

Universidad Aut3noma de Barcelona
Facultad de Veterinaria

Genetic polymorphism in goat

Study of the kappa casein, beta lactoglobulin, and stearyl coenzyme A
desaturase genes

Mohamed Habib Yahyaoui

Bellaterra 2003

“Science is an assault on ignorance..”

Ridley (1991)

ACKNOWLEDGMENTS

This study was carried out in the Department of Animal Sciences (Laboratory of Molecular Genetics) in the Faculty of Veterinary Medicine (Universidad Autónoma de Barcelona).

I want first to thanks Dr. Armand Sanchez, for giving me the opportunity to working in his lab introducing me in the field of molecular genetics. I have been truly privileged to be a part of such group.

I'm also greatly indebted to Dr. Josep M. Folch, co-supervisor of this thesis for his positive attitude, criticism, and valuable comments and advices. Thanks for your support, training, and guidance during these years.

I'm grateful to all my colleagues in the Genetics Lab that I have worked with and shared my time with during theses years, for their daily conversations, their friendship and for putting up with me.

I would also like to thank Antonella Angiolillo (University of Molise, Italia) for her friendship and collaboration that has been very beneficial for me, and I hope to her also.

I wish to acknowledge Dr. Bruce Whitelaw and his group (Gene Expression Lab, Roslin Institute, Scotland) for excellent welcome I received during my stay in Roslin. Special thanks to Romi for her assistance and help.

I would like to thank all my friends who helped me get through these years.

The final acknowledgment goes to all of my family for their support and love; my studies have been in their honor.

This work was supported in part by a research scholarship from AECI (Agencia Española de Cooperación Internacional), Spain.

ABSTRACT

Genetic polymorphism in goat. Study of the kappa casein, the beta-lactoglobulin, and the stearyl coenzyme A desaturase genes. Polymorphism in the goat specie was studied among different Spanish, French, and Italian breeds using the BESS T-Scan method and sequencing. The analysed regions are located in the promoter region and exon 1 of the *beta-lactoglobulin* gene (β -LG), the full coding region of the *kappa casein* (κ -CN) gene (exons 3 and 4), and the exon 5 of the *stearyl coenzyme A desaturase* (SCD) gene.

A total of ten mutations were detected in the kappa casein coding region by using BESS method and sequencing. Four are synonymous mutations (three in exon 4 and one in exon 3) whereas other six produce amino acid changes. All these non-synonymous mutations, located in exon 4, are single nucleotide transitions. The association between the different mutations (*haplotypes*) resulted in seven genetic variants, designated κ -casein A, B, C, D, E, F, and G. A procedure for rapid and simultaneous genotyping for all κ -casein variants was developed. The method is based on primer extension analysis coupled with capillary electrophoresis and fluorescent detection. Kappa casein A and B are the most common variants found in several Spanish, French, and Italian breeds. Variant B is predominant in all these breeds, with the exception of the Canaria breed, where variant A is prevalent. The F variant is predominant in the Spanish wild type goat *Capra pyrenaica sp. hispanica*. Comparative sequence data suggest that the F variant is most likely the original type of kappa casein in caprine specie. Other variants were derived from this allele by successive mutations following two different trunks.

A single nucleotide polymorphism was detected by BESS method in the proximal promoter region of the *b-LG* gene. However, this mutation is not located in the sequence recognition of any known transcription factor. The polymorphism is found in Murciano-Granadina, Malagueña, Payoya, and Saanen, but it absent in the Canaria breed. This difference may be explained by the different origin of this breed.

The goat SCD cDNA was isolated and characterized. The open reading frame contains 1077 nucleotides encoding for a mature protein of 359 amino acids, with a high level of similarity with ovine (98%) and bovine (95%) homologues. The structural organization of the caprine *SCD* gene is similar to those of rodents and human. It spans approximately 15 Kb and consists of six exons and five introns. Several single nucleotide polymorphisms were detected in the coding region (exons 5 and 6) and in the 3'-untranslated region. A PCR-RFLP protocol for genotyping the SNP at position 931 of the cDNA was developed. The polymorphism was used to map the caprine *SCD* gene by linkage analysis to CHI 26q13-26q21.

CONTENTS

REVIEW OF THE LITERATURE

1. INTRODUCTION	1
2. GENETIC POLYMORPHISM	
2.1 Kinds of polymorphism	2
2.2 Origin of SNPs	4
2.3 Polymorphism detection at protein level	5
2.4 Polymorphism detection at DNA level	5
2.5 Scanning methods	6
2.5.1 Conformation-based methods	7
2.5.2 Cleavage-based and enzymatic methods	9
2.6 Genotyping methods	12
2.6.1 Hybridization methods	13
2.6.2 Enzymatic methods	15
2.7 Conclusion	19
3. MILK PROTEINS	
3.1 Kappa casein	21
3.1.1 Kappa casein structure and function	21
3.1.2 Kappa casein gene	24
3.1.3 Polymorphism of the kappa casein gene	26
3.2 Beta lactoglobulin	29
3.2.1 Beta lactoglobulin structure and function	29
3.2.2 Beta lactoglobulin gene	31
3.2.3 Polymorphism of the β -lactoglobulin	32
4- MILK FAT	
4.1 Origins of milk fat	34
4.2 Composition of milk fat	35
4.3 Genetic modification of milk fat	38
4.4 Stearoyl coenzyme A function	39

4.5 Stearoyl coenzyme A mRNA and gene	40
4.6 Regulation of stearoyl coenzyme A gene expression	41
4.7 Polymorphism of the stearoyl coenzyme A gene	43
AIMS OF THIS WORK	44
RESULTS	
1. “Genetic polymorphism of the caprine kappa casein gene” <i>Journal of Dairy Research</i> (2001) 68 : 209-216	45
2. “Short Communication: Characterization of a new genetic variant in the caprine kappa casein gene” <i>Journal of Dairy Science</i> (2002) 85 :2679-2680	59
3. “Characterization and genotyping of the caprine kappa casein variants” <i>Journal of Dairy Science</i> (2003) 86 : 2715-2720	65
4. “Rapid Communication: Polymorphism in the goat beta lactoglobulin proximal promoter region” <i>Journal of Animal Science</i> (2000) 78 : 1100-1101	81
5. “Rapid Communication: Partial nucleotide of the goat stearoyl coenzymeA desaturase cDNA and gene structure” <i>Journal of Animal Science</i> (2002) 80 : 866-867	86
6. “Genetic mapping of the goat stearoyl coenzyme A desaturase gene to chromosome 26 using a biallelic polymorphism” <i>Submitted (Animal Genetics)</i>	93
DISCUSSION	97
CONCLUSIONS	122
REFERENCES	124
ANNEXES	145

LIST OF FIGURES AND TABLES

REVIEW OF THW LITERATURE

Figure 1	Principle of BESS (base excision sequence scanning) technique.	11
Figure 2	Invader assay.	17
Figure 3	Structural organization of the casein locus.	25
Figure 4	Structure of the caprine <i>kappa casein</i> gene.	25
Figure 5	Structural organization of the caprine <i>b-lactoglobulin</i> gene.	33
Table 1	Features of methods for analysis of genetic variation.	20
Table 2	Bovine kappa casein variants.	28
Table 3	Bovine β -lactoglobulin variants.	33
Table 4	Ovine β -lactoglobulin variants.	33
Table 5	Major fatty acids of caprine and bovine milks.	37

DISCUSSION

Figure 6	BESS technique.	99
Figure 7	Mutations detected by BESS T-Scan method.	100
Figure 8	Multiple alignments of the predicted amino acid sequences of κ -casein protein.	106
Figure 9	Genotyping of kappa casein variants by primer extension analysis.	107
Figure 10	Quantitative haplotypes observed at the goat calcium-sensitive <i>loci</i> .	112
Table 6	Frequencies of the kappa casein variants in different goat breeds.	109

ABBREVIATIONS

SNP:	Single nucleotide polymorphism
RFLP:	Restricted fragment length polymorphism
FISH:	Fluorescent in situ hybridization
bp:	Base pair
Kb:	Kilo base pair
dNTP:	Deoxynucleotide triphosphate
ddNTP:	Di-deoxynucleotide triphosphate
ATP:	Adenosine triphosphate
kDa:	Kilodalton
A	Adenine
C:	Cytosine
G:	Guanine
T:	Thymine
PCR:	Polymerase chain reaction
RT-PCR:	Reverse transcription PCR
Taq:	<i>Thermus aquaticus</i>
DNA:	Desoxyribo-nucleic acid
cDNA:	Complementary DNA
RNA:	Ribo-nucleic acid
mRNA:	messenger RNA
ssDNA:	Single stranded DNA
dsDNA:	Double stranded DNA

REVIEW OF THE LITERATURE

1. INTRODUCTION

Many of the traits of interest in animal production are *quantitative traits*. Evaluation of genetic merit of animals is still essentially based on the application of the theory of quantitative genetics. The conceptual basis of this theory is the *polygenic model*, which assumes that quantitative traits result from the action (and interaction) of a large number of genes, each with small effect. The resulting effects are then predicted using powerful statistical methods (animal model), based on pedigree and performance recording of traits from the individual animal and its relatives.

The advances in molecular genetic technology in the past two decades, particularly DNA-based markers, has had a great impact on gene mapping, allowing identification of the underlying genes that control part of the variability of these multigenic traits. Broadly, two experimental strategies have been developed for this purpose: linkage studies and candidate gene approach.

Linkage studies rely on the genetic map knowledge and search for the quantitative trait loci (QTL) by using family materials and comparing segregation patterns of genetic markers (generally microsatellites) and the trait being analyzed. Markers that tend to co-segregate with the analyzed trait provide approximate chromosomal location of the underlying gene (or genes) involved in part of the trait variability determinism.

The second approach focus on the study of the genetic polymorphism of a few genes (candidate genes) suspected, on the basis of the biological and physiological information of the trait, to be implicated for part of the trait variability. Hence, association analysis are carried in order to test whether a particular genotype or *haplotype* (a series of alleles along a stretch of DNA) are stably associated with the analyzed trait (for example the rate of the synthesis of the protein or milk yield). Introduction of such additional molecular information in guiding selection procedures allows to better asses true genetic merit of animals. For example, it allows selection to occur among individuals that do not exhibit the trait in question (e.g. milk protein genotypes in males). This approach is already being employed with regard to bovine leukocyte adhesion deficiency (BLAD; Shuster et al., 1992), and genes with major effects, such as the halotane locus in swine (Rempel et al., 1993) and α s1-casein in goat (Manfredi et al., 1995).

In this work, we have studied the genetic polymorphism of three genes in goat specie: *b-lactoglobulin*, the major whey protein in ruminant milk; *kappa casein*, which

plays a key role in micelle stabilization and milk coagulation; and *stearoyl coenzyme A desaturase*, involved in the biosynthesis of the monounsaturated fatty acids.

2. GENETIC POLYMORPHISM

2.1 Kinds of polymorphism

Genetic polymorphism arises from mutation. It refers to the difference in DNA sequence among individuals, groups, or populations, and can be caused by mutations ranging from a single nucleotide base change to variations in several hundred bases. In a formal sense, there are only two kinds of polymorphisms: those due to replacement of DNA bases and those due to insertion or deletion of base pairs. The simplest type of genetic polymorphism is the single nucleotide polymorphism (SNP). A position is referred to as an SNP when it exists in at least two variants, being the rarer allele more abundant than 1% in the general population. Other types of genetic polymorphisms result from the insertion or deletion of a section of DNA, which include repeat sequences (mini and microsatellites) and gross genetic losses and rearrangement.

Gross alterations are mutations in which substantial portions of DNA sequence (>500 bp) are deleted, duplicated or rearranged. These types of genomic alterations can be detected by high resolution cytogenetics (for extremely large alterations such as chromosomal number and chromosomal translocations) and by fragment analysis of implicated chromosomal regions using southern blot, microsatellites, and fluorescent *in situ* hybridization (FISH).

Hypervariable minisatellites are usually defined as the repetition in tandem of a short (6 to 100 bp) motif spanning 0.5 Kb to several kilobases. They are mostly located between genes, and are dispersed unevenly in the genome preferentially in telomeric locations (Lathrop et al., 1988). Because of their length polymorphism, and the ability to cross-hybridize with similar loci throughout the genome, minisatellites have been used in DNA fingerprinting and forensic for individual identification (Jeffreys et al., 1985).

Microsatellites or short tandem repeats (STRs) are tandem repeats of multiple copies of the same sequence motif composed of 1-4 base pair long units. They are ubiquitous in prokaryotes and eukaryotes and present even in smallest bacterial genomes (Hancock, 1996). These polymorphisms show high levels of allelic variation in the number of repeat units, and are used extensively as markers in linkage studies. A subset of STRs,

namely trinucleotide repeats, are implicated in many human neurodegenerative disorders such as fragile X syndrome and Huntington's disease (Bates and Lehrach, 1994; Reddy and Housman, 1997) and in some human cancers (Wooster et al., 1994). This kind of STRs is often called *dynamic mutations* (Richards and Sutherland, 1992).

In this thesis, the term polymorphism is used and it refers to the SNPs located within the coding sequences or in regulatory regions of the genes. SNPs are the most common type of genetic diversity and are estimated to occur with a prevalence of one SNP per 1300 bases in the human genome (Lander et al., 2001; Venter et al., 2001). In principle, SNPs could be bi-, tri-, or tetra-allelic polymorphisms. However, tri-allelic and tetra-allelic are very rare and SNPs are sometimes simply referred to as bi-allelic markers (Brookes, 1999). SNPs can result from either the transition or transversion of nucleotide bases. Transition substitutions occur between purines (A and G) or between pyrimidines (C and T). Transversions are substitutions between a purine and a pyrimidine. Transition mutations are more common than transversions, with CpG dinucleotides showing the highest mutation rate, presumably due to 5-methylcytosine deamination (Duncan and Miller, 1980). Nucleotide substitutions occurring in protein-coding regions can be classified as synonymous and non-synonymous according to their effect on the resulting protein. A substitution is synonymous if it causes no amino acid change while a non-synonymous substitution results in alteration in the encoded amino acid. The latter type can be further classified into missense and nonsense mutations. A missense mutation results in amino acid changes due to the change of codon used while a nonsense mutation results in a termination codon.

Even within a single chromosome, the SNPs are not uniformly distributed, and some genomic regions have significantly lower or higher diversity than the average. Polymorphisms in the regulatory regions of genes and sequence variants that alter amino acids in the coding regions are generally less common, reflecting a greater selection pressure reducing diversity at these DNA regions. Within coding exons the nucleotide diversity is four-fold lower, with about half resulting in non-synonymous codon changes (Li and Sadler, 1991; Nickerson et al., 1998). In addition, there is enormous diversity in SNP frequency between genes reflecting different selective pressures on each gene as well as different mutation and recombination rates across the genome.

Depending on where a SNP occurs, it might have different consequences at the phenotypic level. SNPs in the coding regions (sometimes termed as cSNPs) or in regulatory

regions are more likely to cause functional differences than SNPs elsewhere. In case of the human monogenic mendelian disorders, SNPs in the coding regions that alter the function or structure of the encoded protein could be the cause of the disease. In animal production, examples of direct impact of SNPs on phenotype include mutations in growth hormone receptor gene in dwarf chicken (Duriez et al., 1993; Huang et al., 1993), myostatin (*GDF-8*) gene in double-muscling cattle (Kambadur et al., 1997; Grobet et al., 1998) and α s1-casein in goat (Grosclaude et al., 1994). In general, association studies have to be performed in order to statistically establish that particular alleles are associated with one or more phenotypic traits. However, most SNPs are located in non-coding regions, and have no direct effect on the phenotype. These SNPs are useful as markers in population genetics and evolutionary studies and to identify genes implicated in complex multigenic traits by using linkage disequilibrium.

2.2 Origin of SNPs

The accuracy of DNA replication is fundamental for the genetic stability of the cell. A key group of enzymes in the replication process are the DNA polymerases, which synthesize new DNA strands by incorporating complementary nucleotides with high fidelity. The enzymes possess proofreading and exonuclease activities to correct for misincorporated bases. However, errors can occur, with frequencies varying between 10^{-9} and 10^{-10} per base replicated (Echols and Goodman, 1991).

Sources of replication errors include both endogenous and exogenous factors. Endogenous reactions consist of transversions, spontaneous depurination of bases (Loeb and Preston, 1986), deamination of cytosine and sometimes adenine residues yielding uracil and hypoxanthine, respectively. Exogenous mechanisms for mutations include dimerization of pyrimidine bases induced by UV light, various chemicals such as alkylating agents forming adducts with DNA bases, reactive oxygen species damaging pyrimidine and purine rings, and ionizing radiation causing DNA strand nicking and breakage. The majority of these modifications are generally recognized and corrected by the DNA repair systems. The crucial nature of these repair mechanisms is evidenced by several inherited human diseases caused by defects in the system (Moses, 2001).

2.3 Polymorphism detection at protein level

Initial identification of polymorphism relied on the study of physiological and biochemical variation at protein level that follow indirectly from variation in DNA sequence. The occurrence of genetic polymorphisms in milk protein was first reported by Aschaffenburg and Drewry (1955). Using paper electrophoresis, they observed the existence of two distinct bands of β -lactoglobulin from individual bovine milk samples. A subsequent analysis demonstrated that these two types were under genetic control, determined by two autosomal codominant alleles (Aschaffenburg and Drewry, 1957). The paper electrophoresis technique was applied for the detection of genetic variants of α -lactalbumin (Blumberg and Tombs, 1958) and β -casein (Aschaffenburg, 1961). Variants of α s1-casein were detected by starch-urea electrophoresis (Thompson et al., 1962), and the use of reducing agents in the sample buffer or in the gel contributed to the characterization of two κ -casein variants (Neelin, 1964; Schmidt, 1964; Woychik, 1964). With the first milk protein (α -lactalbumin) totally sequenced (Brew et al., 1970), the amino acid sequences of the detected variants were subsequently determined in the 1970s. The introduction of analytical techniques with increased resolving power and sensitivity, such as polyacrylamide gel electrophoresis, isoelectric focusing, chromatography, and more recently capillary electrophoresis, enabled detection of more genetic variants in milk of different livestock species.

The main limitation of protein polymorphism detection methods is that are restricted to the resolution of only proteins with differing net charges. However, mutations with amino acid substitution that do not lead to a change in the net charge on proteins should occur in theory three times more frequently than those resulting in an alteration in the charge (Ng-Kwai Hang and Grosclaude, 1992). Furthermore, protein electrophoretic mobility is also affected by post-transcriptional modifications such as different degrees of phosphorylation and glycosylation of the protein.

2.4 Polymorphism detection at DNA level

Monitoring of genetic variation at DNA level became possible with the development of the recombinant DNA technology. Before the polymerase chain reaction (PCR) discovery, it was necessary to clone fragments to get enough DNA and to reduce complexity of the sample. Restriction enzymes and southern blot hybridization (Southern,

1975) were used to identify single-base pair changes in genomic DNA that result in the gain or loss of a restriction site. These nucleotide variants were called restriction fragment length polymorphisms (RFLPs) and were used in early linkage studies (Botstein et al., 1980).

The possibility to amplify specific segments of genomic DNA by PCR (Mullis and Faloona, 1987) has enabled detection of point mutations in large scale. Analysis of genomic DNA is based on amplification of fragments of interest from the genome to increase the copy of the target and to reduce the complexity of the analyzed DNA. Both of these measures are directed to enable sensitive and specific detection of the target of interest. At present, *Thermus aquaticus* (*Taq*) DNA polymerase is widely used for DNA amplification. The error rate of *Taq* polymerase is in the range of 10^{-4} to 10^{-5} per nucleotide (Tindall and Kunkel, 1988), depending on the size of fragment being amplified and reaction conditions (concentration of magnesium chloride and dNTPs, pH, temperature). Depending on the method of analysis, polymerase errors may contribute significantly to unspecific background limiting the level of detection, particularly in hybridization-based methods (Reiss et al., 1990).

The mutation detection methods can be divided into those that scan for unknown mutations in a target region and those that screen for previously described variation. Scanning methods for mutation identification and characterization are usually labor intensive, difficult to interpret and expensive. Once the mutation has been discovered, the genotyping (scoring) methods should provide efficient and straightforward techniques for repetitive testing of the variant in large number of samples.

2.5 Scanning Methods

The most commonly used strategy for detecting point mutations is to amplify fragment of interest by PCR, scan the PCR products for the presence of mutations by a rapid procedure, and then sequence the PCR products that were positive by the scanning techniques.

In addition to direct sequencing and recently developed microarrays, there are two groups of scanning methodologies. The first group is based on the aberrant electrophoretic migration of mutant molecules due to conformation changes (conformation-based methods) which include single-strand conformation polymorphism (SSCP), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis (HA), and denaturing high-performance

liquid chromatography (DHPLC). The second group of techniques relies on enzymatic or chemical cleavage of RNA or DNA molecules.

2.5.1 Conformation-based methods

A common feature of conformation-based methods is that sequence context of a nucleotide change has an important effect on the sensitivity of the detection, and generally optimal conditions need to be assayed for reproducibility. Moreover, these methods do not localize the position of the detected polymorphisms.

