is also only a matter of time before initiatives will be taken to sequence the genomes of farm animals. Recently, several animal geneticists proposed that genome sequencing initiatives should be carried out using DNA samples from two or more divergent populations, such as a beef and a dairy cattle breed, or from an improved breed and the wild ancestor, when possible. This approach will not reduce the efficiency in determining the genome sequence much, but will detect a good proportion of fixed genetic differences between divergent populations. Bioinformatics should provide scores for the likelihood that an observed substitution is functionally important, on the basis of the degree of phylogenetic conservation of the position and the possible functional consequences of the substitution (Andersson, 2001). The functional consequences will then be validated by experimentation in order to unravel the molecular basis for a variety of phenotypic traits of agricultural, biological and medical significance.

1.6 References

Andersson L. (2001). Genetic dissection of phenotypic diversity in farm animals. Nature Revs. Genet 2:130-138.

Andersson, L, C. S. Haley, H. Ellegren, S. A. Knott, M. Johansson, K. Andersson, L. Andersson-Eklund, I. Edfors-Lilja, M. Fredholm, I. Hansson, et al, (1994). Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science 263: 1771–1774.

Band M. R. et al. (2000). An ordered comparative map of the cattle and human genomes. Genome Res. 10, 1359–1368

Barcaccia G. and M. Falcinelli (2007). Analisi genomica Vol. III. Genetica e genomica- Ed. Liguori, Napoli (Italy).

Barton N. H. and P. D. Keightley (2002). Understanding quantitative genetic variation. Nature 2:11-21.

Beavis W. D. (1994). Corn and Sorghum Industry Res. Proc. of the 49th Ann. Conf. ASTA, Washington, DC (USA). pp. 250–266.

Bost B., D. De Vienne, F. Hospital, L. Moreau and Dillmann C. (2001). Genetic and non-genetic bases for the L-shaped distribution of quantitative trait loci effects. Genetics, 157: 1773–1787.

Brascamp E.W., J. A. M. Van Arendonk and A.F. Groen (1993). Economic appraisal of the utilization of genetic markers in dairy cattle breeding. J. Dairy Science, 76: 1204–1214.

Brenneman R. A. et al. (1996). The polled locus maps to BTA1 in a Bos indicus x Bos taurus cross. J. Hered. 87, 156–161.

Burt D. W. et al. (1999). The dynamics of chromosome evolution in birds and mammals. Nature 402, 411–413.

Charcosset A. and Essioux L. (1994). The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. Theor. Appl. Genet. 89, 3336–3343.

Charmet G., N.Robert, M. R. Perretant, G. Gay, P. Sourdille, C. Groos, S. Bernard and M. Bernard (1999). Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. Theor. Appl. Genet. 99, 1143–1148.

Chowdhary B. P., T. Raudsepp, L. Fronicke and H. Scherthan (1998). Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. Genome Res. 8: 577–589.

Comai L., K. Young, B. J. Till, S. H. Reynolds, E. A. Greene, C. A. Codomo, L. C. Enns, J. E. Johnson, C. Burtner, A. R. Odden, S. Henikoff (2004). Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant J. 37: 778–786.

Darvasi A. (1998). Experimental strategies for the genetic dissection of complex traits in animal models. Nature Genet. 18: 19–24.

Darvasi A. and M. Soller (1995). Advanced intercross lines, an experimental population for fine genetic-mapping. Genetics 141: 1199–1207.

Davis G. P. and S. K. DeNise (1998). The impact of genetic markers on selection. J. Animal Science 76:2331–2339.

de Koning D. J., L. L. G. Janss, A. P. Rattink, P. A. M. van Oers, B. J. de Vries, M. A. M. Groenen, J. J. van der Poel, P. N. de Groot, E. W. Brascamp and J. A. M. van Arendonk (1999). Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (Sus scrofa). Genetics 152, 1679–1690

Dekkers J.C.M and J.H.J van der Werf (2007). Strategies, limitations and opportunities for marker-assisted selection in livestock. In: Marker-Assisted Selection. Food and Agriculture organization of the United Nations (FAO), Rome

Dekkers J.C.M. (2004). Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. J. Anim. Sci. 82(E. Suppl.):E313–E328.

Dekkers J.C.M. and F. Hospital (2002). The use of molecular genetics in the improvement of agricultural populations. Nature Revs. Genet. 3: 22–32.

Dekkers J.C.M. and J.A.M. Van Arendonk (1998). Optimum selection for quantitative traits with information on an identified locus in outbred populations. Genet. Res., 71: 257–275.

Edwards J.D. and S.R. McCouch (2007). Molecular markers for use in plant molecular breeding and germplasm evaluation. In: Marker-Assisted Selection. Food and Agriculture organization of the United Nations (FAO), Rome.

Edwards J.D., V.M. Lee, and McCouch, S.R. (2004). Sources and predictors of resolvable indel polymorphism assessed using rice as a model. Mol. Genet. Genomics 271: 298–307.

Eshed Y. and D. Zamir (1995). An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics, 141: 1147–1162.

Fernando R.L. and M. Grossman (1989). Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21: 467–477.

Flint-Garcia, S. A; J. M Thornsberry, ESS.A. Buckler (2003). Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54, 357–374

Georges M. et al. (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139, 907–920.

Gibson J.P. (1994). Short-term gain at the expense of long-term response with selection of identified loci. Proc. 5th World Congr. Genet. Appl. Livestock Prod. 21: 201–204.

Gupta P. K., R. K. Varshney, P.C. Sharma, B. Ramesh (1999). Molecular markers and their applications in wheat breeding. Plant breeding 118:369-390.

Gupta P.K., S. Rustgi, S. Sharma, R. Singh, N. Kumar and H.S. Balyan (2003). Transferable ESTSSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol. Genet.Genomics 270: 315–323.

Hawken R. J. et al. (1999). A first-generation porcine whole genome radiation hybrid map. Mamm. Genome 10:824–830.

Hospital F., I. Goldringer and S. Openshaw (2000). Efficient markerbased recurrent selection for multiple quantitative trait loci. Genet. Res. 75, 357–368.

Hospital F., L. Moreau, F. Lacoudre, A. Charcosset and A. Gallais (1997). More on the efficiency of marker assisted selection. Theor. Appl. Genet. 95: 1181–1189.

Jaccoud D., K. Peng, D. Feinstein, A. Kilian (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. Nucl. Acids Res. 29: E25.

Konieczny A. and F.M. Ausubel (1993). A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 4: 403–410.

Kwok P.Y. (2001). Methods for genotyping single nucleotide polymorphisms. Ann. Rev. Genomics Hum. Genet. 2: 235–258.

Kwok, C. et al. (1998). Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates. Nucleic Acids Res. 26, 3562–3566.

La Rota M., R. V. Kantety, J. K. Yu and M. E. Sorrells, (2005). Non-random distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6: 23.

Lander E. and L. Kruglyak (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics, 11: 241–247.

Larzul C., E. Manfredi and J.M. Elsen (1997). Potential gain from including major gene information in breeding value estimation. Genet. Sel. Evol., 29: 161–184.

Manfredi E., M. Barbieri, F. Fournet and J.M. Elsen (1998). A dynamic deterministic model to evaluate breeding strategies under mixed inheritance. Genet. Sel. Evol., 30: 127–148.

McCouch S.R., G Kochert, Z.H.Yu, Z.Y.Wang, G.S. Khush, W.R. Coffman and S.D. Tanksley. (1988). Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76: 815–829.

McCouch S.R., X. Chen, O. Panaud, S. Temnykh, Y. Xu, Y.G. Cho, N. Huang, T. Ishii, and M. Blair (1997). Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol. Biol. 35: 89–99.

Melchinger A. E.(1999) in The Genetics and Exploitation of Heterosis in Crops (eds Coors, J. G. & Pandey, S.) 99–118 (Crop Science Society of America, Madison, Wisconsin).

Melchinger A.E., H. F. Utz and C.C. Schon (1998). Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics, 149: 383–403.

Meuwissen T.H., B.J. Hayes and M.E. Goddard (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157: 1819–1829.

Moreau L., A. Charcosset, F. Hospital and A. Gallais (1998). Marker-assisted selection efficiency in populations of finite size. Genetics, 148: 1353–1365.

Nielsen R. (2000). Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154: 931–942.

Paran I. and R.W. Michelmore (1993). Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Genetics 85: 985–993.

Powell W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey and A. Rafalsky (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. 2: 225–238.

Raadsma H. W., K. R. Zenger, C. Cavanagh, M. K. Lam, P. C Thomson, M. Jones, G. Attard, D. Palmer and F. W. Nicholas (1999). The mapping of economically important quantitative trait loci in sheep I: male distance framework map and QTL for growth rate and body weight. Proc. Assoc. Advmt. Anim. Breed. Gen., 13: 361-364.

Rafalsky A. and M. Morgante (2004). Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. TRENDS in Genetics Vol.20 No.2.

Rafalsky A., M. Morgante, W. Powell, J. M. Vogel, S. V. Tingey (1996). Generating and using DNA markers in plants. In: B. Birren and E. Lai (eds), Analysis of Non-mammalian Genomes: a Practical Guide, 75-134. Academic Press, Boca Raton.

Rohrer G. A. and Keele, J. W (1998). Identification of quantitative trait loci affecting carcass composition in swine: I. Fat deposition traits. J. Anim. Sci. 76, 2247–2254.

Rothschild M., C. Jacobson, D. Vaske, C. Tuggle, L. Wang, T. Short, G. Eckardt, S. Sasaki, A. Vincent, D. McLaren, O. Southwood, H. van der Steen, A. Mileham and G. Plastow (1996). The estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proc. Natl Acad. Sci. USA 93, 201–205.

Ruan J. and Sonnino A. (2007). Marker-assisted selection as a tool for genetic improvement of crops, livestock, forestry and fish in developing countries: an overview of the issues. In: Marker-Assisted Selection. Food and Agriculture organization of the United Nations (FAO), Rome.

Servin B. and F. Hospital (2002). Optimal positioning of markers to control genetic background in marker assisted backcrossing. J. Hered. 93: 214-217.

Singh M., M. Lam., D. McGill, P.C. Thomson, J.A. Cavanagh, K.R., Zenger and H.W. Raadsma (2007). High resolution mapping of Quantitative Trait loci on Ovine chromosome 3 and 20 affecting protein yield and lactation persistency. Proceedings of the Association for the Advancement of Animal Breeding and Genetics, Armidale, NSW Australia. 23-26 Sept 2007.17: 565.

Smith C. and D.B. Smith (1993). The need for close linkages in marker-assisted selection for economic merit in livestock. Anim. Breed. Abstr. 61: 197–204.

Southern E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-510.

Spelman R. and H. Bovenhuis (1998). Genetic response from marker assisted selection in an outbred population for differing marker bracket sizes and with two identified quantitative trait loci. Genetics, 148: 1389–1396.

Spelman R.J. and D.J. Garrick (1998). Genetic and economic responses for within-family marker-assisted selection in dairy cattle breeding schemes. J. Dairy Science, 81: 2942–2950.

Tanksley S. D. et al. (1989). RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7, 257–264

Van Berloo R. and Stam P. (1998). Marker-assisted selection in autogamous RIL populations: a simulation study. Theor. Appl. Genet. 96: 147–154.

Visscher P. M. and C. S. Haley (1999). On the efficiency of marker assisted introgression. Anim. Sci. 68: 59–68.

Vos P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, (1995). AFLP: a new technique for DNA fingerprinting. Nucl. Acids Res. 23: 4407–4414.

Walling G. A., P. M. Visscher, L. Andersson, M. F. Rothschild, L. Wang, G. Moser, M. A. M. Groenen, J. P. Bidanel, S. Cepica, A. L. Archibald, H. Geldermann, D. J. de Koning, D. Milan and C. S. Haley (2000). Combined analyses of data from quantitative trait loci mapping studies. Chromosome 4 effects on porcine growth and fatness. Genetics 155:1369–1378.

Weber J. L. and P. E. May (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388–396.

Weiss K.M. and A. G. Clark (2002). Linkage disequilibrium and the mapping of complex human traits. Trends Genet. 18:19–24

Weller J., Y. Kashi and M. Soller (1990). Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci. 73, 2525–2537.

Weller, J.I., Song, J.Z., Heyen, D.W., Lewin, H.A. and Ron, M. (1998). A new approach to the problem of multiple comparisons in the genetic dissection of complex traits. Genetics 150: 1699–1706.

Wenzl P., J. Carling, D. Kudrna, D. Jaccoud, E. Huttner, A. Kleinhofs, A. Kilian (2004). Diversity arrays technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci USA 101: 9915–9920.

Yang Y. P. and J. E. Womack (1998). Parallel radiation hybrid mapping: a powerful tool for high-resolution genomic comparison. Genome Res. 8, 731–736.

Ye S., S. Dhillon, X. Ke, A. R. Collins and N. M. Day (2001). An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Research 29:17 e88.

Zane L., L. Bargelloni and T. Patarnello (2002). Strategies for microsatellite isolation: a review. Mol. Ecol. 11: 1–16.

Zhang L.Y., M. Bernard, P. Leroy, C. Feuillet, and P. Sourdille (2005). High transferability of bread wheat EST-derived SSRs to other cereals. Theor. Appl. Genet. 111: 677–687.

Zhang Q. et al. (1998). Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics 149, 1959–1973

Zietkietkiewicz E.A.R., A. Rafalsky and D. Labuda (1994). Genome fingerprinting by simple sequenze repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20: 176–183.

Chapter 2

Genomic DNA fingerprinting of chicken (*Gallus gallus* L.) indigenous breeds with M-AFLP and S-SAP markers designed on interspersed repeats

Abstract: In Italy more than fifty different breeds of chicken (*Gallus gallus* L.) are known to have been locally present in the past. The overall situation is now critical since most of these breeds are becoming extinct or threaten and only a few are object of conservation plans. The adoption of molecular markers for the analysis of chicken populations could help characterizing their genetic variation and preserving them from genetic erosion. Both valuable and irreplaceable sources of chicken germplasm, such as indigenous populations of the Veneto region were analyzed by means of DNA fingerprinting with molecular markers designed on interspersed mini- and microsatellite repeats. The identification of either among-breed discriminant and breed-specific markers was based on the S-SAP and M-AFLP systems derived from the AFLP technology. Genomic DNA fingerprints were generated in 84 individuals belonging to six local breeds (Ermellinata, Padovana, Pépoi, Polverara, Robusta lionata and Robusta maculate) and one commercial line used as reference standard. A number of variation statistics were computed for assessing the genetic variability within and relatedness among breeds: the effective number of alleles per locus (n_e=1.570), total and single-breed genetic diversity (H_T=0.365 and H_S=0.208, respectively) and the fixation index (G_{ST}=0.433). The mean genetic similarity coefficients within and between local breeds were 0.769 and 0.581, respectively. Analysis of the population structure along with individual reassignment tests successfully identified all breed clusters and subclusters. The vast majority of animals were correctly assigned to their breed of origin with a probability threshold of 90%. Markers exploitable for the genetic traceability of breeds revealed significant sequence similarities with either genic or intergenic regions of known chromosome position. Sequence tagged site primers were designed for the most discriminant markers in order to develop multiplex non-radioactive genomic PCR assays. The identification of single local breeds according to sequence-specific SNPs and haplotypes was preliminarily attempted and the polymorphism information content of genomic AFLP-derived markers was also reported and critically discussed.

Keywords AFLP technology, DNA fingerprints, *Gallus gallus*, local populations, SNP, genetic variability and traceability.

2.1 Introduction

The domestic chicken (*Gallus gallus*) provides a main protein source for most human populations throughout the world. Its economic importance has made it the focus of numerous research projects, including a recent effort to sequence the entire chicken genome (*http://genome.wustl.edu/projects/chicken*). The vast majority of the sequence has been anchored to chromosomes, represented by twenty-eight pairs of autosomes 1-24, 26-28 and 32, and two sex chromosomes W and Z (International Chicken Genome Sequencing Consortium, for details see Hillier *et al*., 2004).

The relatively small genome of chicken (1.200 Mpb) has been shown to contain around 15% of repetitive DNA organized as short tandem repeats (*e.g*. centromeric and telomeric tandem repeats) as well as numerous families of interspersed repeats, mainly derived from transposable elements (both transposons and retrotransposons) and located over all chromosomes (autosomes, macro-, intermediate, and micro-chromosomes, and sex chromosomes) even though not uniformly (Wicker *et al*., 2005). It has also been proved that micro-chromosomes contain more single-copy sequences and less repeated sequences than macro-chromosomes, and that sex chromosomes are very rich of highly repetitive DNA (**Figure 2.1**). These genomic findings are in agreement with previous genetic mapping studies in chicken (Schmid *et al*., 2000).

The average repeat and gene content for all chromosomes was also calculated. The most abundant repeated sequence is that known as CR1 (Chicken Repeat One): the chicken genome contains over 90.000 copies of this interspersed element belonging to the class LINE (Long Interspersed Nuclear Element). Each element is about 4,5 kb long and includes two genes, one encoding a reverse transcriptase, responsible for the replication of the element itself, and another encoding for an unknown protein likely involved in the transposition process. Additional repeated elements, very abundant in the chicken genome, are those containing tandem repeats of short nucleotide sequence motifs or microsatellites, also known as SSR (Simple Sequence Repeat). Usually the microsatellite elements are less than 200 bp long and are very useful for population genetics and evolution dynamics, as well as phylogenetic studies because of their high repeatability among laboratories, informativeness of sequences and degree of polymorphisms.

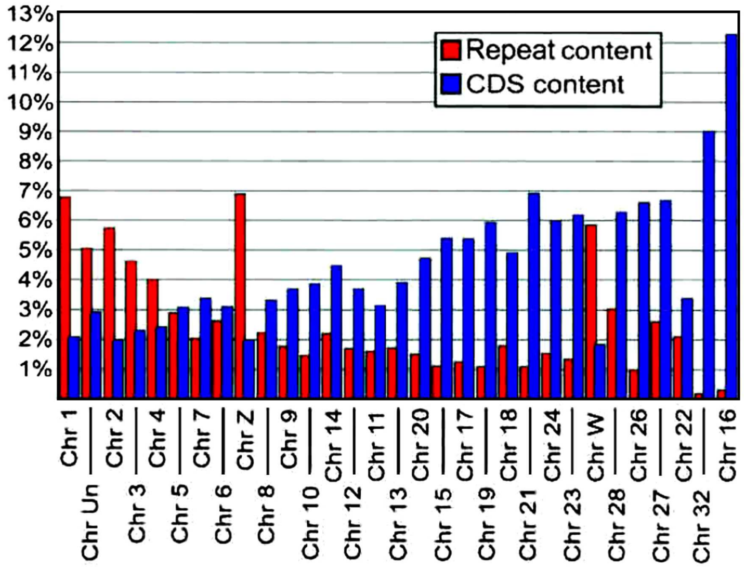


Figure 2.1 Average repeat and gene content for individual chicken chromosomes. The chromosomes are ordered by size with the largest chromosome (n. 1) on the left and the smallest (n. 16) on the right. Chromosomes smaller than 1 Mbp are not included (adapted from Wicker *et al*., 2005).

Genomic DNA sequences of the interspersed hypervariable repetitive elements are considered very useful to set up analysis systems of forensic genetics for a reliable identification of species, breeds and also single individuals on the basis of the banding patterns generated or haplotypes recovered. The generation of multi-locus DNA fingerprints and the detection of sequence polymorphisms in the LINE and SSR regions is usually based on the S-SAP (Sequence-Specific Amplification Polymorphism) and M-AFLP (Microsatellite-Amplified Fragment Length Polymorphism) systems, derived from the more widely known AFLP (Amplified Fragment Length Polymorphism) technology (Vos *et al*., 1995). The approach includes the amplification of genomic cleaved fragments, ligated to specific adaptors and pre-amplified with selective primers, using an AFLP primer in combination with a primer that hybridize to a repeated sequence conserved region, alternatively, to a microsatellite anchored motif. M-AFLP and S-SAP markers derived, respectively, from simple repetitive sequences and transposable elements or minisatellites are also suitable for phylogenetic analyses, since they provide information on genome evolution. Moreover, *in situ* chromosome hybridization experiments and *in silico* bioinformatic investigations have revealed that

DNA repeats and retrotransposons are often localized in euchromatic regions, into or close to functional genes (Hillier *et al*., 2004; Wicker *et al*., 2005). This finding suggests that repetitive and/or transposable elements may be involved in the evolution of animal gene structure and expression, supplying genes with regulatory sequences and facilitating gene duplication and/or exon shuffling (Coullin *et al*., 2005).

In Italy numerous poultry breeds are known to be present. Recently, more than 90 distinct breeds were identified, of which 53 belonging to the chicken. The overall situation of these breeds is nevertheless critical since as much as 61% are becoming extinct, 13% threaten and only 7% are object of conservation plans (Zanon *et al*., 2001). Nowadays, it is generally accepted that the high number of still existing breeds is attributable to the breeding activity based on controlled crosses and selection programs in order to bred very productive lines able to perform and adapt themselves better in a specific territory. To such activity the origin of a consistent biodiversity of the species is mainly owed (Fumihito *et al*., 1996). With the ending of the agriculture in marginal areas and the beginning of chicken breeding at the industrial level, highly specialized lines have been developed in the near past and are currently commercialized, covering the vast majority of the chicken meat and egg market value. This big change is putting to risk of extinction most of the locally spread breeds, particularly those characterized by double attitude which are low productive and not enough competitive compared to commercial broilers. Local breeds of chicken are the populations with the highest genetic variation as well as with the best adaptation to the natural and anthropological environment where they have originated and/or evolved. They contain locally adapted alleles and represent an irreplaceable bank of highly co-adapted genotypes. As a consequence, local breeds are known to posses a good rusticity and many resistance traits to both environmental and biotic stresses, all characteristics that make them of particular interest for the use in biological farms, the recovery of marginal lands and the raising of niche productions. Besides, the autochthonous breeds could be also exploited as alternative to the commercial broilers in case of epidemics. For all the above mentioned reasons, they have recently been object of schemes of protection and maintenance, as well as of studies aimed at the characterization of their gene pools.

Molecular markers are known as a particularly effective and reliable tool for the characterization of genomes and the investigation of gene polymorphisms not only of poultry species, but of several organisms in general (Barcaccia *et al*., 1999; Albertini *et al*. 2003; Soattin *et al*., 2007). In particular, the use of molecular markers allows to measure the genetic variation within single populations and to evaluate the genetic relatedness among populations so that the formulation and implementation of germplasm maintenance programs can be optimized (Cassandro *et al*., 2004). The AFLP and SSR markers along with STS (Sequence Tagged Site) and SNP (Single Nucleotide Polymorphism) markers are the most powerful and robust molecular marker systems for the analysis of genomes and genes, and hence for the molecular characterization of chicken individuals and populations by means of DNA fingerprinting, genotyping or haplotyping (Vanhala *et al*., 1998; Wimmers *et al*., 2000; Zhang *et al*., 2002; Porceddu *et al*., 2002; Hillel *et al*., 2003; Targhetta *et al*., 2003; Cassandro *et al*., 2005). The potentials of AFLP markers in chicken diversity studies were assessed by De Marchi *et al*. (2005). SSR and SNP markers were also applied for investigating genetic variation within and differentiation among chicken breeds (Hillel *et al*., 2003; Twito *et al*., 2007). The analysis based on SSR markers resulted in highly discriminant banding patterns and significant clustering results due to their multi-allelic origin and polymorphism information content (Hillel *et al*., 2003). Nevertheless, SNP markers located in gene regions revealed great advantages in terms of genome coverage, and proved to be an efficient molecular tool for estimating genetic distinctiveness and relatedness in chicken species (Twito *et al*., 2007).

The present research deals with the development of innovative molecular systems of population genomics for chicken DNA fingerprinting based on the M-AFLP and S-SAP analyses of repetitive sequence families (both microsatellites and minisatellites) with the aim of genetically characterizing local breeds and cloning breed-discriminant or breed-specific markers. The identification of novel SNPs is also reported and the polymorphism information content of genomic AFLP-derived markers critically discussed.

2.2 Materials and methods

2.2.1 Animal populations

Twelve individuals for each of the six indigenous chicken breeds under investigation were used for genomic DNA fingerprinting together with a commercial broiler (Golden Comet line) selected for meat production and adopted as reference population, for a total of 84 animals. The indigenous populations at risk of genetic erosion analyzed in this study are the following: Ermellinata, Padovana, Pépoi, Polverara, Robusta lionata and Robusta maculata. The animals were reared in three flocks located throughout the Veneto region, Italy, and their morphological characteristics were previously described by De Marchi *et al*. (2005). The population size for the indigenous breeds have been estimated at 1,500 individuals for Ermellinata, Pépoi, Robusta lionata and at 2,000 individuals for Padovana and Polverara. For each breed, the conservation scheme is based on groups of 34 pure females and 20 males with a breeding scheme that involves a males rotation among conservation units (Cassandro *et al.* 2004).

2.2.2 Molecular markers

Nuclei acids were extracted from whole blood through cell lyses. After purification from RNA residuals and proteins using, respectively, RNase and ammonium acetate, each sample of genomic DNA was precipitated with isopropanol and washed twice with 70% ethanol. Then, all DNA pellets were vacuum dried and redissolved in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The concentration of DNA samples was determined by optical density readings at 260 nm (1 optical density (OD) = 50 µg/ml) and their purity calculated by the OD_{260}/OD_{280} ratio and by the OD_{210}–OD_{310} pattern (Sambrook *et al*. 1989). An aliquot of each genomic DNA was also assayed by electrophoresis on 1% agarose gels.

The detection of polymorphisms on repetitive sequences, such as SSR and CR1 elements, was based on the S-SAP (Sequence-Specific Amplification Polymorphism) and M-AFLP (Microsatellite-Amplified Fragment Length Polymorphism) systems, derived from the more widely known AFLP technology (Vos *et al*. 1995). The approach included the amplification of genomic cleaved fragments, ligated to specific adaptors and pre-amplified with selective primers, using an AFLP primer (*i.e.*, *Eco*RI or *Taq*I rare and frequent cutter-associated primers) in combination with a primer that specifically anneal to the CR1 element or, alternatively, with a primer anchored to a given SSR motif (**Table 2.1**).

Table 2.1 List of conventional AFLP, CR1-specific and SSR-anchored primers.

Primer	Sequence (5′-3′)
EcoRI+A	GACTGCGTACCAATTCA
TaqI+A	GATGAGTCCTGACCGAA
CR1-D1F	TAGTAAATGGGGATGTTGGT
CR1-D2F	TGATCCTCGAGGTCCCTTCC
CR1-S1R	AGCAGCCTTCTGGACCTCTT
CR1-S2R	CAGCAACACTTCACCTCTGG
CR1-InF	AGTTCATGATCTCAAGGGATGTGGGCC
CR1-InR	CAGCCCCCTGATCATCTTTGTGGCCCT
ISSR-6	$(CA)_8GC$
ISSR-13	$CAG(CA)_8$
ISSR-33	$(AGC)_4T$
ISSR-37	$(AGC)_4GT$

M-AFLP and S-SAP fingerprints were generated using the AFLP technology according to Vos *et al*. (1995), as modified by Barcaccia *et al*. (1999) and De Marchi *et al*. (2005). A total of 500 ng of genomic DNA from 84 individuals was digested with a combination of *Eco*RI/*Taq*I restriction enzymes, and ligated to the correspondent adapters with T4 DNA ligase. An aliquot of the restricted-ligated DNA samples was pre-amplified using *Eco*RI and *Taq*I restriction site-specific primers with one selective base each. A radiolabelled specific primer (CR1 or SSR-anchored) were used for the final amplification along with an AFLP primer (*Eco*RI+A o *Taq*I+A). Each 20 µl PCR reaction contained 5 µl of the pre-amplified DNA, 0.2 mM of labelled specific primer and of unlabelled AFLP primer, 2 µl of 10× PCR buffer, 0.2 mM dNTPs and 0.4 U of *Taq* DNA polymerase. The following cycling conditions ensured optimal primer selectivity: 1 cycle of 45 s at 94°C, 30 s at 65°C, 1 min at 72°C followed by 13 cycles of

0.7°C lower annealing temperature each cycle and 18 cycles of 30 s at 94°C, 30 s at 55.9°C, 1 min at 72°C and a final step of 5 min at 72°C. AFLP-derived markers were loaded onto a 6% polyacrylamide gel and electrophoresis was performed at 1.500 V, 40 mA and 40 W. Markers were visualized on autoradiograms after 18 hr exposure at –80°C with intensifying screens.