SSCP

Single-strand conformation polymorphism (Orita et al., 1989) is one of the simplest and the most widely used method for mutation detection. In SSCP, target regions with potential polymorphisms are first amplified by PCR. The amplified fragments are then denatured to generate single-stranded DNA (ssDNA) and separated by electrophoresis on a non denaturing polyacrylamide gel. The single-stranded fragments adopt three-dimensional conformation which is dependent on the primary sequence. If a sequence difference (mutation) exists between wild-type and mutated DNA, this may result in a mobility shift observed during gel electrophoresis. Because the conformation of ssDNA is also affected by electrophoresis conditions (temperature and ionic environment), optimal conditions are largely determined empirically which limit standardization of the method. Under optimal conditions, 70-95% of potential base exchanges are detectable in short (200 bp or less) PCR products (Sheffield et al., 1993; Gross et al., 1999). The sensitivity can be increased by coupling with either dideoxy fingerprinting (ddF; Sarkar et al., 1992) or restriction enzyme fingerprinting (REF; Liu and Sommer, 1995). SSCP has relatively low throughput, although higher capacity methods have been developed using capillary- rather than gel-based detection (Wenz et al., 1999).

DGGE

Denaturing gradient gel electrophoresis utilizes the property that DNA molecules differing by a single base have slightly different melting properties and, therefore, migrate distinctively in a denaturing gel (Fisher and Lermann, 1983). Double-stranded DNA (dsDNA) is electrophoresed through a gradient of increasing concentration of a denaturing agent such as urea or formamide. As dsDNA migrates in the gel, the strands progressively

dissociate in sequence-dependent manner according to their melting temperature, resulting in a decrease in electrophoretic mobility. Base mismatches in heteroduplexes lead to a significant destabilization of domains. For this reason, heteroduplexes between wild-type and mutant fragments are generally used for the analysis of point mutations. Before analysis, optimal conditions must be determined either experimentally or by calculation of theoretical melting profiles. Detection can be improved by adding a GC-rich sequences (GC clamp) to one of the PCR primers (Myers et al., 1985; Sheffield et al., 1989). Close to 100% of point mutations can be detected when heteroduplexes are generated from sense and antisense strands and when GC clamps are attached (Myers et al., 1985; Scholz et al., 1993). Maximum fragment size suited for DGGE is 1000 bp. Other variants of DGGE have been developed, including temperature gradient gel electrophoresis (Wartell et al., 1990) and constant gradient gel electrophoresis (Hovig et al., 1991). The advantages of DGGE are its high accuracy and relatively low costs. However, the major disadvantage is that running conditions must be defined for each PCR product before analysis in addition to the low throughput.

Heteroduplex analysis

The concept of heteroduplex is almost identical to SSCP. The method relies on the detection of heteroduplex formed during re-annealing of the denatured strands of a mixture of wild-type and altered DNA molecules. The heteroduplex can be detected as a band shift on a non denaturing polyacrylamide gel. The optimal fragment length for the SNP detection varies between 200 and 600 bp, and the rate of detection has been estimated to approximately 80% (Cotton, 1993). The method has been shown to be more sensitive in areas of highly repetitive and GC-rich sequences (Korkko et al., 1998). Depending on fragment size, time for electrophoretic separation varies between 14 and 30 h (Nollau and Wagener, 1997) making the method not suitable for automation. Nevertheless, application of capillary electrophoresis may drastically reduce time of electrophoretic separation (Cheng et al., 1994).

A widely used variant of heteroduplex analysis is DHPLC (denaturing high performance liquid chromatography) where the electrophoresis step is replaced by column chromatography (Underhill et al., 1997). The heteroduplexes are separated from the homoduplexes at a temperature that partially denatures the mismatched DNA. The use of DHPLC increases substantially the throughput and improves the rate of mutation detection,

ranging from 95 to 100% (O'Donovan et al., 1998; Gross et al., 1999). Automated instruments for DHPLC analysis are commercially available (Wave, Transgenomics Inc.).

2.5.2 Cleavage-based & enzymatic methods

Cleavage-based methods take advantage of the fact that mismatched bases are sensitive to cleavage by enzymes or chemicals. Heteroduplexes are generated by denaturation and renaturation of PCR products from wild-type and mutant alleles. After incubation with enzymes or chemicals, the products are resolved by gel electrophoresis. Enzymatic methods rely on the use of bacterial mismatch-repair proteins (mismatch binding) and enzymes that cut modified DNA in BESS method (base excision sequence scanning). The most significant advantages of using cleavage-based and enzymatic methods instead of conformation-based methods are that the position of the mutation can be generally localized and longer fragments of DNA can be analyzed.

RNase assay

The RNase cleavage assay has been used to detect mismatches in RNA:DNA (Myers et al., 1985a) and RNA:RNA (Winter et al., 1985) duplexes. RNase, which can recognize single-strand RNA, digests a mismatched hybridized RNA probe. Mutations of purine bases are cleaved with low efficiency or remain uncleaved and only 50% of possible mutations are detected, 70% when both strands are analyzed (Myers et al., 1985a). Because of the low rate detection and the high background due to unspecific cleavage at sites of perfect base pairing, the method appears not to be suited for screening purposes. A variant of the method is commercially available (NIRCA, nonisotopic RNase cleavage assay; Ambion). NIRCA is RNA based, with analysis of RNA-RNA hybrids generated by *in vitro* transcription from reference and polymorphic PCR templates. Cleaved fragments are separated in non denaturing gel electrophoresis (Nash and Inderlied, 1996).

EMC (Enzyme mismatch cleavage)

Mismatches generated in DNA heteroduplexes can be cleaved by bacteriophage resolvases such as T4 endonuclease VII and T7 endonuclease I (Mashal et al., 1995; Youil et al., 1995; Del Tito et al., 1998). These enzymes normally cleave *in vivo* branched DNA intermediates that form during phage replication and packaging, and were shown to cut DNA near the site of base mismatches (Solaro et al., 1993). The cleavage products can then

be analyzed by gel or capillary electrophoresis. The method is known as enzyme mismatch cleavage or EMC. As in RNase cleavage assay, cleavage efficiency varies for different mismatches and it's important to use both strands for detecting mutations. In addition, non-specific background is generally observed probably due to homoduplex cleavage which may pose a problem for the correct interpretation of the results.

CCM (Chemical cleavage of mismatches)

In CCM technique (Cotton et al., 1988; Saleeba et al., 1992), mismatches within heteroduplexes are modified by using hydroxylamine and osmium tetroxide which react with mispaired cytosine and thymine residues, respectively. The modified heteroduplexes (DNA:DNA or DNA:RNA) are then cleaved with piperidine and the reaction products are separated by polyacrylamide gel electrophoresis. DNA fragments up to 2 Kb (Cotton, 1993) can be analyzed and if sense and anti-sense strands are analyzed, all point mutations will be detected. The level of detection can be increased -when few mutants are present in large background of wild-type- by separation and detection of fluorescently labeled fragments on a DNA sequencer. The major drawback of this technique is the use of hazardous chemicals that are not well suited to routine laboratory use.

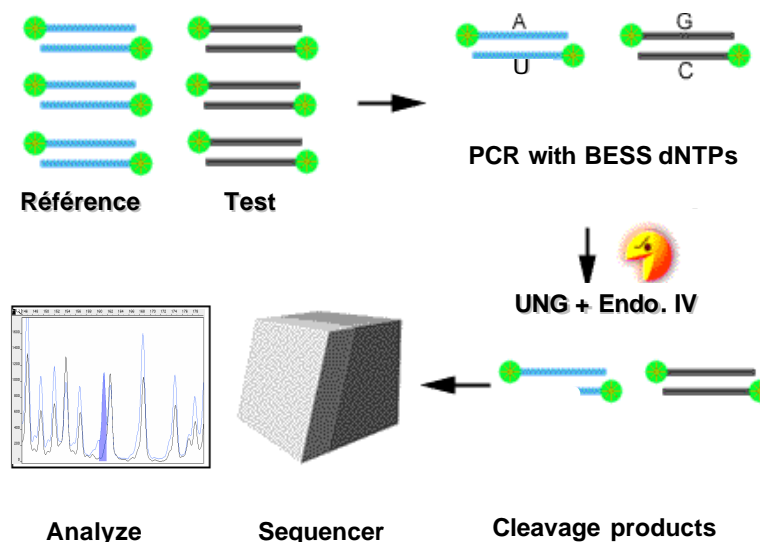
Mismatch binding proteins

Bacterial mismatch-repair proteins (Mut S, and Y) which normally replace misincorporated bases in newly synthesized DNA, have been used in mutation detection (Lishanki et al., 1994; Smith and Modrich, 1996; Xu et al., 1996). In these assays, the protein simply binds to the site of mismatch and do not cleave DNA. Detection is achieved by performing gel shift (retardation) analysis or local protection of DNA from degradation by DNAses. Mismatch-repair proteins bind mismatches with different affinity (some mismatches are poorly recognized) and have unspecific activity. However, immobilization of the protein on nitrocellulose membrane seems to greatly increase the affinity of the protein for mismatched DNA relative to matched DNA (Wagner et al., 1995). The bacterial mismatch repair system has also been used *in vivo* to detect heteroduplex DNA transfected into recipient *Escherichia coli* cells (Faham and Cox; 1995; Faham et al., 2001). The method is referred to as mismatch repair detection (MRD).

BESS (Base excision sequence scanning)

The BESS method allows detection of single nucleotide mutations as well as small insertions and deletions involving deoxythymidine, representing 10 of a total of 12 possible point mutations (Hawkins and Hoffman, 1997). Only C→G and G→C transversions are excluded. The principle of the BESS assay is illustrated in Figure 1. PCR is performed with a limiting amount of dUTP that incorporate in randomly manner in place of dTTP during amplification (analogous to dideoxynucleotide incorporation in Sanger sequencing). The PCR products are then digested with an enzyme mix containing uracil *N*-glycosylase (UNG) and endonuclease IV. UNG removes the uracil bases creating abasic sites at the location of dUTP incorporation, and endonuclease VI cleaves the phosphodiester bond at these abasic sites. Generated fragments are then separated on a sequencing gel or by capillary electrophoresis to obtain a pattern of bands virtually identical to a “T” lane sequencing ladder, representing the banding pattern of dUTP incorporation. The appearance or disappearance of a band, or the change in intensity of a band normalized to a control, indicates the point of mutational change. The band pattern is revealed by using a method appropriate for the radioactive or fluorescent label on the primer. Both strands may be analyzed simultaneously if the two primers are labeled with distinguishable dyes. The method does not require the formation of heteroduplexes, purification of PCR product before analysis, or extensive optimization of reaction conditions. We have used this method for scanning and mutation detection of caprine *k-casein* and *SCD* genes (see results and discussion chapters).

Figure 1. Principle of BESS (base excision sequence scanning) technique.



Microarrays

The microarrays technology is based on hybridization of complementary strands of nucleic acids. Fluorescent-labeled derivatives of analyzed DNA or RNA (the target) are hybridized to oligonucleotides (the probe) immobilized on a solid support, generally glass chip. Variant detection array (VDA) applies this technology to scan large DNA sequences for the identification of unknown polymorphisms (Wang et al., 1998; Halushka et al., 1999). The DNA sample of interest is PCR amplified to incorporate fluorescently labeled nucleotides and then hybridized to the array. The differences in hybridization strength between perfectly matched and mismatched oligonucleotides are quantified by high-resolution fluorescent scanning and analyzed by computer software. The accuracy of VDA is comparable to that of dye-terminator sequencing in large-scale screens (Wang et al., 1998) and allows rapid scanning of large amounts (megabases) of DNA sequences.

Direct sequencing

Currently, direct sequencing is one of the high throughput methods for mutation detection, and is the most accurate method to determine the exact nature of a polymorphism. Sanger dideoxy-sequencing can detect any type of unknown polymorphism and its position, when the majority of DNA contains that polymorphism. Fluorescent sequencing can have variable sensitivity and specificity in detecting heterozygotes because of the inconsistency of base-calling of these sites (Chadwick et al., 1996; Yan et al., 2000). Thus, it has only limited utility when the polymorphism is present in a minor fraction of the total DNA (for example in pooled samples of DNA or in solid tumors) due to low sensitivity. DNA sequencing is usually used as a second step to confirm and identify the exact base altered in the target region previously identified as polymorphic by using scanning methods.

2.6 Genotyping Methods

Although the methods for genotyping known polymorphism are very diverse, they are in general variations on a few standard methodologies. Broadly, each technique involves a method for interrogating SNP (underlying principle), and an appropriate detection system to quantify and display the allele-specific product (fluorescence techniques, mass spectrometry, luminometric detection of pyrophosphate release, or

DHPLC). Most genotyping techniques can be divided into hybridization-based and enzyme-based methods.

2.6.1 Hybridization methods

The specificity of genotyping by hybridization depends strongly on the nucleotide sequence context of the SNPs and the hybridization conditions. Therefore, hybridization techniques are in general more error prone and need carefully designed probes and hybridization protocols. These techniques include DNA microarrays and homogeneous hybridization assays.

DNA microarrays genotyping

DNA microarrays designed to SNP genotyping are generally based on the principle of “sequencing by hybridization” where a set of *tiling* oligonucleotides that walk over each variant of SNP are used (Wang et al., 1998; Mei et al., 2000). This approach is applied in order to circumvent the difficulty of the prediction of hybridization conditions for each SNP that will allow the optimal distinction between two alleles differing at a single nucleotide position. Because the precise sequence of the oligonucleotides at each location in the array is known, the pattern of hybridization can be determined using fluorescently labeled probes (PCR products). The method allows genotyping of large collection of small PCR products in one hybridization, and the limiting step is the preparation of thousands of PCR products for parallel analysis. To reduce this limitation, small PCR products are generated to ensure robust multiplex amplification, and a constant sequence tags at the 5'-end of primers are added to facilitate labeling of pooled products (Wang et al., 1998).

Despite the high levels of scanning capacity, the method is susceptible to any sequence-specific features that affect hybridization efficiency (melting temperature, secondary structure) and fail to distinguish between heterozygote and homozygote genotypes, the correct genotype could only be assigned at 60-70% of the analyzed sites (Wang et al., 1998; Cho et al., 1999). In addition, it is expensive and lacks flexibility in the redesign of polymorphism sets.

An alternative approach in array format that allows more robust SNP genotyping is dynamic allele-specific hybridization (DASH; Howell et al., 1999) where the hybridization between the target DNA and the allele-specific oligonucleotide probe is monitored over a

temperature gradient. Distinct target sequence alleles are then distinguished via fluorescence output by their melting temperature.

Homogeneous hybridization

Homogeneous assays refer to procedures that require no further manipulations after setting up the reaction. These assays monitor the SNP genotype in real time during amplification (real time PCR) and were originally designed for quantitative PCR analysis.

The TaqMan assay (Applied Biosystems) takes advantage of the 5' → 3' exonuclease activity of *Taq* DNA polymerase that cleaves 5' terminal nucleotides of dsDNA (Holland et al., 1991). For SNP genotyping, the PCR is performed with a common pair of PCR primers and two allele-specific oligonucleotide (TaqMan) probes designed to hybridize downstream of one of the primers. TaqMan probes are blocked from extensions at their 3'-terminus and are labeled with a donor-acceptor (reporter-quencher) dye pair that interacts via fluorescence resonance energy transfer (FRET). When the probes are hybridized to the target SNP in annealing step, the reporter dye is quenched due to the physical proximity with the quencher. During the PCR extension phase, the perfectly hybridized probes are cleaved by the 5' nuclease activity of the *Taq* DNA polymerase, leading to an increase of reporter fluorescence. Mismatched probes are displaced without degradation and the fluorophores remain quenched. Specific genotyping is determined by measuring the signal intensity of the fluorescence in real time or after the completion of the PCR. Discrimination of the polymorphism is determined solely by the hybridization of the probe and not by the enzyme activity. Consequently, TaqMan probes must be carefully designed because incorporation into non-specific product will give an apparently positive result.

In another homogeneous hybridization based PCR, molecular beacons are used for alleles discrimination (Tyagi and Kramer, 1996). Molecular beacons are oligonucleotide probes similar to TaqMan probes except that their 3' and 5' ends –flanking the target DNA sequence- are complementary. When not hybridized to the target, the probe adopts a hairpin-loop conformation with the reporter and quencher dyes close together, thereby quenching the donor fluorescence. Upon hybridization to the target, the loop opens out, and the resulting separation of the fluorophore from the quencher produces an increase in signal (Tyagi et al., 1998). Molecular beacons are more sensitive and have higher specificity than linear probes, due to the strong tendency of the probes to self-anneal which destabilizes

mismatched hybrids (Bonnet et al., 1999; Tapp et al., 2000). As in TaqMan assay, the molecular beacon method requires careful design of the probe, since the detection of variant nucleotides is based on allele specific hybridization. The molecular beacon technology has been commercialized by Stratagene.

Both the molecular beacon and TaqMan approaches allow limited multiplexing due to the limited number of compatible reporter-quencher sets (Tyagi et al., 1998; Lee et al., 1999). Because the probes are optimized individually for each SNP, they are most efficient when a limited number of SNPs are analyzed in a large number of samples. The advantages of homogeneous assays include the reduced risk of cross-contamination, the simplicity and rapidity of the methods, since no post-PCR manipulations are required.

2.6.2 Enzymatic methods

Enzymatic methods provide in general more specific allele discrimination, and usually have fewer sequence limitations than hybridization-based methods. These methods include RFLP, allele-specific PCR (AS-PCR), oligonucleotide ligation assay (OLA), invader assays, pyrosequencing, and primer extension analysis (PEA).

RFLP

Due to its simplicity, PCR-RFLP is one of the most commonly used methods for polymorphism genotyping. When restriction enzyme recognition sites in DNA are altered, the enzyme will be able to distinguish between allelic variants, giving rise to different DNA fragments. Initially, the RFLP analysis required a radioactively labeled probe for detection, and now the method is coupled with PCR and simple agarose gel electrophoresis. If the genetic polymorphism produces a gain or loss of the restriction site, a different restriction digestion pattern can be recognized. The drawback of the PCR-RFLP method is that not all polymorphisms alter a restriction site, but this can be sometimes overcome by introducing mismatched primers to create restriction sites (Cohen and Levinson, 1988).

AS-PCR

The method utilizes the discrimination properties of DNA polymerase in extension of a 3'-end mismatch primer. PCR is performed in two parallel reactions. A pair of PCR primers with the 3'-end complementary to either allele at the mutation site is used in combination of a common reverse primer (Newton et al., 1989; Wu et al., 1989). Thus,

assuming that extension occurs only when the primer is perfectly matched, the rate of amplification depends on the allele present in the template. The advantage of the method is that genotype assignment only requires the detection of the amplification signal, for example by agarose gel electrophoresis. To avoid false amplifications, reaction conditions must be extensively and strictly controlled due to the poor discrimination property of DNA polymerases, especially for G/T and A/C mismatches (Kwok et al., 1990). Therefore, the technique has shown to be uncertain for genotyping and needs extensive optimization for individual templates.

OLA

The oligonucleotide ligation assay (OLA) relies on the ability of DNA ligases to discriminate mismatches occurring at the ligation site (Landegren et al., 1988; Alves and Carr, 1988). In this approach, two allele-specific probes are designed to anneal adjacent to the 5'-end of a third common probe on the site of the polymorphism. Generally, the allele-specific probes contain a fluorescent label at the 5'-end. In the presence of DNA ligase, only two perfectly matched oligonucleotide probes are joined by the enzyme. The resulting products are then separated by electrophoresis. By varying the combinations of color dyes and probe lengths, multiple mutations can be detected in a single reaction in an automated sequencer (Baron et al., 1996). Ligation products have been also detected by mass spectrometry (Jurinke et al., 1996). The use of a thermostable ligase (Barany, 1991) allowed amplification of ligated products (ligation chain reaction, LCR).

A recent variation of OLA in single nucleotide polymorphism analysis is the introduction of circular probes (padlock probes) where the PCR reaction is replaced by DNA amplification using strand-displacing DNA polymerase and a circular template at isothermal conditions (Nilsson et al., 1997; Lizardi et al., 1998). Upon binding to the target the probes are circularized by a ligase in a template specific manner. The method was denoted rolling circle amplification (RCA). The main advantage of the ligation-based approaches is the high specificity and sensitivity enabling detection of minority mutations down to 0.2% level (Khanna et al., 1999).

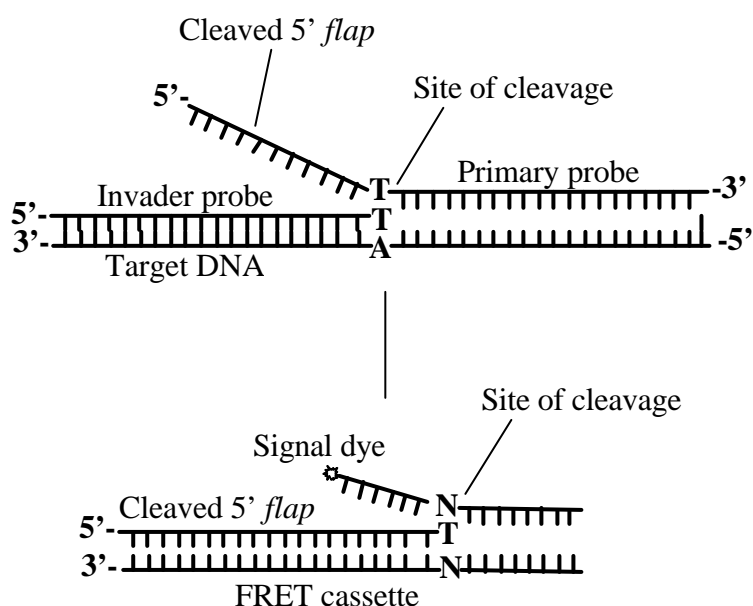
Invader assay

The invader assay (Third Wave Technologies) is based on flap endonuclease (FEN) enzymes, which recognize and cleave specific structures formed by overlapping

oligonucleotides that are annealed to the target DNA strand (Lyamichev et al., 1999). Two oligonucleotides, an allele-specific probe (with a 5'-region that is non complementary to the target sequence) plus an upstream “invader” probe, are used in each reaction. When the allele-specific is perfectly matched at the SNP, the overlapping structure -formed by these probes and the target sequence- is recognized and cleaved by FEN (Figure 2). This cleavage releases the 5' non-complementary region of the allele-specific probe (referred to as *flap*) which is not required for enzyme activity but participates in a second cleavage of a combined labeled FRET probe-template causing release of fluorescent signal (Hall et al., 2000). The invader assay has been coupled with fluorescent polarization (Hsu et al., 2001) and with mass spectrometric detection (Griffin et al., 1999). The major advantage of the method is that allow SNP genotyping directly in genomic DNA without previous PCR amplification. Reactions are carried out isothermally in a single step, and in the common format using fluorescence, only one genotype can be performed per reaction (uniplex). The method has been adapted to a solid-phase format (microarrays) for multiplex genotyping (Wilkins Stevens et al., 2000).

Figure 2. Invader assay.

An *invader* oligonucleotide upstream and a primary probe downstream of the SNP site are annealed to the target. The primary probe has 5' nonspecific tail. When there is a perfect match between the primary probe and the DNA target at the base to be genotyped, an overlapping structure between the invader and primary probe is formed. A thermostable flap endonuclease (FEN) recognizes this structure and cleaves primary probe, releasing the flap. This released flap in turn serves as an invader oligonucleotide in the second invasive cleavage reaction on a FRET oligonucleotide probe. The signal probe is then cleaved and the fluorescent molecule is released.



Pyrosequencing

Pyrosequencing utilizes the principle of “sequencing by synthesis”, and is based on luminometric detection of pyrophosphate which is released during incorporation of dNTPs (Nyren et al., 1993; Ronaghi et al., 1996). The incorporation process releases a pyrophosphate which is converted to ATP (by ATP sulfurylase) in the presence of 5'-phosphosulfate, and the level of ATP is detected by the use of firefly luciferase. The light production in the luciferase-catalyzed reaction is detected by a photon detector or a charge coupled device (CCD) camera. The height of each peak correlates to the light signal and is proportional to the number of nucleotides incorporated. Unincorporated dNTPs and excess of ATP are continuously degraded by apyrase. When degradation is complete, another dNTP is added. The genotype of a SNP is deduced by sequential addition and degradation of the nucleotides. Since the DNA polymerase (polymerization) and the apyrase (nucleotide removal) compete for the same substrate, it is important to optimize the kinetic conditions of the reaction to obtain high-quality data. The method allows sequence determination of short fragments (30-50 bp), however, the sequential identification of bases prevents genotyping of several SNPs per reaction. Pyrosequencing instruments are commercially available (Pyrosequencing AB).

Primer extension analysis

The primer extension analysis (PEA) is one of the simplest and widely used techniques for SNP genotyping. The method is based on the single base extension of an oligonucleotide by the DNA polymerase (Sokolov, 1990; Syvanen et al., 1990). In the common format, an internal unlabeled primer is designed to anneal immediately upstream of the polymorphism site in the presence of DNA polymerase and differentially labeled fluorescent dideoxynucleotides. The polymerase extends the primer one nucleotide, adding a single ddNTP to its 3' end. The fluorescent signal emitted corresponds to the nucleotide incorporated and thus the sequence of the polymorphism. The major advantages of PEA method over hybridization-based methods include its simplicity and accuracy in discrimination between the heterozygous and homozygous genotypes. Moreover, the same reaction conditions can be employed for genotyping any SNP irrespectively of the sequence flanking the variable site. However, the excess PCR reagents must be removed before the reaction in order to obtain specific extension. In addition, a second purification step is usually required after the extension reaction to remove unincorporated labeled nucleotides.

In the early applications of PEA, separated reactions for each allele were performed using radioactively labeled nucleotides (dNTP or ddNTP) with separation by gel electrophoresis and detection by autoradiography or phosphorimager instrument (Kuppuswami et al., 1991; Krook et al., 1992). Different detection systems have been coupled with PEA, in both solution-phase and solid-phase formats (revised in Syvanen, 1999). Particularly, mass spectrometry (MALDI-TOF, matrix-associated laser desorption time-of-flight mass spectrometry) is useful as a read-out for primer extension products because primers of different lengths can be designed to yield allele-specific extension products differing in their mass (Ross et al., 1998; Li et al., 1999, Bray et al., 2001), in which no labeling is necessary. However, extremely clean samples free of ions and other impurities are required for mass spectrometry detection. In addition to SNP genotyping, the PEA method has been adapted for DNA methylation (Gonzalzo and Jones, 1997) and quantitative PCR analysis (Singer-Sam et al., 1992; Greenwood and Burke, 1996; Lombardo and Brown, 1997).

2.7 Conclusion

A variety of technologies and platforms are now available for both known and unknown polymorphism detection (revised in Syvanen, 2001; Kwok, 2001; Kirk et al., 2002; Ahmadian and Lundeberg, 2002). All these techniques have their advantages and limitations (Table 1), and there is presently no ideal technique for all applications. The appropriate choice of a technique and its relative usefulness requires careful consideration of several factors, such as experimental design (limited number of SNPs on large population or a large number of SNPs on a limited number of individuals), kind of sample being queried, throughput, and cost. The level of expertise for a certain procedure and the equipment available are other determining factors.

Regarding accuracy and sensitivity, requirements on mutation analysis are different for known and unknown polymorphisms. Scanning analysis often involve a rapid procedure combined with sequencing, and the level of detection might be critical in some applications (for example when mutants represent a minor fraction of the sample). Conversely, in screening methods, it is crucial to minimize if not eliminate the extent of genotyping errors. Accuracy in range of 99.9% is a basic requirement for clinical applications, and genotyping error rates as small as 3% can have serious effects in association studies and linkage

disequilibrium measures (Akey et al., 2001), leading to biased results and erroneous conclusions.

Table 1. Features of methods for analysis of genetic variation.

Method	Advantages	Limitations
<u>Scanning methods</u>		
Sanger sequencing	High accuracy (close to 100%)	Multiple reactions for larger genes, difficult to detect low level mutations
Microarray (VDA)	High throughput, screen of large sequence blocks	Expensive, frameshift mutations not detected
SSCP	Simplicity, detects low level mutations	Limited automation, missense and silent mutations confounded
DGGE, DHPLC, HA	Simplicity, close to 100% of possible mutations detected	Limited automation, missense and silent mutations confounded
CCM	High sensitivity, identifies position of most mutations	Labor intensive, chemical hazard
EMC, RNase A, Mut Y, Mut S	Localize mutations, identifies missense, nonsense and frameshift mutations	High background, difficult to detect low level mutations
BESS	Simple, localize mutations	G-C mutations not detected, specificity of the enzymes
<u>Genotyping Methods</u>		
Microarrays	High throughput	Limited genotype discrimination, new array for each DNA target
TaqMan, Molecular beacons	Simplicity and rapidity of assay	Expensive probes, limited multiplexing, requires optimization
Primer extension	Robust and rapid, adaptable to mass spectrometry (no labeling)	Purification steps, deletions and insertions not detected
Invader assay	PCR amplification avoided, isothermal, automation	Limited multiplexing, insertions and deletions not detected
Pyrosequencing	Sequencing of up to 50 bp, detection of insertions and deletions	Difficult to multiplex, dedicated instrument
OLA	High accuracy and multiplexing capacity	Multiple detection steps
Padlock probes and RCA	No PCR amplification required	Requires long probes, limited multiplexing capacity
RFLP	Simple, inexpensive	Not suitable for high throughput, limited to enzyme site mutation

VDA, variant detection array; **SSCP**, single strand conformation polymorphism; **DGGE**, denaturing gradient gel electrophoresis; **DHPLC**, denaturing high performance liquid chromatography; **HA**, heteroduplex analysis; **CCM**, chemical cleavage mismatch; **EMC**, enzyme mismatch cleavage; **BESS**, base excision sequence scanning; **OLA**, oligonucleotide ligation analysis; **RCA**, rolling circle amplification; **RFLP**, restriction fragment length polymorphism.

3. MILK PROTEINS

Although both quantitative and qualitative differences occur in milk of different species, the milk proteins of all mammals can be divided into two classes: the caseins and the whey proteins. The caseins (α_s1 , α_s2 , β , and κ) comprise the major protein component of ruminant milk and are secreted in the form of stable calcium phosphate micelles. The assembly of caseins into micelles allows to maintain low viscosity of the milk despite the high protein concentration. The casein micelles are of functional importance for protein and mineral nutrition of the offspring, and in determining the physical properties of milk. The whey proteins of milk correspond to the protein fraction that remains in solution after precipitation of casein micelles and fat globules, and are constituted principally by β -lactoglobulin and α -lactalbumin. Beta lactoglobulin is the major whey protein in ruminant milk and α -lactalbumin is part of the enzyme system involved in lactose synthesis. As well as these major protein classes, many other proteins are present in small or trace amounts, such as serumalbumins, immunoglobulins, lysozymes, lactoferrin, and plasmin.

3.1 Kappa casein

3.1.1 Kappa casein structure and function

Kappa casein (κ -casein) differs from the calcium-sensitive caseins (α_s1 , α_s2 , and β) in its solubility over a broad range of calcium ion concentrations and its low content of phosphate component. The phosphorylation sites (at serine and threonine residues) are confined to the C-terminal region of the protein and are present as single sites rather than the clusters as found in other caseins. Furthermore, phosphorylation of the κ -casein may be prevented due to previous glycosylation at or near the susceptible sites (Swaisgood, 1992). As a consequence, κ -casein does not bind calcium to the same extent as the other caseins and its solubility characteristics are not affected by this ion. Phosphorylation of caseins occurs post-translationally in the Golgi apparatus by the action of casein kinases (Bingham, 1979). The calcium-sensitive caseins are generally thought to be located predominantly within the micelles whereas κ -casein is located primarily on the surface, determining the size of the micelles and stabilizing their structure (Horisberger and Vonlanthen, 1980; Sawyer, 1982). The N-terminal part of κ -casein is involved in interactions with other caseins in micelle while the C-terminal fragment –which has a negative charge due the

presence of acidic amino acids- constitutes the main component of the external layer of casein micelle (Walstra, 1990; Rollema, 1992). Casein micelles are stabilized by steric and electrostatic repulsion of the polar C-terminal domain of the κ -casein protein (Horne, 1992).

Kappa casein is also the only glycosylated member of casein group. Glycosylation occurs during the post-translational modification and is catalyzed by O-glycosyl transferases within the Golgi apparatus of mammary epithelial cells (Takeuchi et al., 1984; Mephram et al., 1992). The extent of glycosylation is variable, and the carbohydrate moieties are attached to κ -casein via O-glycosidic linkages to threonine (and rarely serine) residues within the C-terminal region of the protein. Approximately 36% of caprine κ -casein is glycosylated (Moreno et al., 2001). In bovine κ -casein, the carbohydrate portion consists of N-acetylgalactosamine, N-acetylneuraminic acid (sialic acid), and galactose, arranged as tri- or tetra-saccharides (Jolles and Fiat, 1979). Caprine κ -casein contains N-glycolylneuraminic acid in addition to these sugars (Addeo et al., 1978). Glycosylation degree is higher in colostrums than in milk, and increases during mastitis infection (Dziuba and Minkiewicz, 1996). In addition, genetic variants of bovine κ -casein induce significantly different amounts of glycosylation (Robitaille et al., 1991; Lodes et al., 1996). The influence of glycosidic residues on the stability of casein micelles seems to be insignificant but the level of glycosylation affects the susceptibility of the κ -casein to hydrolysis by chymosin, susceptibility decreasing with increasing of the carbohydrate content (Fox, 1989). On the other hand, glycosylation patterns can result in differential inhibition of gastric pathogens, prevention of bacterial infection and toxin binding (revised in Dziuba and Minkiewicz, 1996; Brody, 2000). Due to these post-translational modifications, caprine κ -casein appears heterogeneous in milk protein electrophoresis assays with at least five forms, the main one being non glycosylated (Addeo et al., 1978; Recio et al., 1997).

Besides being a micelle stabilizer, κ -casein is susceptible to cleavage by the action of chymosin –in the gut or during cheese making- that occurs at a specific labile bond (Phe-Met in ruminants) in the C-terminal region of the protein (Jolles et al., 1968). The products of this cleavage are the hydrophobic (insoluble) N-terminal portion or para κ -casein, and the hydrophilic (soluble) and highly charged and glycosylated C-terminal region, termed caseinomacropptide (CMP) or glycomacropptide. The function of the para κ -casein is not well known. However, CMP is responsible for clotting milk in gut, which increases

retention time and results in more efficient digestion (Mercier et al., 1976). The primary structure of the area surrounding the cleavage site is well conserved between mammalian species. Release of the polar CMP portion from κ -casein –occurring on the surface of micelles- eliminates the electrostatic and steric stabilization of the micelle surface allowing micelles to associate and leading to clot formation (Swaisgood, 1992). Comparisons of CMP of several species suggest that κ -casein could be classified into two groups on the basis of hydrophobicity, carbohydrate content, amino acid composition, and site of proteolytic cleavage (Mercier et al., 1976a). In the first group (cow, sheep, goat, and water buffalo) of κ -caseins the chymosin-sensitive bond is Phe-Met. The cleavage site is specified by Phe-Ile or Phe-Leu in the second group (human, camel, mouse, rat, pig) of κ -caseins. The recent characterized marsupial κ -casein appears to form a separate class with a putative chymosin cleavage site of Phe-Ala, which is different from that found in eutherian mammals (Stasiuk et al., 2000). This divergence may reflect differences in the mechanisms of milk clotting between mammalian species (Ginger and Grigor, 1999).

Ruminant κ -caseins contain two cysteine residues at positions 11 and 88, and the protein normally occurs in polymeric form via disulphide bonds, ranging in sizes from 60 to 600 kDa (Wong et al., 1996). This polymeric nature of κ -casein appears to facilitate its role in covering the micellar surface, thus stabilizing the micelle structure (Rasmussen et al., 1994). During heat treatment, κ -casein is attached through disulphide links to the whey proteins and α s2-casein, which increase the micelle surface and affect technological properties of the milk (Fox, 1992; Dagleish, 1992). Alpha-s1- and β -caseins contain no cysteine residues.

In addition to its function as source for amino acids, phosphate, and calcium, the κ -casein –like other caseins- is precursor of biologically active, opioid-like peptides (revised in Miesel, 1997; Clare and Swaisgood, 2000). These “*bioactive*” peptides appear to act as physiological modulators of various digestive and metabolic processes, such as immune defense, nutrient uptake, and neuroendocrine information transfer.

The caprine κ -casein was first isolated by Zittle and Custer (1966) and its amino acid composition determined by Richardson et al., (1973). Subsequently, the complete amino acid sequence of 171 residues was established (Mercier et al., 1976a, b). The main amino acid differences between caprine and bovine κ -caseins are located in the C-terminal

portion of the protein. Compared to their bovine counterpart, ovine and caprine κ -caseins have in common the insertion of two amino acid residues Val-His between positions 131 and 132.

3.1.2 Kappa casein gene

Caseins are encoded by four tightly linked and clustered genes, covering an area of approximately 250 Kb genomic DNA fragment (Figure 4). The structure and organization of the casein gene locus have been described in human, mouse, and bovine (Fujiwara et al., 1997; Rijnkels et al., 1997a,b) and mapped on chromosome 6 in bovine and caprine species (Threadgill and Womack, 1990; Hayes et al., 1993). Despite some differences in the distance separating casein genes and their number (with an extra α s2-like gene in mouse), the overall genomic organization of the locus is conserved between mammalian species. The order of the casein genes in the cluster is α s1- β - α s2- κ , with the κ -casein located in a region 95-120 Kb downstream of the α s2-gene, and about 200 kb from the α s1-gene in bovine genome (Rijnkels et al., 1997a). A further consequence of casein genes clustering is that the four casein *loci* are considered as one “genetic unit” in which alleles are tightly linked together and transmitted as a *haplotype* rather than individual alleles. Therefore, the existence of genetic linkage have to be considered when selecting for an allele at given casein locus, or when studying the association between polymorphisms and production traits.

Kappa casein cDNA have been characterized in several species, including cattle (Gorodetskii and Kaledin, 1987), sheep (Furet et al., 1990) and goat (Coll et al., 1993). The goat κ -casein mRNA contains an open reading frame of 579 bp coding for 21 amino acids of signal peptide and 171 amino acids of mature protein. The signal peptide of κ -casein is different in both length and amino acid sequence from the consensus sequence of the calcium-sensitive genes. The structure of the gene is also quite different, and has been described in human, bovine, goat, and rabbit (Edlund et al., 1996; Alexander et al., 1988; Coll et al., 1995b; Baranyi et al., 1996). The caprine *k-casein* gene spans approximately 14 Kb and comprises four introns ranging from 1.8 kb to 6.8 Kb (Figure 5). The coding sequence for mature protein is contained in exon three (9 amino acids) and four (162 amino acids). The *k-casein* gene is the only milk protein gene whose signal peptide is encoded by two exons (2 and 3). The 3'-untranslated region is contained within exons 4 and 5.

Figure 3. Overall structural organization of the casein locus in bovine, human, and mouse species. The genes encoding the four casein are depicted by boxes. Orientations of the transcription of genes are indicated by arrows.

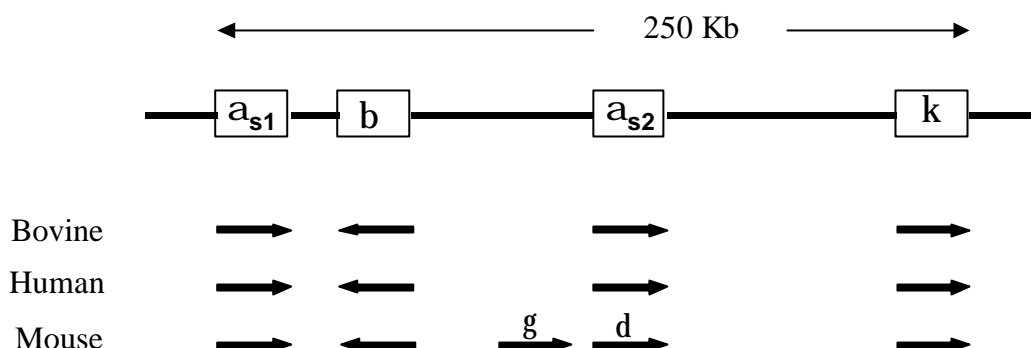
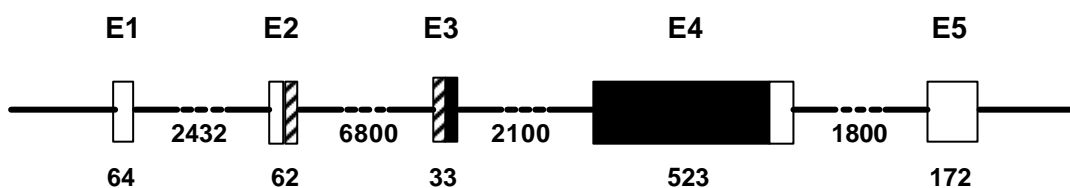


Figure 4. Structure of the caprine *kappa casein* gene. Exons are depicted schematically as boxes, white (5' and 3' untranslated regions), dashed (part of exon encoding the signal peptide), and black (part of exons encoding the mature protein). Exon numbers and sizes are indicated above and below the boxes, respectively.



Repetitive elements have been identified in the 5' flanking region of the *k-casein* gene from different species (Coll et al., 1995a; Gerencser et al., 2002). The caprine κ -casein promoter contains two types of repetitive elements: a 206-bp SINE (short interspersed nuclear element) and a 114-bp fragment of a LINE (long interspersed nuclear element). Similar repetitive elements have also been found in introns 2 and 3 (Coll et al., 1995b). The bovine *k-casein* gene contains a microsatellite repeat in intron 3 with six alleles (Lien and Rogne, 1993; Leveziel et al., 1994).

There is some evidence indicating that the three calcium-sensitive genes have evolved from a common ancestral gene by events such as gene duplication and exon shuffling (Jones et al., 1985; Bonsing and Mackinlay, 1987). In contrast, the *k-casein* gene appears to have evolved along a different pathway, since it does not share any common pattern with other casein genes. The *k-casein* gene was postulated to be evolutionary related to the fibrinogen (γ -chain) gene family whose cleavage by thrombin results in blood

clotting (Jolles et al., 1978). This hypothesis is sustained by the structural and functional similarities between the proteins, and by the nucleotide sequence similarities between κ -casein and γ -fibrinogen cDNAs (Jolles and Henschen, 1982; Thompson et al., 1985).