2.2.3 Genetic diversity and similarity analyses

A preliminarily investigation of diversity was performed computing descriptive statistics such as the observed and effective number of polymorphic loci (n_o and n_e parameters, respectively). The amount of heterozygosity was assessed at two different levels of complexity: single populations or local breeds (H_S) and species as a whole (H_T) according to the formula of Nei (1973) based on marker allele frequency estimates. These statistics of genetic diversity were used to define the genetic structure of populations belonging to single breeds, to estimate the degree of genetic differentiation among different breeds or, equivalently, the fixation index (F_{ST}) as well as the rate of gene flow (Nm). Then, the allele frequency over all marker loci was used to calculate the genetic distance among breeds in all pair-wise comparisons according to Nei (1978). It was also possible to estimate genetic similarities between individuals within single breeds and between different breeds on the basis of genetic fingerprints, adopting the similarity index of Jaccard (1980). UPGMA dendrograms and centroids were constructed using the genetic similarity and diversity matrices. All calculations and analyses were conducted using the software POPGENE (Yeh *et al*., 1997) and NTSYS (Rohlf, 1993).

2.2.4 Population structure and animal assignment tests

The software STRUCTURE (Pritchard *et al*., 2000) was used to analyze the genetic structure of the population and to perform an assignment test on the studied individuals. This program implements a model-based clustering method for inferring population structure using genotype data of unlinked markers. Here it was applied also to assign individuals to each subpopulation or cluster. All AFLP-derived amplicons were treated as haploid markers as suggested by Negrini *et al*. (2007). Analyses were performed using the admixture model with correlated marker allele frequencies. To choose the appropriate number of inferred clusters to model the data, 2 to 12 inferred clusters were performed with 5 independent runs each as suggested by Pritchard *et al*. (2000) and by other authors (Álvarez *et al*., 2004; Glowatzi-Mullis *et al*., 2005). All computations used a burn-in period of 50,000 and 100,000 iterations for data collection.

2.2.5 Subcloning and sequencing of AFLP-derived products

Single discriminant molecular markers that proved to be useful for the traceability of chicken breeds were excised and eluted from the blotted gels, subcloned into plasmid vectors and re-amplified with the same primer combination that yielded the specific genomic DNA fragment. An aliquot of the re-amplified template was sticky-end ligated into a pBluscript II Phagemide. The plasmid DNA was purified from 5 ml of an over-night culture on LB medium of *E. coli* using Plasmid mini prep kit (Sigma Aldrich) following the kit instruction. Plasmid sequences of both strands were performed by the dideoxynucleotide chain reaction termination method using either the M13 forward or reverse primer.

2.2.6 Bioinformatics

The sequence of all discriminant molecular markers was used as query for bioinformatic analyses of the chicken genome database and the major transcript and protein databases. Gene homologues were also searched in public databases by BLASTN and BLASTX applications (Altschul *et al*., 1990) to compare, respectively, nucleotide and translated sequences. Retrievals enabled to attribute given sequences to specific chromosomes and to eventually acquire information on their putative function. All nucleic acid sequences were deposited in the NCBI databases and also recorded in our DNA sequence repertories, including molecular markers conserved within breeds and polymorphic between breeds as well as molecular markers useful to discriminate local breeds from commercial broilers.

For each DNA clone, both strands from at least three distinct animals were aligned to test the veracity of each sequence and to recover their consensus sequence by using the Vector NTI program. Sequences of SCAR (Sequence Characterized Amplified

Region) markers from all chicken breeds were used for multiple sequence alignments in the CLUSTALW program (Higgins *et al.*, 1992) to find out SNP (Single Nucleotide Polymorphisms) and eventually IN/DEL (Insertions/Deletions), and to attempt the identification of breed-specific haplotypes.

2.2.7 SCAR and SNP analyses

The sequence of the most discriminant M-AFLP and S-SAP markers was used for designing primers on their upstream and downstream terminal ends using PerlPrimer program and hence converted into SCAR markers. For each sequence, the analysis of breed-specific SNPs was performed by designing primers with their 3′-end localized on the discriminant point mutation site. PCR of genomic DNA with pairs of sequence-tagged site primers was done using various annealing temperatures (56–66°C) in order to optimize amplification profiles for each selected clone and to visualize polymorphisms for the identification of breeds. The 50 µl reaction volume contained 1× PCR buffer (50 mM EDTA, 1.5 mM $MgCl_2$, 10 mM Tris-HCl), 0.2 mM dNTPs, 0.2 µM of each primer, 200 ng of genomic DNA and 1 U *Taq* DNA polymerase (Sigma Aldrich Red Taq). PCR was carried out with in an initial denaturation step of 94°C for 3 min, followed by 35 cycles of 94°C for 30 s, primers optimal annealing temperature for 30 s, 72°C for 1 min, and a final extension step of 72°C for 10 min. Amplification products were separated by electrophoresis in 2% agarose gels and photographs (DC120 camera, Kodak) were taken after their staining with ethidium bromide.

2.3 Results

2.3.1 Genetic characterization of chicken indigenous breeds by S-SAP and M-AFLP markers

The detection of sequence repeats for fingerprinting the chicken genome was based on the S-SAP and M-AFLP systems, derived from the more widely known AFLP technology. The approach provided for the amplification of genomic cleaved fragments, ligated to specific adaptors and pre-amplified with selective primers, using an AFLP primer in combination with a primer that anneal to a repeated element (*i.e.* CR1) or with a primer anchored to a microsatellite (*e.g.* $(AGC)_n$ and $(CA)_n$) motif. In particular, the S-SAP and M-AFLP marker systems based on the use of either individual or bulked DNA samples enabled reproducible and informative fingerprints and polymorphisms to be obtained within as well as between chicken breeds (Ermellinata di Rovigo, Padovana, Pépoi, Polverara, Robusta lionata and Robusta maculata) and the commercial broiler (Golden Comet line), as shown in **Figure 2.2**.

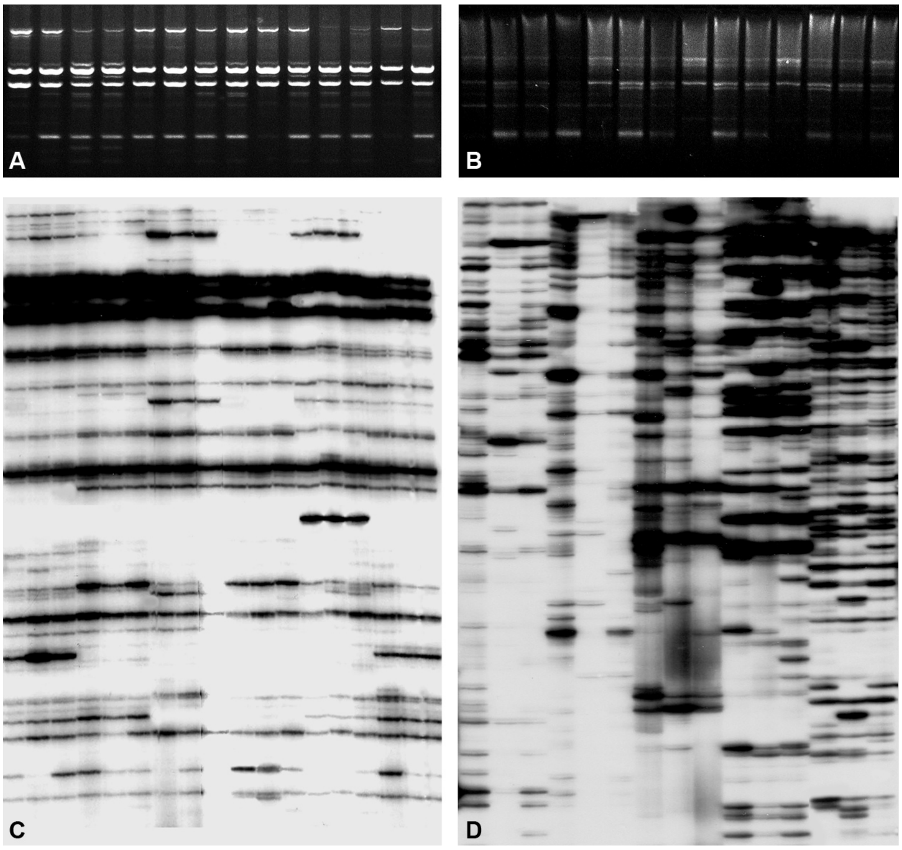


Figure 2.2 Results of primer testing by means of nested PCR experiments using AGC microsatellite-anchored (A) and CR1-internal core (B) primers. Example of genomic DNA fingerprints generated by M-AFLP (C) and S-SAP experiments (D) using a random sample of genomic DNA of animals belonging to different local breeds. Several polymorphic markers are detectable.

Each selected primer combination generated DNA fingerprints showing, on average, from a minimum of 40 to a maximum of 80 fragments. In terms of polymorphism information content, the dinucleotide CA repeat-anchored primers produced the highest number of M-AFLP markers, whereas among the CR1 element-specific primers the highest number of S-SAP markers was yielded by forward ones designed in the most conserved internal region of the chicken repeats. In particular, highly informative and discriminant fingerprints based on microsatellite DNA motifs and repetitive CR1 elements were scored using the primer combinations CAG(CA)$_8$/EcoRI+A and CR1–D2F/TaqI+A, respectively. Polymorphic molecular markers that proved to be useful for the traceability of chicken breeds (*i.e.* both among-breed discriminant and breed-specific markers) were recovered from the blotted gels, subcloned into plasmid vectors and sequenced. The sequence analysis of all selected polymorphic M-AFLP and S-SAP markers allowed us to verify the presence of the specific SSR motif or the partial CR1 element (**Figure 2.3**), thus demonstrating the specificity of the amplification products and, hence, the reliability of the fingerprinting techniques used.

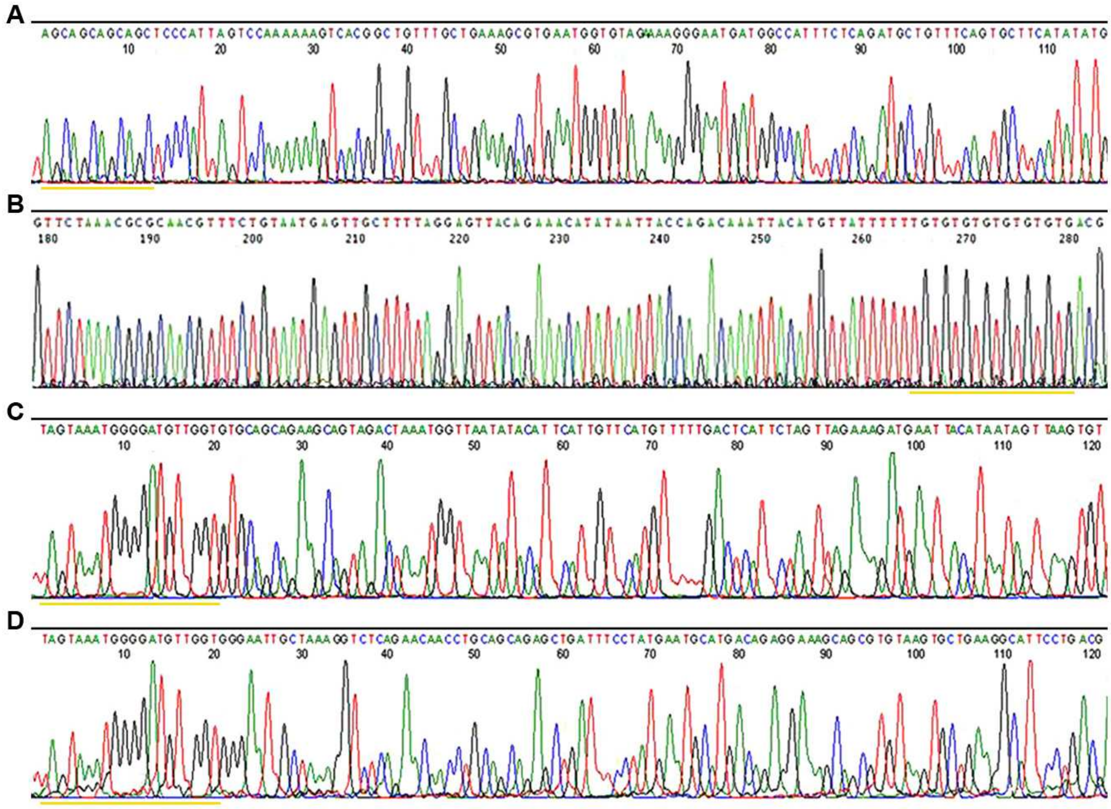


Figure 2.3 Example of chromatograms obtained by sequencing of two amplicons for M-AFLP (A and B) and two for S-SAP primer combinations (C and D). The upstream or downstream region of the sequences includes the repetitive motifs of the microsatellite (GT and AGC) or the CR1 element in which the specific primers were designed.

Both monomorphic and polymorphic DNA markers were scored as present or absent over all chicken DNA fingerprints and used to summarize the M-AFLP and S-SAP data by computing genetic diversity statistics and assessing population relationships. The effective number of alleles per locus was equal to n_e=1.570. Total Nei's genetic diversity was quite similar between M-AFLP markers (H_T=0.334) and S-SAP (H_T=0.381) markers, whereas the mean genetic diversity of single breeds was lower for M-AFLP than S-SAP, measuring H_S=0.162 and 0.232, respectively. Overall, the mean genetic diversity within and between local breeds were 0.365 and 0.208, respectively. Fixation index was around 51% for M-AFLP and 39% for S-SAP, suggesting that the local breeds conserved well-separated their gene pools over time. Variation measured by comparing genetic fingerprints resulted in mean genetic similarity coefficients within and between local breeds equal to 0.769 and 0.581, respectively.

The construction of UPGMA dendrograms and the definition of centroids according to the principal coordinate analysis were also performed using total and mean Dice's genetic similarity matrices. **Figure 2.4** shows centroids plotted using the two principal coordinates separately for M-AFLP markers and S-SAP markers which overall accounted for about 37% of the total genetic variance.

A number of individuals of each breed overlapped the distribution of individuals of other breeds, even though distinct subgroups were clearly discriminated by one or both coordinates. With M-AFLP markers, Robusta lionata and Robusta maculata breeds along with the commercial Broiler were subgrouped separately from the rest of breeds. Moreover, Pépoi and Polverara breeds were clearly distinguishable each other, while Padovana and Ermellinata di Rovigo were clustered very closely. The main distribution difference observed with S-SAP markers concerns the Ermellinata di Rovigo which was clustered apart, together with some individuals of Robusta lionata and Robusta maculata breeds (see Figure 2.4). Such a result can be explained by considering the different chromosome regions assayed by M-AFLP and S-SAP marker systems, since the former was applied with different dinucleotide and trinucleotide repeat-anchored primers whereas the latter was mainly based on the genome-wide spread CR1 elements. As a matter of fact, the phylogenetic relationships among breeds can also be influenced by the polymorphisms information content of DNA markers.

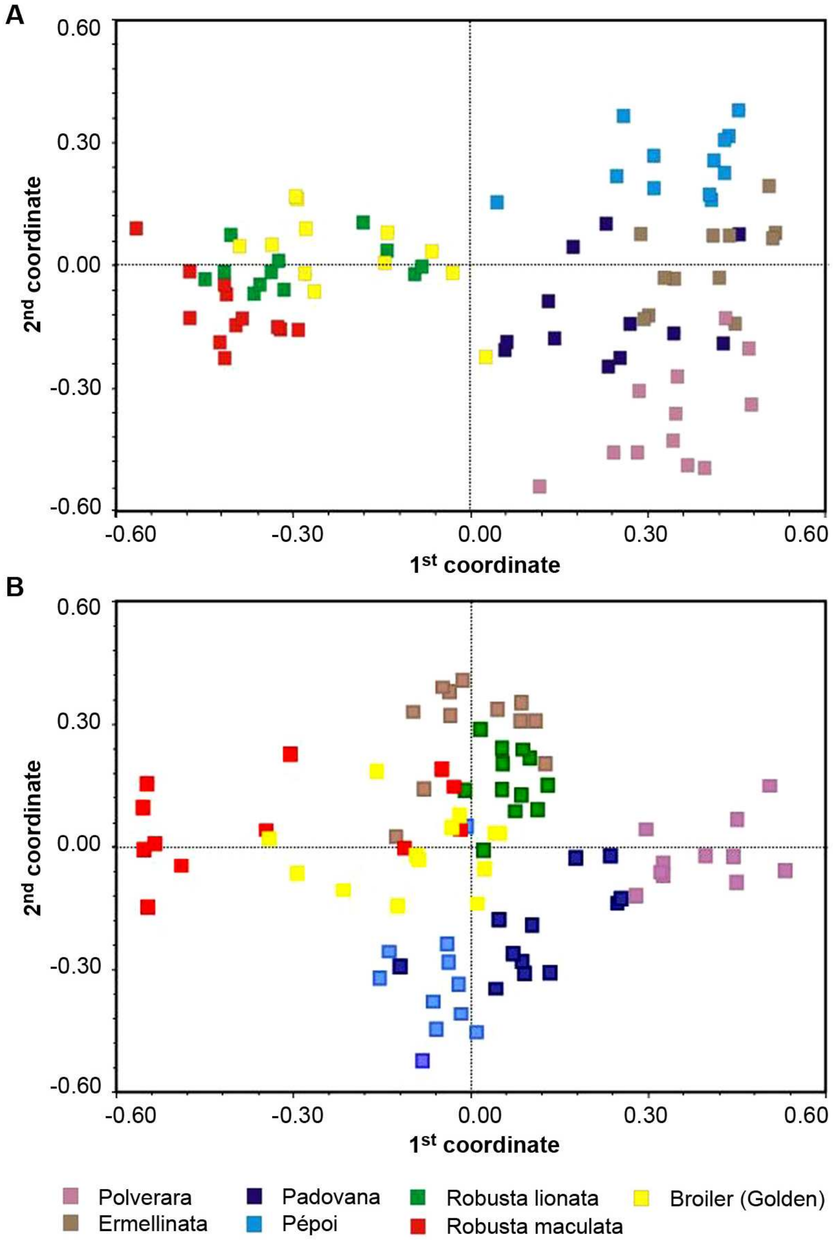


Figure 2.4 Centroids identified from total Dice's genetic similarity matrices using polymorphisms obtained by M-AFLP (A) and S-SAP (B) molecular markers.

Analysis of molecular marker data performed by using the STRUCTURE software enabled the identification of eight clusters out of the total population, each of them corresponding to a single breed, except for clusters 3 and 4 (**Table 2.2**). In fact, the Padovana breed proportions of membership were rather high for these two clusters, reaching altogether more than 90%. It is worth mentioning that Padovana animals analyzed here belonged to two different varieties: Dorata and Camosciata, depending on the colour of their feathers. Actually, in cluster 3 all Dorata animals were identified, whereas Camosciata ones were grouped in cluster 4, showing the ability of the chosen markers to detect this substructure. The average proportion of membership of each breed to the cluster it belongs to was always grater than 90% with the exception of Robusta Maculata (88%).

Table 2.2 Percentage of individuals correctly assigned to their breed of origin considering different threshold in each chicken breed.

Breeds	> 99%	> 95%	> 90%	no threshold
Ermellinata di Rovigo	0	41.7	75.0	100
Pépoi	0	66.7	91.7	100
Robusta Lionata	0	83.3	91.7	100
Padovana	0	25.0	75.0	100
Robusta Maculate	0	50.0	66.7	100
Polverara	0	58.3	75.0	100
Broiler	0	50.0	75.0	100
Total	0	53.6	78.6	100

Results of the individual assignment test, considering different thresholds, are shown in **Table 2.3**. All animals were correctly assigned to their breed of origin. On the whole, about 80% of the samples were correctly allocated considering a threshold of 90%. Pépoi and Robusta Lionata evidenced the best results with more than 90% of individuals correctly assigned at 90% probability rate.

Table 2.3 Proportion of membership of each of the 7 chicken breeds in the inferred clusters.

Clusters	1	2	3	4	5	6	7	8
Ermellinata di Rovigo	0.016	0.014	0.009	0.009	0.008	0.923	0.015	0.007
Pépoi	0.935	0.004	0.006	0.010	0.004	0.031	0.004	0.006
Robusta Lionata	0.007	0.006	0.004	0.007	0.008	0.006	0.004	0.958
Padovana	0.039	0.011	0.532	0.378	0.012	0.006	0.014	0.007
Robusta Maculata	0.005	0.010	0.005	0.006	0.879	0.005	0.006	0.083
Polverara	0.010	0.006	0.014	0.019	0.004	0.006	0.934	0.007
Broiler	0.007	0.912	0.027	0.012	0.017	0.008	0.007	0.009

Overall reassignment estimates demonstrate the discriminant ability of AFLP-derived markers for detecting population substructure and tracing individual breeds.

2.3.2 Polymorphism information content of discriminant AFLP-derived markers

The sequence of the most among-breed discriminant markers as well as the sequence of the breed-specific markers were used as queries for public database interrogations. Chicken genome retrievals revealed significant similarities with either genic or intergenic sequences of known chromosome position and primary structure homologies with known gene products (**Table 2.4**). All nucleotide sequences recovered from the indigenous chicken breeds as AFLP-derived markers were deposited in the NCBI databases (accession numbers EF417921- EF417932).

Table 2.4 List of breed-specific clones with relative length, position on chicken chromosome, accession number, sequence type and amino acid homology.

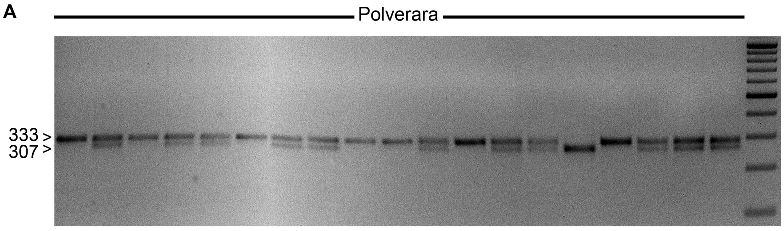
Clone	Chr.	Locus ID	Type	Gene product
#37	1	NW001471534	Intergenic	5′ Hypothetical protein – 3′ Monoamine oxidase B
#38	15	NW001471459	Genic	Hypothetical protein
#39	3	NW001471677	Genic	Acetyl-CoA synthetase 2
#40	2	NW001471654	Intergenic	5′ Syntrophin beta 1 – 3′ Hyaluronan synthase 2
#42	3	NW001471671	Genic	Sorting nexin 14
#44	2	NW001471639	Intergenic	5′ Hypothetical protein – 3′ Tolloid-like protein 1
#48	4	NW001471687	Genic	Fibroblast growth factor receptor-like 1
#50	1	NW001471525	Genic	Hypothetical protein
#51	26	NW001471609	Genic	Potassium voltage-gated channel protein
#52	1	NW001471534	Genic	Interferon alpha/beta/omega receptor 1
#55	9	NW001471743	Intergenic	5′ Cek6 protein – 3′ Acetylglucosaminyltransferase
#56	17	NW001471503	Genic	Hypothetical protein

To attempt the setting up of a reliable PCR-based molecular reference system suitable for the precise identification of the single breeds, sequence-tagged site primers were designed on the 12 most discriminant clones in order to convert the among-breed discriminant and breed-specific M-AFLP and S-SAP markers into easily detectable non-radioactive SCAR markers. The sequence of the forward and reverse primers is reported in **Table 2.5**.

Table 2.5 Forward and reverse sequences of the SCAR primers designed for the most discriminant AFLP-derived clones.

Clone	Expected size	Primer sequence (5′-3′)
#37	486	For TAGTAAATGGGGATGTTGGTGGTGA Rev GAATTCACCAACATCCCCATT
#38	307	For TAGTAAATGGGGATGTTGGTCACCA Rev GAATTCACATCAATATAAAGCAA
#39	300	For TAGTAAATGGGGATGTTGGTGGGAA Rev GAATTCAGCACGTTTCACTACA
#40	302	For GAATTCAAACAGACAAAATAAATG Rev TAGTAAATGGGGATGTTGGTATAAA
#42	272	For GAATTCAGTAAGAAAGACCA Rev TAGTAAATGGGGATGTTGGTCACT
#44	201	For TAGTAAATGGGGATGTTGGTGTGCA Rev GAATTCACCTGCCTATCAAATT
#48	630	For GAATTCAGCTATGGGACCAT Rev AGCAGCAGCAGCTCACACTAA
#50	597	For AGCAGCAGCAGCTCCCATTAG Rev GAATTCAGGGAGCTTGCAGA
#51	493	For GAATTCACAGGCCTTGGTTC Rev AGCAGCAGCAGCTAGGAAAG
#52	422	For GAATTCAGAAGGAATAGCTTTA Rev AGCAGCAGCAGCTGGCAGT
#55	295	For AGCAGCAGCAGCTCAGCACAG Rev GAATTCAGGGTTATCATTTCC
#56	209	For GAATTCACACAGAAACGTCT Rev AGCAGCAGCAGCTCAGTATGG

When these primers were used in PCR experiments using chicken genomic DNA as templates, in most cases amplification products showed to be shared among individuals over all breeds and thus not useful for discriminating single breeds (data not shown). This result is most likely attributable to the origin of AFLP polymorphisms usually based on single nucleotides on the restriction endonuclease action site and/or oligonucleotide primer annealing region. Nevertheless, certain polymorphisms could be generated using very stringent PCR conditions but they proved to be not fully reliable with a few exceptions. The most robust result with SCAR markers was obtained for the clone #38 preliminarily selected as specific for the Polverara breed. When the correspondent SCAR primers were used to analyze genomic DNA samples from all breeds, two distinct amplification products of 307 and 333 bp were detected. The upper marker proved to be shared among the vast majority of individuals over all local breeds and the commercial broiler, whereas the lower marker was detected only in the Polverara individuals (**Figure 2.5**), with a relative frequency of 52% (32 individuals out of 62).



B

```
  1 TAGTAAATGGGGATGTTGGTCACCAGAAAGTAGGAAAAATGCTATTAGATGAAGTGAAAA 60
  1 TAGTAAATGGGGATGTTGGTCACCAGAAAGTAGGAAAAATGCTATTAGATGAAGTGAAAA 60

 61 TATATACATACACACACACACAAAATAAACTAAAAAGTTTTCTTCTATTTTCAGATGGTT 120
 61 TATATATATACACACACACACAAAATAAACTAAAAAGTTTTCTTCTATTTTCAGATGGTT 120

121 CAGCTCTTCAATTAAAATGACATATTTACATCCTTGTTTGCCCTAAAGATGTACTCTTCA 180
121 CAGCTCTTCAATTAAAATGACATATTTACATCCTTGTTTGCCCTAAAGATGTACTCTTCA 180

181 CTGAACACACCATTTTTGCCAAGAATTTCAGCCAAACTTAGGTCCTAATGATCTTTAGCT 240
181 CTGAACACACCATTTTTGCCAAGAATTTCAGCCAAACTTAGGTCCTAATGATCTTTAGCT 240

241 ATAAATG--------------------------GAAAGAGGGAGCTATATGACATTTCAC 274
241 ATAAATGCAACAATGAGGCTTTTCCTACTATTGGAACGAGGGAGCTATATGACATTTCAC 300

275 ATAAATAGTTTTGCTTTATATTGATGTGAATTC 307
301 ATAAATAGTTTTGCTTTATATTGATGTGAATTC 333
```

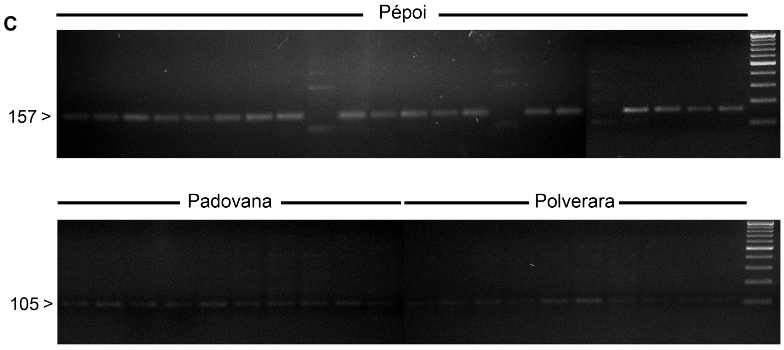


Figure 2.5 Amplification products of 333 bp and 307 bp generated with SCAR primers specific for clone #38: the lower band is specific of the Polverara breed and 52% of individuals shares it (A). Simple alignment of the two nucleotide sequences showing a 26 bp-long insertion/deletion: SNP sites at nucleotide positions 67 and 251 are marked by underlined letters. (B). Amplification products of 157 bp and 105 bp generated with a forward SNP site-specific primer in combination with a reverse SCAR marker-specific primer: the upper band is specific of the Pépoi breed and it is present in 78% of individuals (C).