Although the *k-casein* gene is not evolutionarily related to the genes encoding the calcium sensitive caseins it is physically and functionally linked to them, and the four genes are coordinately expressed at high levels in a tissue- and stage-specific fashion. Thus, the expression pattern of the κ -casein seems to be similar to that of the other caseins in spite of the different organization of its 5' flanking region. Nevertheless, κ -casein genomic clones (from goat, cow, and rabbit) were either nonfunctional (Ninomiya et al., 1994; Rijnkels et al., 1995) or were poorly expressed (Persuy et al., 1995; Baranyi et al., 1996) in transgenic mouse lines under their own regulatory sequences. In contrast, *k-casein* gene has been shown to be expressed in the mammary gland of transgenic mice (Persuy et al., 1995; Gutierrez et al., 1996) and transgenic cattle (Brophy et al., 2003) when linked to the β -casein regulatory sequences. These observations suggest that regulatory elements might be involved in the expression of the entire casein gene locus (may be located in the 5' proximal region of the cluster), analogous to the locus control region (LCR) described for the *b-globin* gene cluster (Grosveld et al., 1987; revised in Li et al., 2002).

3.1.3 Polymorphism of the *k-casein* gene

Owing to the importance of κ -casein in the technological properties of milk, the polymorphism in the *k-casein* gene have been extensively studied in ruminants. Kappa casein is widely polymorphic in bovine specie with nine genetic variants characterized to date (Table 2). The most common variants A and B were early detected by paper electrophoresis and by using reducing agents such as β -mercaptoethanol in order to break disulphide bonds and to reduce polymeric forms to monomers (Neelin, 1964; Schmidt, 1964; Woychik, 1964). The variants were found with variable frequency in all analyzed populations including in zebu (*Bos indicus*) and in yak (*Bos grunniens*) (Ng-Kwai-Hang and Grosclaude, 1992). Variant A differs from variant B by two amino acid substitutions at positions 136 (Thr \rightarrow Ile) and 148 (Asp \rightarrow Ala), both occurring in the CMP region of the protein (Grosclaude et al., 1972). The first mutation affects glycosylation site. The two variants can be genotyped by PCR-RFLP using *Hind*III or *Hinf*I endonucleases (Medrano and Aguilar-Cordova, 1990a; Pinder et al., 1991).

Variants C and E have been characterized at amino acid level (Miranda et al; 1993), and other alleles were identified by sequencing the corresponding PCR product (Prinzenberg et al., 1996; 1999; Mahe et al., 1999) (Table 2). Variant H has been previously described and characterized in zebu (Grosclaude et al., 1972). The six κ -casein variants (A, B, C, E, F, and G) can be genotyped by PCR-RFLP using *Hind*III (or *Hinf*I), *Hae*III, *Hha*I, and *Mae*II endonucleases (Schielben et al., 1991; Prinzenberg et al., 1996). Alternatively, a PCR-SSCP procedure has also been described (Barroso et al., 1998; Prinzenberg et al., 1999).

As aforementioned, the most diffused κ -casein alleles are A and B. Variant A is predominant in a large majority of breeds while B variant is prevalent in beef cattle breeds. Kappa casein C is less common but was found in many breeds. Other alleles are rare and their presence is often limited to local breeds.

The impact of bovine κ -casein variants A and B on milk production traits have been extensively studied, and most results agree in indicating that milk from cows genotyped κ -casein BB contains higher proportions of fat, proteins, and caseins than milk derived from κ -casein AA cows (revised in Ng-Kwai Hang, 1998; Di Stasio and Mariani, 2000). The B allele is significantly associated with higher casein and lower whey protein contents, resulting in a higher ratio of caseins to total proteins (casein number). The bovine κ -casein BB genotype has been also associated with the production of milk with superior manufacturing properties, e.g. shorter rennet coagulation time, formation of a firmer curd, and in higher cheese yield (Schaar, 1984; Marziali and Ng-Kwai Hang, 1986). This effect is associated with the milk casein micelles. The κ -casein B is characterized by a more homogenous micellar pattern with a larger proportion of small micelles (Morini et al., 1975) resulting in a larger micelle surface area which allows the formation of a firmer and more consistent curd.

Quantification of CMP from A and B alleles in the milk of heterozygous animals has revealed a differential content of the two allele-products, with more protein variant encoded by allele B than that encoded by allele A (Van Ennenman and Medrano, 1991). The differential expression could be related to variants within potential regulatory sites in the 5' flanking region, possibly involved in the expression of the *k-casein* gene (Shild et al., 1994; Martin et al., 2002).

Table 2. Bovine kappa casein variants.

Variants	Amino acid positions							References
	10	97	104	135	136	148	155	
A	Arg	Arg	Ser	Thr	Thr	Asp	Ser	Neelin (1964), Schmidt (1964), Woychik (1964)
B					Ile	Ala		
C		His			Ile	Ala		
E							Gly	DiStasio and Merlin (1979) Erhardt (1989)
F	His							
G		Cys						Ikonen et al. (1996)
H				Ile				Erhardt (1996)
I			Ala					Prinzenberg et al. (1999)
J					Ile	Ala	Arg	Prinzenberg et al. (1999) Mahe et al. (1999)

The κ -casein seems to be monomorphic in sheep and no variants were detected in protein electrophoresis (Alais and Jolles, 1961; Soulier et al., 1974). Possible polymorphism was suggested using chromatographic techniques (Addeo et al., 1992) but no variants have been characterized. At DNA level, restriction analysis showed a two-allele polymorphism with *HindIII* and *PvuII* enzymes (Di Gregorio et al., 1991) and three-allele polymorphism with *PstI* endonuclease (Leveziel et al., 1991).

Although several studies using different techniques have reported polymorphism in the caprine κ -casein, no variants had been characterized at the beginning of the present work. Recently, κ -casein variants have been characterized in different goat breeds (Yahyaoui et al., 2001; Caroli et al., 2001; Angiolillo et al., 2002). First evidence of genetic polymorphism was reported by Mercier et al. (1976) when sequencing the CMP from the milk of an Alpine-Saanen goat. They found either valine or isoleucine at position 119 and considered that the animal sequenced was heterozygous at this position. Isoleucine residue was postulated to be predominant since it was found in the sequence of bovine and ovine species. This neutral substitution was also observed by the analysis of the amino acid composition of κ -casein prepared from milk of Saanen goat (Addeo et al., 1978).

The occurrence of goat κ -casein polymorphism has also been demonstrated by protein electrophoresis (Russo et al., 1986; Di Luccia et al., 1990), chromatographic techniques (Jaubert and Martin, 1992; Law and Tziboula, 1993) and capillary electrophoresis (Recio et al., 1997). However, none of these polymorphisms has been

characterized. Therefore, they might correspond to the same polymorphism as well as different polymorphisms or heterogeneous forms due to post-translational modifications. The variant reported -in local Italian goat- by Di Luccia et al. (1990) effectively results from amino acid substitution in the protein since the κ -casein electrophoresis was carried out after chymosin treatment and phosphate content was analyzed. This variant is probably identical to the polymorphism detected in British Saanen and local Greek goats by using chromatography and polyacrylamide gel electrophoresis (Law and Tziboula, 1993). It has been localized in the N-terminal region of the protein and it occurs at similar frequencies in the two analyzed populations ($\approx 12\%$). This polymorphism was later confirmed and characterized in Italian and German goats (Caroli et al., 2001). In Spanish goat breeds, a polymorphism detected by protein electrophoresis has been described in Murciano-Granadina and Payoya goats (Diaz-Carillo, 1993) with a frequency of 0.29 ($n = 76$). The polymorphism seems to be associated with effects on milk production traits (Angulo et al., 1994). The analysis of caprine caseins by capillary electrophoresis above mentioned (Recio et al., 1997) was carried out on milk from Granadina goat.

At DNA level, RFLP genetic markers were found in goat genomic DNA region containing *k-casein* gene after digestion with *Bam*HI, *Eco*RV, and *Pvu*II endonucleases and hybridization with a bovine κ -casein cDNA probe (Di Gregorio et al., 1989).

During this thesis, we have reported first characterization of genetic variants in the caprine *k-casein* gene in Spanish and French goat breeds (Yahyaoui et al., 2001). Additional variants are reported and protocols for rapid and easy genotyping for all known variants are described (see results and discussion chapters).

3.2 Beta-lactoglobulin

3.2.1 Beta-lactoglobulin structure and function

Beta-lactoglobulin (β -LG) is the major whey protein in the milk of ruminants. It is also found in the milk of other mammals, but is absent from the milk of rodents, lagomorphs, humans, and probably camels (Hambling et al., 1992). In ruminant milk, the native protein was found as a dimer with a molecular weight of 36.4 kDa corresponding to 162 amino acids. In most other species in which it has been found β -LG appears to be monomeric (Hambling et al., 1992). Under physiological conditions (pH 6.5), the protein presents a globular and compact structure and it tends to dissociate into monomers at low

and high pH (below 3.5 and above 7.5). β -LG was among the first proteins to be crystallized and its secondary and tertiary structures are well characterized (Papiz et al., 1986; Monaco et al., 1987).

Comparison of the amino acid sequences of bovine, ovine, and caprine β -LGs shows that the three proteins are highly homologous with sequence identity greater than 95%. The latter two proteins differ from bovine β -LG only at six positions. All three proteins contain five cysteine residues, four of which form intrachain disulphide bridges with the fifth. The primary sequence of the caprine β -LG was established by Preaux et al. (1979).

β -LG belongs to the lipocalin protein family, constituted by small secreted proteins which are characterized by their affinity to bind hydrophobic molecules. Although lipocalins have been classified primarily as transport proteins, they are implicated in several biological processes such as retinol and pheromone transport, synthesis of prostaglandins, immune response, and cell homeostasis (revised in Flower, 1996). Members of the lipocalin family show a great diversity at the protein sequence level (only 20-30% of similarity), however, stretches of sequence that are important in defining the three-dimensional structure of the proteins are highly conserved. The lipocalin family include, among others, apolipoprotein D, retinol-binding protein, prostaglandin synthase, and protein pregnancy 14 (Ali and Clark, 1988).

The true physiological function of the β -LG is still unknown although several suggestions have been made (revised in Sawyer and Kontopidis, 2000). The fact that the protein is quite resistant to gastric proteases *in vitro* and *in vivo* suggest that the primary function of β -LG is not nutritive (Reddy et al., 1988). The protein remains mostly intact after digestion in the stomach and trace amounts can be detected in human milk after consumption of bovine milk (Yvon et al., 1984; Monti et al., 1989). Because β -LG is a lipocalin, a transport role of retinol and fatty acids has been suggested. However, binding of β -LG to these molecules appears to be due to the general affinity of the protein to hydrophobic molecules and does not constitute a specific role (Puyol et al., 1991). β -LG could play role in lipid metabolism through activation of lipases by uptaking free fatty acids (Perez and Calvo, 1995). Nevertheless, monomeric β -LG from non ruminant species contain no fatty acids and appears to be unable to bind with them (Perez et al., 1993). Other possible functions for the protein are transfer of passive immunity from mother to offspring

(Ouweland et al., 1997) and source of active peptides similar to those produced by caseins (Mullally et al., 1997).

3.2.2 *Beta lactoglobulin gene*

The cDNA encoding for β -LG has been sequenced in ruminant species: bovine (Alexander et al., 1989), ovine (Gaye et al., 1986) and caprine (Folch et al., 1993). As already noted the coding sequence (486 bp coding for 162 amino acids) is highly conserved in the three species. The signal peptide is 18 amino acids in goat and sheep and 16 amino acids in cow. The complete sequence for *b-LG* gene is also known in bovine (Alexander et al., 1993), ovine (Ali and Clark, 1988; Harris et al., 1988), and caprine (Folch et al., 1994) species. The structural organization of the gene is very conserved in the three species. The transcription unit spans 4.7 Kb arranged in seven small exons (42 to 178 bp in goat) and six introns ranging from 213 bp to 1116 bp (Figure 6). The first exon contains part of the 5' untranslated region and the sequence encoding for the signal peptide and the first fourteen amino acids. The 3' untranslated region is included in exons 6 and 7. The goat *b-LG* gene has been assigned by *in situ* hybridization to chromosome 11q23 (Hayes and Petit, 1993) and is located in a GC-rich region of the genome (Bernardi et al., 1985); the GC content of the gene is 60.4% (Folch et al., 1996). *b-LG* pseudogenes have been described in both bovine (Passey and Mackinlay, 1995) and caprine (Folch et al., 1996) species. The main alteration in the unprocessed caprine pseudogenes is an insertion of 29 bp in exon two generating a downstream stop codon in the same exon. In the milk of cat, horse, and donkey several related forms of β -LG protein (β -LG I, β -LG II, β -LG III), probably produced by different genes, have been isolated and characterized (Godovac-Zimmerman et al., 1990; Halliday et al., 1991; Pena et al., 1999). Comparison of the predicted ancestral protein translated from the caprine pseudogene shows a higher amino acidic similarity of the β -LG II forms reported in these species than to the ruminant β -LG. This finding suggests the β -LG pseudogene and the genes encoding for the β -LG II forms have probably evolved from a common ancestral gene.

3.2.3 *Beta lactoglobulin polymorphism*

Bovine β -LG is the first milk protein in which polymorphism was detected (Aschaffenburg and Drewry, 1955). The protein is highly polymorphic in bovine species with a dozen variants detected to date (Godovac-Zimmerman et al., 1996), mostly by their

different electrophoretic mobility. Table 3 shows all known and characterized variants. By far, the most common genetic variants in *Bos taurus* and *Bos indicus* are A and B forms that differ by two amino acid substitutions at positions 64 (Asp-Gly) and 118 (Val-Ile). The B variant is considered as the ancestral type on the basis of sequence and frequency criteria (Ng-Kwai Hang and Grosclaude, 1992). The two variants can be genotyped by PCR-RFLP using *Hae*III endonuclease (Medrano and Aquilar-Cordova, 1990), by AS-PCR (Lindersson et al., 1995), and by DGGE (Tee et al., 1992). Apart from A and B variants, other alleles are less common and were found in general at low frequencies and only in one breed or local populations. Variant Dr, detected in the Australian Draughtmaster breed (Bell et al., 1970), is the only glycosylated form observed in β -LG. Carbohydrate moieties are attached to the aspartic acid residue at position 28, and contain, in addition to glycosidic residues found in kappa casein, acetyl-glucosamine and mannose (Eigel et al., 1984).

Bovine β -LG polymorphism has been associated with an important effect on milk composition and cheese yield. For instance, allele B is associated with higher casein and fat contents in milk, while milk from animals genotyped AA was shown to contain more whey proteins and total protein than those of the genotypes AB and BB (revised in Ng-Kwai Hang, 1998). The higher content of protein in milk from AA animals is due to higher amounts of β -LG. In fact, a higher expression of allele A has been described in heterozygous animals AB where quantification of β -LG protein in milk shows a relative proportion of 60 and 40% for A and B forms, respectively (Graml et al., 1989; Lum et al., 1997). This differential expression is attributed to allele-specific nucleotide mutations within the AP-2 (activator protein 2) transcription factor in the promoter region of the gene (Lum et al., 1997; Folch et al., 1999).

In sheep, the β -LG is polymorphic in all analyzed breeds with two variants, A and B, variant A is normally predominant (King et al., 1969; Chiofali and Micari, 1987). The two variants differ by a single amino acid substitution (Tyr-His) at position 20 of the protein (Gaye et al., 1986) (Table 4). A third variant (β -LG C) was detected in German and Spanish merino breeds (Erhardt, 1989; Recio et al., 1995) which differ from A variant by a

Figure 5. Structural organization of the caprine *b-lactoglobulin* gene. Exons are depicted schematically as boxes, white (5' and 3' untranslated regions), dashed (sequence encoding the signal peptide), and black (sequence encoding the mature protein). Exon and intron sizes are indicated below.

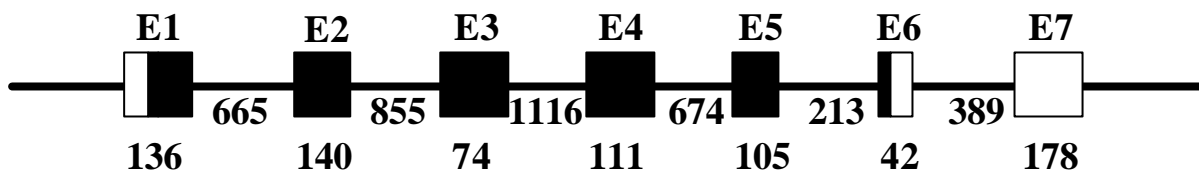


Table 3. Bovine β -lactoglobulin variants.

Variant	Amino acid positions											References
	28	45	50	56	59	64	78	108	118	129/ 130	158	
A						Asp			Val			Ascheffenburg and Drewry (1955)
B	Asp	Glu	Pro	Ile	Gln	Gly	Ile	Glu	Ala	Asp	Glu	
C					His							Bell (1962)
D		Gln										
Dr	Asn					Asp			Val			Grosclaude et al. (1966)
E											Gly	
F			Ser							Tyr	Gly	Grosclaude et al. (1976)
G							Met				Gly	
W				Leu								Bell et al. (1981)
I								Gly				Krause et al. (1988)
												Baranyi et al. (1993)

Table 4. Ovine β -lactoglobulin variants.

Variant	Amino acid positions		References
	20	148	
A	Tyr	Arg	King (1969)
B	His		
C		Gln	Erhardt (1989)

single exchange Arg-Glu at position 148. A protocol for genotyping variants A and B by PCR-RFLP with *RsaI* endonuclease have been described (Schlee et al., 1993). As regards effect of β -LG polymorphism on milk composition and dairy performances, several studies have been carried out (revised in Moiola et al., 1998) but the results are somewhat contradictory and no clearly association have been established as found in bovine specie.

In goat, the β -LG have been considered to be monomorphic and no variants were reported. A polymorphism in the 3' untranslated region (exon 7) has been recently reported (Pena et al., 1999). We have detected a single nucleotide polymorphism in the proximal promoter region of the gene (see results and discussion chapters).

4. MILK FAT

In mammalian species, the major sites of lipogenesis are the intestinal mucosal cells, the liver, adipose tissue and, in lactating animals mammary gland. Intestine cells synthesize triglycerides (TG) from ingested fat while the liver has a central clearing role, uptaking and esterifying plasma fatty acids and synthesizing fatty acids *de novo* from acetyl-CoA derived from the catabolism of dietary carbohydrates, and to a lesser extent, amino acids (Vernon et al., 1999). Synthesized TG are secreted as chylomicrons (intestinal mucosa) and very low density lipoproteins (liver) and are used by other tissues. These TG can be used by adipose and mammary tissues after hydrolysis by the action of lipoprotein lipase which they synthesize and secrete. Like liver, these tissues can synthesize fatty acids *de novo*. In ruminants, adipose tissue is the major site for fatty acids synthesis, because most dietary carbohydrate is fermented to volatile fatty acids (acetate, propionate, and hydroxybutyrate) in the rumen, whereas the liver metabolism is dominated by glucose synthesis (Vernon, 1980; Ponce-Castanada et al., 1991).

4.1 Origins of milk fat

Milk fat is composed primarily of triglycerides (TG) which represent 97-98% of the total fat (Dils, 1986). The remaining components comprise small amounts of diglycerides, phospholipids, unesterified fatty acids, cholesterol, and glycolipids. Fatty acids, esterified into TG by mammary epithelial cells of lactating animals, are derived either from blood lipids or by *de novo* biosynthesis from small molecule precursors. Glucose, fatty acids, and glycerol, as well as acetate and β -hydroxybutyrate in ruminants, are the major substrates for fatty acid synthesis (revised in Neville and Picciano, 1997; Barber et al., 1997; Clegg et al.,

2001). Short- and medium-chain fatty acids are synthesized *de novo* within the mammary epithelial cells. Thus, the proportion of these fatty acids in milk reflects the contribution of mammary fatty acid synthesis to total milk fat content. In ruminants, the principal precursor of this synthesis is acetate (and at less extent β -hydroxybutyrate), whereas glucose is the principal carbon source in monogastric animals. The main metabolic pathways involve two key enzymes, acetyl-CoA carboxylase (ACC) and fatty acid synthase (FAS). ACC catalyses the conversion of acetyl-CoA to malonyl-CoA, and FAS catalyses a sequence of reactions that add a two-carbon unit derived from malonyl-CoA to a growing fatty acyl chain (Smith, 1994). The chain-termination reaction is catalyzed by the intrinsic thioesterase activity of FAS (thioesterase I) and results in the formation of C16:0 fatty acid (palmitic), the major product of the FAS reaction in the majority of eukaryotic cells (Barber et al., 1997). Fatty acids with 16 or more carbons cannot be further elongated by mammary gland and are susceptible to hydrolysis by thioesterase I. Almost all C_{4:0} to C_{14:0} fatty acids and approximately one-half of C_{16:0} fatty acid in milk are derived from *de novo* fatty acid synthesis (Grummer, 1991). Long-chain fatty acids are derived from the second major source of fatty acids, circulating blood lipids. They are released from TG in chylomicra (absorbed dietary fat) or very low density lipoprotein (mobilized from adipose tissue) by the enzyme lipoprotein lipase. Non-esterified fatty acids are also incorporated into milk fat, however, their contribution is typically low (Moore and Christie, 1981). Other minor fatty acids found in milk, such as branched-chain, odd-numbered chain, and trans-fatty acids are contributed by the rumen microorganisms or by partial biohydrogenation of dietary unsaturated fatty acids in the rumen.

Fatty acids, whether formed by *de novo* synthesis in the mammary cells or imported from the plasma, are esterified to the hydroxyl group of glycerol molecules by transacylases located in the endoplasmic reticulum to form TG. The TG are incorporated into microlipid droplets that coalesce into larger cytoplasmic lipid droplets, coated with a membrane of phospholipids and proteins, and secreted in the milk as fat globules (Mather and Keenan, 1998).

4.2 Composition of milk fat

Total lipid content of milk is widely variable among species, it ranges from 1.9% in horse milk, to 4% in humans and ruminants, to as much as 50% in seals and whales (Dils, 1986). Despite these enormous variations, the predominance of TG fraction (97-98%) in

milk lipids is remarkably consistent across mammalian species (Christie, 1983; Dils, 1986). However, there are wide differences among species in the fatty acid composition of TG fraction in both degree of unsaturation and chain lengths of fatty acids. The proportion of saturated fatty acids varies for example from 30% in seal milk to 98% in elephant milk, whereas ruminant milk contains 70-80%. The major constraint on fatty acid composition seems to be related to the need to ensure a sufficiently low melting point of TG to be readily secreted as liquid droplets and delivered to the offspring (Dils, 1986; Gibson, 1991). Increased degree of unsaturation and increased proportion of short- and medium-chain fatty acids tend to cause lower melting points.

Ruminant milk is characterized by the presence of high proportions of short-chain fatty acids (fewer than eight carbons), probably resulting from efficient extraction of their precursors from the plasma where high concentrations are due to rumen fermentation (Moore and Christie, 1979). Tissue and milk fatty acids of ruminants are more saturated than those of diet as a result of hydrogenation of unsaturated fatty acids by ruminal microorganisms. Ruminant biohydrogenation of unsaturated fatty acids is extensive, ranging from 60 to 90% for polyunsaturated fatty acids (Murphy et al., 1987). However, desaturase activity in mammary gland functions to convert saturated fatty acids to monounsaturated fatty acids, thus limiting effects of extensive hydrogenation and ensuring sufficient fluidity of milk fat for efficient secretion. Since milk is quite high in C_{18:1}, a considerable proportion of the predominantly saturated fatty acids supplied by the rumen are desaturated in the mammary gland by the action of stearoyl coenzyme A desaturase (SCD) enzyme. Ratio of C_{18:1} to C_{18:0} is typically between 2:1 to 3:1 in milk, whereas the ratio of TG-rich lipoproteins circulating in plasma is 1:2 (Grummer, 1991).

As a result of fatty acids synthesis, uptake, and desaturation, a wide range of fatty acids (more than 400 in bovine milk) have been identified in milk (Jensen, 2002). Table 5 shows major fatty acids found in bovine and caprine milks. As already mentioned, ruminant milk fat is highly saturated, and saturated fatty acids represent more than 70% of total. Caprine and bovine milks contain low proportion of polyunsaturated fatty acids (linoleic C_{18:2} and linolenic C_{18:3}) due to the biohydrogenation of unsaturated fatty acids by the rumen microorganisms. In goat milk, the five most important fatty acids in quantitative terms (C_{16:0}, C_{18:1}, C_{10:0}, C_{14:0}, and C_{18:0}) account for more than 75% of total fatty acids. Caprine milk fat is also characterized by higher values of caprylic (2.7%) and capric (10%)

acids compared to bovine milk. The C_{12:0}:C_{10:0} ratio is usually used to detect the authenticity of caprine milk (Alonso et al., 1999) with a value of 0.5 in caprine dairy products whereas it > 1 in those from bovine origin. On the other hand, the caprine milk contain a high number of minor branched-chain fatty acids, mostly monomethylates, that have potential implications for the flavor of the dairy products (Alonso et al., 1999; Morgan and Gaborit, 2001).

Table 5. Major fatty acids of caprine and bovine milks (weight %).

Fatty acids		Bovine milk ^a	Caprine milk ^b
<u>Saturated fatty acids</u>			
Butyric	C _{4:0}	4,5	2,2
Caproic	C _{6:0}	2,3	2,4
Caprylic	C _{8:0}	1,3	2,7
Capric	C _{10:0}	2,7	10,0
Lauric	C _{12:0}	3,0	5,0
Myristic	C _{14:0}	10,6	9,8
Palmitic	C _{16:0}	28,6	28,2
Stearic	C _{18:0}	12,6	8,9
Arachidic	C _{20:0}	-	0,15
<u>Unsaturated fatty acids</u>			
Palmitoleic	C _{16:1}	1,6	1,6
Oleic	C _{18:1}	21,4	19,3
Linoleic	C _{18:2}	2,9	3,2
Linolenic	C _{18:3}	0,3	0,4

^a Adapted from Neville and Piccini (1997).

^b Adapted from Alonso et al. (1999).

4.3 Genetic modification of milk fat

Several investigations have demonstrated that consumption of saturated fatty acids tend to increase plasma total and low density lipoprotein cholesterol concentrations, and constitute a risk factor for coronary and heart diseases (Ney, 1991; Hu et al., 2001). Hypercholesterolemic effects of saturated fats in human diets are largely due to lauric (C_{12:0}), myristic (C_{14:0}), and palmitic (C_{16:0}) fatty acids (Ney, 1991; Grummer, 1991; Hu et al., 2001). These three fatty acids constitute more than 40% of total fatty acids in ruminant milk (Table 5). From the point of view of human health, the ideal nutritional fatty acid composition for milk fat would contain 82% of monounsaturated fatty acids rather than 25%; the increase would occur at the expense of saturated fatty acids (Grummer, 1991). Thus, the increase of unsaturated fatty acids proportion in milk constitutes an important selection objective, to make a healthier product. It also offers the additional benefit of a softer and more marketable butter arising from the change in the fatty acid profile of milk fat.

The total concentration of milk TG, although highly variable between species, is relatively constant in a given specie. On the contrary, the composition of the fatty acids of TG is highly variable among individuals and depends on several factors that can be divided into two groups, animal and feed factors (Palmquist et al., 1993; Demeyer and Doreau, 1999). Animal factors includes genetics (breed and selection), stage of lactation, mastitis; and feed factors comprises grain and energy intakes, dietary fats, and seasonal and regional effects (Palmquist et al., 1993). Particularly, the composition of the diet (lipid content and grain/forage ratio) has marked effects on the fatty acid composition of milk. Nevertheless, unsaturated fat in diet should be protected against rumen biohydrogenation, generally by heat and/or chemical treatments (Chilliard et al., 2000). Change in milk fat composition resulting from supplemental fats is often accompanied by other changes, including protein content, urea and calcium concentrations that can affect cheese yield and heat stability of milk (Palmquist et al., 1993).

As regards genetic component, there are a few data on the amount of within-breed genetic variation in fatty acid composition (revised in Gibson, 1991; Palmquist et al., 1993). These data indicate modest variation among bovine breeds in fat composition. The molar proportion of fatty acids components of total fat present a moderate heritability (around 0.3) with a coefficient of variation in the range of 0.05-0.2 indicating that genetic

response should be possible. Genetic correlations with fat percentage were positive for the proportion of short fatty acids and negative for long-chain fatty acids. In addition, there is a negative correlation between short-chain and unsaturated C₁₈ fatty acids. Thus, fatty acid composition of milk can be changed through selection. However, changes in fat composition that would improve the quality of milk would often be detrimental to other products.

An alternative approach to the conventional breeding (quantitative genetics) that can contribute to the modification of milk fat composition is the study of candidate genes that are known to be involved in the determinism of this trait. In this purpose, a prime candidate gene is the gene coding for mammary enzyme stearoyl coenzyme A desaturase (SCD).

4.4 SCD function

SCD, also known as Δ^9 desaturase, is an iron-containing enzyme involved in the biosynthesis of the monounsaturated fatty acids. It belongs to the fatty-acyl desaturases super-family of enzymes, representatives of which are found in all animals and plants, which function is the introduction of double bond (or unsaturation) into fatty acyl chains (Tocher et al., 1998; Los and Murata, 1998). Desaturation of fatty acids in animals takes place in the endoplasmic reticulum via an aerobic process. SCD catalyses the Δ^9 -*cis* desaturation of a spectrum fatty acyl-CoA substrates with a preference for palmitoyl- and stearoyl-CoA, which are converted into palmitoleoyl- and oleoyl-CoA, respectively (Enoch et al., 1976). The resulting monounsaturated fatty acids are substrates for incorporation into membrane phospholipids, triglycerides, and cholesterol esters. The introduction of the *cis* double bond is catalyzed by a set of microsomal electron-transport proteins composed sequentially of NADH-cytochrome *b5* reductase, cytochrome *b5*, and the terminal SCD which is rate-limiting in the reaction.

The activity of SCD is reflected on the fatty acid composition of phospholipids and triglycerides and is affected by several factors including diet, hormones, temperature, metals, development processes, and peroxisomal factors (Ntambi, 1995; Zhang et al., 1999). Unsaturated fatty acids are key components of cellular membranes and are involved in many processes, including energy metabolism, cell growth and differentiation, apoptosis, and signal transduction (De Vries et al., 1997; Gyorfy et al., 1997). In ruminants, oleic acid is the major unsaturated fatty acid found in adipose tissue lipid stores and in milk triglycerides. The proportion of saturated and monounsaturated fatty acids in membrane

phospholipids directly affects the membrane fluidity of the cell and its physical properties. Alterations in the ratio of stearic acid to oleic acid have been implicated in a variety of diseases states, such as cancer, neurological diseases, diabetes, obesity, immune disorders, and cardiovascular diseases (Ntambi, 1999).

4.5 SCD mRNA and gene

The nucleotide sequence of SCD have been isolated and characterized from many organisms and species (Tocher et al., 1998; Zhang et al., 2001) including mouse (Ntambi et al., 1988; Kaestener et al., 1989; Zheng et al., 2001), bovine (Chung et al., 2000), ovine (Ward et al., 1998), caprine (Bernard et al., 2001; Yahyaoui et al., 2002), and human (Zhang et al., 1999). At the levels of nucleotide and amino acid sequences, the *SCD* gene family shows a high level of identity of the open reading frame among the orthologous genes and presents a conserved amino acid residues (and their spacing) involved in the catalysis of the $\Delta 9$ desaturation reaction (Shanklin et al., 1994; Zhang et al., 2001). However, the major difference between species is the variable number of *SCD* genes. While sheep (Ward et al., 1998), human (Zhang et al., 1999), and *Saccharomyces cerevisiae* yeast (Stukey et al., 1989) have a single *SCD* gene, other species such as *Drosophila melanogaster* (Dallerac et al., 2000), *Caenorhabditis. elegans* (Watts and Browse, 2000), *Mortierella alpina* yeast (Wongwathanarat et al., 1999), mouse, and rat (Thiede et al., 1986; Mihara, 1990) present more than one structural *SCD* gene.

The physiologic significance of having many SCD isoforms is currently unknown but could be related to the lipid metabolism characteristics in each specie. There is evidence to suggest that SCD proteins have different substrate specificity (in mouse and *C. elegans*) and are regulated through tissue-specific expression. In mouse, the SCD activity results from the transcription of three related genes (*SCD1*, *SCD2*, and *SCD3*) expressed in a tissue specific fashion (Ntambi et al., 1988; Kaestener et al., 1989; Zheng et al., 2001). Nucleotide sequence comparison of the three mouse SCD cDNAs revealed that the homologous region is limited only to the protein coding sequence (more than 88% identity), with no significant homology in either the 5' or the 3' noncoding regions. The SCD cDNA sequence from human, mouse, rat, and goat contain unusually long 3' untranslated regions (3'-UTR) that are derived from single exons. The role of such a long 3'-noncoding stretch is currently unknown but it may influence the mRNA stability and thereby plays a role in the regulation of expression of *SCD* genes (Jackson, 1993). In human, the 3'-UTR region contain two

functional polyadenylation signals resulting in two mRNAs transcripts (of 3.9 and 5.2 Kb) that are differentially expressed in different tissues (Zhang et al., 1999).

Ruminant SCD proteins consist of 359 amino acid residues with a high level of similarity (more than 94%) between the three species (see results and discussion chapters). The structural organization of the *SCD* genes has been described in rodents (Ntambi et al., 1988; Kaestener et al., 1989, Mihara et al., 1990), human (Zhang et al., 1999), and recently in goat (Bernard et al., 2001; Yahyaoui et al., 2002), and is conserved within these species. The caprine *SCD* gene spans approximately 15 Kb and contains six exons and five introns with all intron-exon junctions conforming to GT/AG splicing rule.

The cluster of *SCD* genes has been assigned to mouse chromosome 19 and it is overlapped by an identified QTL which is associated with elevated plasma levels of free fatty acids and triglycerides (Tabor et al., 1998). Human *SCD* gene is located in HSA10, and a processed pseudogene in chromosome 17 has been identified (Zhang et al., 1999). In ruminant species, *SCD* genes have been mapped to ovine chromosome 22q21 and bovine and caprine chromosomes 26q21 (Campbell et al., 2001; Bernard et al., 2001).

4.6 Regulation of SCD gene expression

The expression of *SCD* genes has been mainly studied in rodents (rat and mouse). Under normal dietary conditions, mouse and rat *SCD1* genes are expressed constitutively in adipose tissue and at very low levels in liver, whereas *SCD2* genes are expressed principally in brain (Thiede et al., 1986; Ntambi et al., 1988; Kaestener et al., 1989). Under these conditions, fatty acids derived from dietary sources or *de novo* synthesis are desaturated for storage in the form of unsaturated fatty acids. Like *SCD2*, *SCD1* mRNA is expressed at lesser extent in kidney, spleen, heart, and lung in response to feeding a fat-free and high carbohydrate diet. However, hepatic *SCD1* expression is dramatically induced by this feeding schedule, correlating positively with mRNA levels (Thiede et al., 1986; Ntambi, 1992). On the contrary, *SCD2* is not expressed in liver, and brain *SCD2* is little altered by dietary conditions (Kaestener et al., 1989). In all cases, *SCD1* gene, but not *SCD2* is expressed in the liver, whereas *SCD2* gene, but not *SCD1* is expressed in the brain. The major function of *SCD2* in brain is to supply unsaturated fatty acids for the synthesis of membrane phospholipids and to ensure continuing activity during brain development (Kaestener et al., 1989; Cook and McMaster, 2002). The *SCD2* mRNA is also expressed in B cells, whereas mature T cells do not express either isoforms (Kim and Ntambi, 1999).

The mouse *SCD3* gene expression is restricted to skin indicating a unique function of this gene (Zheng et al., 2001).

The *SCD1* enzyme activity and mRNA levels are significantly increased in liver when feeding rats with various diets supplemented with saturated fatty acids and cholesterol (Garg et al., 1986; Landau et al., 1997). The *SCD1* expression is increased in order to provide oleoyl-CoA as a substrate for cholesterol esterification to prevent the toxic accumulation of free cholesterol in liver, and increases the availability of esterified cholesterol for export to other tissues in form of VLDL (very low density lipoproteins). Like the majority of genes that encode lipogenic enzymes (acetyl-CoA carboxylase, fatty acid synthase, malic enzyme, GLUT4), *SCD* mRNA levels are reduced in all tissues in response to dietary polyunsaturated fatty acids (PUFA) with the exception that *SCD2* is not repressed in the brain (Kaestener et al., 1988; Jump and Clarke, 1999). The PUFA responsive elements have been characterized in the proximal promoter regions of mouse *SCD1* and *SCD2* genes (Waters et al., 1997; Tabor et al., 1999) but they are absent in the *SCD3* promoter (Zheng et al., 2001). Inhibitory action of PUFAs on *SCD1* gene is exerted principally through mRNA stability in mouse adipocytes (Sessler et al., 1996) and in yeast (Gonzalez and Martin, 1996). The *SCD* expression is also under hormonal control, being upregulated by insulin and downregulated by glucagon in mouse (Waters et al., 1994) and chicken (Lefevre et al., 1999) livers, and in sheep adipose tissue (Ward et al., 1998). On the other hand, *SCD1* gene transcription is increased by peroxysome proliferators (compounds that cause hypolipidemia and hypocholesterolemia) such as colfibrate (Miller and Ntambi, 1996). The enzymatic activity and *SCD* mRNA levels increase significantly during differentiation of 3T3-L1 pre-adipocytes (Ntambi et al., 1988; Kaestener et al., 1989). In these cells, glucose availability directly increase *SCD* gene transcription (Jones et al., 1998). Mice lacking *SCD1* gene (*SCD1* ^{-/-}) have reduced body adiposity, increased insulin sensitivity, and are resistant to diet-induced weight gain (Ntambi et al., 2002). Compared with wild-type, *SCD1*^{-/-} mice have reduced levels of leptin and plasma insulin. Leptin was found to specifically repress RNA levels and enzymatic activity of hepatic *SCD1* (Cohen et al., 2002).

In sheep, the *SCD* gene is expressed principally in adipose tissue, liver, and in mammary gland of lactating animals (Ward et al., 1998). Transcripts are also detected in lung, muscle, kidney, heart, pancreas, and brain. The *SCD* expression in adipose tissue is

repressed during late pregnancy and lactation; mRNA transcripts are reduced from 678 ± 271 copies per cell in non-pregnant non-lactating animals to 76 ± 35 copies per cell during lactation. This suppression is thought to be due to hormonal status change (insulin and growth hormone) and insulin resistance of adipose tissue (Vernon et al., 1991; Ward et al., 1998). Monounsaturated fat content of ovine adipose tissue has been correlated to depot-specific expression of the *SCD* gene (Barber et al., 2000).

4.7 Polymorphism of SCD gene

In a study of DNA variability of lipogenesis genes on four selected lines of turkeys, Sourdioux et al. (1996) found a restriction fragment length polymorphisms with the *MspI* enzyme. The probes used for southern analysis were generated from chicken genomic DNA and include promoter and exon 1 (1st probe) and exons 4, 5, and part of exon 6 (2nd probe) of *SCD* gene. The polymorphisms seem not be significantly associated with fatness variability in analyzed lines.

A bi-allelic RFLP polymorphism using *TaqI* endonuclease has been identified in pure breed Japanese black cattle (Wilson et al., 1993). The probe used contains the complete coding sequence of the rat *SCD1* gene (Thiede et al., 1986). Frequencies of the two alleles are 0.69 and 0.39 (n=16). When sequencing bovine *SCD* gene, Medrano et al. (1998) found three single nucleotide substitutions in exon 5 resulting in two haplotypes designated A and B. Two mutations are silent base-pair substitutions corresponding to proline and tyrosine, and the third substitution produces a codon change of valine to alanine. Variants are found in Holstein, Jersey, and Brown Swiss breeds. In goat, a three base pair (TGT) deletion polymorphism in the 3' noncoding region have been found in French Alpine breed (Bernard et al., 2001). We have detected polymorphism in exon 5 and the 3' noncoding region of the *SCD* gene in Spanish and French goats (see discussion chapter).

AIMS OF THIS WORK

AIMS OF THIS WORK

The aim of this study was to analyze, using molecular biology techniques, the polymorphism of genes of interest in milk production and composition in the caprine specie. The specific aims of this study were:

- 1) To search for genetic polymorphisms in the full coding sequence of the caprine *kappa casein* gene.
- 2) To analyze the genetic polymorphisms in the proximal promoter region and exon one of the caprine *b-lactoglobulin* gene.
- 3) To isolate and characterize the caprine stearoyl-coenzyme A desaturase mRNA and gene.
- 4) To develop methods for genotyping the characterized genetic variants.

RESULTS

1. “*Genetic polymorphism of the caprine kappa casein gene*”

Yahyaoui, M.H., Coll, A., Sanchez, A. and Folch, J.M.

Journal of Dairy Research (2001) **78**: 209-216

Genetic polymorphism of the caprine kappa casein gene

M. HABIB YAHYAOU, AGUSTINA COLL, ARMAND SANCHEZ AND JOSEP M.
FOLCH*

Unitat de Genètica i Millora, Departament de Ciència Animal i dels Aliments, Facultat de
Veterinària, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain

Heading: Genetic polymorphism of the caprine kappa casein gene

* For correspondence

Josep M. Folch

Unitat de Genètica i Millora, Departament de Ciència Animal i dels Aliments, Facultat de
Veterinària, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain

Tel: + (34) 93-581 20 87

Fax: + (34) 93-581 21 06

e-mail: JosepMaria.Folch@uab.es

SUMMARY.

Polymorphism in the goat kappa casein gene was studied using the Base Excision Sequence Scanning (BESS) method and sequencing. Seven polymorphic sites, corresponding to single nucleotide transversions were detected. Three of these were silent mutations while the other four produced amino acid substitutions. The association between these polymorphic sites was investigated, which resulted in the identification of three goat kappa casein alleles, designated A, B, and C. Protocols for rapid genotyping of the C variant were developed by PCR-RFLP using *Alw44I* and *BseNI* restriction endonucleases. The occurrence of this allele was found to be very low in Spanish breeds but more frequent in the French Saanen goat. Further studies among different goat populations are necessary to establish the distribution of these alleles and their effects on the quality and functional properties of milk.