It is interesting to note that the two marker sequences differed only for an insertion/deletion of 26 bp in length found in the internal region. Owing to their high nucleotide sequence similarity (92%) and apparent co-segregational patterns in the population (*i.e.* the longer, the shorter or both amplification products scored over all assayed DNA samples), a total of 62 Polverara individuals were molecularly characterized to study the marker allele frequencies and to verify the allelism relationships between marker alleles. The three possible genotypes at the marker locus being tested were found in the following relative proportions: $M_{303}M_{303}$=0.500, $M_{303}M_{307}$=0.403 and $M_{307}M_{307}$=0.097. Since the observed proportions proved to be not statistically deviating from the proportions expected in case of Hardy-Weinberg equilibrium (χ^2=0.0919), the two amplification products of 307 bp and 333 bp can be most likely considered as alleles of the same gene. On the basis of genebank retrievals, the core sequence of clone #38 confirmed a significant similarity with the gene encoding for an hypothetical protein of locus NW_001471459 located on chicken chromosome 15.

To verify the occurrence of single nucleotide polymorphisms in the selected clones, as main source of AFLP polymorphisms, several amplification products yielded with each of the SCAR primer combinations were recovered from the gels, subcloned into plasmid and sequenced. A multiple alignment of the consensus sequences of each of the 12 clones from all local breeds and the commercial broiler allowed us to discover SNPs and IN/DELs, not only in the terminal ends of the SCAR markers as expected, but also in their internal regions. The sequence information was used to define haplotypes to be preliminarily adopted for the recognition of single breeds (**Table 2.6**).

In the clone #38, the detection of a SNP at nucleotide position 67 was useful not only to test the reliability of the IN/DEL polymorphism found in the cosegregating sequence, but also to confirm its utility for the identification of individuals belonging to the Polverara breed. A forward SNP site-specific primer, designed with its discriminant 3′-end nucleotide in the point mutation position (p#38_SNP_{67}: GCTATTAGATGAAGTGAAAATATATAC), when used in combination with a reverse SCAR marker-specific primer yielded a doublet of amplicons, as expected. The nucleotide sequences of the two markers confirmed complete identity except for the 26 bp insertion/deletion, as well as the segregation patterns proved to fully match the segregation ratios previously scored as SCAR analysis (see Figure 2.5).

Table 2.6 Main single nucleotide polymorphims detected by sequencing of the breed-specific clones #38, #39, #50, and #56 amplified using SCAR primers (for each clone, the numbers indicate the nucleotide position of the SNP).

Breed	Clone #38		Clone #39		Clone #50			Clone #56		
	67	**251**	**189**	**191**	**23**	**223**	**312**	**51**	**70**	**96**
Ermellinata	C/T (A)	A	C	T	-	-	-	A	C	T
Padovana	C/T	A	C	T	C	G	G	A	C	T
Pépoi	C/T	A/C	T	C	T	A	A	C	G	C
Polverara	C/T	A/C	-	-	C	A	A	A	C	T
Robusta Lionata	C or C/T	A	C	T	C	A	A	A	C	T
Robusta Maculata	C or C/T	A	C	T	C	A	A	C	C	T
Broiler	C/T (A)	A	C	T	C	G	A	C	C	T

- Not determined; ( ) indicates rare nucleotides.

For the identification of the Pépoi individuals a number of breed-specific SNPs were discovered, for instance, in the clones #39, #50 and #56. In particular, each of two different SNPs found in the sequence of clone #56 proved to be useful for the traceability of individuals belonging to this breed. For the detection of single nucleotide polymorphisms, a forward SNP site-specific primer was designed with its discriminant 3′-end in the point mutation position (p#56_SNP$_{69}$: GGAGGTTCCCAAGCCCG and p#56_SNP$_{95}$: CTGTAGAGTTTTCAGCC). When either one of these two primers was used in combination with a common reverse SCAR marker-specific primer, a single amplicon was specifically detected in most of the individuals of the Pépoi breed (see Figure 2.5). However, the same primer combinations yielded a second amplification product, shorter in length and also fainter as signal intensity, which was shared by all individuals tested of the Padovana and Polverara breeds. On the whole, the upper marker of 157 bp was scored in 67 individuals out of 86 (78%) totally analyzed, whereas the remaining 19 (22%) individuals of the breed showed the lower marker of 105 bp, as all Padovana and Polverara individuals analyzed in this study. This finding support the occurrence of gene flow among Padovana, Polverara and Pépoi populations for an allele encoding for an hypothetical protein (locus NW_001471503).

It is worth mentioning that in the clone #38 two different SNPs showed either a homozygote or heterozygote state in some of the analyzed individuals. For instance, at nucleotide position 67 individuals with CC, TT or CT were observed (**Figure 2.6**).

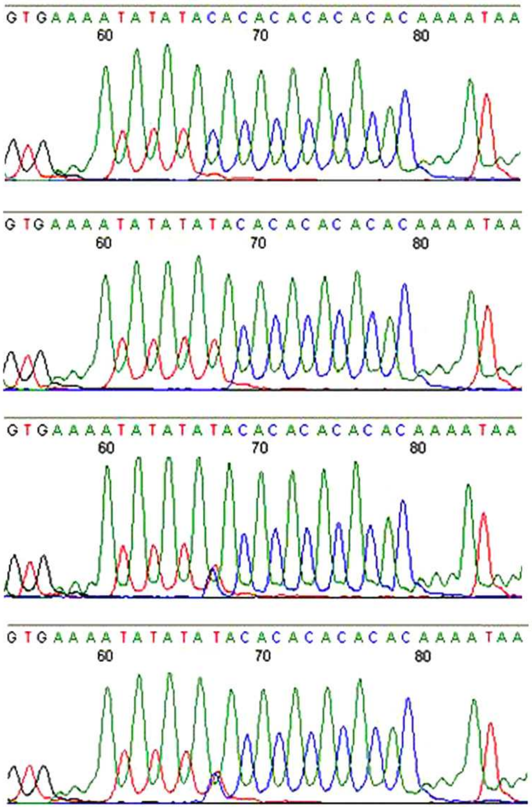


Figure 2.6 Chromatograms of clone #38 showing individuals in a homozygote (CC or TT) and heterozygote state (CT) at nucleotide position 67.

2.4 Discussion

Molecular markers have revolutionized and modernized our ability to characterize genetic variation and to rationalize genetic selection, being effective and reliable tools for the analysis of genome architectures and gene polymorphisms in animals (Lanteri and Barcaccia, 2006). Until now, the area of chicken genomics that has shown the greatest development with respect to the use of molecular marker technology is that of population genetics. For instance, SSR and AFLP markers have been already exploited for assessing genetic diversity in chicken breeds (Hillel *et al*., 2003; De Marchi *et al*., 2005). Although not yet investigated at the population level, SNP markers should also be very useful for surveying genetic variation and differentiation of poultry breeds (Wong *et al*., 2004). One strategy for genetic variability and traceability studies could be based on the identification of informative AFLP-derived markers and the exploitation of the SNPs contained in their sequences as source of among-breed discriminant or breed-specific polymorphisms.

On the basis of gained results, the S-SAP and M-AFLP marker systems were shown to be suitable to visualize reproducible multi-locus DNA fingerprints of chicken breeds: they can be potentially exploited for investigating genetic variation within and assessing genetic relatedness among populations on the basis of polymorphic interspersed repeats. Highly informative and discriminant genetic fingerprints were obtained assaying the variation for short tandem repeats (*e.g.*, CA/TG) and longer unique repeats (*i.e.*, CR1) by using repeat-anchored primers in combination with *Eco*RI+A or *Taq*I+A primers. These two AFLP technology-derived molecular marker systems can be used in genetic characterization studies including the determination of the main genetic variability statistics, such as marker allele frequency, degree e of expected heterozygosity of single breeds, genetic distance and gene flow among breeds. As far as our case-study in concerned, the genetic variation among the considered indigenous populations was measured to be around 45%, meaning that 55% of genetic variation was due to polymorphisms within single populations at the assayed marker loci. Overall, the combination of M-AFLP and S-SAP data and their comparison with previously obtained SSR data (unpublished results) confirmed the high genetic variation detectable within breeds of the Veneto region and the clear genetic differentiation still present among their gene pools. Our results also revealed substantial differences concerning the genetic distance estimates among local breeds and the commercial broiler adopted as reference line. This finding is most likely due to the fact that the three molecular marker systems were applied to different individual sample sizes and the analysis was based on different genomic loci numbers. Additional molecular analyses are needed to corroborate the primary results and to confirm the statistics related to the genetic diversity within and the phylogenetic relationships among breeds native to and locally reared in north-eastern Italy.

On the basis of our preliminary results, the polymorphic markers isolated from genomic interspersed repeats can be applied not only to assess genetic variability estimates of any chicken germplasm resource, but also to develop a genetic traceability system for the identification of the different regional chicken breeds and the commercial valorization of their meat and egg niche products. In fact, the analysis of the population structure along with individual assignment tests successfully identified all breed clusters. In particular, six main clusters out of the eight totally inferred were found to correspond to as many single breeds, with proportions of membership for each breed around 90% on average. Two additional smaller clusters identified two varieties of the Padovana breed, showing an overall proportion of membership higher than 90%. The vast majority of animals were correctly assigned to their breed of origin (*i.e.*, about 79% of the individuals were correctly allocated considering a probability threshold of 90%), thus demonstrating the suitability and reliability of the chosen AFLP-derived marker systems for detecting population substructure and tracing individual breeds.

The final aims of our research at regional scale are the following: i) to characterize the gene pools of indigenous chicken breeds using multi-locus DNA-based assays; ii) to generate chicken DNA barcodes using single-locus unique marker tags. Knowledge of genetic variation within local breeds and genetic differentiation among breeds is expected to have a significant impact on the preservation and valorization of regional chicken germplasm resources. As a matter of fact, AFLP-derived markers anchored to interspersed mini- and microsatellite repeats are exploitable to find out multiple polymorphisms per assay and to investigate genetic variability levels. SCAR analysis seems not reliable for genetic traceability since a given discriminant marker isolated by AFLP-based systems proved to be reproducible only in some cases. This can be due to the presence of single nucleotide polymorphisms at the restriction site level not detectable using clone-specific primers. Moreover, the sequencing of SCAR markers of the different breeds underscored the presence of SNPs and IN/DELs both in terminal and internal regions of the clones. The sequence information was used in an attempt of defining breed-specific haplotypes to discriminate individuals belonging to a given autochthonous population. Preliminary results on breed-specific SNPs and haplotypes obtained for each clone, although very promising, need to be deeply investigated and finally validated by increasing the number of animals for single breed and by analyzing additional local breeds. Once informative AFLP-derived sequences and discriminant SNPs have been selected and tested, the use of high-throughput methods will be essential to develop a robust PCR-based and low cost genetic traceability system.

Wong *et al*. (2004) reported that the majority of SNPs identified are common among most of the poultry breeds and that they are highly abundant, being their average frequency equal to 1/200 bp when two chromosomes are compared. Thus, the identification of polymorphisms due to single nucleotide substitutions in the AFLP-derived sequences is not surprising if one considers that when a random fragment of genomic DNA is sequenced in a sample of 10 or more animals one SNP will be theoretically identified every 80-100 bp (M. Groenen, pers. comm.). Our major goal was, however, not that of discovering additional SNPs but, rather, that of assessing the type and nature of polymorphisms in the most informative AFLPs visualized in our genomic DNA fingerprints. In fact, if it is true that almost 3 million SNPs are now available for genetic studies in chicken (Wong *et al*., 2004), it is also true that in chicken no AFLP-derived sequences are retrievable from the NCBI databases (http://www.ncbi.nlm.nih.gov/). Furthermore, in our case-study, the detection of SNPs into among-breed polymorphic and breed-specific S-SAP and M-AFLP markers was preferred to the random investigation of publicly available SNPs: the former were expected to be useful for the genetic traceability of indigenous chicken breeds because preliminarily selected as shared within and polymorphic between breeds, whereas for the latter any information in terms of polymorphism information content was available in chicken local populations.

To the best of our knowledge, this is the first survey of the polymorphism information content of AFLP-derived sequences belonging to chicken breeds. Despite the increasing number of publications based on the use of AFLP markers for genome fingerprinting and gene mapping purposes as well as for quantitative genetics and population genetics studies, the number of AFLP-derived sequences deposited in the GenBank is very low for animals (less than 100 records!). It means that AFLP markers are worldwide exploited by many researchers as anonymous polymorphisms without performing any sequence content analysis or without recording any sequence in public databases.

Moreover, it is interesting to note that breed-specific markers were shown to be highly similar to genic regions of known chromosome position. Most of the breed-specific clones isolated has shown homology with genes implicated in metabolic processes. For instance, the clone #39 scored high similarity with an acetyl-CoA synthetase whereas clone #50 with a putative mitochondrial carrier protein. Both proteins are molecules utilized in various metabolic pathways, including fatty acid and cholesterol synthesis. The implementation of a genomic database for Italian chicken local breeds, including many more molecular markers conserved within breeds and polymorphic among breeds as well as molecular markers useful to discriminate local breeds from commercial broilers will be one of the future goals of our project. An increasing number of agro-alimentary companies provides for an internal traceability system, though it is necessary to document the whole food product chain from the producers to the consumers. The European Union, with the approval of the Reg. CE no. 178, 28.01.2002, makes obligatory from the January 1st, 2005 the traceability of any food item, as the possibility to reconstruct and recover the pathway followed by a given product across all the steps of production, transformation and distribution. The traceability must be referred to every single product portion. As a consequence, for specific products like, for instance, the avian carcasses which reach the consumer as disjuncted parts, the setting of a sequence-tagged molecular marker-based genetic traceability system is extremely useful. Such a molecular system is the only one that can offer, at any time of the food product chain, the possibility to assess the origin and reveal the nature of meat products, for improving reliability of traditional labelling systems.

Acknowledgements

This work was supported by grant from Veneto Agricoltura with regional project CO.VA. titled “Piano regionale di intervento per la conservazione e la valorizzazione di razze avicole locali”. Thanks are due to Dr. Chiara Dalvit for providing genomic DNA samples of local chicken breeds and for performing assignment tests of chicken individuals using AFLP-derived markers.

2.5 References

Albertini E., Porceddu A., Marconi G., Barcaccia G., Pallottini L. and Falcinelli M. (2003). Microsatellite-AFLP for genetic mapping of complex polyploids. Genome, 46: 824-832.

Altschul S.F. Gish W., Miller W., Meyers E.W. and Lipman D.J. (1990). Basic local alignements tool. J. Molecular Biology, 215: 403-410.

Barcaccia G., Albertini E., Tavoletti S., Falcinelli M. and Veronesi F. (1999). AFLP fingerprinting in Medicago spp.: its development and application in linkage mapping. Plant Breeding, 118: 335-340.

Barcaccia G., Albertini E. and Falcinelli M. (1999). AFLP fingerprinting in Pelargonium peltatum: its development and potential in cultivar identification. J. Horticultural Science Biotechnology, 74: 243-250.

Cassandro M., De Marchi M., Targhetta C., Dalvit C., Ramanzin M. and Baruchello M. (2004). An in situ marker-assisted conservation scheme of 11 Italian avian breeds. Proc. of the 55th annual meeting of the European Association for Animal Production, September 5-9, 2004. Bled, Slovenia. p. 30.

Cassandro M., C. Targhetta, M. De Marchi, C. Dalvit, G. Barcaccia and G. Bittante (2005). DNA fingerprinting characterization of four Italian poultry species. Proc. of Plant and Animal Genome XIII Conference. January 15-19, 2005. San Diego, CA .(USA). p. 210.

Coullin P., B. Bed’Hom, J. J. Candelier, D. Vettese, S. Maucolin, S. Moulin, S. A. Galkina, A. Bernheim and V. Volobouev (2005). Cytogenetic repartition of chicken CR1 sequences evidenced by PRINS in Galliformes and some other birds. Chromosome Research, 13: 665–673.

De Marchi M., Cassandro M., Targhetta C., Baruchello M. and Notter 8 D.R. (2005). Conservation of poultry genetic resource in the Veneto region of Italy. Animal Genetic Resources Information, 37: 63-74.

De Marchi M., Dalvit C., Targhetta C. and Cassandro M. (2005). Assessing genetic diversity in indigenous Veneto chicken breeds using AFLP markers. Animal Genetics, 37: 101-105.

Fumihito A., Miyake T., Takada M., Shingu R., Endo T., Gojobori T., Kondo N. and Ohno S. (1996). Monophyletic origin and unique dispersal patterns of domestic fowls. Proc. Natl. Academy Science USA, 93: 6792-6795.

Higgins D.G., Bleasby A.J. and Fuchs R. (1992). ClustalW: a package for performing multiple sequence alignments on a micro-computer. Computational Applied Bioscience, 8: 189-191.

Hillel J., Groenen M.A.M., Tixier-Boichard M., Korol A.B., David L., Kirzhner V.M., Burke T., Barre-Dirie A., Crooijmans R.P.M.A., Elog K., Feldman M.W., Freidlin P.J., Mäki-Tanila A., Oortwijn M., Thomson P., Vignal A., Wimmers K. and Weigend S. (2003). Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. Genetics Selection Evolution 35: 533-557.

Hillier L.W., Miller W., Birney E., Warren W., Hardison R.C., Ponting C.P., Bork P., Burt D.W., Groenen M.A.M, Delany M.E. and Dogson J.B. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature, 432: 695-716.

Lanteri S. and Barcaccia G. (2006). Molecular markers based analysis for crop germplasm preservation. In: The role of biotechnology in exploring and protecting agricultural genetic resources (J. Ruane and A. Sonnino, eds.). FAO, Rome. pp. 105-120.

Porceddu A., Alberini E., Barcaccia G., Marconi G., Bertoli F.B. and Veronesi F. (2002) Development of an S-SAP technique based on a LTR-like sequence of Medicago sativa L. Molecular Genetics Genomics, 267: 107-114.

Sambrook J., Fritsch E.F. and Maniatis T. (1989). Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schmid, M., Nanda, I., Guttenbach, M., Steinlein, C., Hoehn, M., Schartl, M., Haaf, T., Weigend, S., Fries, R., Buerstedde, J.M., et al. (2000). First report on chicken genes and chromosomes 2000. Cytogenet. Cell Genetics, 90: 169–218.

Soattin, M., Barcaccia G., Cassandro M. and Bittante G. (2007). Development of sequence-specific DNA markers for the genetic traceability of autochthonous chicken breeds. Plant and Animal Genome XV Conference. January 13-17, 2007. Town & Country Convention Center. San Diego, California (USA). p. 240.

Targhetta C., De Marchi M., Cassandro M., Barcaccia G., Ramanzin M. and Bonsembiante M. (2003). Use of AFLP markers in poultry genome characterisation. Proc. of the 54th annual meeting of the European Association for Animal Production, August 31th - September 3rd, 2003. Roma, p. 12.

Twito T., Weigend S., Blum S., Granevitze Z., Feldman M.W., Perl-Treves R., Lavi U., Hillel J. (2007). Biodiversity of 20 chicken breeds assessed by SNPs located in gene regions. Cytogenet Genome Research, 117: 319-326.

Vanhala T., Tuiskula-Haavisto M., Elo K., Vilkki J. and Maki-Tanila A. (1998). Evaluation of genetic variabilità and genetic distances between eight chicken lines using microsatellite markers. Poultry Science, 77: 783-790.

Vos P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper and M. Zabeau (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research, 12: 3396-3303.

Wicker T., Robertson J.S., Schulze S.R., Feltus F.A., Magrini V., Morrison J.A., Mardis E.R., Wilson R.K., Peterson D.G., Paterson A.H. and Ivarie R. (2005). The repetitive landscape of the chicken genome. Genome Research, 15: 126-136.

Wimmers K., Ponsuksili S., Hardge T., Valle-Zarate A., Mathur P.K. and Horst P. (2000). Genetic distinctiveness of African, Asian and South American local chickens. Animal Genetics, 31: 159-165.

Wong G.K. et al. (International Chicken Polymorphism Map Consortium) (2004). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature, 432: 717-722.

Zanon A. and Sabbioni A. (2001). Identificazione e salvaguardia genetica delle razze avicole italiane. Annuali Facoltà Medicina Veterinaria, Università di Parma, 21: 117-134.

Zhang X., Leung F.C., Chan D.K.O., Yang G. and Wu C. (2002). Genetic diversity of Chinese native chickens breeds based on protein polymorphism, randomly amplified polymorphic DAN and microsatellite polymorphism. Poultry Science, 81: 1463-1472.

Chapter 3

Investigating SNPs and haplotypes for the calpain (CAPN1) gene in cattle (*Bos taurus* L.) populations: Preliminary analysis of linkage disequilibrium

Abstract: A reliably tender meat product is one of the most important attributes for maintaining consumer satisfaction with beef steaks. Selection programs for genetic improvement in tenderness might be implemented in cattle if candidate genes and appropriate markers could be found. Calpain 1 (*CAPN1*), a gene encoding the enzyme μ-calpain which degrades myofibrillar proteins post-mortem, is thought to be one of the most promising candidates for beef tenderness. Three distinct single-nucleotide polymorphisms (SNPs) in the *CAPN1* gene were studied in a population of cattle to determine their frequency and the frequency of their haplotypes. A total of 170 animals were subjected to genotyping for the *CAPN1* gene-associated SNP markers 316 (G>C), 530 (G>A) and 4751 (T>C). The detection of nucleotides in the three sequence positions of the *CAPN1* gene was performed by using the ARMS-PCR and the PCR-RFLP methods. Genetic diversity statistics and inbreeding coefficients were computed, and population structure and animal ordination analyses were also performed. Moreover, Linkage Disequilibrium (LD) values were estimated using the standardized disequilibrium coefficient (D′) and the correlation coefficient (r^2). The relative proportions of the single nucleotides at each SNP site were equal to p(G)=0.9176 and q(C)=0.0824 for the marker 316, to p(G)=0.5710 and q(A)=0.4290 for the marker 530, and to p(T)=0.5178 and q(C)=0.4822 for the marker 4751. The observed proportions of single SNP markers were not significantly deviating from proportions expected in case of Hardy-Weinberg equilibrium. However, an excess of heterozygous animals was found for SNP markers 316 and 530, whereas an excess of homozygous animals was found for the SNP marker 4751. The pair-wise comparisons between SNP markers 316 and 530, and 316 and 4751 revealed linkage equilibrium in the population, whereas a highly significant (P<1‰) linkage disequilibrium was scored between SNP markers 530 and 4751. In particular, a marked excess was found for three haplotypes, *i.e.* GG/AA/TT, GG/GG/CC, and GG/GA/TC, while deficiency was found for most of the remaining haplotypes. The availability of preliminary data on SNP haplotypes can be useful to plan experiments aimed at testing the relationship between a candidate gene, like *CAPN1*, and a number of tenderness parameters, such as shear force and driploss.

Keywords: Cattle *CAPN1* gene, SNP markers, haplotypes, LD, MAS for tenderness

3.1 Introduction

A reliably tender meat product is one of the most important attributes for maintaining consumer satisfaction with beef steaks (Morris *et al.*, 2006). If it is true that beef tenderness is a critical component of palatability, it is also true that the difficulty in obtaining phenotypic data until after harvest has made it difficult to breed for this trait (White *et al.*, 2005). Therefore, selection for genetic improvement in tenderness has rarely been attempted in cattle. Marker-assisted selection would allow to bypass this obstacle if appropriate markers could be found.

Quantitative analysis of crossbred populations revealed the presence of two loci with effects on meat tenderness, measured as Warner-Bratzler shear force of longissimus muscle. One mapped to the central portion of BTA15 (Keele *et al.*, 1999) and the other to the telomeric end of BTA29 (Casas *et al.*, 2000). Comparative mapping indicates that both BTA15 and BTA29 have substantial homology with human chromosome 11 (Solinas-Toldo *et al.*, 1995; Barendse *et al.*, 1997; Kappes *et al.*, 1997). HSA11 was therefore examined for genes with known effects on muscle metabolism or development to identify potential candidate genes underlying the effect on meat tenderness. Since the calcium-dependent protease system has been implicated in the post-mortem tenderization process (Koohmaraie, 1996) and because the gene coding for the large subunit of micromolar activated calpain (*CAPN1*) maps to HSA11, this gene was identified as a potential candidate gene for the observed QTL (**Figure 3.1**).

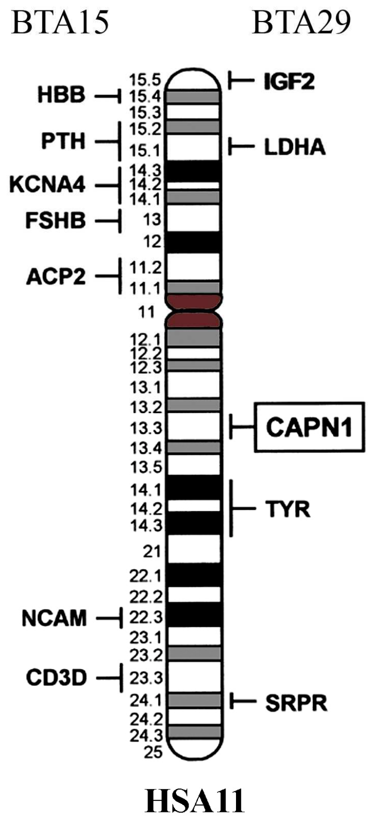


Figure 1 Interruptions in conserved synteny between HSA11 chromosome and the genes assigned to BTA15 and BTA29 linkage groups (adapted from Smith *et al.*, 2000).

The cDNA full-length was obtained in cattle by Smith *et al.* (2000) allowing to shed light on the structure of this gene: a total of 19 exons and 17 introns were revealed and sequenced. The polymorphisms detected in one of the introns were used to map the

CAPN1 gene to the telomeric end of the BTA29 linkage group. It was then demonstrated that this region approximately coincides with the position of the QTL, hence suggesting that *CAPN1* gene is a positional candidate encoding for a protease potentially affecting variation in meat tenderness in cattle (Smith *et al.*, 2000).

Processing and cooking practices are important for tenderness, along with activity levels of calpain (micro- and milli-calpain) and their inhibitor calpastatin (Koohmaraie, 1996). Recently, single-nucleotide polymorphism (SNP) markers in the *CAPN1* gene have been suggested to fill this role (Page *et al.*, 2002; 2004; White *et al.*, 2005; Morris *et al.*, 2006). The *CAPN1* gene encodes the protease μ-calpain, which degrades myofibrillar proteins post-mortem and is thought to be one of the most important enzymes involved in beef tenderness (Koohmaraie, 1996). It is known that the calpain and calpastatin proteins include distinct subunits encoded by four genes: calpain 1 (*CAPN1*), calpain small subunit 1 (*CAPNS1*) and calpastatin (*CAST*) have been mapped on chromosomes BTA29 (Smith *et al.*, 2000), BTA18 (Band *et al.*, 2000) and BTA7 (Bishop *et al.*, 1993), respectively, whereas calpain 2 (*CAPN2*) has not been mapped yet in cattle.