INTRODUCTION

Ruminant milk contains six major proteins that can be classified into two groups: caseins (α_{S1} , α_{S2} , β , and κ) and whey proteins (β -lactoglobulin and α -lactalbumin). The four caseins are the main component, accounting for 76-86 % of total milk protein (Swaisgood, 1992). Kappa casein (κ -Cn) is essential for micelle formation and stabilisation, and influences the manufacturing properties of milk. Cheese making is based on the cleavage of the κ -Cn Phe₁₀₅-Met₁₀₆ peptide bond by enzymes or heat. This liberates the polar domain and leads to micelle coagulation. Kappa casein, which is considered to be monomorphic in sheep, is widely polymorphic in cattle with six variants characterised being variants A and B the most common (reviewed in Kaminski *et al.* 1996). It has been reported that the B variant is associated with a higher total casein content and a higher ratio of casein to whey protein, which has a direct impact on the manufacturing properties of milk and results in a higher cheese yield (reviewed in Ng-Kwai Hang, 1998).

Caprine κ -Cn was first isolated by Zittle and Custer (1966) and its complete amino acid sequence of 171 residues was established subsequently by Mercier *et al.* (1976a, b). Due to post-translational modifications, κ -Cn appears heterogeneous in milk protein electrophoresis assays with at least 5 forms, the main one being non glycosylated (Addeo *et*

al. 1978). The nucleotide sequence of the cDNA encoding goat κ -Cn and the promoter region of the gene have been reported (Coll *et al.* 1993, 1995). The κ -Cn gene comprises 5 exons of which more than 90% of the coding region for mature protein is contained in the penultimate exon.

In goats, genetic polymorphism has been reported in the α_{S1} -Cn, α_{S2} -Cn, and β -Cn genes. Variants of the α_{S1} -Cn have been associated with differences in the protein content of milk, the ratio of casein/total protein and cheese yield (Grosclaude *et al.* 1994). Owing to the importance of κ -Cn in the technological properties of milk, the polymorphism in the κ -Cn gene has been extensively studied. Several groups have described the existence of polymorphism in the goat κ -Cn gene (Di Luccia *et al.* 1990; Jaubert & Martin, 1992; Law & Tziboula, 1993; Angulo *et al.* 1994; Recio *et al.* 1997). For instance, polymorphism has been localised to the N-terminal region of the protein (Di Luccia *et al.* 1990; Jaubert & Martin, 1992; Law & Tziboula, 1993). However, differences in the amino acid sequence between variants have not been established. At the DNA level, Di Gregorio *et al.* (1989) found genetic polymorphism in the region containing the κ -Cn gene, by digesting with *Bam*HI, *Eco*RV, and *Pvu*II endonucleases and hybridising with a bovine κ -Cn cDNA probe.

The aim of this work was to study the genetic polymorphism in the caprine κ -Cn gene. We describe, for the first time, genetic variants in exon 4 of the goat κ -Cn gene. Protocols for genotyping polymorphic sites from DNA samples have been developed, and the allelic frequencies have been evaluated in French and Spanish breeds.

MATERIALS AND METHODS

Animal samples

A total of 147 goat DNA samples were used in this study, including the Canaria, Malagueña, Murciano-Granadina, and Payoya Spanish breeds and the French Saanen breed. In addition, one DNA sample from the Spanish wild goat (*Capra pyrenaica sp. hispanica*) was analysed. DNA was isolated from blood samples following standard protocols (Ausubel *et al.* 1987).

Amplification of the caprine κ -Cn gene

A 459 bp fragment of goat κ -Cn exon 4 was amplified by PCR from DNA genomic samples using primers Kb1 and Kb2 (see Figure 1) which were designed relative to the caprine κ -Cn cDNA (Coll *et al.* 1993). Both forward and reverse primers were fluorescently labelled at the 5' end with Hex and 6-Fam, respectively. The PCR reaction was performed in a 25 μ l final volume containing 0.625 units of *Taq* DNA polymerase (Life Technologies, Rockville, MD 20849-6482, USA), 1x PCR buffer, 1.5 mM-MgCl₂, 200 μ M of each dNTP, 0.4 μ M of each primer and 100 ng of goat genomic DNA. Thermal cycling conditions were: 95°C for 5 min, 10 cycles of 97°C for 15 s, 63°C for 1 min and 72°C for 1 min 30 s, followed by 25 cycles of 95°C for 30 s, 63°C for 1 min and 72°C for 1 min 30 s, with a final extension at 72°C for 5 min.

Polymorphism detection with the BESS method

Polymorphism was detected using the BESS assay (BESS T-Scan; Epicentre Technologies, Madison, WI 53713, USA). In this method, limiting amounts of dUTP are incorporated during PCR amplification after which the product is treated with Uracil *N*-glycosylase and Endonuclease IV enzymes. By comparing the pattern of bands produced with a control, mutations where deoxythymidine is involved can be detected. Ten DNA samples from different animals were analysed per breed. The PCR reaction was performed using 200 μ M-BESS T-Scan dNTP Mix (containing 16 μ M-dUTP). For the excision reaction, 5 μ l of the PCR product were mixed with 0.5 μ l-BESS T-Scan Excision Enzyme Mix and 0.6 μ l of BESS T-Scan Excision Enzyme Buffer. The reaction was incubated for 20 min at 37°C and then placed on ice. Twenty five μ l of formamide and 1 μ l of GeneScan 500 (ROX) size standard (Perkin Elmer Applied Biosystems, Foster City, CA 94404, USA) were added before denaturing at 95°C for 3 min. Products were separated and analysed by capillary electrophoresis and fluorescent detection using a Genetic Analyzer 310 and the accompanying GeneScan software (Perkin Elmer). Since both amplification primers were labelled with different fluorescent dyes, the sense and antisense strands were screened simultaneously using this method.

DNA sequencing reactions

PCR products were purified and sequenced to obtain preliminary data. In samples where polymorphism was detected, the PCR products were cloned into plasmid vectors (pTZ18U and pTZ19U) and clones for the different alleles were isolated and sequenced. Both strands were sequenced by the dideoxy method using an automated DNA sequencer (ABI Prism 310, Perkin Elmer). Several clones of each allele were sequenced in order to eliminate any errors in the PCR reaction.

Genotyping by restriction fragment length polymorphism (RFLP)

Ten μ l of PCR product was digested with 10 U of *Alw44I* endonuclease (Roche Molecular Diagnostics, Indianapolis, IN 46250, USA) at 37°C overnight and with *BseNI* (Fermentas, Hanover, MD 21076, USA) at 65°C for 6 hours. After digestion, the resultant fragments were separated by electrophoresis in a 2 % agarose gel stained with ethidium bromide.

RESULTS

Polymorphism detection by the BESS method

The 459-bp region of κ -Cn exon 4 was amplified by PCR in 10 animals per breed and scanned with the BESS method using capillary electrophoresis. Three potential polymorphic sites were detected from which 2 were confirmed by sequence analysis at positions 471 and 509 of the cDNA sequence (Figure 1). The third detected mutation was considered to be an artefact, since it was located close to fluorescent primer where signal and background intensities are very high. To overcome this problem, the primer should not be located proximal to the analysed region. Indeed, non-specific peaks could be produced by primer-dimers and other PCR artefacts. Signal peak intensity decreases with fragment length, so peaks longer than 300-350 bp were generally weak and difficult to detect.

Nucleotide sequence analysis of κ -Cn exon 4

To confirm the polymorphism detected with the BESS TScan kit, PCR products from 9 animals of different breeds were sequenced. Previously detected mutations at positions 471 and 509 were confirmed. Moreover, sequence analysis showed five additional mutations at positions 245, 284, 309, 583, and 591 (Figure 1). Since they correspond to deoxythymidine on one of the DNA strands, these mutations are theoretically detectable by the BESS technique. However, no differences in the peak profiles for these additional mutations were detected in the analysed animals. Therefore, the sensitivity of this technique seems to have limitations in the detection of mutations, especially at the ends of the analysed region.

In total, seven polymorphic sites were detected within the amplified 459-bp region of κ -Cn exon 4 (Figure 1). All were point mutations, corresponding to base transversions, where T and G are substituted by C and A, respectively. Three sites are silent base pair substitutions corresponding to amino acid residues tyrosine, leucine, and threonine. The other four mutations produce codon changes of valine to isoleucine (at positions 309 and 471), alanine to valine (583), and serine to proline (591) (Figure 1).

Molecular characterisation of new variants for the κ -Cn gene

In order to establish the haplotypes, PCR products from 5 animals representing the different detected mutations were cloned into the plasmids pTZ18U and pTZ19U. Two independent clones from each animal were sequenced along both strands.

Table 1 shows the κ -Cn variants and corresponding amino acid differences deduced from the sequence analysis. The novel variants were designated κ -Cn A, κ -Cn B, and κ -Cn C. The κ -Cn A corresponds to the previously reported sequence of the goat κ -Cn exon 4 (Coll *et al.* 1993). The exon 4 of the κ -Cn was also sequenced in a Spanish wild goat, showing a different sequence from these three alleles of the goat κ -Cn.

The difference between κ -Cn A and κ -Cn B involves only one amino acid substitution at position 119, where the valine in variant A is substituted by isoleucine in variant B. The κ -Cn A differs from the C variant in the following amino acid substitutions: valine for isoleucine at positions 65 and 119, alanine for valine at position 156, and serine for proline at position 159. The first amino acid substitution (position 65) occurs in the N-

terminal region of the protein (para- κ -Cn), while the three other substitutions are located in the C-terminal region (caseinomacropeptide).

Genotyping by PCR-RFLP

PCR-RFLP protocols were developed for detecting the mutations that occur at positions 309 and 591, both of which produce amino acid changes. The polymorphism at position 309 (G to A) affects the *Bse*NI target site which is present in variants A and B but not in the C variant. Therefore, digestion of the 459-bp region of κ -Cn exon 4 produced three fragments of 54, 51, and 354 bp for alleles A and B whereas allele C generated only two fragments of 54 and 405 bp (Figure 2a). The single nucleotide substitution at position 591 (T to C) resulted in the appearance of an *Alw*44I recognition site in allele C. Consequently, the cleavage of allele C by this endonuclease resulted in two fragments (78 and 381 bp) while the PCR products from the A and B alleles remained intact (Figure 2b).

Genotyping 147 animals belonging to French and Spanish breeds using these two PCR-RFLP methods revealed that the following two polymorphic sites were always linked (G/A at position 309 with T/C at position 591). Thus, variants A and B can be distinguished from variant C using either of these two PCR-RFLP methods. Table 2 summaries the allele frequencies for these mutations in the goat breeds studied.

DISCUSSION

The 459-bp region of exon 4, which contains the major part of coding sequence for the goat κ -Cn (141 amino acids out of a total of 171), was amplified by PCR and screened for polymorphism using the BESS method and sequencing. Seven polymorphic sites were identified and characterised, of which two were detected by BESS and five by sequencing. The possibility of DNA polymerase errors was excluded by sequencing two independent PCR products and by the use of restriction enzyme analysis. Although all the identified mutations involved deoxythymidine on sense or antisense strands, and therefore were potentially detectable by the BESS technique, only two were identified by this method. This may be explained by the non-specific activity of the BESS excision enzyme, whose activity depends on nucleotide composition and sequence context.

Although several studies using different techniques have reported polymorphism in the caprine κ -Cn, none of the variants have been characterised. Mercier *et al.* (1976a) sequenced the caseinomacropeptide from the milk of one Alpine-Saanen goat and found either valine or isoleucine at position 119. They considered that the animal sequenced was heterozygous at this position. However, isoleucine was postulated to be the predominant residue at position 119 of the goat κ -Cn because it was found in the sequence of related species. Conversely, a valine is present at this position when the mature protein is deduced from the cDNA sequence (Coll *et al.* 1993). This amino acid sequence corresponds to the κ -Cn A reported here, while the substitution of valine by isoleucine at position 119 corresponds to the B variant. Furthermore, a variant of κ -Cn has been reported by Di Luccia *et al.* (1990) in a local Italian goat breed, and by Law and Tziboula (1993) in the British Saanen breed and in a commercial herd in Greece. It probably arose by substitution of one or more amino acids in the N-terminal region of the protein, presumably an arginine substitution for a neutral amino acid residue (Di Luccia *et al.* 1990). This putative variant is different from the alleles which have been described here. Of these alleles, only the C variant has an amino acid change in the N-terminal region, but this has no effect on the net charge of the protein.

The C variant is present at a low frequency (1 %) in the Spanish Murciano-Granadina and Canaria breeds, but is not found in either the Malagueña or the Payoya breeds, which could be due to the small number of animals analysed from these two breeds. However, the frequency of this allele is higher (11 %) in the French Saanen breed (Table 2). It is possible that the distribution of κ -Cn alleles has been influenced by selection pressure for milk production, being the French Saanen a highly milk selected breed, but it seems more plausible that genetic drift and founder effects have played a determinant role in the allelic distribution of the milk proteins genetic variants, including κ -Cn. Nevertheless, the distribution of the novel goat κ -Cn alleles remains to be evaluated in a large-scale study among different goat breeds.

Amino acids at positions 119 (isoleucine) and 159 (proline) in the κ -Cn are well conserved within different species, including cattle, sheep, pig, rat, mouse, and human. Nucleotide sequences from the different caprine alleles and from the Spanish wild goat were aligned to the sequences of κ -Cn exon 4 from related species, including sheep, cattle, pig, mouse, and human. From this alignment a consensus sequence was deduced for the

seven goat polymorphic sites. The sequence from the Spanish wild goat is identical to the obtained consensus sequence and may correspond to κ -Cn ancestral sequence. This ancestral type was not found in the animals from the goat breeds studied here. Caprine κ -Cn B exhibits the highest similarity (at five nucleotide positions) with the ancestral allele, while κ -Cn C is more divergent. The A variant shows an intermediate similarity. This is in accordance with the lower frequency of the C allele found in the different goat breeds analysed here.

Several studies have demonstrated the effects of particular genetic variants on the composition and physio-chemical properties of milk. These include variants of the goat α_{S1} -Cn, and of the bovine β -lg and κ -Cn genes. The bovine κ -Cn genotype is already listed in some artificial insemination catalogues and selection for κ -Cn B is practised in several countries (Ng-Kwai Hang, 1998). The influence of the novel caprine κ -Cn genetic variants reported here on the quality and technological properties of milk still remain to be evaluated in further studies. The PCR-RFLP method reported here should be useful in genotyping large numbers of animals.

This work was supported in part by Comissionat per a Universitat i Recerca (SGR 00329) and by a research scholarship to M.H.Y. from CIHEAM (Instituto Agronómico Mediterráneo de Zaragoza, Spain). We are thankful to S. Boa for critical reading of the manuscript, and to M. Amills and J. Capote for providing some of the goat DNA samples analysed in this study.

REFERENCES

- ADDEO, F., SOULIER, S., PELISSIER, J. P., CHOBERT, J. M., MERCIER, J. C., & RIBADEAU-DUMAS, B. 1978 Preparation and fractionation of goat kappa-casein: analysis of the glycan and peptide components. *Journal of Dairy Research* **45** 191-196
- ANGULO, C., DIAZ-CARILLO, E., MUÑOZ, A., ALONSO, A., JIMENEZ, I., & SERRADILLA, J. M. 1994 Effect of electrophoretic goat's κ -casein polymorphism on milk yield and main components yield. *V World Congress on Genetic Applied to Livestock Production Guelfh* **19** 333-336
- ALEXANDER L. J., STEWART, A. F., MACKINLAY, A. G., KAPELINSKAYA, T. V., TKACH, T. M., & GORODETSKY, S. I. 1988 Isolation and characterisation of the bovine κ -casein gene *European Journal of Biochemistry* **178** 395-401
- AUSUBEL, F. M., BRENT, R., KINGSTON, R. E., MOORE, D. D., SEIDMEN, G. G., SMITH, J. A., & STRUHL, K. (Ed.) 1987 *Current Protocols in Molecular Biology* Green publishing Associates and Wiley-Interscience
- COLL., A., FOLCH, J. M., & SANCHEZ, A. 1993 Nucleotide sequence of the goat κ -casein cDNA *Journal of Animal Science* **71** 2833
- COLL., A., FOLCH, J. M., & SANCHEZ, A. 1995 Structural features of the 5' flanking region of the caprine κ -casein gene *Journal of Dairy Science* **78** 973-977
- DI GREGORIO, P., RANDO, A., RAMUNNO, L., MASINA, P & PIERAGOSTINI, E. 1989 (Polymorphism of the DNA region containing casein genes in sheep and goat) *XXIV Simposio Internazionale Zootecnica* pp 275-282
- DI LUCCIA, A., MAURIELLO, R., CHIANESE, L., MOIO, L., & ADDEO, F. 1990 Kappa casein polymorphism in caprine milk *Scienza e Tecnica Lattiero-Casearia* **41** 305-314.
- GROSCLAUDE, F., RICORDEAU, G., MARTIN, P., REMEUF, F., VASSAL, L., & BOUILLON, J. 1994 (From gene to cheese: the polymorphism of the caprin α s1-casein, its effects and evolution) *INRA Productions Animales* **7** 3-19
- JAUBERT, A. & MARTIN, P. 1992 Reverse phase HPLC analysis of goat caseins. Identification of α s1 and α s2 genetic variants. *Lait* **72** 235-247.

- KAMINSKI, S. 1996 Bovine κ -casein gene: molecular nature and application in dairy cattle breeding *Journal of Applied Genetics* **37** 179-196
- LAW, A. J. R., & TZIBOULA A. 1993 Fractionation of caprine κ -casein and examination of polymorphism by FPLC *Milchwissenschaft* **48** 68-71
- MERCIER, J. C., ADDEO, F., & PELISSIER J. P. 1976a (Primary structure of the caprine caseinomacropptide) *Biochimie* **58** 1303-1310
- MERCIER, J. C., CHOBERT, J. M., & ADDEO, F. 1976b Comparative study of the amino acid sequences of the caseinomacropptides from seven species *FEBS Letters* **72** 208-214
- NG-KWAI-HANG, K. F. 1998 Genetic polymorphism of milk proteins: relationships with production traits, milk composition and technological properties *Canadian Journal of Animal Science* **78** (Supplement) 131-147
- RECIO, I., PEREZ-RODRIGUEZ, M. L., AMIGO, L., & RAMOS, M. 1997 Study of the polymorphism of caprine milk caseins by capillary electrophoresis *Journal of Dairy Research* **64** 515-523
- SWAISGOOD, H. E. 1992 Chemistry of the caseins. In *Advanced Dairy Chemistry- 1. Proteins*, pp. 63-110 (Ed. P. F. Fox). London: Elsevier Applied Science
- ZITTLE, C. A., & CUSTER, J. H. 1966 Identification of the κ -casein among the components of whole goat casein *Journal of Dairy Science* **49** 788-791

Figure 2

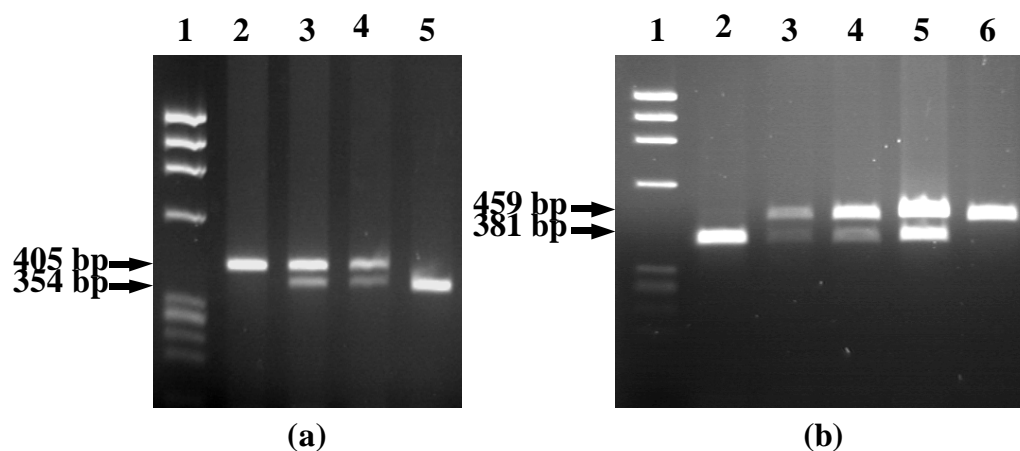


Fig.1. (a) Structure of the caprine κ -Cn gene. Exons are depicted schematically as boxes. Exon numbers and sizes are indicated above and under the boxes, respectively. The primers used for amplification of exon 4 are shown (arrows). □ Non coding region, ▨ Signal peptide, ■ mature protein.

(b) Nucleotide sequence of the caprine κ -Cn exon 4 (GeneBank accession number X60763). The numbering on the left of the sequence corresponds to that of the GeneBank database (referred in the text). The location of amplification primers Kb1 and Kb2 is indicated. Boldface letters indicate the mutations; sites and the corresponding substitutions are shown in parenthesis. Amino acid substitutions are indicated below the sequence.

Fig. 2. Genotyping of the mutations detected in the κ -Cn exon 4 by PCR-RFLP. **(a)** With *Bse*NI at position 309: Lane 1: ϕ X174 *Hae*III marker, lane 2: homozygous for adenine, lanes 3 and 4: heterozygous for adenine/guanine, lane 5: homozygous for guanine. **(b)** With *Alw*44I at position 591: Lane 1: ϕ X174 *Hae*III marker, lane 2: homozygous for cytosine, lanes 3, 4, and 5: heterozygous for cytosine/thymine, lane 6: homozygous for thymine.

Table 1. Variants of the goat κ -Cn gene.

Nucleotide Position†	Amino acid position†	Goat variants‡			Wild goat	Genotyping§
		A	B	C		
245	43	T	T	C	C	
284	56	G	G	A	G	
309	65	G (Val)	G (Val)	A (Ile)	G (Val)	<i>Bse</i> NI
471	119	G (Val)	A (Ile)	A (Ile)	A (Ile)	
509	131	A	A	G	A	
583	156	C (Ala)	C (Ala)	T(Val)	C (Ala)	
591	159	T (Ser)	T (Ser)	C (Pro)	C (Pro)	<i>Alw</i> 44I

† Position of polymorphic nucleotides and the corresponding amino acid residues.

‡ Nucleotides present at polymorphic positions in each variant. Amino acid changes are indicated in parenthesis.

§ Restriction enzymes used for PCR-RFLP genotyping.

Table 2. Distribution of allelic frequencies for the goat κ -Cn gene in Spanish and French breeds.

Breed	Number of animals	Allele frequencies†	
		A+B	C
Malagueña	17	1	0
Payoya	11	1	0
Canaria	48	0.99	0.01
Murciano-Granadina	38	0.99	0.01
Saanen	33	0.89	0.11

† Allele frequencies were genotyped by PCR-RFLP with *Bse*NI and *Alw*44I. The A and B alleles represent guanine and thymine at positions 309 and 591 respectively, whereas allele C has adenosine and cytosine at the same positions, respectively.

2. *“Short Communication: Characterization of a new genetic variant in the caprine kappa casein gene”*

Angiolillo, A., Yahyaoui, M.H., Sanchez, A., Pilla, F. and Folch, J.M.
Journal of Dairy Science (2002) **85**:2679-2680

SHORT COMMUNICATION

Characterisation of a new genetic variant in the caprine *k*-casein gene

A. Angiolillo¹, M.H. Yahyaoui², A. Sanchez², F. Pilla¹, J.M. Folch²

¹Department S.A.V.A., Università del Molise, Campobasso, Italy,

²Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain

Corresponding author:

Antonella Angiolillo. e-mail: angiolillo@unimol.it

ABSTRACT

A new polymorphism has been identified in the goat *k*-casein gene by evaluating genomic DNA of the local Italian breed Montefalcone. The polymorphic site consists of a single nucleotide substitution A to G at position 242 of the exon 4 and produces an amino acid substitution Asp/Gly. A PCR-RFLP protocol for rapid genotyping of the variant has been developed, using the *Hae*III enzyme. Samples from animals of Italian, Spanish and French breeds have been analysed to investigate the occurrence of the allele, which resulted exclusively present in the Montefalcone breed.

(Key words : genetic polymorphism, goat, *k*-casein)

The polymorphism of milk proteins has been investigated for more than 40 yr, and a great number of genetic variants have been identified. The influence of such variants on composition and technological properties of milk are largely confirmed (Grosclaude et al., 1994; Ng-Kway-Hang K. F., 1998).

Several studies, using various techniques, have reported polymorphisms in the caprine *k*-casein gene, but only recently have genetic variants been characterized. Seven polymorphic sites have been detected (Yahyaoui et al., 2001) that may result in three *k*-casein alleles. Discrimination of two alleles was also obtained by DNA single-strand conformation polymorphism analysis (Caroli et al., 2001).

In this study, a new polymorphism producing an amino acid substitution has been identified and characterized in exon 4 of the *k*-casein gene in Montefalcone goats in Italy. Montefalcone is one of the small native goat populations still raised in the mountainous region of Southern Italy. Montefalcone goats are small in number and in danger of extinction. The officially recorded population consists of 442 goats in 19 herds, but approximately 600 animals are believed to live in the area (Pilla et al., 2000). In recent years, concern over this population has increased, due to the peculiarity of its products—mainly cheese—and for the exploitation of marginal resources and preservation of rural landscape.

DNA samples used in this study have been isolated from blood, following standard protocols. The complete coding sequence of *k*-casein was amplified by PCR. Primers I2F

(intron 2) and I3R2 (intron 3) have been used for the exon 3 amplification and primers I3F (intron 3) and Kb2 (exon 4) for the exon 4.

I2F: 5'-ATGTATCTGTCATTTCTTGAGGTTTC-3',

I3R2: 5'-CTCATGAAAATCAACACAACCTTAGCC-3',

I3F: 5'-TCCCAATGTTGTACTTTCTTAACATC-3'

Kb2: 5'-GCGTTGTCCTCTTTGATGTCTCCTTAG-3'.

The PCR was performed in a 25µl reaction mixture consisting of 100ng goat genomic DNA, 0.625U *Taq* DNA polymerase (Life Technologies, Rockville, USA), 2mM MgCl₂, 200µM of each dNTP, 0.4µM of each primer, and 1X PCR buffer. Thermal cycling conditions were: a denaturation step of 95°C for 5 min, 10 cycles of 97°C for 15s, 63°C for 1 min and 72°C for 1 min 30sec, followed by 25 cycles of 95°C for 30 s, 63°C for 1 min and 72°C for 1 min 30sec, with a final extension of 72°C for 10 min, using a GeneAmp PCR System 9700 (Applied Biosystems, CA).

The PCR products were purified by Microcon (Millipore Corporation, MA) and sequenced using the Big Dye Terminator DNA Sequencing kit, on an ABI PRISM 310 automated sequencer (Applied Biosystems, CA).

Comparison of the obtained sequence (accession number AF486523) with the previously published sequences (Coll et al., 1993; Caroli et al., 2001; Yahyaoui et al., 2001), revealed a nucleotide substitution at position 242 of the exon 4 (A to G). The mutation produces an amino acid substitution Asp to Gly. The polymorphism creates a *Hae*III target site in the new variant, so that a protocol for direct genotyping of the locus by PCR-RFLP can be developed. For this analysis, 10 µl of the PCR product were restricted at 37°C, overnight, using 10 U of *Hae*III endonuclease (MBI Fermentas). All PCR products were resolved by agarose gel electrophoresis and visualized by ethidium bromide staining.

The digestion of the allele 242A produced two fragments (416 and 229 bp), while allele 242G generated three fragments (366 and 50 and 229 bp). The fragments 416 and 366 bp have been useful for the identification of the variant, while the 229 bp fragment is a control for the digestion. The 50-bp band was not visible on the gel (Figure 1).

The occurrence of the polymorphism in a population of 62 Montefalcone goats by PCR-RFLP was high, with an allele frequency of about 16%. The presence of the 242G allele was further investigated in Italian (Teramana, Garganica, Girgentana, Sarda, Cilentana nera; 31, 28, 27, 25, 28 animals, respectively), Spanish (Murciano- Granadina

and Canaria; 25 and 23 animals, respectively) and French (Saanen and Alpine, 21 and 20 animals, respectively) breeds. The allele was absent in the goats from the other breeds.

The economic value of Montefalcone goat is significant because of its consistent performance, productivity, and potential in alternative uses (Pilla et al., 2000). Several studies reported the identification of milk protein genetic variants in the Montefalcone population (Bevilacqua et al., 1999; Bevilacqua et al., 2001; Lagoni-gro et al., 2001). The exclusive presence of a *k*-casein variant confirms the genetic peculiarity of this goat, making its preservation a worthwhile effort. Moreover, a breed-specific allele could be used as molecular marker for breed identification and certification of a specific product. Investigations are evaluating the influence of the new variant on properties of milk from this breed.

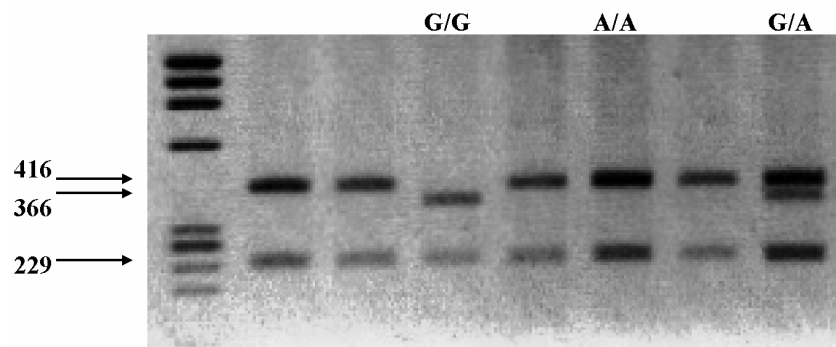
ACKNOWLEDGEMENTS

We would like to thank the Agencia Española de Cooperación Internacional (AECI) for the research scholarship of M.H.Y.

REFERENCES

- Bevilacqua, C., C. Veltri, and F. Pilla. 1999. Genotyping of α -s1-casein locus in Montefalcone goat population. Proc. XIII Congresso A.S.P.A., Piacenza, Italy. 179–181.
- Bevilacqua, C., P. Ferranti, G. Garro, C. Veltri, R. Lagonigro, C. Leroux, E. Pietrola` , F. Addeo, F. Pilla, L. Chianese, and P. Martin. 2001. Interallelic recombination is likely responsible for the occurrence of a new rare α s1-casein variant in the goat species. Eur. J. Biochem. 269:1293–1303.
- Caroli, A., O. Jann, E. Budelli, P. Bolla, S. Ja` ger, and G. Erhardt. 2001. Genetic polymorphism of goat *k*-casein (CSN3) in different breeds and characterisation at DNA level. Anim. Genet. 32:226–230.
- Coll, A., J. M. Folch, and A. Sanchez 1993. Nucleotide sequence of the goat kappa casein cDNA. J. Anim. Sci. 71:2833.
- Grosclaude, F., G. Ricordeau, P. Martin, F. Remeuf, L. Vassal, and J. Bouillon. 1994. Du gene au fromage: le polymorphisme de la caseine α s1 caprine, ses effets, son evolution. INRA Prod. Anim. 7:3–19.
- Lagonigro, R., E. Pietrola` , M. D'Andrea, C. Veltri, and F. Pilla 2001. Molecular genetic characterisation of the goat α s2-casein E allele. Anim. Genet. 32: 390–393.
- Ng-Kway-Hang, K. F. 1988. Genetic polymorphism of milk proteins: Relationships with production traits, milk composition and technological properties. Can. J. Anim. Sci. 78:131–147.
- Pilla, F., L. Mastronardi, E. Pietrola` , and G. Cannata. 2000. An Italian local goat population: Zootechnical and economic aspects. Page 956 in Proc. 7 International Conference on Goats, Tours, France
- Yahyaoui, M. H., A. Coll, A. Sanchez, and J. M. Folch. 2001. Genetic polymorphism of the caprine kappa casein gene. J. Dairy Res. 68:209–216.

Figure 1. Agarose gel electrophoresis showing the different PCR *Hae*III genotypes detected in the *k*-casein exon 4 of Montefalcone goat. Marker *FX174Hae*III.



3. *“Characterization and genotyping of the caprine kappa casein variants”*

Yahyaoui, M.H., Angiolillo, A., Pilla, F., Sanchez, A. and Folch, J.M.

Journal of Dairy Science (2003) 86: 2715-2720

(GOAT KAPPA CASEIN POLYMORPHISM)

Characterization and Genotyping of the Caprine Kappa Casein Variants

M. H. Yahyaoui,* A. Angiolillo,† F. Pilla,† A. Sanchez,* and J. M. Folch*

(*) Departament de Ciència Animal i del Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

(†) Dipartimento S.A.V.A, Università del Molise, Campobasso, Italy

Corresponding author:

M. Habib Yahyaoui

Departament de Ciència Animal i del Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

Tel: (34) 935812087

Fax: (34) 935812106

E-mail: Habib.Yahyaoui@uab.es

ABSTRACT

Kappa casein (κ -CN) is the milk protein that determines the size and specific function of milk micelles, and its cleavage by chymosin is responsible for milk coagulation. We have previously detected and characterized four variants of the goat κ -CN in Spanish and French breeds by screening the major part of the coding region in exon 4. Here we have sequenced and analyzed the full coding region of the κ -CN gene which includes exons 3 and 4. No additional mutations were found, with exception of a single nucleotide substitution in exon 3, with no amino acid change. However, the analysis of the association between the different mutations resulted in two new variants designated κ -CN F and G. The novel variants are present in the Italian breeds Teramana, Girgentana, and Sarda (variant F). A protocol for rapid simultaneous genotyping of all known κ -CN variants using the primer extension method was described, and a total of 210 animals from nine European breeds were genotyped. Alleles A and B are the most frequent variants occurring in the majority of breeds with prevalence of the B variant, except for the Canaria breed where the A allele is more frequent. Sequence data suggest that the F variant is the original type of caprine κ -CN, other alleles being derived from this type following two different trunks by successive mutations.

(Key words : genetic polymorphism, genotyping, goat, kappa casein)

Abbreviation key: κ -CN = kappa casein, PEA = primer extension analysis, RFLP = restriction fragment length polymorphism

INTRODUCTION

Kappa casein (κ -CN) is the milk protein that allows the formation and stabilization of milk micelles and determines their size and function (Gutiérrez et al., 1996). It differs from other caseins in its solubility over a broad range of calcium ion concentrations and contains a hydrophilic C-terminal region. Besides these characteristics, the mature κ -CN protein has a labile peptide bond whose cleavage by chymosin (or rennin) produces a soluble hydrophilic glycopeptide (caseino-macropptide) as well as an insoluble peptide or para- κ -CN. The caseino-macropptide is responsible for milk coagulation while the

function of the para- κ -CN is not well known. The study of genetic polymorphisms of the caseins is of interest, since some variants could be more beneficial from the point of view of human nutrition (Boland et al., 2001) or be associated with milk quality, composition and technological characteristics. In sheep, the κ -CN is considered to be monomorphic (revised in Moioli et al., 1998) whereas bovine κ -CN has six variants with A and B the most common variants (revised in Kaminski, 1996). The milk of cows carrying the B allele of κ -CN contains a smaller and more homogeneous micelle size (Schaar, 1984), with an elevated concentration of κ -CN in milk resulting in higher cheese yield (revised in Ng-Kwai Hang, 1998).

The caprine κ -CN protein contains 171 amino acid residues (Mercier et al., 1976a, b) and the nucleotide sequence of the cDNA and the promoter region of the gene have been reported (Coll et al., 1993; 1995). The caprine κ -CN gene comprises five exons with the coding region for mature protein contained in exons 3 (9 amino acids) and 4 (162 amino acids). We have previously detected and characterized three variants of the goat κ -CN gene (designated A, B, and C) in Spanish and French breeds, and reported the frequency of the C allele (Yahyaoui et al., 2001). The analyzed region (459 bp of exon 4) contains the major part of the coding sequence for mature protein (141 amino acids over a total of 171). Two other variants present in German and Italian goat breeds have been reported recently (Caroli et al., 2001; Angiolillo et al., 2002).

Due to the confusion now present in the literature caused by the assignation of the same name to different variants, and prior to the association studies in order to include genetic variants in goat selection schemes, it is necessary to clearly define the different alleles and develop methods for their rapid discrimination. The aim of this work was to describe a method for rapid and cost efficient genotyping of all known caprine κ -CN variants. A nomenclature for the different alleles is proposed. In addition, we analyzed the full coding region of the κ -CN gene and report two new genetic variants as well as the allele distribution among different European goat breeds.

MATERIALS AND METHODS

Animal Samples

A total of 210 animals belonging to different Spanish (Murciano-Granadina n=30, Canaria n=30, Malagueña n=11), Italian (Teramana n=28, Montefalcone n=17, Girgentana n=19, Sarda n=19) and French (Alpine n=28, Saanen n=28) goat breeds were analyzed. In addition, 23 DNA samples from the Spanish wild goat (*Capra pyrenaica sp. hispanica*) were included in the study. DNA was extracted from blood samples using a standard procedure (Ausubel et al., 1987).

Amplification of the Caprine κ -CN Gene

For the amplification of the DNA region encoding the mature κ -CN protein, two sets of primers were used: I2F (5'-ATG TAT CTG TCA TTT CTT GAG GTT TC-3') - I3R2 (5'-CTC ATG AAA ATC AAC ACA ACT TAG CC-3') for exon 3 amplification (469 bp), and I3F (5'-TCC CAA TGT TGT ACT TTC TTA ACA TC-3') - Kb2 (5'-GCG TTG TCC TCT TTG ATG TCT CCT TAG-3') for amplification of exon 4 (645 bp). Primers I2F, I3F, and I3R2 were designed over the sequenced introns 2 and 3 of the caprine κ -CN gene (unpublished data) while Kb2 was designed relative to the caprine κ -CN cDNA (Coll et al., 1993). The PCR reaction was performed in a 25 μ l final volume containing 0.625 units of *Taq* DNA polymerase (Life Technologies, Rockville, MD), 1x PCR buffer, 1.5 mM-MgCl₂, 200 μ M of each dNTP, 0.4 μ M of each primer and approximately 100 ng of goat genomic DNA. Thermal cycling conditions were: 95°C for 5 min, 10 cycles of 97°C for 15 s, 63°C for 1 min and 72°C for 1 min 30 s, followed by 25 cycles of 95°C for 30 s, 63°C for 1 min and 72°C for 1 min 30 s, with a final extension at 72°C for 5 min.

DNA Sequencing Reactions

A total of 16 PCR products from different breeds (Murciano-Granadina, Canaria, Saanen, and Teramana) were sequenced for exon 3 using a nested primer I3R1 (5'-GAT TCA TAC GAT TGG ATG AAA TG-3') and for exon 4 with primers I3F and Kb2. In addition, three DNA samples from the Spanish wild goat were sequenced for exon 4. In

order to establish haplotypes, five DNA samples from Teramana and Murciano-Granadina breeds were amplified (exon 4) with Expand High Fidelity PCR System (Roche Molecular Diagnosys, Indianapolis, IN), purified with Concert kit (Life Technologies) and cloned using TOPO-TA cloning kit (Invitrogen, Carlsbad, CA). Two clones per sample were sequenced by the dideoxy method using forward and reverse M13 primers and an ABI Prism 310 automated DNA sequencer (Perkin Elmer). Nucleotide sequences were analyzed by Multialin software (Corpet, 1988).

Single Nucleotide Primer Extension Analysis

All animal samples (210 goats from different breeds and 23 Spanish wild goats) were genotyped for the κ -CN variants by the primer extension analysis method using the Multiplex SNaPshot ddNTP kit (PE Applied Biosystems, Foster City, CA). The assay utilizes internal unlabeled primer which binds to a complementary PCR-generated template in the presence of AmpliTaq DNA Polymerase and fluorescently labeled ddNTPs. The polymerase extends the primer one nucleotide, adding a single ddNTP to its 3' end. Primers were designed to allow size- and color- discrimination between the different alleles ([Table 1](#)) and were optimized to be used simultaneously.

The extension reactions were performed in a final volume of 10 μ l, containing 4 μ l of the PCR product purified with EXOSAP-It enzyme mix (USB Corporation, Cleveland, O), 5 μ l of the SNaPshot mix and 1.2 μ l of pooled primers (final concentrations: 0.24 μ M of primers E4R166, E4R328, E4F328, and 0.12 μ M of other primers). Thermal cycling consisted in 25 cycles of 10 sec at 96°C, 5 sec at 50°C and 30 sec at 60°C. To remove unincorporated ddNTPs, 1 unit of alkaline phosphatase (Roche Molecular Diagnosys) was added to the reaction mixture and incubated at 37°C for one hour followed by a denaturation step at 72°C for 15 min. One microliter of the extension reaction was mixed with 25 μ l of formamide and analyzed by capillary electrophoresis and fluorescent detection using a Genetic Analyzer ABI 3100 and the accompanying Genescan software (PE Biosystems).

Genotyping by PCR-RFLP

All animal samples were also genotyped by the PCR-RFLP method. Ten microliters of the PCR product (exon 4) were digested with 10 units of the restriction endonucleases *Alw44I* and *HaeIII* (Roche Molecular Diagnostics) overnight at 37°C, and with *BseNI* (Fermentas, Hanover, MD) at 65°C for 6 h. The resultant fragments were separated by electrophoresis in a 2% agarose gel stained with ethidium bromide.

RESULTS

Nucleotide Sequence Analysis of the κ -CN full coding region

We have previously detected four variants of the κ -CN gene by screening the major part of the coding sequences for mature protein (Yahyaoui et al., 2001; Angiolillo et al., 2002). In order to investigate whether the characterized variants are associated with other mutations in the remaining coding region, a total of 16 DNA samples from different breeds were directly sequenced for exons 3 and 4. Among them, nine samples were previously genotyped and correspond to the alleles A, B and C. Sequence analysis of exon 3 (33 bp from which 27 are coding for mature protein) revealed a single nucleotide substitution (A to G) at position 27 (from the start of the exon) in three samples from Murciana-Granadina and Girgentana breeds. This mutation is located at the third base of the codon Gln and therefore does not produce any amino acidic change.

The obtained sequences for exon 4 showed no additional mutations to those previously described. However, the existence of new haplotypes was suggested. For instance, some samples were heterozygotes A/G at position 166 and homozygotes C/C at position 448 ([Table 2](#)), indicating a new G-166/C-448 haplotype.