The human *CAPN1* gene includes 22 exons, spanning approximately 30 Kb, and the bovine *CAPN1* homolog has a similar structure (**Figure 3.2**). Four SNPs in the *CAPN1* gene revealed significant effects on tenderness of cooked beef (Page *et al.*, 2002; 2004; White *et al.*, 2005; Morris *et al.*, 2006). Page *et al.* (2002) found two non-synonymous SNP in the *CAPN1* gene that produce amino acid substitutions at positions 316 (Glycine/Alanine) and 530 (Valine/Isoleucine) in the protein. The two SNP markers consistently identified favourable alleles at the QTL for tenderness in two distinct resource families (Casas *et al.*, 2000; Morris *et al.*, 2001) and were later shown to be associated with tenderness in a wide range of crossbred *Bos taurus* cattle (Page *et al.*, 2004). Nonetheless, these SNP markers were also shown to not segregate at appreciable frequencies in Brahman cattle (Casas *et al.*, 2005) and to be homozygous in a Brahman × Hereford population, which nonetheless was found to segregate at the QTL for tenderness located on BTA 29 (Casas *et al.*, 2003), indicating that SNP 316 and SNP 530 do not identify all variation at the *CAPN1* gene affecting tenderness. White *et al.* (2005) found one additional SNP in the *CAPN1* gene at position 4751 on the intron 17 correlated with tenderness in a large, multisire American Brahman population, being associated with shear force on post-mortem. This finding demonstrated that genetic variation important for tenderness segregates in *Bos indicus* cattle at or near the *CAPN1* locus (White *et al*., 2005). More recently, Morris *et al*. (2006) demonstrated that one more SNP at position 947 (Alanine/Glycine) of the *CAPN1* gene is significantly associated with shear force at intermediate stages of ageing and with tenderness of cooked strip loins (M. longissimus dorsi steaks).

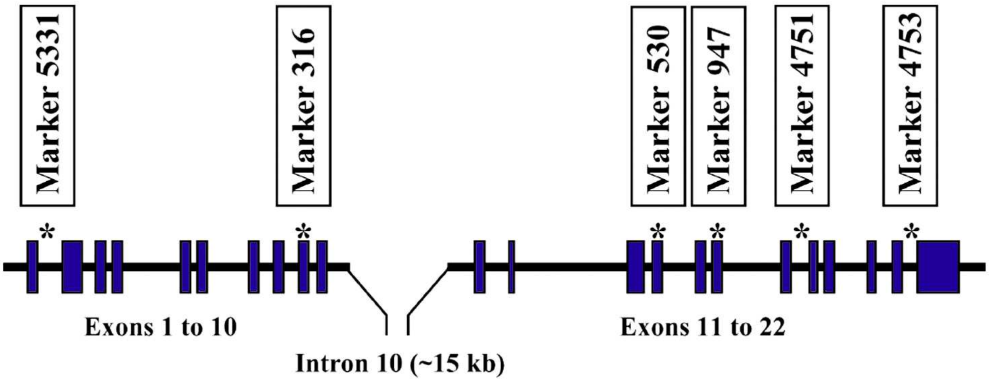


Figure 3.2 Organization of the *CAPN1* gene in the cattle (*Bos taurus* L.) genome and localization of SNP markers (adapted from White *et al*., 2005).

Interest in the study of linkage disequilibrium (LD) (*i.e*., non-random association of alleles) has increased dramatically in recent years because of two factors. First, genomic technology enables rapid identification of haplotypes at many genetic loci, either by direct DNA sequencing-by-synthesis or by high-throughput single-nucleotide polymorphism (SNP) analysis. Second, in the presence of significant LD, of the order of tens of kilobases or more, it can be possible to identify genetic regions that are associated with a particular trait of interest (e.g., Mendelian genes and QTLs) by a systematic, high-density genome scan of individuals from an existing population (Rafalski and Morgante, 2003). By contrast, if LD declines rapidly around the causative gene, the identification of genes responsible for the trait of interest is possible by screening a limited number of candidate genes. Individual SNPs or haplotypes within a candidate gene are systematically tested for association with the phenotype of interest.

In some cases, the identification of causative variants, known as QTNs (quantitative trait nucleotides) might be possible.

Whole-genome scanning and candidate gene mapping methodologies are fundamentally similar, and differ primarily in the scale at which the analysis is performed. Understanding the patterns of LD across the genome in the available populations as well as measuring LD between markers within a candidate gene will facilitate the choice of the appropriate methodology for genetic association mapping and marker-assisted breeding. In particular, SNP markers that are tightly linked to a QTL as well as SNP markers that are directly located on a candidate gene can be in partial or complete population-wide linkage disequilibrium with the trait, such that some SNP marker-QTL allele haplotypes are more frequent than expected by chance. In this case, selection can be performed on the basis of the marker genotype. The probability of population-wide LD is higher for closely linked markers and in selected populations of small effective size, which is the case for livestock species and breeds. Breeding programs assisted by molecular markers has rarely been attempted in cattle. Genetic improvement through marker-assisted selection would facilitate phenotypic evaluation by progeny testing if appropriate markers could be found.

This study deal with the investigation of single-nucleotide polymorphisms in the calpain (*CAPN1*), a gene with great potential importance for meat tenderness in cattle (*Bos taurus* L.). In particular, the main objectives were to test the existence of genetic variation at the calpain locus in the population under study by assessing SNP markers and haplotypes. Preliminary measurements of LD between site pears of the *CAPN1* gene were also performed. The characterization of the population on the basis of animal genotypes is reported and implications for marker-assisted selection are discussed.

3.2 Materials and methods

3.2.1 Experimental populations

A total of 170 animals belonging to a larger population of Piedmontese breed (*Bos taurus* L.) were used for detecting SNP markers and genotyping analysis. In particular, animals were chosen randomly from a population made of 1,080 young bulls selected within the progeny of 148 AI sires. At slaughtering animals were 523±73 d old and their average carcass weight was equal to 417±45 kg. Two samples of *Longissimus thoracis* muscle (LM) were collected from each animal 24 h after slaughtering and immediately stored at −20°C. All samples were used for the evaluation of meat traits according to EU grading system and for the definition of animal genotypes based on SNP markers. Meat tenderness data were collected using the method proposed by ASPA in 1996 measuring the shear force on 5 cylindrical cores (1.13 cm in diameter), obtained from the steaks at aged 7 d post-mortem, by a TA.HDi Texture Analyzer, equipped with a Warner-Bratzler shear device and a texture expert software (Boukha *et al.*, 2007).

3.2.2 Genomic DNA preparations

Meat samples deriving from the 170 Piedmontese individuals were used for total genomic DNA isolation. The DNA extraction was performed using 100 mg of stored LM through cell lyses and protein precipitation steps by ammonium acetate followed by purification using isoproponanol.

The concentration of genomic DNA samples was determined by optical density spectrophotometer readings at 260 nm (1 O.D.=50 μg/ml) and their purity calculated by the $O.D._{260}/O.D._{280}$ ratio and by the $O.D._{210}/O.D._{310}$ pattern (Sambrook *et al.*, 1989). An aliquot of each genomic DNA sample was also assayed by electrophoresis on 1% agarose gel.

3.2.3 SNP markers and genotyping of animals

Markers 316 and 530 in the *CAPN1* gene have been previously described (Page *et al.*, 2002), while marker 4751 has been more recently characterized by White *et al.* (2005). Marker names for these SNPs were derived from U.S. Meat Animal Research Center primer numbers, and have no meaning in regard to *CAPN1* sequence at either the DNA or protein levels. Marker CAPN316 is a C/G substitution (Ala/Gly) polymorphism and lies in exon 9 (base 5,709 of AF252504), marker CAPN530 corresponds to a A/G substitution (Ile/Val) in exon 14 (base 4,558 of AF248054) while CAPN4751 is equivalent to a C/T substitution in intron 17 (base 6,545 of AF248054).

Development and assessment of SNP markers in the sequence of the CAPN1 gene were performed as reported by Rincòn and Medrano (2006) with some changes to adapt their protocols to available instruments and reagents. The two authors used an alternative system for genotyping animals, on the basis of the above mentioned SNP markers, using the tetraprimer amplification refractory mutation system-polymerase chain reaction (ARMS-PCR) and the polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) methods. The SNP genotyping based on the ARMS-PCR technique, previously described by Ye *et al*. (2001), was carried out employing two primer combinations to amplify the two different marker alleles of each SNP site in a single PCR reaction. More specifically, two allele-specific amplifications occur in opposite directions with two outer primers that assay the region of the SNP and two inner allele-specific primers that have an allele-specific mismatch at the 3'-terminal base and a second deliberate mismatch at position −2 from the 3'-end (Rincòn and Medrano, 2003; 2006).

The *Bos taurus CAPN1* accession numbers AF248054 and AF252504 were used for designing primers by Rincòn and Medrano (2006). In particular, the SNP markers 316 and 4751 were analyzed by tetra-primer ARMS technique using two primer pairs to amplify the two correspondent alleles in a single PCR reaction. The nucleotide variations of SNP marker 530 were assayed using a PCR-RFLP technique adopting the *Psy*I restriction enzyme to digest amplicons. The primer sequences and fragment sizes are presented in **Table 3.1**.

For SNP markers 316 and 4751, tetra-primer ARMS-PCR reactions were performed in total volumes of 25 µl, containing 50 ng of genomic DNA, 10 pmol each inner primer, 1 pmol each outer primer, 200 µM dNTPs, 1× *RedTaq* PCR buffer (10 mM Tris-HCL, pH 8.3, 50 mM KCl, 1.1 mM MgCl2) and 1 U *RedTaq* DNA polymerase (Sigma-Aldrich). The amplifications were performed using a touchdown PCR with annealing of primers set at 64°C for the first cycle, decreasing 1°C every two cycles until 60°C. Then the steps of denaturation at 95°C, annealing at 60°C and extension at 72°C were performed for 27 cycles. PCR products were resolved by electrophoresis in 2.5% agarose gels, stained with ethidium bromide and visualized under UV light.

The PCR products for the SNP marker 530 were generated using a standard PCR protocol in a total volume of 25 µl, containing 50 ng of genomic DNA, 1 pmol of each primer, 200 µM dNTPs, 1× *RedTaq* PCR buffer and 1 U *RedTaq* DNA polymerase (Sigma-Aldrich). The PCR products were then digested with *Psy*I following the instructions of manufacturer (Fermentas) through an incubation at 37°C for 2 h. The RFLP profiles were analyzed in 3% agarose gels, stained with ethidium bromide and visualized under UV light.

Table 3.1 Sequences (5′ to 3′) of *CAPN1*-related primers used for genotyping animals.

Markers	Primer sequence*	Product size (bp)
SNP 316	For inn TTTCCTGCAGCTCCTCGGAGTGGAAGGG	269 (G allele)
	Rev inn GCTCCCGCATGTAAGGGTCCAGGG	228 (C allele)
	For out GCTGTGCCCACCTACCAGCATC	446
	Rev out CAGGTTGCAGATCTCCAGGCGG	
SNP 530	For CGTTTCTTCTCAGAGAAGAGCGCAGGGA	341 (A allele)
	Rev CCTGCGCCATTACTATCGATCGCAAAGT	195 and 146 (G allele)
SNP 4751	For inn GCATCCTCCCCTTGACTGGGGGGAAACCC	158 (C allele)
	Rev inn GTCACTTGACACAGCCCTGCGCCGCA	231 (T allele)
	For out CCTGGAGTCCTGCCGCAGCATGGTCAAC	334
	Rev out AAGCTGCAGGAGCTGCCCAAAGCCAGGC	

*The underlined nucleotides represent the mismatched bases in the inner amplification refractory mutation system primers.

3.2.4 Genetic diversity statistics

SNP data from each animal DNA sample were recorded according to the dual nature of the marker system by assigning a nucleotide correspondence to each band identified in the gel. Standard genetic diversity (H) statistics of Nei (1973) and the inbreeding (F) coefficients of Wright (1965) were used to summarize the population structure on the basis of SNP markers for the *CAPN1* gene. Let p_i denote the frequency of the *i* SNP marker at a given site, the total genetic diversity of the subpopulation was computed as $H_S=1-\Sigma p_i^2$ which corresponds to the expected heterozygosity (Nei, 1973). Once directly determined the observed heterozygosity (H_I) of individuals, the fixation index was calculated as $F_{IS}=(H_S-H_I)/H_S$ to measure the deficit (inbreeding) or excess (outbreeding) of heterozygotes in the subpopulation (Wright, 1978). The observed number of markers (n_o) and the effective number of markers (n_e) per SNP site were also calculated according to Kimura and Crow (1964).

All calculations and analyses were conducted using the software PopGene version 1.21 (Yeh *et al.*, 1997).

3.2.5 Population structure and animal ordination analyses

The population structure and animal ordination analyses were performed according to the unweighted pair-group arithmetic average method (UPGMA) clustering algorithm (Sneath and Sokal 1973), and the centroids of all SNP haplotypes were constructed from the symmetrical nucleotide similarity matrix determined using the Simple Matching coefficient. The principal coordinate analysis was then applied to compute the first two components out of the similarity coefficient estimates: the triangular matrix of nucleotide similarity was double-centered and then bi-dimentionally plotted according to the extracted Eigen-vectors (Rohlf, 1972).

All calculations and analyses were conducted using the appropriate routines of the software NTSYS version 1.80 (Rohlf 1993).

3.2.6 Statistical analysis of LD estimates

For each of the three single-nucleotide polymorphisms investigated in the CAPN1 gene, the frequency of SNPs and genotypes were computed over all animals.

The frequency of haplotypes in the population were also computed for all possible combinations of the three SNPs. The Linkage Disequilibrium (LD) for all pair-wise comparisons of SNPs was estimated by using the standardized disequilibrium coefficient (D′) of Hedrick (1987) and the squared marker frequency correlation (r^2) of Weir (1996). In particular, $D'=D/D_{max}$, where $D=f_{ab}-p_a p_b$ is equivalent to the difference, in absolute value, between the frequency of a given marker haplotype in the population and the product of the frequency of single markers (Lewontin, 1972). The coefficient of correlation, $r^2=D^2/(p_a q_a p_b q_b)$ is given by the ratio between the squared standardized disequilibrium coefficient and the product of the frequency of all possible markers (Hill and Robertson, 1968). D′ is affected solely by recombination events between SNP sites, whereas r^2 is also affected by differences in SNP frequency at the two sites being analyzed (Remington *et al*., 2001). The probabilities of obtaining LD estimates at least as extreme as values observed under an hypothesis of linkage equilibrium (*P* values) were then calculated by using the exact test of Fisher (1918) for SNP site pairs with two nucleotide variants each. A probability less than 0.001 was arbitrarily chosen to indicate a statistically significant amount of linkage disequilibrium.

All calculations and analyses were conducted using the software Genetic Data Analysis (GDA) version 1.0 (Lewis and Zaykin, 1999).

3.3 Results

3.3.1 Detecting SNPs and haplotypes for the calpain (*CAPN1*) gene

To test nucleotide variability in the *CAPN1* gene within the cattle population under study and its utility for association mapping, SNP markers 316 (G>C), 530 (G>A) and 4751 (T>C) were individually analyzed and all 170 animals subjected to genotyping. The relative proportions of the single nucleotides at each SNP site were equal to $p(\mathrm{G})=0.9176$ and $q(\mathrm{C})=0.0824$ for the marker 316, to $p(\mathrm{G})=0.5710$ and $q(\mathrm{A})=0.4290$ for the marker 530, and to $p(\mathrm{T})=0.5178$ and $q(\mathrm{C})=0.4822$ for the marker 4751.

Table 3.2 reports the number and frequency of observed animal genotypes for the three SNP sites. The observed proportions of single SNP markers were not significantly deviating from proportions expected in case of Hardy-Weinberg equilibrium, as demonstrated by the chi-square values. Concerning *CAPN1* gene polymorphisms, it is worth noting that G was already shown to be the most common nucleotide for both SNP markers 316 and 530 in *Bos taurus* populations (Page *et al*., 2004). Moreover, White *et al*. (2005) reported C as a rare nucleotide (<10%) for the SNP marker 4751 in a population of *Bos indicus*.

Table 3.2 Number and frequency of observed and expected genotypes for the three SNPs.

SNPs	Observed		Expected		χ^2
Marker 316	**No.**	**Frequency**	**No.**	**Frequency**	
GG	143	0.8412	143.2	0.8419	
GC	26	0.1529	25.7	0.1514	
CC	1	0.0059	1.2	0.0067	0.0141 n.s.
Marker 530					
GG	53	0.3136	55.1	0.3260	
GA	87	0.5148	82.8	0.4900	
AA	29	0.1716	31.1	0.1840	0.3864 n.s.
Marker 4751					
TT	50	0.2959	45.3	0.2681	
TC	75	0.4438	84.4	0.4994	
CC	44	0.2604	39.3	0.2325	2.2069 n.s.

A number of genetic diversity coefficients were estimated using the formulas of Nei (1987). The observed and effective number of SNP markers were computed along with the heterozygosity statistics for all SNP sites. Markers 530 and 4751 revealed very high and similar values of variation, whereas marker 316 scored the lowest variation statistics (**Table 3.3**).

Table 3.3 Summary of heterozygosity (H) statistics along with fixation index (F) values.

SNP	Size	n_a	n_e	H_I	H_S	F_{IS}
Marker 316	340	2.0000	1.1781	0.1529	0.1511	−0.0119
Marker 530	338	2.0000	1.9605	0.5148	0.4899	−0.0508
Marker 5751	338	2.0000	1.9975	0.4438	0.4994	0.1113
Mean	339	2.0000	1.7120	0.3705	0.3801	0.0486
St. Dev.	–	0.0000	0.4628	0.1917	0.1984	0.0846

The observed (H_I) and expected heterozygosity (H_S) estimates for the single SNP markers and the mean values over all SNP markers are summarized in **Table 3.3**. Moreover, the fixation index (F_{IS}) was computed for each of the SNP markers as an estimate of heterozygote deficiency (+ values) or excess (– values) within the population. It is worth mentioning that an excess of heterozygous animals was found for SNP markers 316 and 530, whereas an excess of homozygous animals was found for the SNP marker 4751 (see also **Table 3.2**).

3.3.2 Analysis of linkage disequilibrium: preliminary data

The non-random association of nucleotides (*i.e.*, linkage disequilibrium) at the three different SNP sites of the *CAPN1* gene was investigated as a key step for association mapping purposes.

A highly significant (P<1‰) linkage disequilibrium was scored between SNP markers 530 and 4751, whereas the pair-wise comparisons between SNP markers 316 and 530, and 316 and 4751 revealed linkage equilibrium in the population under study (**Table 3.4**).

Table 3.4 Single population linkage disequilibria between pairs of SNP markers.

SNP markers	**No.**	**D′**	**r^2**	**χ^2**	***P***
316-530	168	0.0440	0.1045	0.514	n.s.
316-4751	168	0.0422	0.0934	1.103	n.s.
530-4751	168	0.8616	0.4109	28.020	<0.00001

When pair-wise combinations of point mutation sites of the *CAPN1* gene with two nucleotides each were compared, nine different two-site SNP haplotypes were expected at frequencies resulting from the product of frequencies of single-site SNP markers. Concerning the haplotypes based on SNP markers 316-530 and 316-4751, any significant difference was found between observed and expected proportions.

In particular, **Table 3.5** reports the relative proportions of haplotypes and the differences between observed and expected estimates for SNP markers 316 and 530. Very small absolute and relative differences were found between observed and expected proportions over all haplotypes.

Table 3.5 Proportions of haplotypes and differences between observed and expected estimates for SNP markers 316 and 530.

SNP 316-530	**Observed proportions**		**Expected proportions**		**Differences**	
Haplotypes	**No.**	**Freq.**	**No.**	**Freq.**	**No.**	**Freq.**
GG/GG	46	0.2738	44	0.2638	2	0.0100
GG/GA	71	0.4226	73	0.4330	−2	−0.0104
GG/AA	25	0.1488	24	0.1443	1	0.0045
GC/GG	7	0.0417	8	0.0479	−1	−0.0062
GC/GA	14	0.0833	13	0.0787	1	0.0046
GC/AA	4	0.0238	4	0.0262	0	−0.0024
CC/GG	0	0.0000	0	0.0019	0	−0.0019
CC/GA	1	0.0060	1	0.0030	0	0.0030
CC/AA	0	0.0000	0	0.0010	0	−0.0010
Total	168	1.0000	168	1.0000	7	0.0044

Table 3.6 reports the relative proportions of haplotypes and the differences between observed and expected estimates for SNP markers 316 and 4751.

Table 3.6 Proportions of haplotypes and differences between observed and expected estimates for SNP markers 316 and 4751.

SNP 316-4751	**Observed proportions**		**Expected proportions**		**Differences**	
Haplotypes	**No.**	**Freq.**	**No.**	**Freq.**	**No.**	**Freq.**
GG/TT	42	0.2500	42	0.2489	0	0.0011
GG/TC	62	0.3690	63	0.3733	−1	−0.0043
GG/CC	38	0.2262	37	0.2190	1	0.0072
GC/TT	7	0.0417	8	0.0452	−1	−0.0035
GC/TC	13	0.0774	11	0.0679	2	0.0095
GC/CC	5	0.0298	7	0.0398	−2	−0.0100
CC/TT	0	0.0000	0	0.0017	0	−0.0017
CC/TC	1	0.0060	0	0.0026	1	0.0034
CC/CC	0	0.0000	0	0.0015	0	−0.0015
Total	168	1.0000	168	1.0000	8	0.0422

It is important to emphasize that absolute differences between observed and expected proportions of haplotypes for SNP markers 316 and 530 as well as for SNP markers 316 and 4751 proved to be very low (≤2 animals). In case of numerical discrepancies attributable to the animal sample size, markedly higher differences would have been scored between observed and expected haplotype frequencies.

The haplotypes based on SNP markers 530 and 4751 revealed consistent differences between observed and expected proportions that were shown to be highly significant (**Table 3.7**). In particular, haplotypes for SNP markers 530 and 4751 revealed very large differences, ranging from 7 to 27 animals per single haplotype. Marked excesses and deficiencies were discovered for certain haplotypes. For instance, haplotypes GG/CC (+27, 3-fold), AA/TT (+20, 3-fold), and GA/TC (+26, 2-fold) scored frequencies much higher than expected, while haplotypes GA/CC (−20), GG/TT (−15), AA/TC (−13) and GG/TC (−11) scored frequencies much lower than expected (about 2-fold each).

Table 3.7 Proportions of haplotypes and differences between observed and expected estimates for SNP markers 530 and 4751.

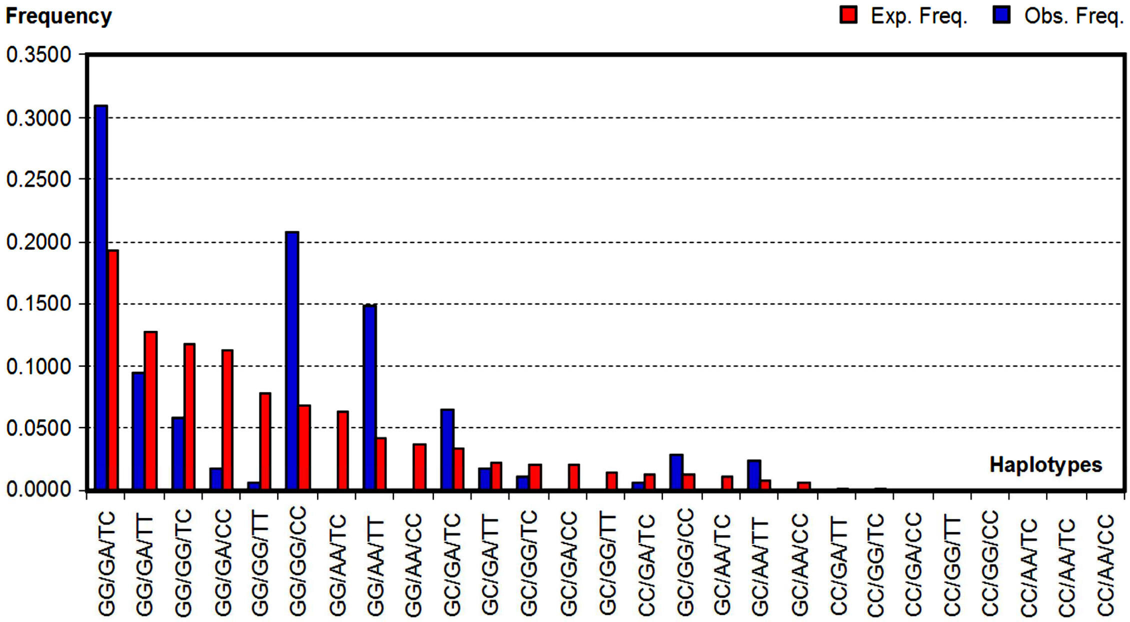
SNP 530-4751	**Observed proportions**		**Expected proportions**		**Differences**	
Haplotypes	**No.**	**Freq.**	**No.**	**Freq.**	**No.**	**Freq.**
GG/TT	1	0.0060	16	0.0928	−15	−0.0868
GG/TC	12	0.0714	23	0.1392	−11	−0.0678
GG/CC	40	0.2381	13	0.0817	27	0.1564
GA/TT	19	0.1131	26	0.1523	−7	−0.0392
GA/TC	64	0.3810	38	0.2285	26	0.1525
GA/CC	3	0.0179	23	0.1341	20	−0.1162
AA/TT	29	0.1726	9	0.0508	20	0.1218
AA/TC	0	0.0000	13	0.0762	13	−0.0762
AA/CC	0	0.0000	8	0.0447	−8	−0.0447
Total	168	1.0000	168	1.0000	147	0.8616

As a matter of fact, the discrepancies between observed and expected haplotype proportions proved to be very small for the combinations of SNP markers 316 and 530 and SNP markers 316 and 4751, and much larger for the haplotypes deriving from SNP markers 316 and 4751. Moreover, skewed distributions of animals among the nine different haplotypes were expected to be randomly distributed over the three pair-wise SNP site comparisons.

Overall SNP marker frequencies and haplotype proportions could have been biased by the relatively small sample size, being the number of animals equal to 168. As a consequence, for association mapping purposes additional animals have to be investigated to avoid skewed distributions of SNP markers and possible haplotypes. Preliminary analysis of SNP haplotypes can however be very useful to plan experiments aimed at testing the relationship between a candidate gene such as *CAPN1* and a number of meat quality traits, such as tenderness.

The analysis of haplotypes was then extended at the three SNP markers and their proportions were estimated in the whole population. **Figure 3.4** reports the observed and expected proportions of the 27 possible SNP haplotypes in the analyzed population of cattle. As few as three of them proved to be the most abundant genotypes, whereas six genotypes were never recorded.

Figure 3.4 Histograms of observed and expected SNP haplotype frequencies.



A marked excess was found for haplotypes GG/AA/TT (+18, 257%), GG/GG/CC (+24, 200%), and GG/GA/TC (+19, 59%) over all SNP markers, whereas deficiency was found for most of the remaining haplotypes (particularly, GG/GA/CC −16, 84%; GG/GG/TT −12, 92%; and GG/AA/TC −11, 100%).

Principal components analysis allowed the definition of centroids of SNP haplotypes using the Simple Matching genetic similarity matrix. All animals were grouped into three main subgroups, as can be seen from the scatter diagram plotted according to the first two components (**Figure 3.5**). The most numerically represented haplotypes, namely GG/GA/TC (51 animals), GG/GG/CC (36 animals), and GG/AA/TT (25 animals), accounting for two third of the whole population, were plotted within distinct quadrants.

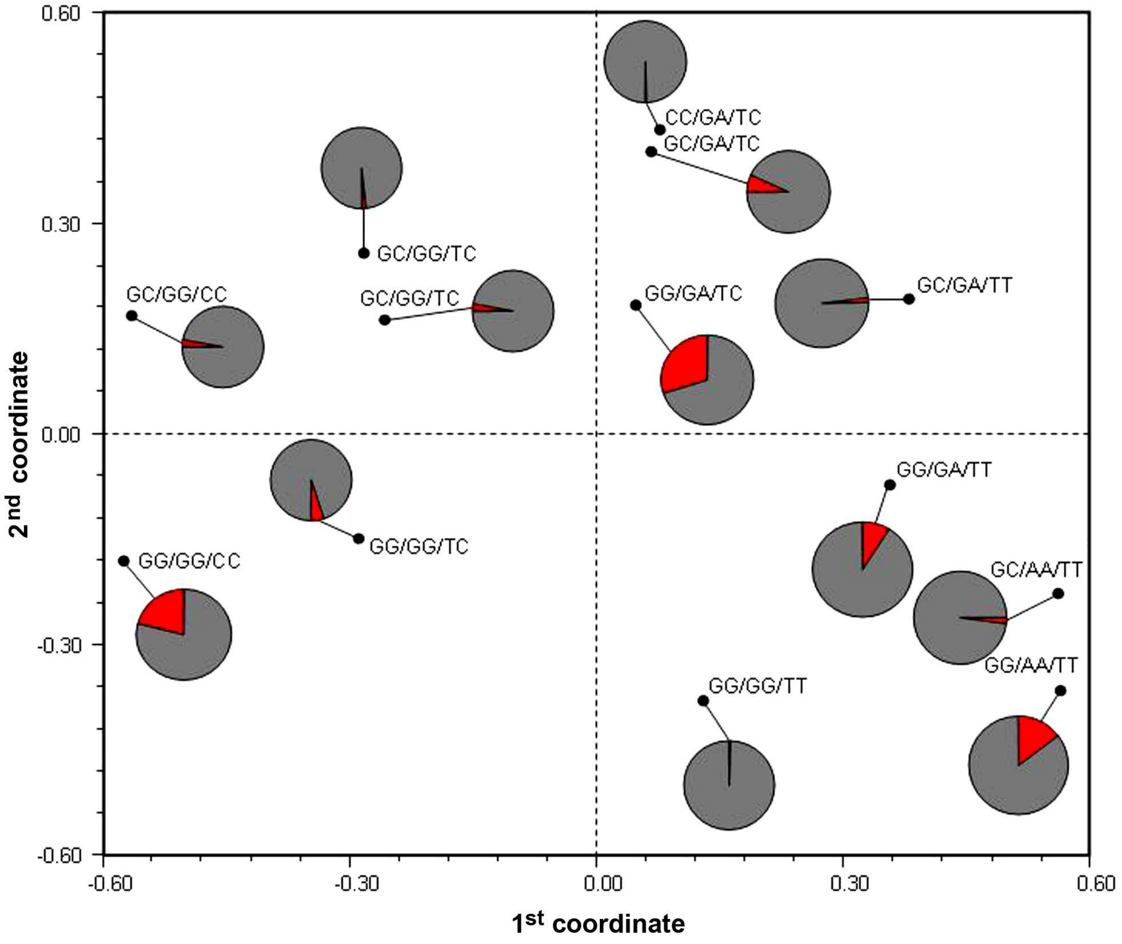


Figure 3.5 UPGMA centroids of animal SNP haplotypes defined according to the first two coordinates: the most numerically represented haplotypes were plotted within distinct quadrants.

It is interesting to note that all haplotypes share the same nucleotide for SNP marker 316, whereas are polymorphic and show both nucleotides for either SNP marker 530 or SNP marker 4751. Combinations GA/TC, GG/CC and AA/TT are therefore responsible for the strong LD found between SNP markers 530 and 4751, being their frequency significantly different from that expected in case of random assortment of nucleotides.

The first three components with Eigenvalues of 26.7, 11.3 and 7.5 were able to explain 90% of the total nucleotide variation found at the SNP sites of the *CAPN1* gene. In particular, the first component, which explains 52.6% of the total variation, was positively associated with nucleotides A and T and negatively associated with nucleotide G and C of markers 530 and 4751, respectively. The second component, which explains 22.4% of the total nucleotide variation, was positively associated with the C and negatively associated with the G of marker 316 (Figure 3.5).

3.4 Discussion

3.4.1 Linkage disequilibrium mapping

In recent years, linkage disequilibrium (LD) has received considerable attention as it may be exploited to effectively map genes underlying both simple and complex traits (Terwilliger and Weiss, 1998; Ron and Weller, 2007). The potential advantage of LD mapping over conventional genetic analysis of linkage performed within families lies in the use of “historical” recombinants, thereby increasing mapping resolution (*e.g*., Hästbacka *et al*., 1992; Talbot *et al*., 1999; Farnir *et al*., 2000).

As the fundamental aim of animal genetics is to connect the genotype to the phenotype, the association mapping strategy based on LD estimates seeks to identify specific functional genic variants such as alleles linked to phenotypic differences in a given trait. Association mapping can therefore facilitates the detection of Mendelian or quantitative trait-causing DNA sequence polymorphisms and/or the selection of genotypes that closely resemble the phenotype (Oraguzie and Wilcox, 2007). As a matter of fact, many Mendelian genes and Polygenes and affecting disorders, diseases and other economically important productive traits in livestock have now been identified (Andersson, 2001; Dekkers and Hospital, 2002). For instance, several mutations that cause simple monogenic disorders in farm animals are catalogued in the OMIA (online Mendelian inheritance in animals) database. Most traits of interest, such as growth, milk yield and meat quality, have a polygenic inheritance being controlled by an unknown number of quantitative trait loci (QTL) and influenced by environmental factors. As a result, there is a considerable interest in finding genetic markers or diagnostic tests that can be applied in breeding programs. Genome scans for detecting QTL have been attempted for several traits in cattle. The confidence interval of individual QTL as determined by linkage mapping analysis often spans tens of map units, containing hundreds of genes. Linkage disequilibrium (LD) mapping can increase the precision of tagging single traits in terms of map units, but the identified chromosome region will still contain tens of genes. To successfully implement genetic selection programs for complex traits, the identification of specific polymorphism(s) associated to the QTL responsible for the observed effect is needed. However, methods suitable for model animals to find out candidate genes (CGs) and validate specific quantitative trait nucleotides (QTNs) underlying quantitative individual phenotypes are now available but they are not always easily applicable to livestock species because of their long generation intervals, the cost of maintaining each animal and the difficulty of producing transgenics or knock-outs (Ron and Weller, 2007). Nevertheless, the most important practical benefit that genomics can have in animal breeding is the discover and use of molecular markers as tools for marker-assisted selection or pre-selection of animals. This is particularly useful to achieve early selection of a trait or a combination of traits.

Liu *et al*. (2001) claimed that SNP haplotype analysis is more efficient than single SNP marker analysis for mapping traits. In fact, simply looking at the marginal dependency between marker alleles and antagonist phenotypes may be inefficient. For a LD mapping strategy to be optimal in fine genetic mapping, it is essential to consider the information observed in a set of contiguous, tightly linked markers (*i.e.*, haplotypes). Linkage mapping should be carried out in a structured pedigree to obtain LD between putative QTL and SNP markers. The optimal experimental designs, with respect to maximal statistical power and minimal number of individuals genotyped, are based on BC_1 or F_2 populations generated from crosses between inbred lines. If QTL-detection experiments are based on crosses between different breeds, it can generally be assumed that all individuals from one breed inherited the same QTL allele. Although livestock breeds are not isogenic, differences between the breeds will generally be large relative to differences within breeds. In these designs, all individuals will be informative for marker genotype, and the QTL contrast will be maximal (Ron and Weller, 2007).

Moreover, to be effective, QTL mapping by means of LD requires a marker density compatible with the distances across which LD extends in the population of interest. Kruglyak (1999) estimated by simulation that useful levels of LD were unlikely to extend beyond an average distance of a few kb in the human. Several studies have later shown that the distribution of meiotic crossover events, which contribute to the breakdown of LD, is distinctly non-random in humans (Cardon and Abecasis, 2003). As a result, extended segments containing a few haplotypes and characterized by strong LD are interspersed with recombination hot spots with little LD (reviewed by Rafalski and Morgante, 2003). For example, Jeffreys *et al*. (2001) found several 1–2 kb long hot spots separated by 20–100 kb long sequences showing high levels of LD. On average, the length of high LD haplotypes and chromosome blocks is known to range from 5 to 60 kb in human (Reich *et al*., 2001). Although large sets of experimental LD data have been accumulated in the human and some primate species, little is known about the extent of LD in most of the other mammals, including domestic species.

In cattle, and extensive genome-wide LD analysis revealed surprisingly high levels of LD that, although more pronounced for closely linked marker pairs (<10 cM), extended over several tens of centiMorgans (Farnir *et al*., 2000; see also Rafalski and Morgante, 2003). A considerable drop in D′ values was however observed between markers associated 1 to 5 cM apart, suggesting that it should be possible to achieve genetic resolution down to the centiMorgan level (Farnir *et al*., 2000). Overall data indicated that LD mapping has the potential to be very effective in livestock populations not only to select markers tightly linked to important genes useful for selecting desirable genotypes, but also to discover candidate genes through positional cloning strategies.

3.4.2 Marker-assisted breeding

Marker-assisted selection has great potential to improve traits for which selection has been historically difficult, such as meat tenderness. A reliably tender meat product is one of the most important attributes for maintaining consumer satisfaction with beef steaks. Selection programs for genetic improvement in tenderness might be implemented in cattle if candidate genes and appropriate markers could be found. Calpain 1 (*CAPN1*), a gene encoding the enzyme μ-calpain which degrades myofibrillar proteins post-mortem, is thought to be one of the most promising candidates for beef tenderness.

The current challenge is to find out gene markers for such quantitative trait that will be useful in many populations, especially those with the greatest opportunity for improvement through marker-assisted breeding programs. The majority of DNA markers developed for quantitative traits in the next few years is not likely to be the causative nucleotide variation because identification and proof of nucleotide substitutions with low to moderate effects is extremely difficult (White *et al*., 2005). However, markers that effectively track functional alleles in a reliable way have substantial value because they permit relatively accurate assessment of genetic merit at a locus based solely on the genotype of the individual animal, without the need for extensive testing of pedigree material to determine the presence of variation and phase with respect to marker alleles.

Examples of this type of marker are the SNP markers 316, 530 and 4751 previously shown to have predictive merit in breeding populations of cattle (Page *et al*., 2004; White *et al*., 2005; Rincòn and Medrano; 2006). Neither marker is likely to represent the nucleotide difference causing the influence on the trait, but the data so far indicate that they are useful for tracking functional alleles with respect to tenderness in major *Bos taurus* beef cattle breeds. Cattle of *Bos indicus* descent are widely used for their heat tolerance and disease resistance, but tenderness has been problematic in many of the animals (Crouse *et al*., 1989). Because previously released SNP markers 316 and 530 for the *CAPN1* gene are almost fixed in *Bos indicus* cattle (Casas *et al*., 2005), one of initial goals of researchers was to provide markers segregating in *Bos indicus* cattle associated with effects on meat tenderness. White *et al*. (2005) focused on the development of an additional SNP marker 4751 for the *CAPN1* gene and the assessment of its predictive merit in populations of *Bos indicus*, *Bos taurus*, or *Bos indicus* × *Bos taurus* crossbred cattle.

Overall results suggest that a high sequence variation can occur at the *CAPN1* gene because of the non-synonymous distinct point mutations so far discovered. Haplotypes and combinations of SNP markers showing a significant effect on meat tenderness have been documented and the genetic information acquired is likely to be useful in making selection decisions for such a trait in cattle.

In this study, three distinct SNP markers in the *CAPN1* gene were investigated in a population of cattle to determine their frequency and the frequency of their haplotypes. A total of 170 animals were subjected to genotyping for the *CAPN1* gene-associated SNP markers 316 (G>C), 530 (G>A) and 4751 (T>C). The observed proportions of single SNP markers were not significantly deviating from proportions expected in case of Hardy-Weinberg equilibrium. However, an excess of heterozygous animals was found for SNP markers 316 and 530, whereas an excess of homozygous animals was found for the SNP marker 4751. The pair-wise comparisons between SNP markers 316 and 530, and 316 and 4751 revealed linkage equilibrium in the population, whereas a highly significant linkage disequilibrium was scored between SNP markers 530 and 4751. If is worth mentioning that a marked excess was found for three haplotypes, *i.e.* GG/AA/TT, GG/GG/CC, and GG/GA/TC, while deficiency was found for most of the remaining haplotypes, particularly GG/GA/CC, GG/GG/TT and GG/AA/TC.

For association mapping purposes additional animals have to be investigated to avoid skewed distributions of SNP markers and possible haplotypes. However, if it is true that SNP marker frequencies and haplotype proportions could have been biased by the relatively small population size, it is also true that skewed distributions of animals among the different haplotypes were expected to be randomly distributed over the three pair-wise SNP site comparisons.

In conclusion, ARMS-PCR and PCR-RFLP techniques set for the development and assessment of SNP markers in the cattle *CAPN1* gene and information on their frequency, along with the availability of preliminary data on SNP haplotypes can be very useful to plan experiments aimed at testing the relationship between a candidate gene, like *CAPN1*, and a number of beef tenderness parameters, such as shear force and driploss.

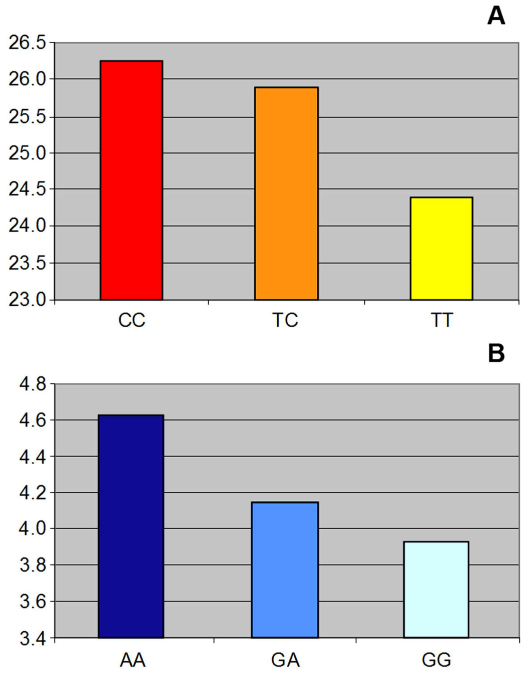


Figure 3.6 Histograms of shear force (A) and driploss (B) estimates of animals grouped according to genotypes of SNP markers 4751 and 530, respectively.

Although appropriate and conclusive statistical analyses were not performed because of the small sample size, preliminary observations suggest that individual SNP markers may have influence on the trait and may be useful for tracking functional alleles with respect to tenderness in the analyzed cattle population, as already reported by other authors (Page *et al.*, 2004; White *et al*. 2005). **Figure 3.6** reports histograms of shear force and driploss estimates of animals grouped according to genotypes scored for SNP markers 4751 and 530.

On the basis of the rather small *Bos taurus* breed differences in tenderness compared with the large breed differences in allele frequency for *CAPN1*, it is also likely that there may be other genes awaiting to be identify. The utility of cloning genes responsible for tenderness and finding markers exploitable in marker-assisted breeding programs mainly will depend on future methods of payment that farmers receive for guaranteed tender beef, compared with cattle consigned to slaughter with no guarantee of tenderness.

3.5 References

Andersson L. (2001). Genetic dissection of phenotypic diversity in farm animals. Nature Genetics Reviews, 2: 130–139.

Cardon L.R. and Abecasis G.R. (2003). Using haplotype blocks to map human complex loci. Trends Genetics, 19: 135–140.

Dekkers J.C.M. and Hospital F. (2002). The use of molecular genetics in the improvement of agricultural populations. Nature Genetics Reviews, 3: 22–32.

Farnir F, Coppieters W., Arranz J.C., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Mni M., Nezer C., Simon P., Vanmanshoven P., Wagenaar D. and Georges M. (2000). Extensive genome-wide linkage disequilibrium in cattle. Genome Research, 10: 220–227.

Fisher R.A. (1918). The correlations between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh, 52: 399–433.

Geesink G.H., M. Koohmaraie (1999). Effect of calpastatin on degradation of myofibrillar proteins by µ-calpain under postmortem conditions. J. Animal Science, 77: 2685–2692.

Haegeman A., J.L. Williams, A. Law, A. Van Zeveren, L.J. Peelman (2003). Mapping and SNP analysis of bovine candidate genes for meat and carcass quality. Animal Genetics, 34: 349–353.

Hedrick, P.W. (2000). Genetics of populations. Jones and Bartlett Publ., London.

Lewis P.O. and Zaykin D. (1999). GDA Version d12. University of New Mexico, Albuquerque, NM.

Lewontin R.C. 1972. The Genetic Basis of Evolutionary Change. Columbia University Press.

Liu J.S., Sabatti C., Teng J., Keats B.J.B. and Risch N. (2001). Bayesian analysis of haplotypes for linkage disequilibrium mapping. Genome Research, 11: 1716–1724.

Kimura M. and Crow J.F. (1964). The number of alleles that can be maintained in a finite population. Genetics 49: 725–738.

Koohmaraie M. (1992). The role of Ca2+-dependent proteases (calpains) in postmortem proteolysis and meat tenderness. Biochimie, 74: 239–245.

Koohmaraie M. (1994). Muscle proteinases and meat aging. Meat Science, 36: 93-104.

Koohmaraie M. (1996). Biochemical factors regulating the toughening and tenderization process of meat. Meat Science, 43: S193–S201.

Nei M. (1973). Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA, 70: 3321–3323.

Nei M. (1978). Estimation of average heterozigosity and genetic distance from a small number of individuals. Genetics, 89: 583–590.

Oraguzie N.C. and Wilcox P. (2007). An overview of association mapping. In: Association mapping in plants (Edited by N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner and H.N. De Silva). Spinger, New York. pp. 1–9.

Page B.T., Casas E., Heaton M.P., Cullen N.G., Hyndman D.L., Morris C.A., Crawford A.M., Wheeler T.L., Koohmaraie M., Keele J.W., Smith T.P.L. (2002). Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. J. Animal Science, 80: 3077–3085.

Reich, D.E. et al. (2001). Linkage disequilibrium in the human genome. Nature, 411: 199–204.

Rincòn G. and J.F. Medrano (2003). Single nucleotide polymorphism genotyping of bovine milk protein genes using the tetra-primer ARMS-PCR. J. Animal Breeding Genetics, 120: 331–337.

Rincòn G. and J.F. Medrano (2006). Assays for genotyping single nucleotide polymorphisms in the bovine CAPN1 gene. Animal Genetics, 37: 293–307.

Rohlf F.J. (1993). NTSYS-pc Numerical Taxonomy and Multivariate Analysis System. State University of New York, Stony Brook, NY, Version 1.8.

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

Smith T.P.L., E. Casas, C.E. Rexroad III, S.M. Kappes, J.W. Keele (2000). Bovine CAPN1 maps to a region of BTA29 containing a quantitative trait locus for meat tenderness. J. Animal Science, 78: 2589–2594.

Sneath P.H.A and Sokal R.R. (1973). Numerical taxonomy. W.H. Freeman and Company, San Francisco, CA.

Vignal A., Milan D., San Cristobal M., Eggen A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution, 34: 275–305.

Weir B.S. (1996). Genetic Data Analysis, II. Sinauer, Sunderland, MA (USA).

White S.N., Casas E., Wheeler T.L., Shackelford S.D., Koohmaraie M., Riley D.G., Chase C.C., Johnson D.D., Keele J.W. and Smith T.P.L. (2005). A new single nucleotide polymorphism in CAPN1 extends the current tenderness marker test to include cattle of Bos indicus, Bos taurus, and crossbred descent. J. Animal Science, 83: 2001–2008.

Wright S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution, 19: 395–420.

Wright S. (1978). Evolution and the Genetics of Populations. Variability within and among natural populations. University of Chicago Press, USA.

Ye S., Dhillon S., Ke X., Collins A.R. and Day I.N.M. (2001). An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Research, 29: 17–88.

Yeh F.C., Yang R.C., Boyle T. (1997). POPGENE. CIFOR and University of Alberta, Canada, Version1.21.

Chapter 4

Discovering of novel leptin gene-based SNPs and sequence-tagged site genotyping of dairy sheep (*Ovis aries* L.) populations

Abstract: The improvement of molecular marker technologies and the implementation of bioinformatics and statistical tools allow the detection of QTL for economically important traits in livestock, including sheep. This information is crucial not only for performing marker-assisted selection programs, but also for discovering and testing candidate genes on the basis of combined genetical, biochemical and physiological data. Individual SNPs within a candidate gene can then be molecularly assayed and statistically tested for association with the quantitative phenotype of interest in order to identify the causative variants. Leptin is considered important in animal production and reproduction in livestock species. The gene encodes for a 16 kDa protein hormone that plays a key role in regulating energy intake and energy expenditure, including important physiological functions, such as the body weight regulation, reproduction, bone formation and growth as well as immunological functions. This study deals with the cloning and analysis of the sheep leptin gene in order to reconstruct its molecular structure and to discover its polymorphism content. Genomic DNA samples of four sires deriving from an extreme mating Awassi × Merino were used as template with the aim of amplifying the sheep leptin gene using primers designed on conserved regions of homologous genes. A sequence of the sheep leptin gene as long as 4,883 bp was recovered, including 714 bp of the promoter, 1,189 bp of exon 1 and intron 1 regions, 539 bp of exon II, 1,262 bp of intron II, 570 bp of exon III and 609 bp of the 3′-UTR. Multiple sequence analysis allowed to discover two SNPs in exon III at position 170 (G>A) and 332 (G>A). Both polymorphisms determined a missense mutation (Arg>Gly). The SNP G(170)A was adopted for the genotyping of animals by means of PCR-RFLP markers. A number of 255 animals belonging to the backcross and the double backcross Awassi × Merino progenies were analyzed in order to evaluate the frequency of the SNP in single families and in the population as a whole. The leptin gene-related SNP marker discovered in this study along with the PCR-RFLP protocol set for individual DNA genotyping will be useful for association mapping purposes and discovering quantitative trait nucleotides correlated to milk yield and quality components in sheep.

Keywords: leptin gene, SNPs, PCR-RFLP markers, association mapping

4.1 Introduction

4.1.1 Lactation curve and persistency

A novel and simple definition for persistency of lactation yield was proposed by Grossman *et al*. (1999). Taking into account the milk production on the time after calving, it is possible to obtain a typical lactation curve. This curve can be described as increasing from initial yield at calving to peak yield, maintaining peak yield and decreasing from it to the end of lactation (Grossman *et al*., 1999). An animal showing a flatter lactation curve is more persistent than another animal with the same total yield but with a curve that decreases rapidly after the peak. In other words, individuals are persistent if able to maintain their peak yield within the lactation period (Cupps, 1966). Animals that yield a flatter lactation curve (*i.e.* more persistent) are characterized by more constant nutrient requirements than animals that are characterized by a less persistent lactation curve (Gengler, 1995). The same lactation trend allows animals to consume adequate amounts of nutrient during first steps just after calving resulting in a positive energy balance. Animals with high peak milk yields are more subjected to metabolic stress during the first three weeks of lactation. Moreover, persistent animals are less likely to experience metabolic stress as the energy requirements are constant throughout the lactation. Other advantages of animals with a constant milk production are their higher reproduction efficiency and more desirable health status (Dekkers *et al*., 1998).

High yielding cows have been selected to produce more milk in large part through their ability to mobilize fat and muscle to support milk production in early lactation (Wathes *et al*., 2007). This results in a loss of body condition and is associated with alterations in blood metabolite and hormone profiles which in turn influence fertility (Pryce *et al*., 2001; Taylor *et al*., 2003). In late gestation and early lactation the nutrient requirements for foetal growth and milk synthesis increase dramatically and the cow is unable to meet these energetic demands from her feed intake. Most cows therefore enter a period of negative energy balance (NEB) in the postpartum period from which it may take them many weeks to recover (Taylor *et al*., 2003). NEB is thus a metabolic disorder affecting high yielding cows that can impair health and decrease fertility. The length and depth of NEB vary according to the genetic merit, pre-calving body condition, milk yield, feed-intake and diet (Bonczek *et al*. 1988; Garnsworthy *et al*., 1982; Grummer 1995).

During the transition period (*i.e*. three weeks before parturition and nine weeks later), marked alterations in the general partitioning of nutrients and metabolism of the whole animal occur to accommodate energetic demands (*i.e*. glucose, amino acids and fatty acids) of the mammary gland. Maternal tissues have adapted to meet foetal needs during pregnancy but these adaptations get more pronounced in support of lactation (Bauman *et al*., 1980). During lactation there are several metabolic processes in which the body energetic reserves are mobilized and simple molecules mainly deriving from lipids utilization are free in blood circulation. All metabolic changes take place in different organs of animals, such as mammary, adipose tissue, liver, muscle, kidney. Consequently, the lactation is not just a function of mammary gland but involves all vital organs of the animal allowing the regulation of nutrient partitioning that is important for milk production. Moreover, the main energetic district is the adipose tissue in which synthesis of storage lipids decreases and lipid reserves are mobilized (Bauman *et al.*, 1980; Clark *et al.*, 2005; Lake *et al.*, 2007). A variety of small and simple molecules are generated from fatty acid degradation. For example, non-esterified fatty acids (NEFA) can be oxidized to carbon dioxide for producing energy or partially oxidized to provide ketone bodies or acetate which are transported throughout the body for their utilization. NEFA can undergo an esterification turning into triacylglycerols (TAG) that accumulates in the liver, where reach the highest concentration at 7-13 days after calving and then gradually decreasing (Bauman *et al.*, 1980; Grummer, 1995). Another example is the beta-hydroxybutyrate (BHB) that represents the predominant form of circulating ketone body and its concentration is an index of fatty acid oxidation (Wathes *et al*., 2007).

Besides molecules deriving from fatty acid metabolism, several metabolic hormones alter in concentration over the critical peripartum period. Lucy *et al*. (2001) described that independent changes occur in the GH-insulin-IGF-I-glucose signalling pathway during this period. In particular, IGF-I seems to be the main mediator of the growth hormone on milk production, regulating milk synthesis by the mammary gland (Etherton *et al*., 1998).

It is well known that reproduction, characterized in mammals by the ability to produce viable gametes and to support pregnancy and lactation, is sensitive to changes in energy reserves (Tena-Sempere, 2007). Situations of extended energy deficit, such as starvation or extreme physical exercise, being invariantly coupled to impaired reproductive function. The physiologic base that plays a role on energy balance regulation is a phenomenon that involves multiple common regulatory signals acting at different levels of the reproductive system (Tena-Sempere, 2007).

The equation of energy balance is defined by the equilibrium between food/feed intake and energy expenditure (Casanueva *et al*., 1999). In order to keep constant body energy stores, a state of energy abundance (*e.g*., by excess of food/feed intake) should activate a series of homeostatic events leading to maintenance of energy balance for example, decreasing in food consumption and/or increasing in energy expenditure. In other words, the energy balance in mammalians is controlled by a feedback loop where the amount of stored energy is sensed by the hypothalamus, which in turns adjusts food intake and energy expenditure to maintain a constant body weight (Brobeck *et al*., 1948; Kennedy *et al*., 1953). Many experimental evidences suggest that this phenomenon is regulated by endocrine factors in which hormones and neuropeptides are involved. In particular, one of the key molecules that is involved in the regulation of energy balance is the adipocyte-derived hormone leptin (Friedman *et al.*, 1998; Rosenbaum *et al*., 1998; Casanueva *et al*., 1999; Ahima *et al*., 2000). Plasma leptin levels are strongly correlated with body condition score (BCS) and decrease in late pregnancy (Ehrhardt *et al*., 2000; Wathes *et al*., 2007a). Furthermore, leptin concentrations are still low during the postpartum period even when energy balance has improved and it seems to influence voluntary feed intake (Ingvartsen *et al*., 2000; Wathes *et al*., 2007a;). In multiparous cows, levels of leptin increase just before calving and they seem to be strong predictors of a delayed first ovulation and associated with prolonged intervals to first service and to conception (Wathes *et al*., 2007). Since at this prepartum stage, the leptin concentration is correlated with the BCS (Ehrhardt *et al*., 2000; Wathes *et al*., 2007a), it can be indicative of the amount of adipose tissue available for subsequent mobilization to support lactation (Wathes *et al*., 2007). Likely, the elevated prepartum leptin can reduce appetite and therefore contribute to greater BCS loss via reduced dry matter intake (Kadokawa *et al*., 2006).

During early lactation much attention should be taken for monitoring the energy balance with the aim to prevent becoming negative (Jack *et al*., 2006). This is mainly important for sheep since they show a high genetic variability that leads large variability in milk yield (Cannas *et al*., 2002). The only practical on farm indicators for assessing NEB in dairy sheep, in order to prevent the several associated disorders, are changes in body condition score (BCS) and in live weight (LW). These indicators do not seem sufficiently accurate for the identification of NEB in the short-period. Among the metabolites that might be appropriate to assay the energy condition of animals during the first weeks of lactation, it is possible to include leptin, non-esterified fatty acids (NEFA) and the beta-hydroxybutyric acid (BHBA). Milk yield and its composition (protein, lipids, lactose) can also be used to monitor EB in dairy sheep (Jack *et al*., 2006). Studying a population of multiparos Awassi-Merino cross ewes, milked for 100 days, Jack *et al*. (2006) found that NEFA and BHBA are correlated and have similar patterns throughout the experimental period. Instead, plasma circulating leptin proved to be negatively correlated to EB. According to Jack *et al*. (2006), milk yield, lactose, protein concentration, BCS and LW are the best indicators for monitoring the EB variance in sheep. Moreover, Clark *et al*. (2005) demonstrated that milk acetone is an accurate and practical indicator for estimating EB of cows in early lactation. Besides, they found that plasma glucose and plasma BHBA together explain a large proportion of the variation in EB. Despite this, Clark *et al*. (2006) established that milk acetone concentration may be useful in experimental situations and not as a practical indicator of EB due to its variability over time.

4.1.2 Leptin: genetic and physiological aspects

Until 1994 the molecular pathogenesis of the commonest nutritional disorder in modern societies, obesity, was unknown. Before this data five single gene mutations in mouse that resulted in an obese phenotype have been described (Friedman *et al*., 1992). Already in 1950 a first recessive obesity mutation, the obese mutation (*ob*), was identified by Ingails *et al*. (1950). *Ob* is a single gene mutation that results in severe obesity and type II diabetes as part of syndrome that resembles morbid obesity in humans (Friedman *et al*., 1991). Walther *et al*. (1991) analyzed populations of mutant and wild-type mice, finding that *ob* mice were deficient for a blood factor that regulates nutrient intake and metabolism, but the nature of this putative factor was still unknown. The milestone of the discovery of *ob* gene was the experiment run by Zhang *et al*. (1994) which led to the identification the *ob* gene by positional cloning. *ob*/*ob* mice were severely obese due to a non-sense mutation at codon 105 in the *ob* gene that resulted in the lack of secretion of functional leptin protein, the product encoded by this gene.

Zhang *et al*. (1994) reported the sequence of mouse *ob* gene and its human homologue. The *ob* gene encodes a 4.5 kb messenger RNA and a 16 kDa protein with a highly conserved 167, in mouse, and 166, in human, amino acid residues. The two primary structures differ in the presence vs. absence of a glutamine residue at position +49 of protein chain (Zhang *et al*., 1994; Isse *et al*., 1995). These variants come from alternative splicing of mRNA, transcribed from the unique leptin gene as demonstrated by Isse *et al*. (1995) and Oberkofler *et al*. (1997). The human leptin gene is made of three exons. The coding region is contained in exons 2 and 3, whereas exon 1 is not translated. The initial region of exon 2 encodes a signalling peptide of 21 amino acid residues that is not represented in the mature protein. Leptin revealed 67% sequence identity among vertebrates, such as human, gorilla, chimpanzee, orangutan, dog, cow, pig, rat and mouse, suggesting a highly conserved function (Zhang *et al*., 1997).

Leptin is expressed predominantly in adipose tissue and is thought to act as a satiety signal in a feedback mechanism involving a target receptor in the hypothalamus (Zhang *et al*., 1994; Masuzaki *et al*., 1995). Moreover, leptin is a cytokine-like hormone that is also secreted by other tissues as placenta (Green *et al*., 1995; Hassink *et al*., 1997; Masuzaki *et al*., 1997; Senaris *et al*., 1997), stomach (Bado *et al*., 1998), mammary epithelium (Chelikani *et al*., 2003; Smith-Kirwin *et al*., 1998), skeletal muscle (Wang *et al*., 1998; 1999), brain and pituitary (Morash *et al*., 1999., Wiesner *et al*., 1999; Jin *et al*., 2000).

Leptin has several important physiological functions, such as the regulation of body weight (Pelleymounter *et al*., 1995), reproduction (Henson and Castracane, 2003), immune functions (Lord *et al*., 1998), bone formation and growth (Steppan *et al*., 2000; Cornish *et al*., 2002; Harmick *et al*., 2004). Besides its association with obesity, leptin has drawn much attention in biomedical research for other health disorders such as non-insulin-dependent diabetes mellitus, cardiovascular disorders and hypertension (Peelman *et al*., 2004; Rahmouni and Haynes, 2004). Leptin turns out to be important in animal production and reproduction. In livestock, the identification of genetic markers that are positively associated with economically important traits has the potential to be used in breeding programs to significantly alter the rate of genetic improvement (Van der Lende *et al*., 2005). Traits that can be improved by this approach are feed intake, feed efficiency, energy balance, fast lean growth, fertility and reproductive efficiency. Their employment can be exploited for the profitability of milk and meat production enterprise (Hossner, 1998).

Comparing the available information regarding the leptin gene in livestock species, the bovine leptin gene has been extensively studied at the molecular level. **Figure 4.1** shows the position and name of the main SNP markers discovered in the bovine leptin gene.

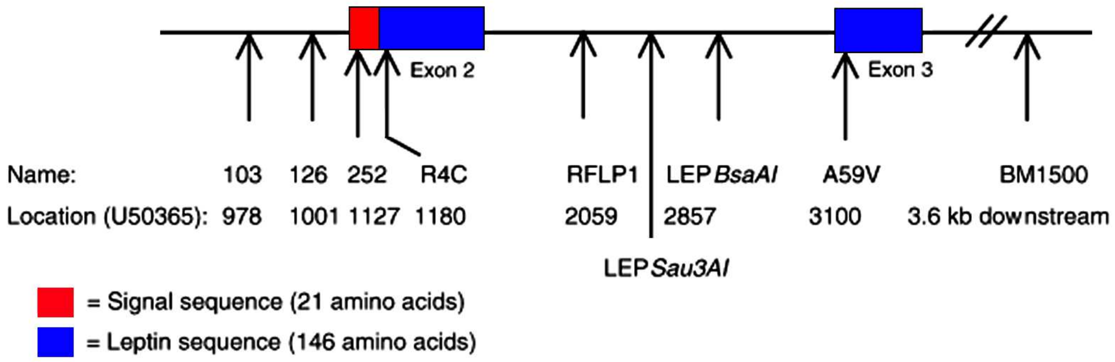


Figure 4.1 Organization of the bovine leptin gene with names and sites of the polymorphisms used in genetic association studies. Location numbers refer to Genbank accession no. U50365 (Liefers *et al*., 2005).

The leptin gene was mapped to bovine chromosome 4 (Stone *et al*., 1996). Up to now numerous polymorphisms have been described in the bovine leptin gene (Pomp *et al*., 1997; Konfortov *et al*.,1999; Haegeman *et al*., 2000). In particular, Fitzsimmons *et al*. (1998) reported a positive association between a microsatellite marker (*BM1500*;

located approximately 3.6 kb away from the leptin gene) and body fatness in cattle. Polymorphisms have also been reported in the coding regions of the leptin gene that show considerable associations with feed intake (Liefers *et al*., 2002; Lagonigro *et al*., 2003; Oprzadek *et al*., 2003), carcass merit (Buchanan *et al*., 2002; Nkrumah *et al*., 2004), milk quantity and quality (Buchanan *et al*., 2003; Liefers *et al*., 2003a), and serum leptin (Liefers *et al*., 2003b; see also references in Nkrumah *et al*., 2005).

An overview of polymorphisms detected in the coding region of the bovine leptin gene and their employment in association studies is given in **Table 4.1**.

Table 4.1 Summary of polymorphisms detected in bovine leptin gene and its receptor up to date. Location numbers refer to Genebank accession no. U50365 (*****Polymorphisms used in association studies. †Mutation results in amino acid change. ††Polymorphism changes transcriptions-factor affinity at promoter region level).

	No. SNPs	SNP ID	References
Intron 1	4	46 (C/T), 103***** (T/C), 126***** (C/G), 143 (C/T)	Konfortov *et al*. (1999); Liefers *et al*. (2003b).
EXON 2	2	252*****† (A/T), 305*****† (C/T)	Konfortov *et al*. (1999); Lagonigro *et al*. (2003).
Intron 2	13	536 (C/T), 538 (C/A), 600 (A/G), 644 (G/C), 726 (C/T), 744 (C/T), 852 (T/C), 860 (C/T), 867 (G/A), 964 (G/A), 1185***** (C/T),1560 (C/T), 1620***** (G/A)	Konfortov *et al*. (1999); Lien *et al*. (1997).
EXON 3	5	140*****† (C/T), 297 (C/T), 300 (T/C), 312 (T/C), 396 (C/T)	Konfortov *et al*. (1999).
Promoter	25	207***** (C/T), 528***** (C/T), 1759***** (C/G), -1457***** (A/G), -1452 (A/G), -1446 (T/C), -1392 (G/A), -1255 (AG/del), -1198 (G del), -1066 (T/A), -963***** (C/T), -901 (A/T), -578* (C/G), -498 (C del), -483 (G del), -415 (G del), -292 (T/C), -282***** (G/T), -272 (G/A), -211 (A/G), -201 (C/T), -197***** (A/C), -170 (C/T), -147***** (C/T), -105*****†† (C/G)	Taniguchi *et al*. (2002); Nkrumah *et al*. (2005); Liefers e*t al*. (2005b).
Receptor	1	115* (C/T)	Liefers *et al*. (2004).

Studies in human and other species have shown that polymorphisms in the leptin promoter may be of major importance in genetic association studies (Fukuda. and Iritani 1999; Hoffstedt *et al*., 2002; Nkrumah *et al*., 2005). The bovine promoter region has been sequenced by Taniguchi *et al*. (2002) for a total length of 3,000 bp. Liefer *et al*.

(2005b) investigated a region 1,600 bp in length located in the 5′-end flanking promoter region using a different source population (Holstein-Friesian heifers) and discovered a total of 20 SNPs, mainly due to substitutions and only two to deletions. Fourteen of them were extensively analyzed to establish their potential association with some productive and reproductive traits. The detected SNPs revealed genetic association with commencement of luteal activity after calving, change in live weight, feed intake, dry matter intake, energy balance, milk protein percentage, serum leptin levels during late gestation and, in some cases in early gestation (Hoffstedt *et al*., 2002; Fukuda and Iritani 1999; Nkrumah *et al*., 2005). In particular, three gene-specific SNPs located in the promoter of the bovine leptin gene revealed genetic association with serum leptin concentration, growth rate, body weight, feed intake, feeding behaviour, and measures of carcass merit (Nkrumah *et al*., 2005). The mutation at position 1,759 of the promoter affecting some productive traits, detected by Nkrumah *et al*. (2005), was further investigated for its variability by Lisa *et al*. (2007) and Di Stasio *et al*. (2007) using both beef and dairy cattle breeds.

Mutations of the promoter region are important because the resulting polymorphisms can alter the transcription factor binding domains leading to an alteration of the leptin gene transcription. Some of these polymorphisms affect domains for CCAAT/enhancer binding proteins (Gong *et al*., 1996; Hwang *et al*., 1996; Miller *et al*., 1996; Taniguchi *et al*., 2002), cAMP response element-binding proteins (Gong *et al*., 1996), SP1 (Fukuda and Iritani, 1999) and LP1 binding proteins (Mason *et al*., 1998). It is worth mentioning that CCAAT/enhancer binding protein is an important factor for the transcription of most genes expressed in adipose tissue and for other genes involved in energy metabolism (Darlington *et al*., 1995).

According to Liefers *et al*. (2005b), the most important regulation sites such as SP1, C/EBP, and TATA are located in the first 200 bp upstream from the transcription start site of the bovine leptin gene that is an evolutionarily conserved region. In cattle the SP1 transcription factor has a putative binding site to nucleotides from 114 to 102 (Taniguchi *et al*., 2002). The effect of the C/G mutation at position 105 on leptin gene expression in cattle was investigated by Adamowicz *et al*. (2006): the C(105)G substitution was found to change SP1-binding affinity and gene expression level in bovine liver. This result is in agreement with another investigation that showed how one point mutation in the SP1 site can be responsible for a decreased expression of the leptin gene in rat liver (Mason *et al*., 1998).

In the pig, most of the polymorphisms of the leptin gene has been studied and discovered within exon and intron regions. Three SNPs [C(867)T, A(2845)T, T(3469)C] out of seven totally identified were found associated with backfat thickness in Duroc breed (Chen *et al*., 2004), feed intake during growing period and genetic merit in Landrace pigs (Kennes *et al*., 2001). Moreover, the SNP T(3469)C has been extensively investigate in association studies. The most important finding is that the C-allele in comparison to the T-allele is responsible for a reduced backfat thickness (Jiang and Gibson, 1999; Urban *et al*., 2002), an increased lean meat percentage (Kulig *et al*., 2001; Urban *et al*., 2002), a reduced intramuscular fat and an increased loin weight (Szydlowski *et al*., 2004). The T-allele was also suggested to be a favourable allele for reproduction (Korwin-Kossakowska *et al*., 2002; Kmieć *et al*., 2003; Chen *et al*., 2004). Recently, Stachowiak *et al*. (2007) investigated a 245 bp long region of the promoter in swine leptin gene discovering four SNPs segregating as two haplotypes. One of identified SNPs [C(113)G] was found in the putative consensus site for the AP2 transcription factor. Nevertheless, any relationship was found between SNP genotypes and level of leptin mRNA in subcutaneous fat and leptin protein concentration in serum. Moreover, no association of the leptin promoter polymorphisms with the analyzed fatness traits was detected evaluating a population made of Duroc breed, Polish Large White, Polish Landrace, Pitrain and a synthetic line (Stachowiak *et al*., 2007).

Information about the leptin gene in sheep is scanty. So far, only Boucher *et al*. (2006) carried out a research in order to detect polymorphisms in the sheep leptin gene. They were able to discover a total of three SNPs in a population of Dorset and Suffolk breeds. Two of these SNPs, namely A(103)G and C(154)T, were detected in the intron 2 and the other one, C(617)G, in the 3′-UTR. In the Suffolk breed, the A(103)G polymorphism revealed association with reduced muscle thickness, loin eye area and with increased shear forces and pH. In particular, this SNP apparently played a negative effect on the muscle growth for Suffolk lambs (Boucher *et al*., 2006).

This chapter deals with the sequencing of the sheep leptin gene using available information for the homolog in human and other livestock species in order to discover single nucleotide polymorphisms in the promoter, exons, introns and the 3′-UTR region. The major goal was that of exploiting the point mutations detected in this study along with the three SNPs identified by other authors (Boucher *et al.*, 2006) for genotyping ewes of back-cross and double back-cross populations deriving from an initial mating between Awassi and Merino. Genotyping data were then used to perform preliminary association studies aimed at estimating potential correlations between SNPs and certain milk traits and feed intake parameters.

4.2 Material and methods

4.2.1 Animal population

The source population employed in this research was obtained at the Centre for Advanced Technologies in Animal Genetics and Reproduction, ReproGen (Faculty of Veterinary Science, University of Sydney, NSW, Australia). It was based on an extreme mating between Awassi fat-tail sheep and Merino superfine and medium wool sheep (Raadsma *et al.*, 1999). This population was developed for the identification of QTL putatively associated with traits of economical importance. The Awassi breed has been extensively selected for its high milk yield, producing around 1,000 litres of milk per lactation (Epstein, 1982). It originates from the Middle East (mainly Israel) and was imported into Australia in 1986. Conversely, Merino breed is good in producing high quality wool and fleece but poor in milk production (Bencini and Dawe, 1998). Because of their different aptitudes, these two breeds represent extreme types for the production traits such as wool and milk. Merino and Awassi breeds also show marked differences in frame size and fat distribution. The former sheep breed has a much smaller frame size than latter, and there is also a major difference in fat distribution between Merino and the fat-tail Awassi (Raadsma *et al.*, 1999).

In detail, the source population was generated by employing four F_0 Awassi sires mated to thirty Merino females. The F_1 Awassi × Merino (AM) sires were crossed with Merino ewes to obtain the (Awassi × Merino) × Merino backcross (AMM) progeny.

From each family of the four F_0 founder sires one sire was mated to 750, 300, 300 and 300 Merino ewes, respectively. In total 519 AMM F_2 progeny of the first family, 201 of the second, 132 of the third and 175 of the fourth family were produced. The ewes of the AMM BC progeny were further mated with the four F_1 sires in order to obtain a (Awassi × Merino) × Merino ×(Awassi × Merino) double backcross (AM_AMM). The number of animals belonging to these last crosses was 256, 233, 121 and 212 for the four AM_AMM_DB families. Animals were reared in the same flock under similar conditions. Detailed information of the structure of the populations can be found in Raadsma *et al*. (1999). A schematic representation of the mating structure is given in **Figure 4.2**.

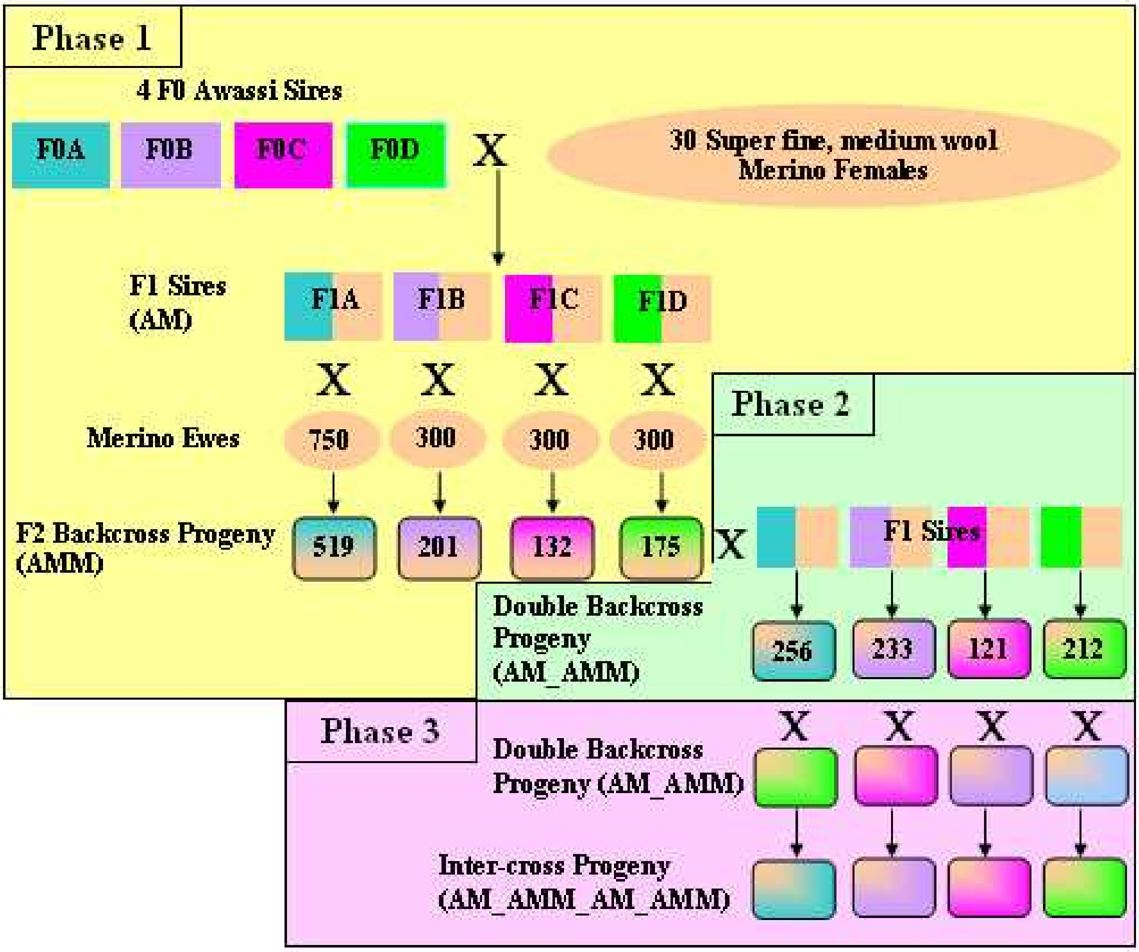


Figure 4.2 Schematic representation of source population analyzed in the present research (Raasdma *et al*., 1999).

The four F_1 sires derived from the above mentioned mating system were used for leptin gene sequencing in order to detect SNPs. A total of 255 ewes from AMM_BC and AM_AMM_DB populations was employed for animal genotyping on the basis of newly discovered SNPs and the three SNPs previously detected by Boucher *et al*. (2006).

The ewes were chosen on the basis of available feed intake and milking data collected at ReproGen during the 2005 and 2006 experimental years. As described by Jack *et al*. (2006), all the analyzed sheep were monitored since the day immediately after parturition which was classified as day 1 of days in milk (DIM). Ewes were milked once a day in 2005 (in the morning) and twice a day (morning and afternoon) in 2006 for a period of 100 days. Some animals were also milked in both years. A milk sample was collected weekly and measured for its fat, protein and lactose content. Ewes were fed a pelleted concentrate diet (ME 12.5 MJ/kg DM and 16% crude protein content) *ad libitum* through automatic feeders that automatically recorded feed intake (VFI) and live weight (LW).

4.2.2 Genomic DNA extraction

The genomic DNA of each animal was extracted from blood following the protocol described by Montgomery and Sise (1990) and modified by Raadsma *et al*. (1999). The purity of all extracted DNA was inspected using Eppendorf BioPhotometer. The purity of the DNA was assessed by calculating the ratio between the 260 nm/280 nm readings. Any sample deviating from the purity criterion (*i.e*., a ratio of 1.8) was re-extracted.

4.2.3 Leptin gene primer design

The sequence information of sheep, human, cattle and pig leptin gene was used for designing primers to be used for the PCR amplification of the sheep leptin gene. All homologous sequences belonging to these species were analyzed by multiple alignment in order to detect the most conserved regions of the leptin gene. In particular, primer sequences related to the promoter region were designed depending on available *Bos taurus* sequences and comparing them to *Sus scrofa*. The intron 1 anchored primer sequences were deduced exploiting the intron 1 region of *Sus scrofa* compared to *Homo* *sapiens* leptin gene, except for two sequences that were obtained by comparing the most conserved regions between *Sus scrofa* and *Bos taurus* sequences. Primer sequences associated to exon 2, intron 2 and exon3, and to the 3′-UTR were designed using *Ovis aries* known sequences and by comparing them to *Bos taurus* available sequences.

On the basis of serial multi-alignments, 16 primer combinations were designed using PerlPrimer and Primer3 software. Primer binding sites were positioned on the leptin gene so that each combination could yield a fragment of 600 bp in length on average. The sequence of designed primers, their melting temperature and expected size of the correspondent PCR products are reported in **Table 4.3**, along with the position of primers compared to the complete bovine leptin gene (>Chr4 from nt 85556158 to nt 85576157) deducted by the Bovine Assembly v. 3.1 Scaffolf Browser (http://genomes.tamu.edu/bovine/).

4.2.4 Optimization of PCR conditions

All primer combinations were tested in preliminary experiments at different annealing temperatures and $MgCl_2$ concentrations using two sheep and one cattle genomic DNA samples as template. The PCR setting up was performed in a volume of 10 or 20 μl containing 50 or 100 ng of genomic DNA, 1× Qbuffer (Qiagen), 200 μM dNTPs mix, 3 pmol of each primer, 1 U Hotstart *Taq* DNA polymerase (Qiagen) and 1.5 mM $MgCl_2$. In order to find out the optimum PCR conditions, different additives such as Qsolution (Qiagen), DMSO and Tween 20 were also used.

The temperature profiles for PCR consisted of an initial activation of the *Taq* DNA polymerase at 95°C for 15 min, followed by 35-40 cycles of denaturation at 95°C for 30 sec, primer pear-specific annealing temperature for 30 sec, extension at 72°C for 1 min and a final step at 72°C for 10 min.

Quality of amplification products were assayed by electrophoresis using 2% agarose gels visualized by staining with ethidium bromide and their size analyzed by Molecular Imager Gel Doc XR System (Biorad). For sequencing purposes, the amount of DNA of each single amplicon was estimated by using the Quantity One software (Biorad).

Table 4.3 Primer sequences for SNP discover in sheep leptin gene along with GeneBank accession numbers used for the design of primer pairs, expected amplicon size (bp), melting temperature of primers (Tm) and comparative bovine leptin gene position.

	ID	GB accession No.	Sequence (5'-3')	Bt gene position	Tm	Exp size
Promoter						
Pr 2 F	#1	AB070368	TGTGAAATGAAACATACGCTGAC	1492	60	607
Pr 2 R			GGTGTATCCTTCCATGGATATTCT	2099	62	
Pr 3 F	#2	AB070368	CCATAAACTGAGAAGAACAACTAC	2098	59	589
Pr 3 R			GTCTGCAACATCTTTGTAACTTG	2687	60	
Pr 4 F	#3	AB070368	TAGGTAGGTGATGGGTAGATAGG	2477	62	599
Pr 4 R			CTATTGGGAGTAGGATCAGGA	3076	60	
Pr5 F	#4	AB070368	CTTAGCCACTGGACCACCAG	2975	64	557
Pr5 R			TCACAAGACCATTACCACAC	3532	60	
Pr 6 F	#5	AB070368	CAACCAGGCTCAAACAAAGCA	3416	64	719
Pr 6 R			GCTGGGCAGGTGTGAGAAAT	4135	65	
					Tot	**3,071**
Intron 1						
Pr-In1 F	#6	AB070368	CGAGGATTTCTCACACCTGC	4111	63	649
Pr-In1 R		AF492499	CCTCAGAAACTCAACGCACA*	4760	63	
In-1 F1	#7	AF492499	CCAAAGGAGCGTCCTCCTAATA	5686	64	557
In-1 R1			AGGCATCAGGTGTGGTGTTT	6243	64	
In-1 F2	#8	AF492499	CCAGTACTAGCAAATCAGCA	10105	60	526
In-1 R2			AGGAGGTGGCATTTGATAAG	10631	60	
In-1 F3	#9	AF492499	ATCAAGACCCAGATGAAACAG*	15111	60	718
In-1 R3			ATAATATGTCAGATGCCGTGC	15829	60	
In-1 F4	#10	AF492499	CTGTAAGCTACTTGAGAGCAG	15753	60	403
In-1 R4			AGGTCTGGTTTGAATCCAG	16156	59	
					Tot	**2,853**
Exon 2						
Ex2 F 1	#11	U50365+DQ496248	TCTTTGAGGAGATGATAGCCA	16297	60,4	540
Ex2 R 1			GACCTTTGTGACTCCCTCTG	16836	62	
					Tot	**540**
Intron 2						
In2a F1	#12	U50365+DQ496248	CAGAGGGAGTCACAAAGGTC	16817	62	639
In2a R1			TCTTCACAAGGTCTTATGGGT	17427	61	
In2b F2	#13	U50365+DQ496248	TCATTTCCTTCTCCCAACGA	17310	61	591
In2b R2			AATCAATCCTGGCACCACAC	17850	61	
In2c F1	#14	U50365+DQ496248	CAGTCTTTCAACAAGCTGTACAC	17784	62	618
In2c R1			GGACATCAAGAGCTCAGTGG	18349	63	
					Tot	**1,848**
Exon 3						
Ex3 F 1	#15	EF534370+U50365	CCACTGAGCTCTTGATGTCC	18330	62,5	572
Ex3 R 1			TATATCTTTGGTTTCCTGCCTC	18902	60	
					Tot	**572**
3'-UTR						
3'Utr F1	#16	EF534370+U50365	TATGTGGGCATCCTTTATGCAG	18800	63	606
3'Utr R1		EF534370+U50365	AAACAACTTGGTCCTTCGAG	19405	61	
					Tot	**606**

4.2.5 Sequencing of PCR amplicons

All reliable and reproducible PCR products were recovered from the gels and purified using EXOSAP (Amersham Pharmacia Biotech). Sequencing reactions were performed with 50-100 ng of purified PCR product as template using the BigDye terminator v3.1 cycle sequencing kit (Applied Biosystems) following the manufacturer' instructions. Samples were purified by ethanol/EDTA/Na acetate precipitation and finally loaded on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems). Each of the amplicons was bi-directionally sequenced using either the forward or the reverse primer.

4.2.6 Sequence analysis and multiple alignments

The sequences obtained were analyzed by Sequencing Analysis v. 5.1.1 software (Applied Biosystems). In order to test if the sequences overlapped part of the leptin gene, each sequence was used as query to search for homologous gene sequences in the NCBI databases by BLASTN application (Altschul *et al*., 1990). The sheep sequences obtained from each F_1 sire were then used for multiple alignments using the SeqScape v. 2.5 software (Applied Biosystems) in order to detect single nucleotide polymorphisms in the sheep leptin gene.

4.3 Results

4.3.1 Setting up of PCR conditions for leptin gene amplification

The optimization of PCR conditions was done testing each of 16 primer combinations at different annealing temperatures, $MgCl_2$ concentrations, number of PCR cycles, reaction volume and different additives.

Primer combinations that generated PCR products of the expected size in sheep were used for the amplification of the four F_1 sires DNA samples, whereas the combinations that generated PCR product only in cattle DNA samples were discarded. The three SNPs identified by Boucher *et al.* (2006) were not found analysing the four sequences used by the authors. Consequently, the two PCR products generated with the two primer combinations designed by Boucher *et al.* (2006) were sequenced. The resulting amplicons were named Fragment 1 (FR1 of 260 bp), that includes part of exon

2 and part of intron 2, and Fragment 2 (FR2 of 926 bp), that covers exon 3 and part of 3′ UTR. The amplification of both target sequences were obtained using the following primers: FR1for 5′-CGCAAGGTCCAGGATGACACC-3′ in combination with FR1rev 5′-GTCTGGGAGGGAGGAGAGTGA-3′; and FR2for 5′-CTCTTGATGTCCCCTTCCTC-3′ in combination with FR2rev 5′-TGGTCCTTCGAGATCCATTC-3′, after optimization of the PCR protocol.

Ten primer combinations generated good PCR products that were then sequenced. Nonetheless, the six primer combinations (#2, #3, #4, #7, #9, and #10) that failed amplification were tested several times using different temperature profiles and additives. Single PCR products were undetectable while multiple PCR products were often visible either in cattle and sheep samples, demonstrating that designed primers were not specific for the target regions. Additional primers covering the same regions were developed but they still provided unreliable results. Most primer combinations needed a $MgCl_2$ concentration equal to 1.5 mM, except for primer combinations #1 and #5 that were employed using a $MgCl_2$ concentration of 1.6 mM. The optimal annealing temperature ranged from 57°C to 64°C for primer combination #1, and #13 and #14, respectively. All amplified fragments were detected in a reaction volume of 20 µl using 35 to 40 PCR cycles without any additive. The only exception was the amplicon #6 for which the reaction was carried out in a volume of 10 µl adding 1× Q solution (Qiagen) to each sample.

4.3.2 Sequencing and SNP detection

The adoption of the 10 selected primer combinations enabled the amplification of one single amplicon for each of the four F_1 sires. Every single PCR-derived fragment was used sequencing both strands using either the forward or the reverse primer. A number of 8 reliable sequences were generated for eight of the ten amplicons. The amplicons generated by primer combinations #1 and #13 produced unclear chromatograms, suggesting that all PCR-derived fragments for these two primer combinations were most likely made up of two amplicons of the same molecular weight and hence co-migrating in the agarose gel.

Overall, a total of 4,883 bp of the leptin gene were sequenced. The whole DNA fragment could be subdivided depending on its covered regions as follow: 714 bp for the promoter, 1,189 bp for exon I and intron I, 539 bp for the exon II, 1,262 bp for the intron II, 570 bp for exon III and 609 bp for the 3′-UTR. **Table 4.4** reports the expected and observed size of amplicons on the basis of the designed primer pairs and sequencing analysis, respectively.

Table 4.4 Features of PCR reactions and amplicons related to the sheep leptin gene: annealing temperature (Ta) of primers, expected and observed size (bp) of fragments.

Primer ID	Ta	Expected size	Observed size
# 1	57	606	?
# 2	-	-	-
# 3	-	-	-
# 4	-	-	-
# 5	58	720	714
# 6	62	649	659
# 7	-	-	-
# 8	60	550	530
# 9	-	-	-
# 10	-	-	-
# 11	60	540	539
# 12	58	639	641
# 13	64	591	?
# 14	64	618	621
# 15	60	572	570
# 16	62	606	609
FR1	63	260	261
FR2	60	926	1,060

On the whole, no appreciable differences were found between the expected and observed fragment size using the newly designed primers. However, substantial differences were found using the FR2 primer combination previously proposed by Boucher *et al*. (2006): the generated amplicon was long 1,060 bp instead of 926 nt, with a discrepancy of 134 bp. More specifically, this fragment is composed by amplicons #15 and #16 detected in this study. Conversely, the fragment generated by FR1 primer combinations was shown to be 260 bp long, as reported by Boucher *et al*. (2006). This amplicon represents part of the fragment #11 analyzed in the present study.

The sequencing and alignment of the PCR products enabled the detection of two SNPs at positions 206 and 368 of the fragment #15 of the F_1 sire #463. These two point mutations, both of the G>A type, were placed at positions 170 and 332 of the exon III of the leptin gene. Using the free EMBOSS Transeq tool of EMBL-EBI (freely available at http://www.ebi.ac.uk/emboss/transeq/), the nucleotide regions corresponding to exon III were translated in order to verify the occurrence of differences in terms of putative amino acid sequences. Both single nucleotide variations were shown to determine a missense mutation in which the presence of G leads to the amino acid Arginine while the presence of A determines the Glycine (*i.e.*, missense mutation).

On the basis of the sequencing of FR1 and FR2 fragments, none of the SNPs A(103)G, C(154)T (in intron II) and C(617)G (in 3'-UTR) described by Boucher *et al*. (2006) was detected in the four sires used in the present study. As a consequence, they were not adopted for the genotyping of animals.

4.3.3 PCR-RFLP development and its employment for SNP genotyping

The two G>A substitutions detected at positions 170 and 332 of the exon III of leptin gene were shown to produce two distinct missense mutations (Arg>Gly) and to introduce two different restriction site variations. The SNP G(170)A was found to eliminate one recognition site of the *Msp*I restriction enzyme (cut bases: C$^{\blacktriangledown}$CG$_{\blacktriangle}$G), whereas the SNP G(332)A was found to generate a recognition site for the *Hpy*CH4V restriction enzyme (cut bases: T$^{\blacktriangledown}$GC$_{\blacktriangle}$A).

On the basis of this knowledge, a PCR-RFLP method was developed for rapidly identifying the two SNPs on a total of 255 ewes. Since the two point mutations were found to be only 162 pb from each other, a situation of linkage disequilibrium was assumed so that the two SNP markers could most likely be inherited together. Consequently, the amplification products obtained using as template the DNA samples belonging to the 255 animals were analyzed only by the *Msp*I restriction enzyme that specifically identify the SNP G(170)A. This polymorphism was chosen because of the easy interpretation of the electrophoretic banding patterns generated by amplified-restricted DNA fragments. A total of 10 µl of the amplified fragment resulting from primer combinations #15 were digested using 5 U of *Msp*I enzyme (BioLab) at 37°C for two hours. The resulting DNA fragments were separated by electrophoresis in a 2% agarose gel. An example of PCR-RFLP profiles is shown in **Figure 4.3**.

The analysis revealed that the population was composed of only two genotypes, GG and AG, for the mutation at position 170 of exon III of the leptin gene, whereas the genotype AA was never found. The GG genotype showed three clearly distinguishable bands, while the AG genotype revealed four bands because of the additional *Msp*I restriction site generated by the A mutation.

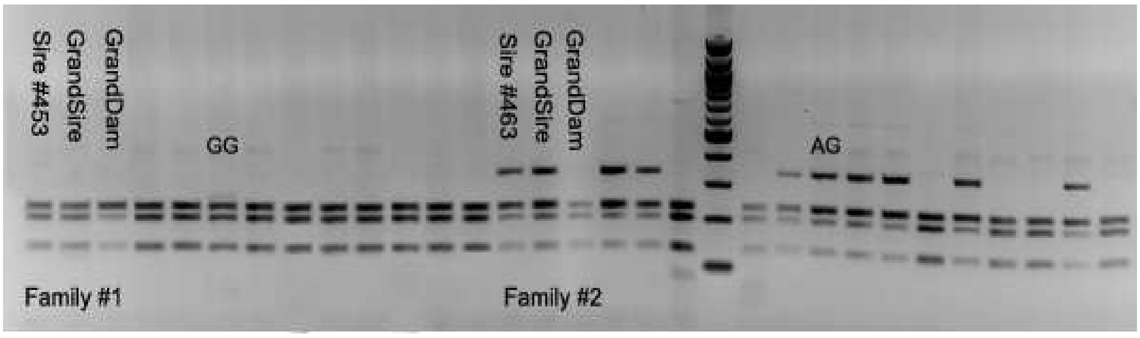


Figure 4.3 Example of PCR-RFLP markers produced by *Msp*I digestion of amplicons #15 generated in some of the 255 animals used in this study (different individuals belonging to the families generated by the two sires #453 and #463 were used).

All animals were classified on the basis of their genotypes at G(170)A mutation site as GG or AG. The frequencies of the two genotypes in the whole sample of 255 animals were equal to 79% for GG and 21% for AG. Investigating the single families, it was possible to score the genotype AG in the family #2 only with a frequency of 51%. The A allele was found to be transmitted to the progeny by the sire #463 where the mutation was initially discovered. The presence of three animals showing the genotypes AG in the family #1 and #4 is most likely attributable to sampling errors. It is important to note that the frequency of 51% for both genotypes (*i.e.*, GG and AG in a 1:1 ratio) would have been expected if the analyzed individuals were made only of AMM backcross animals. Despite, the animals belonging to the family #2 derived both from backcross (41 animals AMM) and double backcross (57 animals AM_AMM) populations. This means that a proportion of AA genotype should be detected in the family #2 because the AM_AMM individuals derived from a cross between AMM individuals and sire #463, both presenting the A allele. On the basis of calculated SNP frequencies, the expected genotype frequencies were measured for the population as a whole and for each family as well (**Table 4.5**).

Table 4.5 Number and frequency of observed and expected genotypes for the G(170)A leptin SNP sorted by family and overall.

Family	**N.**	**Obs. genotypes**			**Genotype freq.**			**Allele freq.**		**Exp. genotypes**			χ^2
		GG	**AG**	**AA**	**GG**	**AG**	**AA**	**G**	**A**	**GG**	**AG**	**AA**	
1 (#453)	103	101	2	0	0.98	0.02	0	0.99	0.01	100.95	2.04	0.01	1.04 n.s.
2 (#463)	98	48	50	0	0.49	0.51	0	0.74	0.26	54.38	37.24	6.38	11.2***
3 (#473)	32	32	0	0	1.00	0.00	0	1.00	0.00	32.00	0.00	0.00	0.00 n.s.
4 (#474)	22	21	1	0	0.95	0.05	0	0.98	0.02	21.01	0.98	0.01	0.54 n.s.
Overall	**255**	**202**	**53**	**0**	**0.79**	**0.21**	**0.00**	**0.90**	**0.10**	**204.8**	**47.5**	**2.8**	**3.43 n.s.**

The observed number of genotypes in the whole sample of 255 animals proved to be not significantly deviating from the number of genotypes expected in case of Hardy-Weinberg equilibrium. Conversely, the differences became significant analyzing the family #2 in which the expected and observed number of genotypes GG, AG and AA were 54,4, 37,2 and 6.4 *versus* 48, 50 and 0, respectively.

4.4 Discussion

In the modern agriculture, sheep play an important role for wool, meat and milk production among the livestock species. Their value is also due to the possession of many heritable traits that are of economic importance. Furthermore, features like size, adaptability, temperament and productive life, make them an appropriate model for studying a variety of biological functions such as immunology, endocrinology, reproduction, embryology and fetal development (Maddox *et al.*, 2001).

4.4.1 QTL mapping in sheep

Sheep are often used as diseases model for inherited diseases, such as asthma (Wright *et al.*, 1999), muscular dystrophy (McGavin, 1974), McArdle's disease (Tan *et al.*, 1997), Batten's disease (Broom *et al.*, 1998; Houweling *et al.*, 2006) and for different infectious diseases. During the last decades, several studies have been successfully performed to develop a sheep linkage map using different populations allowing the construction of low-density autosomal and X chromosome genetic maps (Crawford *et al.*, 1995; Broom *et al.*, 1996; Galloway *et al.*, 1996; Lord *et al.*, 1996; McLaren *et al.*, 1997; Cockett et al., 2001). An additional sheep linkage map has been developed by de Gortari *et al.* (1998) including overall 512 loci, even though centromeric and telomeric regions of some chromosomes proved to be not covered. Maddox *et al.* (2001) developed a medium-density linkage map composed of 1,062 unique loci (941 anonymous plus 121 genes). More recently, mapping studies led to the identification of some QTL associated with milk yield and milk quality traits in dairy sheep (Barillet *et al.*, 2005). Comparative mapping between cattle and sheep was also useful for the identification of a QTL for milk protein content on BTA3/OAR1 (Calvo *et al.*, 2004; Barillet *et al.*, 2005). Moreover, Walling *et al.*, 2004 identified a total of 24 suggestive QTLs by analyzing Suffolk and Texel commercial sheep. They discovered one significant QTL affecting fat depth in chromosome 2 (only 26 cM distal to myostatin) on Texel family. The most significant QTL, which affected muscle depth, was located on chromosome 18 near the callipyge and Carwell loci.

A candidate gene approach has been followed to investigate OAR1 containing a QTL tightly linked to five genes known to be involved in fatty acid metabolism and transport (Calvo *et al.*, 2006). More recently, an extreme breed-cross of Awassi and Merino sheep has been employed for detecting QTL associated with lactation performance for protein yield and lactation persistency on OAR3 and OAR20 using linkage analysis approach. In this study, markers already discovered and many novel markers located in these two chromosomes were exploited for genotyping purposes and for increasing map resolution in order to discover positional candidate genes affecting milk quality and yield (Singh *et al.*, 2007).

In recent years, an increasing number of projects aimed at understanding the genome constitution and gene function in main livestock species have been carried out. The improvement of molecular technology and the implementation of statistical tools allowed the detection of quantitative trait loci (QTL) for economically important traits in livestock (Georges *et al.*, 1995; Andersson 2001; Dekkers and Hospital, 2002). This information is crucial not only for performing marker-assisted selection programs, but also for discovering and testing candidate genes on the basis of genetical, biochemical and physiological data. Individual SNPs within a candidate gene can then be molecularly assayed and statistically tested for association with the phenotype of interest in order to identify causative variants, such as QTNs (quantitative trait nucleotides).

4.4.2 SNP discover in the sheep leptin gene

The leptin hormone has a crucial role in the central nervous system (hypothalamus) in which controls the food intake and energy expenditure. Leptin could also be involved in regulation fat mobilization (Halaas *et al.*, 1995). Moreover, it has been demonstrated that the acute or long-term changes in food composition or food restriction caused changes in plasma leptin in ruminants (Thomas *et al.*, 2001). In particular, complete food deprivation caused a rapid fall in plasma leptin (Marie *et al.*, 2001) and long-term food restriction decreased plasma leptin concentration in sheep (Morrison *et al.*, 2001; Delavaud *et al.*, 2000). Leptin concentration in plasma is affected by the presence of specific binding proteins, variation in nutrition and adiposity, and by changes in physiological stages like pregnancy and lactation (Liefer *et al.*, 2005a). Furthermore, in sheep circulating leptin levels increased and remained elevated starting from early pregnancy through the mid and late pregnancy (Forhead *et al.*, 2002). The level or circulating leptin decreases rapidly towards parturition (Liefers *et al.*, 2003b) and it seems to be due to several mechanisms such as: decline in adiposity, decreased expression of leptin in adipose cells and the decline in insulin concentration (Liefers *et al.*, 2005a). It is possible to declare that this hormone is involved in several correlated function like feed intake, energy balance, lactation, pregnancy, fertility. Because of its importance, many studies have been carried out to discover leptin gene structure (Zhang *et al.*, 1994) and the polymorphism extent of the gene that may explain the cause of the hormone variability during different physiological events (Konfortov *et al.*, 1999).

Most of the literature concerning leptin gene polymorphisms is represented by studies conducted in cattle (Pomp *et al.*, 1997, Lien *et al.*, 1997; Fitzsimmons *et al.*, 1998; Buchanan *et al.*, 2002; Liefers *et al.*, 2002; 2003a/b; Lagonigro *et al.*, 2003). Only one research has been done on the leptin gene in sheep by Boucher *et al.* (2006): they cloned and sequenced a total of 1,186 bp, comprising exon II, part of intron II, exon III and part of 3’-UTR. Along this region three SNPs were detected, namely A(103)G and C(154)T in intron II and C(617)G in 3′-UTR.

In the present study the DNA of four sires deriving from an extreme mating Awassi × Merino was employed to analyze the sheep leptin gene. A total of 4,883 bp of the leptin gene were sequenced and their regions subdivided as follows: 714 bp for the promoter, 1,189 bp spanning exon 1 and intron 1 regions, 539 bp covering exon II, 1,262 bp comprising intron II, 570 bp covering exon III and 609 bp representing part of the 3′-UTR. The most problematic regions for the amplification of the PCR products were the promoter and intron I regions. In particular, it was possible to obtain only two good PCR products for the promoter region (one including also the exon I) and one for the intron I. This finding suggests a low conservation of the 5′ region of the leptin gene. Even during the preliminary step of primer design based on heterologous sequences, these two upstream genic regions showed a high divergence with respect to the leptin gene of human, cattle and pig. The low sequence conservation of the promoter and the intron I among the species considered for primer design might be the cause for the lack of successful amplifications. In some cases, it was possible to get the amplicons #1 and #13 but they did not result in reliable sequences most likely because of multiple PCR products and non-specific amplifications.

The sequencing enabled to discover two SNPs in exon III at position 170 (G>A) and 332 (G>A). Both polymorphisms determined a missense mutation (Arg>Gly). For the genotyping of animals, only the SNP G(170)A was adopted since the SNP G(332)A was found to be in linkage disequilibrium during preliminary experiments performed with a subset of animals. The three SNPs identified by Boucher *et al.* (2006) were not detected in our animal set likely due to the analysis of sires and not of dams. The selected SNP G(170)A was analyzed by PCR-RFLP exploiting the variation of *Msp*I restriction site introduced by the point mutation. A number of 255 animals belonging to the backcross and the double backcross Awassi × Merino progenies were subjected to genotyping in order to evaluate the frequency of the SNP in single families and in the population as a whole.

The detected SNPs allowed two distinct non-conservative amino acid variations to be discovered which might induce protein structural changes. If these variations belong to domains of the protein they would influence the proper recognition between leptin and its receptor, hence influencing its biological role. Another critical point is the absence, in the analyzed population, of animals with genotype AA at G(170)A SNP. The whole population of 255 individuals turn out to be in Hardy-Weimberg (HW) equilibrium. Despite, the lack of AA genotype in the family #2 was significant deviant ($P<0.05$). The absence of HW equilibrium could be explained by taking into account non-random mating and/or selection effect. It is obvious that this cannot be the final consideration because other animals and at least one more SNP need to be analyzed to give a reasonable explanation about the lack of AA genotype.

As future perspectives, statistical analyses should be performed to test the potential association between each single leptin-gene related SNP and a number of quantitative traits represented by the Wood’s parameters curve (Wood 1967), such as persistency. The persistency represents the rate of milk yield decreasing after the lactation peak. The ideal lactation curve has a high peak and a moderately flat trend afterwards. Many authors proved that more persistent lactation is desirable because it is related to better animal health and reduction of feeding costs (Sölkner and Fuchs, 1987; Pryce *et al.*, 1997; Dekkers *et al.*, 1998; Grossman *et al.*, 1999; Pulina *et al.*, 2007). Moreover, persistency was shown to have a certain degree of genetic variation, with moderate heritability (0.15 - 0.20), being selection for this trait feasible and worth (Muir 2004; Macciotta *et al.*, 2006).

Some measures of persistency were found to be highly correlated with total lactation yield, even though some authors declare that a robust measure of persistency should be independent from total yield (*i.e.*, real persistency) (Gengler, 1996) or that the total lactation yield should be included as a (co)variate in the genetic model used to estimate genetic parameters and breeding values for lactation persistency (Swalve, 1995). Up to date, SNP association to lactation persistency has not been investigated yet. Interesting researches aimed mainly at identifying QTL associated with milk yield and quality. Two of these were carried out by Barillet *et al*. (2005) and Singh *et al*. (2007) in two different sheep populations. In particular, QTLs associated to fat content and milk yield were mapped on chromosomes OAR3, OAR9 and OAR20 in European sheep cross-breeds (Barillet *et al*., 2005), whereas additional QTLs located on chromosomes OAR3 and OAR20 proved to be associated with lactation persistency and protein yield by high resolution mapping approach (Singh *et al*., 2007). None of these QTLs is attributable to chromosome OAR4 where the leptin gene is known to be located (Pomp *et al*., 1997). It is worth mentioning that the population used by Singh *et al*. (2007) derived from an extreme breed backcross and double backcross between Awassi and Merino sheep (Raadsma *et al*., 1999). Since this population is the same used in the present study, most likely polymorphisms of the leptin gene alone cannot be taken into account to explain the variation for persistency. Therefore, other genes located in the mapped QTL regions of OAR3 and OAR20 could be involved in the determination of phenotypic variation for traits such as persistency. In addition to leptin, other candidate genes should be searched in these chromosome windows.

Association between the leptin gene polymorphisms and quantitative traits such as the plasma circulating leptin and the energy milk content should be precisely assessed. As a matter of fact, one of the main role of the leptin is its involvement in the regulation of feed intake and energy expenditure. The additive genetic relationship between animals (*i.e*., pedigree) needs to be fitted as a random effect to take into consideration the genetic background. Significance of the genotype effects could be estimated using the approximated F-statistic provided by ASREML (Gilmour *et al*., 2001).

In conclusion, this research have drawn the discovery of two novel sheep leptin gene-related SNP markers. One of the discovered SNPs in this study along with the PCR-RFLP protocol set for individual DNA genotyping will be useful for association mapping purposes. However, additional animals and phenotypic measurements have to be investigated to avoid any spurious significant correlation between leptin gene polymorphisms and milk yield traits and qualitative components.

Acknowledgements

This is study has been carried out at the REPROGEN (Faculty of Veterinary Science, Sydney University, NSW - Australia) under the supervision of Prof. Herman Raadsma and Dr. Pietro Cieli. I would like to thank Dr. Elizabeth Jones for her precious reviews and suggestions.

4.5 References

Adamowicz T., K. Flisikowski, R. Starzy´nski, L. Zwierzchowski and M. Switonski (2006). Mutation in the Sp1 motif of the bovine leptin gene affects its expression. Mamm. Genome 17:77–82.

Ahima R. S., C. B. Saper, J. S. Flier, J. K. Elmquist (2000). Leptin regulation of neuroendocrine systems. Front Neuroendocrinol 21: 263–307.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignments search tool. J. Mol. Biol., 215: 403-410.

Andersson L. (2001). Genetic dissection of phenotypic diversity in farm animals. Nature Revs. Genet 2:130-138.

Bado A., S. Levasseur, S. Attoub, S. Kermorgant, J.-P. Laingneau, M. N. Bortoluzzi, L. Moizo, T. Lehy, M. Guerre-Millo, Y. Le Marchand-Brustel, M. J. M. Lewin (1998). The stomach is a source of leptin. Nature 394:790-793.

Barillet F., Arranz J.J. & Carta A. (2005). Mapping quantitative trait loci for milk production and genetic polymorphisms of milk proteins in dairy sheep. Genetics Selection Evolution 37(Suppl. 1):109–123.

Bauman D. E., W. B. Currie (1980). Partitioning of nutrients during pregnancy and lactation: a review of mechanisms involving homeostasis and homeorhesis. J Dairy Sci 63:1514–29.

Bencini R. and S. Dawe (1998). Sheep milking. In the new rural industries. A handbook for farmers and investors published by the Rural Industries Research & Development Corporation, 69-75.

Boucher D., M. F. Palin, F. Castonguay, C. Gariépy and F. Pothier (2006). Detection of polymorphisms in the ovine leptine (LEP) gene: Association of a single nucleotide polymorphism with muscle growth and meat quality traits. Canadian Journal of Animal Science, 86: 31-35.

Bonczek R. R, C. W. Young, J.E. Wheaton, K. P. Miller (1988). Responses of somatotropin, insulin, prolactin and thyroxine to selection for milk yield in Holsteins. J Dairy Sci 71:2470–9.

Brobeck J. R. (1948). Food intake as a mechanism of temperature regulation. Yale J. Biol. Med. 20:545-552.

Broom J.E., M.L. Tate and K.G. Dodds (1996). Linkage mapping in sheep and deer identifies a conserved pecora ruminant linkage group orthologous to two regions of HSA16 and a portion of HSA7Q. Genomics 33:358-364.

Broom M. F., C. Zhou, J.E. Broom, K. J. Barwell, R. D .Jolly and D. F. Hill (1998). Ovine neuronal ceroid lipofuscinosis: A large animal model syntenic with the human neuronal ceroid lipofuscinosis variant CLN6. J. Med. Genet. 35:717-721.

Buchanan F. C., C. J. Fitzsimmons, A. G. Van Kessel, T. D. Thue, D. C. Winkelman-Sim and S. M. Schmutz (2002). Association of a missense mutation in the bovine leptin gene with carcass fat content and leptin mRNA levels. Genetics Selection and Evolution, 34: 105–16.

Buchanan F. C., A. G. Van Kessel,C. Waldner, D. A. Christensen, B. Laarveld, and S. M. Schmutz (2003). Hot Topic: An Association Between a Leptin Single Nucleotide Polymorphism and Milk and Protein Yield. J. Dairy Sci. 86:3164–3166.

Calvo J. H., A. Martı´nez-Royo, A. E. Beattie, K. G. Dodds, A. Marcos-Carcavilla and M. Serrano (2006). Fine mapping of genes on sheep chromosome 1 and their association with milk traits. Animal Genetics 37:205–210.

Calvo J.H., S. Marcos, A. E. Beattie, C. Gonzalez, J. J. Jurado and M. Serrano (2004). Ovine alpha-amylase genes: isolation, linkage mapping and association analysis with milk traits. Animal Genetics 35, 329–32.

Cannas A., A. Nudda and G. Pulina, G. (2002) Great Lakes Dairy Sheep Symposium, November 7-11 2002. Cornell University, Ithaca, New York

Casanueva F. F., Dieguez C. (1999). Neuroendocrine regulation and actions of leptin. Front Neuroendocrinol 20: 317–363.

Chelikani P. K., D. R. Glimm, J. J. Kennelly (2003). Short communication: tissue distribution of leptin and leptin receptor mRNA in the bovine. J. Dairy Sci. 86:2369:2372.

Chen C. C., T. Chang and H. Y. Su (2004). Genetic polymorphisms in porcine leptin gene and their associaition with reproduction and production traits. Austr. J. Agric. Res. 55:699-704.

Clark C. E. F., W. J. Fulkerson, K. S. Nandra, I. Barchia, K.L. Macmillan (2005). The use of indicators to assess the degree of mobilisation of body reserves in dairy cows in early lactation on a pasture-based diet. Live. Prod. Sci. 94, 199-211.

Clark C. E. F., W. J. Fulkerson, K. S. Nandra, H. Smith, K.L. Macmillan (2006). Pattern of change in the concentration of milk acetone and its association with ovarian activity for pasture-based cows in early lactation. Live. Sci. 105, 144-150.

Cornish J., K. E. Callon, U. Bava, C. Lin, D. Naot, B. L. Hill, A. B. Grey, N. Broom, D. E. Myers, G. C. Nicholson, I. R. Reid (2002). Leptin directly regulates bone cell function in vitro and reduces bone fragility in vivo. J. Endocrinology 175:405-415.

Cockett N. E., T. L.Shay and M. Smit (2001). Analysis of the sheep genome. Physiol. Genomics 7:69–78.

Crawford A. M., K. G. Dodds, A. J. Ede, C. A. Pierson, G. W. Montgomery, H. G Garmonsway, A. E. Beattie, K. Davies, J. F. Maddox, S. W. Kappes (1995). An autosomal genetic linkage map of the sheep genome. Genetics 140: 703-724.

Cupps P. T. (1966). Breeds of dairy cattle. Page 149 in Introduction to livestock Production. 2nd ed. H. H. Cole, ed W. H. Freeman & Co., San Francisco, CA.

Darlington G. J., N. Wang and R. W. Hanson (1995). C/EBP alpha: a critical regulation of genes governing integrative metabolic processes. Curr. Opin. Gen. Dev. 5:565-570.

de Gortari M.J., B. A. Freking, R. P. Cuthbertson, S. M. Kappes, J. W. Keele, R. T. Stone, K. A. Leymaster, K. G. Dodds, A. M. Crawford and C. W. Beattie (1998). A second-generation linkage map of the sheep genome. Mammal. Genome 9:204-209.

Delavaud C., F. Bocquier, Y. Chilliard, D. H. Keisler, A. Gertler, G. Kann (2000). Plasma leptin determination in ruminants: effect of nutritional status and body fatness on plasma leptin concentration assessed by a specific RIA in sheep. J Endocrinol, 165(2): 519-526.

Dekkers J. C. M., J. C. M. Ten Hag, A. Weersink (1998). Economic aspects of persistency of lactation in dairy cattle. Livestock Science, 53:237-252.

Dekkers J. C. M. and F. Hospital (2002). The use of molecular genetics in the improvement of agricultural populations. Nature Revs. Genet. 3:22–32.

Di Stasio L, A. Brugiapaglia, M. Galloni, G. Destefanis and C. Lisa (2007). Effect of the leptin c.73T>C mutation on carcass traits in beef cattle. Anim. Genet. 38(3):316-317.

Ehrhardt R. A., R. M. Slepetis, J. Siegal-Willott, M. E. Van Amburgh, A.W. Bell AW, Y. R. Boisclair (2000). Development of a specific radioimmunoassay to measure physiological changes of circulating leptin in cattle and sheep. J. Endocrinol 166:519–28.

Etherton T. D, D. E. Bauman (1998). Biology of somatotropin in growth and lactation of domestic animals. Physiol Rev. 78:745–61.

Epstein, (1982). Awassi sheep. World Animal Reviews, 44: 9-18.

Fitzsimmons C. J., S. Shmutz, R. D. Bergen and J. J. McKinnon (1998). A potential association between the BM 1500 microsatellite and fat deposition in beef cattle. Mammalian Genome, 9:432-434.

Forhead A. J., L. Thomas, J. Crabtree, N. Hoggard, D. S. Gardner, D. A. Giussani, et al. (2002). Plasma leptin concentration in fetal sheep during late gestation: ontogeny and effect of glucocorticoids. Endocrinology;143(4): 1166-1173.

Friedman J. M. (1992). Tackling a weighty problem. Cell 69:217-220.

Friedman J. M., R. L. Leibel, D. S. Siegel, J. Walsh and N. Bahary (1991). Molecular mapping of the mouse ob mutation. Genetics 11,(4):1054-1062.

Friedman J. M., Halaas J. L (1998). Leptin and regulation of body weight in mammals. Nature 395:763–770.

Fukuda H. and N. Iritani (1999). Transcriptional regulation of leptin gene promoter in rat. FEBS letters, 455: 165-169.

Galloway S. M., V. Hanrahan, K. G. Dodds, M. D. Potts, A. M. Crawford and D. F. Hill (1996). A linkage map of the ovine X chromosome. Genome Res. 6: 667-677.

Garnsworthy P. C, J. H. Topps (1982). The effect of body condition of dairy cows at calving on their food intake and performance when given complete diets. Anim Prod 35:113–9.

Gengler N. (1995). Use of mixed models to appreciate the persistency of yields during the lactation of milk cows. Ph. D. Diss. Facultè Univ. Sci. Agron., Gembloux, Belgium.

Gengler N. (1996). Persistency of lactation yields: A review. Interbull Bull., 12: 87-96.

Georges M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorensen, M. R. Steele, X. Zhao, J. E. Womack, I. Hoeschele, (1995). Mapping quantitative trait loci controlling milk production by exploiting progeny testing. Genetics 139:907–920.

Gilmour A. R., B. R. Cullis, S. J. Welham and R. Thomson (2001). ASRELM, Reference Manual.

Gong D. W., S. Bi, R. E. Pratley and B. D. Weintraub (1996). Genomic structure and promoter analysis of the human obese gene. J Biol. Chem 271:3971-3974.

Grossman M., S. M., Hartz, W. J. Koops (1999). Persistency of lactation yield: a novel approach. Journal of Dairy Science, 82: 2192-2197.

Grummer R. R. (1995). Impact of changes in organic nutrient metabolism on feeding the transition dairy cow. J Anim Sci 73:2820– 33.

Haegeman, A., A. Van Zeveren, and L. J. Peelman. 2000. New mutation in exon 2 of the bovine leptin gene. Anim. Genet. 31:79.

Halaas J. L., K. S. Gajiwala, M. Maffei, S. L. Cohen, B. T. Chait, D. Rabinowitz et al. (1995). Weight-reducing effects of the plasma protein encoded by the obese gene. Science, 269(5223):543-546.

Hamrick M. W., C. Pennington, D. Newton, D. Xie and C. Isales (2004). Leptin deficiency produces contrasting phenptypes in bones of the limb and spine. Bone 34:376-383.

Hassink S. G., E. De Lancey, D. V. Sheslow, S. M.Smith-Kirwin, D. M. O’Connor, R. V. Considine, I. Opentanova, K. Dostal, M. L. Spear, K. Leef, M. Ash, A. R. Spitzer and V. L. Funanage (1997). Placental leptin: an important new growth factor in intrauterine and neonatal development? Pediatrics 100, e1 (6pp).

Henson M. C. and V. D. Castracane (2003). “Leptin and Reproduction.” Kluwer Academic/Plenum, New York.

Hossner K. L. (1998). Cellular, molecular and physiological aspects of leptin: potential application in animal production. Can. J. Anim. Sci. 78:463-472.

Hoffstedt J., P. Eriksson, S. Mottagui-Tabar, P. Arner (2002). A polymorphism in the leptin promoter region (-2548 G/A) influences gene expression and adipose tissue secretion of leptin. Horm. Metab. Res., 34: 355-359.

Houweling P. J., J. A. L. Cavanagh, D. N. Palmer, T. Frugier, N. L. Mitchell, P. A. Windsor, H. W. Raadsma, I. Tammen (2006). Neuronal ceroid lipofuscinosis in Devon cattle is caused by a single base duplication (c.662dupG) in the bovine CLN5 gene. Biochimica et Biophysica Acta 1762:890–897.

Hwang C. S. S. Mandrup, O. A. Mac Dougald, D. E. Geiman and M. D. Lane (1996). Transcriptional activation of the mouse obese (ob) gene by CCAAT/enhancer binding protein α. Proc. Natl. Acad. Sci. USA 93:873-877.

Ingvartsen K. L., Boisclair Y. R. (2001). Leptin and the regulation of food intake, energy homeostasis and immunity with special focus on periparturient ruminants. Domest Anim Endocrinol 21:215–50.

Ingalls A. M., M. M. Dickie and G. D. Snell (1950). Obese, a new mutation in the house mouse. J Hered 41: 317-318.

Isse N., Y. Ogawa, N. Tamura, H. Masuzaki, K. Mori, T. Okazaki, N. Satoh, M. Shigemoto, Y. Yoshimasa, S. Nishi, K. Hosoda, J. Inazawa and K. Nakao (1995). Structural organization and chromosomal assignment of the human obese gene. J. Biol. Chem. 270:27728-27733.

Jack C., P. Cieli, D. McGill, K. Fullard, P.C. Thompson, M. Lam, D. Blache, G.B. Martin, P.C. Wynn and H.W. Raadsma (2006). The use of biomarkers to assess energy balance in ewes milked once a day. Current Topics in Dairy Production, Australia, University Printing Service, Sydney, 11: pp. 142-144.

Jiang Z. H. and J. P. Gibson (1999). Genetic polymorphisms in the leptin gene and their association with fatness in four pig breeds. Mamm. Genome 10:191-193.

Jin L., S. Zhang, B. G. Burguera, M. E. Couce, R. Y. Osamura, E. Kulig and R. V. Lloyd (2000). Leptin and leptin receptor expression in rat and mouse pituitary cells. Endocrinology 141:333-339.

Lagonigro R, P. Wiener, F. Pilla, J. A. Woolliams, J. L. Williams (2003). A new mutation in the coding region of the bovine leptin gene associated with feed intake. Anim Genet 34:371–374.

Lake S. L., T. R. Weston, E. J. Scholljegerdes, C. M. Murrieta, B. M. Alexander, D. C. Rule, G. E. Moss and B. W. Hess (2007). Effects of postpartum dietary fat and body condition score at parturition on plasma, adipose tissue, and milk fatty acid composition of lactating beef cows. J. Anim Sci., 85: 717-730.

Liefers, S. C., M. F. W. te Pas, R. F. Veerkamp, and T. van der Lende (2002). Associations between leptin gene polymorphisms and production, liveweight, energy balance, feed intake and fertility in Holstein heifers. J. Dairy Sci. 85:1633–1638.

Liefers S. C., R. F. Veerkamp, M. F. W. te Pas, C. Delavaud, Y. Chilliard, T. van der Lende (2003a). Leptin concentrations in relation to energy balance, milk yield, intake, live weight, and estrus in dairy cows. J Dairy Sci., 86(3): 799-807.

Liefers, S. C., M. F. W. te Pas, R. F. Veerkamp, Y. Chilliard, C. Delavaud, R. Gerritsen, and T. van der Lende (2003b). Association of leptin gene polymorphisms with serum leptin concentration in dairy cows. Mamm. Gen. 14:657–663.

Liefers S. C., R. F. Veerkamp, M. F. W. te Pas, C. Delavaud, Y. Chilliard and T. van der Lende (2004). A missense mutation in the bovine leptin receptor gene is associated with leptin concentrations during late pregnancy. Animal Genetics, 35, 138–141

Liefers S. C., R. F. Veerkamp, M. F. W. te Pas, C. Delavaud, Y. Chilliard, T. van der Lende (2005a). Genetics and physiology of leptin in periparturient dairy cows. Domestic Animal Endocrinology, 29: 227-238.

Liefers S. C., R. F. Veerkamp, M. F. W. te Pas, C. Delavaud, Y. Chilliard, M. Platje and T. van der Lende (2005b). Leptin promoter mutations affect leptin levels and performance traits in dairy cows. Animal Genetics, 36, 111–118.

Lien S., H. Sundvold, H. Klungland, D. I. Vage (1997). Two novel polymorphism in the bovine obesity gene (OBS). Animal Genetics, 28: 238-246.

Lisa C., S. Sartore, L. Di Stasio (2007). Variabilità of Leptin gene promoter in cattle. Ital. J. Anim. Sci. 6 (suppl.1):150-152.

Lord E.A., J. M. Lumsden, K. G. Dodds, H. M. Henry, A. M. Crawford, H. A. Ansari, P. D. Pearce, D. W. Maher, R. T. Stone, S. M. Kappes (1996). The linkage map of sheep chromosome 6 compared with orthologous regions in other species. Mammal. Genome 7: 373-376.

Lord G. M. G. Matarese, J. K. Howard, R. J. Baker, S. R. Bloom and R. I. Lecheir (1998). Leptin modulates the T-cell immune response and reverses starvation-induced immunosupression. Nature 394:897-901.

Lucy M. C, H. Jiang, Y. Kobayashi (2001). Changes in the somatotrophic axis associated with the initiation of lactation. J. Dairy Sci. 84:E113–9.

Kadokawa H., D. Blache, G.B. Martin (2006). Plasma Leptin Concentrations Correlate with Luteinizing Hormone Secretion in Early Postpartum Holstein Cows. J. Dairy Sci., 89: 3020-3027.

Kennedy G. C. (1953). The Role of Depot Fat in the Hypothalamic Control of Food Intake in the Rat. Proc. R. Soc. B, Vol. 140, No. 901, pp. 578-592.

Kennes Y. M., B. D. Murphy, F. Pothier and M. F. Palin (2001). Characterization of swine leptin (LEP) polymorphisms and their association with production traits. Anim. Genet. 32:215-218.

Kmieć M., H. Kulig and A. Konik (2003). Preliminary results on associations between leptin gene (LEP) and some reproduction performance traits of boars. Arch. Tierz. Dummerstorf 46:63-70.

Konfortov B. A., V. E. Licence, J. R. Miller (1999). Re-sequencing of DNA from a diverse panel of cattle reveals a high level of polymorphism in both intron and exon. Mammaliam Genom., 10: 1142-1145.

Korwin-Kossakowska A., M. Kamyczek, D. Cieśilak, M. Pierzchala and J. Kuryl (2002). The effect of the polymorphism of leptin (LEP), leptin receptor (LEPR) and osteopontin (OPN) genes on selected reproduction traits of synthetic line 990 sows. Anim. Sci. Pap. Rep. 20:159-168.

Kulig H., W. Grzesiak and I. Szatkowska (2001). Effect of leptin gene polymorphism on growth and carcass traits in pigs. Arch. Tierz. Dummerstorf 44:291-296.

Maddox, J. F., K. P. Davis, A. M. Crawford, D. J. Hulma, D. Vaiman, E. P. Cribiu, B. A. Freking, K. J. Beh, N. E. Cockett, N. Kang, C. D. Rifkin, R. Drinkwater, et al., (2001). An enhanced linkage map of the sheep genome comprising more than 1000 loci. Genome Research 11: 1275-1289.

Macciotta N. P. P, D. Vicario, A. Cappio-Borlino (2006). Use of multivariate analysis to extract latent variables related to level of production and lactation persistency in dairy cattle. J. Dairy Sci., 89: 3188-3194.

McGavin M.D. (1974). An animal model of human disease: Progressive ovine muscular dystrophy. Comp. Pathol. Bullet. 6: 3-4.

McLaren R.J., G. R. Rogers, K. P. Davies, J. F. Maddox and G. W. Montgomery (1997). Linkage mapping of wool keratin and keratin-associated protein genes in sheep. Mammal. Genome 8:938-940.

Marie M., P.A. Findlay, L. Thomas and C.L. Adam (2001). Daily patterns of plasma leptin in sheep: effects of photoperiod and food intake. J. Endocrinol., 170(1): 277-286.

Mason M. M., Y. He, H. Chen, M. J. Quon and M. Reitman (1998). Regulation of leptin promoter function by Sp1, C/EBP, and a novel factor. Endocrinology 139:1013-1022.

Masuzaki H., Y. Ogawa, N. Isse, N. Satoh, T. Okazaki, M. Shigemoto, K. Mori, N. Tamura, K. Hosoda, Y. Yoshimasa, H. Jingami, T. Kawada, K. Nakao (1995). Human obese gene expression adipocyte-specific expression and regional differences in the adipose tissue. Diabetes 44:855-858.

Masuzaki H., Y. Ogawa, N. Sagawa, K. Hosoda, T. Matsumoto, H. Mise, H. Nishimura, Y. Yoshimasa, I. Tanaka, T. Mori, K. Nakao (1997). Non-adipose tissue production of leptin: leptin as a novel placenta-derived horomone in humans. Nat. Med. 3:1029-1033.

Miller S. G., P. De Vos, M. Guerre-Millo, K. Wong, T. Hermann, B. Stales, M. R. Briggs and J. Auwerx (1996). The adipocyte specific transcription factor C/EBPα modulates human ob gene expression. Proc. Natl. Acad. Sci. USA 93:5507-5511.

Montgomery G. W. and J. A. Sise (1990). Extraction of DNA from sheep white blood cells. New Zeeland Journal of Agriculture Research, 33: 437-441.

Morash B. A. li, P. R. Murphy, M. Wilkinson and E. Ur (1999). Leptin gene expression in the brain and pituitary gland. Endocrinology 140:5995-5998.

Morrison C. D., J. A. Daniel, B. J. Holmberg, J. Djiane, N. Raver, A. Gertler, et al (2001). Central infusion of leptin into well-fed and undernourished ewe lambs: effects on feed intake and serum concentrations of growth hormone and luteinizing hormone. J Endocrinol., 168(2):317-24.

Muir B. L. (2004). Genetics of lactation persistency and relationships with reproductive performance in holsteins. PhD thesis, University of Guelph.

Nkrumah, J. D., C. Li, J. A. Basarab, S. Guercio, Y. Meng, B. Murdoch, C. Hansen, and S. S. Moore (2004). Association of a single nucleotide polymorphism in the bovine leptin gene with feed intake, growth, feed efficiency, feeding behaviour and carcass merit. Can. J. Anim. Sci. 84:211–219.

Nkrumah J. D., C. Li, J. Yu, C. Hansen, D. H. Keisler and S. S. Moore (2005). Polymorphisms in the bovine leptin promoter associated with serum leptin concentration, growth, feed intake, feeding behavior, and measures of carcass merit. J. Anim. Sci. 83:20–28

Oberkofler H., A. Beer, D. Breban, E. Hell, F. Krempler, W. Patsch (1997). Human obese gene expression: alternative splicing of mRNA and relation to adipose tissue localization. Obesity Surgey 7:390-396.

Oprzadek J., K. Flisikowski, L. Zwierzchowski and E. Dymnicki (2003). Polymorphisms at loci of leptin (LEP), Pit1 and STAT5A and their association with growth, feed conversion and carcass quality in Black-and-White bulls. Anim. Sci. Papers. Rep 21:135-145.

Peelman F., W. Waelput, H. Iserentant, D. Lavens, S. Eyckerman, L. Zabeau and J. Tavernier (2004). Leptin: linking adipocyte metabolism with cardiovascular and autoimmune diseases. Prog. Lipid Res. 43:283-301.

Pelleymonunter M. A., M. J. Cullen, M. B. Baker, R. Hetch, D. Winters, T. Boone and F. Collins (1995). Effects of the obese gene product on body weight regulation in ob/ob mice. Science 269:540-543.

Pomp D., T. Zou, A. C. Clutter and W. Barendse (1997). Rapid communication: mapping of leptin to bovine chromosome 4 by linkage analysis of a PCR-based polymorphism. J. Anim. Sci. 75:1427.

Pryce J. E., R. F. Veerkamp, R. Thomoson, W. G. Hill, G. Simm (1997). Genetic aspects of common health disorders and measures of fertility in Holstein Friesian dairy cattle. Anim. Sci., 65: 353-360.

Pryce J. E., M. P. Coffey, G. Simm (2001). The relationship between body condition score and reproductive performance. J Dairy Sci 84:1508–15.

Pulina G., A. Nudda, N. P. P. Macciotta, G. Battacone, S. P. G. Rassu, A. Cannas (2007). Non-nutritional factors affecting lactation persistency in dairy ewes: a review. Ital. J. Anim. Sci., 6: 115-141.

Raadsma H. W., K. R. Zenger, C. Cavanagh, M. K. Lam, P. C Thomson, M. Jones, G. Attard, D. Palmer and F. W. Nicholas (1999). The mapping of economically important quantitative trait loci in sheep I: male distance framework map and QTL for growth rate and body weight. Proc. Assoc. Advmt. Anim. Breed. Gen., 13: 361-364.

Rahmouni K. and W. G. Haynes (2004). Leptin and the cardiovascular system. Recent Prog. Horm. Res. 59:225-244.

Rosenbaum M., R. L. Leibel (1998). Leptin; a molecule integrating somatic energy stores, energy expenditure and fertility. Trends Endocrinol Metab. 9:117–124.

Senarìs R., T. Garcia-Caballero, X. Casabiell, R. Callego, R. Castro, R. V. Considine, C. Dieguez, F. F. Casaneuva (1997). Synthesis of leptin in human placenta. Endocrinology 138:4501-4504.

Singh M., M. Lam, D. McGill, P.C. Thomson, J.A. Cavanagh, K.R. Zenger and H.W. Raadsma (2007). High resolution mapping of quantitative trait loci on ovine chromosome 3 and 20 affecting protein yield and lactation persistency. Proc. of AAABG, Association for the Advancement of Animal Breeding and Genetics, Armidale, NSW Australia. 23-26 Sept 2007.17: 565.

Smith-Kirwin S. M., D. M. O’Connor, J. Johnston, E. de Lancey, S. G. Hassink and V. L. Funanage (1998). Leptin expression in human mammary cells and breast milk. J. Clin. Endocrinol. Metabol. 83:1810-1813.

Sölkner J., W. Fuchs (1987). A comparison of different measures of persistency with special respect to variation of test day yields. Livest. Prod. Sci., 16: 305-319.

Stachowiak M., M. Mackowski, Z. Madeja, M. Szydlowski, A. Buszka, P. Kaczmarek, B. Rubis, P. Mackowiak, K. W. Nowak and M. Switonski (2007). Polymorphism of the porcine leptin gene promoter and analysis of its association with gene expression and fatness traits. Biochemical Genetics 45:245-253.

Steppan C. M., D. T. Crawford, K. L. Chidsey-Frink, H. Z. Ke and A. G. Swick (2000). Leptin is a potent stimulator of bone growth in ob/ob mice. Regul. Pept. 92:73-78.

Stone R. T., S. M. Kappes, C. W. Beattie (1996) The bovine homolog of the obese gene maps to chromosome 4. Mamm Genome 7, 399–400

Swalve, H.H. (1995). Genetic relationship between dairy lactation persistency and yield. J. Anim. Breed. Genet., 112: 303-311.

Szydlowski M., M. Stachowiak, M. Mackowski, M. Kamyczek, R. Eckert, M. Rozycki and M. Switonski (2004). No major effect of the leptin gene polymorphism on porcine production traits. J. Anim. Breed. Genet. 121:149-155.

Taylor V. J., D. E. Beever, D. C. Wathes (2003). Physiological adaptations to milk production that affect fertility in high yielding dairy cows. In: Dairying, using science to meet consumer needs. Br. Soc. Anim. Sci. Occasional Publication No. 29. Nottingham University Press; p. 37–71.

Tan P., J. G. Allen, S. D. Wilton, P. A. Akkari, C. R. Huxtable and N. G. Laing (1997). A splice-site mutation causing ovine McArdle's disease. Neuromuscular Disorders 7: 336-342.

Taniguchi Y., I. Tomohiro, Y. Takahisa and S. Yoshiyuki (2002). Genomic Structure and Promoter Analysis of the Bovine Leptin Gene. IUBMB Life, 53: 131–135.

Tena-Sempere M. (2007). Roles of Ghrelin and Leptin in the Control of Reproductive Function. Neuro-endocrinology. 86:229-241

Thomas L., J. M Wallance, R. P. Aitken, J. G. Mercer, P. Trayhurn, N. Hoggard (2001). Circulating leptin during ovine pregnancy in relation to maternal nutrition, body composition and pregnancy outcome. J. Endocrinol., 169(3): 465-76.

Urban T., J. Kuciel and R. Mikolášová (2002). Polymorphism of genes encoding for ryanodine receptor, growth hormone, leptin and MYC protooncogene protein and meat production in Duroc pigs. Czech. J. Anim. Sci. 47:411-417.

Walling G. A., P. M. Visscher, A. D. Wilson, B. L. McTeir, G. Simm and S. C. Bishop (2004). Mapping of quantitative trait loci for growth and carcass traits in commercial sheep populations. J. Anim. Sci. 82:2234–2245.

Walther C., J. L. Guenet, D. Simon, U. Deutsch, B. Jostes, M. D. Goulding, D. Plachov, R. Balling and P. Gruss (1991). Pax: A murine multigene family of paired box-containing genes. Genomics 11(2):424-434.

Wang J., R. Liu, M. Hawkins, N. Barzilai and L. Rosetti (1998). A nutrient-sensing pathway regulates leptin gene expression in muscle and fat. Nature 393:684-688.

Wang J., R. Liu, L. Liu, R. Chowdhury, N. Barzilai, J. Tan and L. Rosetti (1999). The effect of leptin on Lep expression is tissue-specific and nutriotionally regulated. Nat. Med. 5:895-899.

Wathes D. C., Z. Cheng, N. Bourne, V. J. Taylor, M. P. Coffey, S. Brotherstone (2007a). Differences between primiparous and multiparous dairy cows in the inter-relationships between metabolic traits, milk yield and body condition score in the periparturient period.. Dom Anim Endocrinol 33(2):203-225.

Wathes D. C, M. Fenwick, Z. Cheng, N. Bourne, S. Llewellyn, D. G. Morris, D. Kenny, J. Murphy, R. Fitzpatrick (2007). Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. Theriogenology 68S: S232–S241.

Wiesner G., M. Vaz, G. Collier, D. Seals, D. Kaye, G. Jennings, G. Lambert, D. Wilkinson and M. Esler (1999). Leptin is released from the human brain: influence of adiposity and gender. J. Clin. Endocrinol. Metabol. 84:2270:2274.

Wright C. D., A. M. Havill, S. C. Middleton, M. A. Kashem, P. A. Lee, D. J. Dripps, T. G. O'Riordan, M. P. Bevilacqua and W. M. Abraham (1999). Secretory leukocyte protease inhibitor prevents allergen-induced pulmonary responses in animal models of asthma. J. Pharmacol. Exp. Therapeut.. 289: 1007-1014.

Wood P. D. P. (1967). Algebraic Model of the Lactation Curve in Cattle. Nature 216:164-165.

Zhang Y., R. Proenca, M. Maffei, M. Barone, L. Leopold, J. M. Friedman (1994). Positional cloning of the mouse obese gene and its human homologue. Nature 372: 425-432.

Zhang F., M. B. Basinski, J. M. Beals, S. L. Briggs, L. M. Churgay, D. K. Clawson, R. D. Di Marchi, T. C. Furman, J. E. Hale, H. M. Hsiung, B. E. Schoner, D. P. smith, X. Y. Zhang, J. P. Wery and R. W. Schevitz (1997). Crystal structure of the obese protein leptin-E100. Nature 387:206-209.

A compimento di questo triennio di Scuola di Dottorato, desidero ringraziare tutti coloro che hanno reso possibile il raggiungimento di questo traguardo. In primis, il Prof.Giovanni Bittante per avermi accolta nel suo gruppo di ricerca e avermi dato la possibilità di intraprendere questa nuova esperienza formativa. Ringrazio il Prof. Martino Cassandro per il continuo incoraggiamento e supporto soprattutto durante i "momenti bui" del ricerca scientifica! Un sentito ringranziamento è rivolto al Prof. Gianni Barcaccia riuscito, di anno in anno, nell'impresa di fugare miei radicati dubbi genetici!

Ringrazio il Dr. Pietro Cieli e il Prof. Herman Raadsma per avermi accolta presso il centro REPROGEN della Facoltà di Veterinaria dell'Università di Sydney (NSW, Australia) e per avermi inserito nel progetto di ricerca riguardante lo studio del gene della leptina in pecora. Grazie al Prof. Gianfranco Gabai per aver contribuito al mio viaggio-soggiorno oltre-oceano in Australia e Mary per aver portato laggiù un'ulteriore ventata di "Italianità".

Ringrazio i miei compagni di avventura appartenti al gruppo dei "Bottoniani Biotech" del settore arboreo della Facoltà di Agraria, con i quali ho condiviso tre anni di ricerca, confronto e conforto reciproco. Tra essi sono annoverati: Mauretta, Pippo, Saretta, FabioMassimo, Francuzzo e il capitano della squadra...Botton! Grazie ragazzi, siete stati il carburante di questo mio Dottorato.

Un dovuto ringraziamento è rivolto ai colleghi/amici del Dip. di Scienze Animali con i quali, se pur con reciproca iniziale diffidenza, ho instaurato una invidiabile collaborazione e amicizia. Un grazie speciale è dedicato ai due forestali DOC, Maresciallo Rocca (Giampaolo) e Refo (Luca) con i quali ho condiviso ufficio (il loro!), chiacchiere, pranzi e continui dibattiti più o meno scientifici! Ringrazio le due dottorande per eccellenza, Arianna e Aziza che hanno allietato questi ultimi tre anni della mia vita consolidando giorno dopo giorno l'amicizia che all'inizio sembrava più miraggio che realtà.

Ringrazio, inoltre, mamma, papà, Roberto e nonna Caterina per aver supportato (e sopportato) questo mio ultimo (lo prometto!) percorso formativo.

Grazie a tutti coloro che formalmente non ho ringraziato ma che nel bene e nel male hanno contribuito a rafforzare (se mai ce ne fosse stato bisogno!) il mio spirito e il mio carattere.