Molecular Characterization of the new κ -CN Variants

To ascertain these new haplotypes, PCR products were generated using a high fidelity Taq DNA polymerase and cloned into plasmid vectors. Two independent clones per sample were sequenced. [Table 2](#) shows the two new κ -CN variants and corresponding amino acid differences deduced from the sequence analysis. The two novel haplotypes were

designated κ -CN F and κ -CN G according to their order of detection. The difference between F and G variants involves only one amino acid substitution (valine for isoleucine) at position 65 of the κ -CN protein. The F variant presents valine and proline at positions 65 and 159 respectively, and this combination has not been observed in the previously reported goat variants. Conversely, this variant is identical to the exon 4 sequence obtained from the Spanish wild goat *Capra pyrenaica*. The G allele differs from C and D alleles by only one amino acid change at positions 440 (Ala/Val) and 104 (Gln/Arg), respectively.

Genotyping and Allelic Frequencies of the κ -CN Variants

Using PCR-RFLP for the endonucleases *Alw44I*, *BseNI*, and *HaeIII*, we have reported the frequency of the C and E alleles in different European goat breeds (Yahyaoui et al., 2001; Angiolillo et al., 2002). Samples indicated in Table 3 were also genotyped by this method, however, these endonucleases do not discriminate between alleles A and B or between C and the new allele G (Table 2). The primer extension analysis (PEA) allows genotyping of known SNP alleles by using appropriate primers and fluorescent detection. We have developed a set of primers that allow to distinguish unambiguously the different goat κ -CN alleles (Figure 1). All samples were evaluated with this method using the six primers simultaneously. However, alleles A, C, D, and E can be individually differentiated from others by single genotyping using PEA-328, PEA-440, PEA-104, and PEA-242, respectively. Allele E can also be distinguished by PCR-RFLP with *HaeIII* endonuclease, being this allele the only one with guanine (Gly) at position 242. Using the same technique, the *BseNI* allows discrimination of alleles A, B, E, and F from others, whereas *Alw44I* distinguishes only A, B, and E alleles. Consequently, the F allele can be differentiated combining these two endonucleases. Table 3 shows the frequencies of the different κ -CN genetic variants in nine European goat breeds. The novel variants F and G are found only in the Italian breeds Teramana, Girgentana and Sarda, with a relatively high frequency. On the contrary, the F variant is so far the most frequent allele in the wild goat *Capra pyrenaica*. κ -CN A and B are the most common variants in the majority of breeds with a prevalence of the B variant, with the exception of the Canaria breed where variant A is the most frequent (0.6 vs. 0.4). κ -CN B is the unique variant found in all analyzed breeds with frequencies ranging from 0.4 (Canaria) to 0.7 (Teramana). The C variant is particularly present in

Saanen (13 %) and the E variant is specific of Italian Montefalcone breed. The D variant is not found in the analyzed breeds, except in Teramana and Girgentana and at low frequencies (5 - 10 %).

DISCUSSION

This study describes the polymorphism of the caprine κ -CN gene. The full coding region which encompasses exons 3 and 4 was screened for polymorphisms by DNA sequencing. Although no additional mutations to those previously reported were detected, two new genetic variants resulting from the association between the different mutations are observed. To avoid the confusion now present in nomenclature, we suggest that caprine κ -CN variants would be designated according to the order in which they were reported: A, B, C (Yahyaoui et al., 2001), D (Caroli et al., 2001), E (Angiolillo et al., 2002), F and G (described in this paper). The new variants F and G are found only in the Italian breeds (Teramana, Sarda, and Girgentana) at frequencies ranging from 0.04 to 0.14. Among the seven genetic variants characterized, only the B variant was found in all studied populations and is predominant over other alleles, except in Spanish Canaria breed. It could therefore be considered as the original variant on the basis of frequency criterion; the original variant is expected to have usually the highest frequency over a large number of populations. However, when considering comparative sequence data, variant F is identical at position 159 to the κ -CN of other species (sheep, cattle, pig, mouse, rat, rabbit, camel and human), all having proline, while the B variant presents serine at this position. In addition, the F variant is identical to the sequence of the κ -CN exon 4 obtained from the Spanish wild goat. The similarity of sequences indicates that this variant is likely the original type. The reasons of the divergence in the conclusions of the frequency and of the sequence conservation criteria are unclear. Because the κ -CN plays a critical role in several important physiological processes, it seems unlikely that this gene would be completely free of the selective pressure, and there is an evidence for positive selection within the family *bovidae* (Ward et al., 1997). Thus, the low frequency of variant F could have come through selection and it was very early supplanted by type B. In this regard, it is important to notice that the chemical and nutritional properties of the amino acid residues serine and proline (in F and B variants, respectively) are very different. However, genetic drift and founder effect

could also explain the prevalence of the B variant. Until further developments by typing more goat populations either wild type or goat breeds especially non-European breeds, the F variant could be considered as the ancestral type. The different alleles were most likely derived from this original type by successive mutations following two distinct trunks ([Figure 2](#)): A, B, E which occur -when found- at high frequencies, and C, D, G which are observed at low frequencies.

In goat, the three calcium-sensitive casein genes ($\alpha s1$, $\alpha s2$, and β) are polymorphic and alleles are associated with strong differences in their level of expression. Alleles with a null amount of protein have been found for the three genes. In this regard, the high level of genetic variability observed at the casein loci and its association with milk traits allows to obtain groups of animals producing types of milk which could be more suitable for specific technologies of transformation or for specific needs of human nutrition (Rando et al., 2000). Due to the linkage between casein genes it is necessary to take into account κ -CN when considering haplotype studies in order to include milk protein variants in breeding programmes. The technique described here allows the rapid and simultaneous genotyping of all known κ -CN variants, and because an automated instrument was used, it was suitable for large screening and genotyping in selection schemes and industrial purposes.

Further studies involving different goat breeds are required to elucidate whether these κ -CN polymorphisms have a physiological role as observed in bovine, and to identify their linkage within the casein cluster.

ACKNOWLEDGMENTS

We would like to thank the Agencia Española de Cooperación Internacional (AECI) for the research scholarship of M.H.Y. This research was partially funded by Cofin MIUR 2001.

REFERENCES

- 1
2
- 3 Angiolillo, A., M. H. Yahyaoui, A. Sanchez, F. Pilla, and J. M. Folch. 2002. Characterization of a
4 new genetic variant in the caprine κ -casein gene. *J. Dairy Sci.* 85:2679-2680.
- 5 Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, G. G. Seidman, J. A. Smith, and K. Struhl
6 (Eds). 1987. *Current protocols in molecular biology*. Green Publishing Associates and Wiley-
7 Interscience, New York, NY.
- 8 Boland, M., A. MacGibbon, and J. Hill. 2001. Designer milks for the new millenium. *Livest. Prod.*
9 *Sci.* 72:99-109.
- 10 Caroli, A., O. Jann, E. Budelli, P. Bolla, S. Jäger, and G. Erhardt. 2001. Genetic polymorphism of
11 goat κ -casein in different breeds and characterization at DNA level. *Ani. Gen.* 32:226-230.
- 12 Coll, A., J. M. Folch, and A. Sanchez. 1993. Nucleotide sequence of the goat κ -casein cDNA. *J.*
13 *Anim. Sci.* 71:2833
- 14 Coll, A., J. M. Folch, and A. Sanchez. 1995. Structural features of the 5' flanking region of the
15 caprine κ -casein gene. *J. Dairy Sci.* 78:973-977.
- 16 Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.*
17 16:10881-10890.
- 18 Gutierrez, A., E. A. Maga, H. Meade, C. F. Shoemaker, J. F. Medrano, G. B. Anderson, and J. D.
19 Murray. 1996. Alterations of the physical characteristics of milk from transgenic mice producing
20 bovine κ -casein. *J. Dairy Sci.* 79:791-799.
- 21 Kaminski, S. 1996. Bovine κ -casein gene: molecular nature and application in dairy cattle breeding.
22 *J. Appl. Gen.* 37:176-196.
- 23 Mercier, J. C., F. Addeo, and J. P. Pelissier. 1976a. Structure primaire du caséinomacropeptide de la
24 caséine kappa caprine. *Biochimie* 58:1303-1310.
- 25 Mercier, J. C., J. M. Chobert, and F. Addeo. 1976b. Comparative study of the amino acid sequences
26 of the caseinomacropeptides from seven species. *FEBS Letters* 72:208-214.

27 Muioli, B., F. Pilla, and C. Tripaldi. 1998. Detection of milk protein genetic polymorphisms in
28 order to improve dairy traits in sheep and goats: a review. *Small Rum. Res.* 27:185-195.

29 Ng-Kwai-Hang, K. F. 1998. Genetic polymorphism of milk proteins: relationships with production
30 traits, milk composition and technological properties. *Can. J. Anim. Sci.* 78 (Suppl):131-147.

31 Rando, A., L. Ramuno, and P. Masina. 2000. Mutations in casein genes. *Zoot. Nutr. Anim.* 26:105-
32 114.

33 Schaar, J. 1984. Effect of κ -casein genetic variants and lactation number on the renneting properties
34 of individual milks. *J. Dairy Res.* 51:397-406.

35 Ward T. J., R. L. Honeucutt, and J. N. Derr. 1997. Nucleotide sequence evolution at the κ -casein
36 locus: evidence for positive selection within the family bovidae. *Genetics* 147:1863-1872.

37 Yahyaoui, M. H., A. Coll, A. Sanchez, and J. M. Folch. 2001. Genetic polymorphism of the caprine
38 kappa casein gene. *J. Dairy Res.* 68:209-216.

39

Table 1. Primers used in the extension analysis.

Primer [§]	Position (exon 4)	Size (bp)	Sequence (5'-3')
E4F-104	104	44	(AT) ₅ AGTAGGTATCCTAGTTATGGACTCAATTACTATC
E4R-166	166	40	(AT) ₉ AGGGCAGGTGACCTAACTGCAA
E4F-242	242	36	(AT) ₈ ACCTGCCAAGTCCTGCCAAG
E4R-328	328	23	AGCAATGGTATTGATGGCAGGGA
E4F-440	440	48	(AT) ₁₃ GAAGCTTCCTCAGAATCGATTG
E4R-448	448	28	ATAGTAACTTGGGCTGTGTTGGTCTCAG

[§] F: Forward, R: Reverse

Table 2. Variants of the caprine κ -CN gene. Nucleotides present at polymorphic positions and corresponding amino acid changes (in parentheses) in each variant are indicated. Positions of nucleotides are relative to the start of exon 4. Amino acid positions in the protein are also indicated. The nucleotide sequences of the different κ -CN variants were deposited to GenBank under the accession numbers AF485339, AF485340, AF485341, AY090465, AF486523, AY090466, and AY090467.

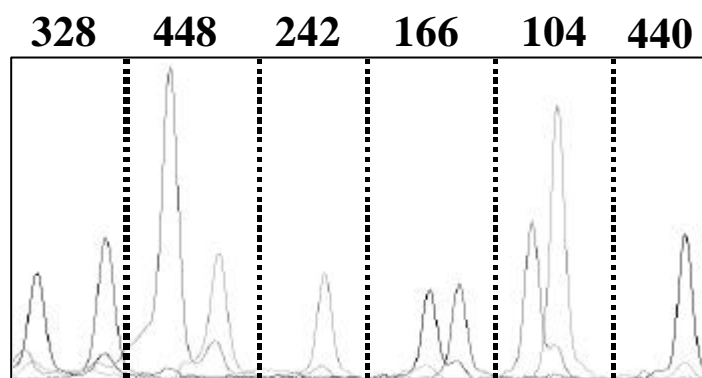
Position (Exon 4)	Position (Protein)	Variants							Genotyping [§]
		A	B	C	D	E	F	G	
104	44	A (Gln)	A (Gln)	A (Gln)	G (Arg)	A (Gln)	A (Gln)	A (Gln)	
166	65	G (Val)	G (Val)	A (Ile)	A (Ile)	G (Val)	G (Val)	A (Ile)	<i>BseNI</i>
242	90	A (Asp)	A (Asp)	A (Asp)	A (Asp)	G (Gly)	A (Asp)	A (Asp)	<i>HaeIII</i>
328	119	G (Val)	A (Ile)	A (Ile)	A (Ile)	A (Ile)	A (Ile)	A (Ile)	
440	156	C (Ala)	C (Ala)	T (Val)	C (Ala)	C (Ala)	C (Ala)	C (Ala)	
448	159	T (Ser)	T (Ser)	C (Pro)	C (Pro)	T (Ser)	C (Pro)	C (Pro)	<i>Alw44I</i>

[§] Restriction enzymes used for PCR-RFLP genotyping.

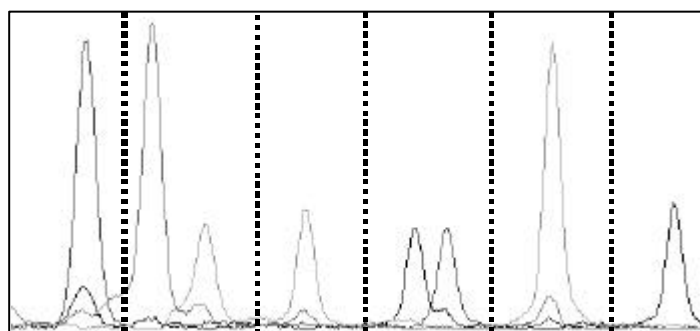
Table 3. Frequencies of the κ -CN variants in different European goat breeds.

Breeds	N° animals	Variants						
		A	B	C	D	E	F	G
Murciano-Granadina	30	0.37	0.63	-	-	-	-	-
Malagueña	11	0.45	0.55	-	-	-	-	-
Canaria	30	0.58	0.42	-	-	-	-	-
Teramana	28	-	0.70	0.02	0.10	-	0.14	0.04
Montefalcone	17	-	0.59	-	-	0.41	-	-
Girgentana	19	0.34	0.45	-	0.05	-	0.05	0.11
Sarda	19	0.31	0.61	-	-	-	0.08	-
Alpine	28	0.34	0.66	-	-	-	-	-
Saanen	28	0.39	0.48	0.13	-	-	-	-
Wild goat	23	-	0.02	-	-	-	0.98	-

Figure 1. Genotyping of the κ -CN variants by primer extension analysis. The six mutations were analyzed simultaneously using primers shown in Table 1. The six positions relative to exon 4 are indicated below. Shown genotypes are AD, BG in (a), (b), respectively. Allele discrimination is based on both primer size and color (not shown in the figure) assigned to the fluorescent ddNTPs incorporated in the extension reaction.

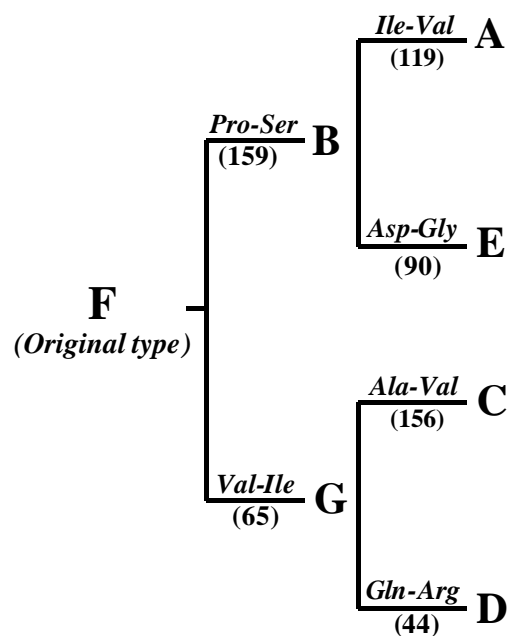


(a)



(b)

Figure 2. Proposed phylogenetic tree of the goat κ -CN variants. Amino acid changes and their position (in parentheses) are indicated.



4. *“Rapid Communication: polymorphism in the goat beta-lactoglobulin proximal promoter region”*

Yahyaoui, M.H., Pena, R.N., Sanchez, A. and Folch, J.M.

Journal of Animal Science (2000) **78**:1100-1101

**Rapid Communication: Polymorphism in the goat *b-Lactoglobulin*
proximal promoter region¹**

M. H. Yahyaoui², R. N. Pena, A. Sánchez, and J. M. Folch

Departament de Patologia i Producció Animals, Facultat de Veterinària, Universitat
Autònoma de Barcelona, Bellaterra 08193, Spain

¹This work was supported by a research scholarship to M.H.Y. from CIHEAM (Instituto Agronomico Mediterraneo de Zaragoza, Spain).The authors thank M. Amills and J. Capote for providing some goat DNA samples.

²To whom correspondence should be addressed. Phone: (34) 935812087; fax: (34) 935812106; E-mail: habib@guara.uab.es.

Polymorphism. Polymorphism was detected within a 710-bp-long PCR amplification product of the goat ***b-Lactoglobulin (BLG)*** gene, comprising 588 bp of the proximal promoter region and 122 bp of exon 1, using the Base Excision Sequence Scanning method (BESS T-Scan; Epicentre Technologies, Madison, WI). A total of forty genomic DNA samples from animals of different Spanish (Canaria, Malagueña, Payoya, and Murciano-Granadina) and French (Saanen) breeds were analyzed.

Primer description. Primers were designed from published nucleotide sequence of the goat *BLG* gene (Folch et al., 1994) and were located in the proximal promoter region (-588 to -569) and in the first exon (+100 to +122). Forward and reverse primers were fluorescently labeled with HEX and 6-FAM, respectively.

Primer sequences. Forward primer: 5'-GTCACCTTCCCGTCCTGGGG-3'; reverse primer: 5'-GGCCTTTCATGGTCTGGGTGACG-3'.

Method of detection. For the detection of polymorphism, the PCR reaction was performed in a 25- μ L final volume containing .625 unit of *Taq* DNA polymerase (Life Technologies, Rockville, MD) with 1X PCR buffer; 2 mM $MgCl_2$; 50 μ M BESS T-Scan dNTP Mix; 150 μ M each dNTP; .4 μ M of each labeled primer and 100 ng of goat genomic DNA. Thermal cycling conditions were 95 °C for 5 min, 10 cycles of 97 °C for 15 s, 63 °C for 1 min and 72 °C for 1 min 30 s, then 25 cycles of 95 °C for 30 s, 63 °C for 1 min and 72 °C for 1 min 30 s, followed by a final extension at 72 °C for 5 min. For the excision reaction, 5 μ L of the PCR product were mixed with .5 μ L of BESS T-Scan Excision Enzyme Mix and .6 μ L of BESS T-Scan Excision Enzyme Buffer. The reaction was incubated for 20 min at 37 °C and then placed on ice. Twenty-five μ L of formamide and 1 μ L of GeneScan 500 (ROX) size standard (Perkin Elmer, Foster city, CA) were added before denaturation at 95 °C for 3 min. Products were analyzed by capillary electrophoresis and fluorescent detection using a Genetic Analyzer 310 (Applied Biosystems, Perkin Elmer, Foster city, CA). A protocol for the rapid genotyping of this polymorphic site was developed by PCR-RFLP procedure using the *Sma*I endonuclease: Ten μ L of the PCR product was digested overnight and separated by electrophoresis in a 2 % agarose gel.

Description of polymorphism. The polymorphic site consists of a single nucleotide substitution (C to T) at position -60 of the goat *BLG* promoter region (Folch et al., 1994). The *Sma*I digestion of the PCR product from allele -60C produces four fragments (472+181+50+7 bp), whereas allele -60T generates only three fragments (472+231+7 bp) (Figure 1). Fragments 181 and 231 bp allow the identification of the -60C and -60T variants, respectively. Fragment 472 is useful as a control of the digestion.

Chromosomal location. The goat *BLG* gene has been assigned by in situ hybridization to chromosome 11q28 (Hayes et al., 1993).

Frequency. Sixty-nine Murciano-Granadina, 42 Canaria, 11 Payoya, 18 Malagueña, and 20 Saanen goats were genotyped by PCR-RFLP analysis. The frequencies of the -60C allele were .86, 1, .73, .75, and .73, respectively, in the five breeds.

Comments. *b-Lactoglobulin* is the major whey protein in the milk of ruminants. In cattle, eight coding region variants are known, with variants A and B the most frequent. The quantitative effects of these common variants on milk composition and cheese making properties have been reported (reviewed in Ng-Kwai Hang and Grosclaude, 1992). In addition, a higher expression of allele A has been described in heterozygous (AB) animals (Graml et al, 1989). This differential allelic expression has been explained by nucleotide differences in the promoter regions associated with these two alleles (Lum et al, 1997; Folch et al., 1999). In contrast, no polymorphism has been previously characterized in the goat *BLG* gene, although variation of mobility was observed by protein electrophoresis (reviewed in Moili et al. 1998). Several transcription factors are known to bind to recognition sequences of the *BLG* promoter (Watson et al., 1991; Folch et al., 1994), but the polymorphism reported in this work is not located in these binding sites. Further studies are required to evaluate the effect of these variants in the expression of the *BLG* gene. Allele frequencies for the five breeds were similar with the exception of Canaria breed where allele -60T was not found. This difference can be explained by the different origin of this breed (Sánchez-Belda, 1986).

Literature Cited.

- Folch, J. M., A. Coll, and A. Sanchez. 1994. Complete sequence of the caprine β -lactoglobulin gene. *J. Dairy Sci.* 77:3493-3497.
- Folch, J. M., P. Dovc, and J. F. Medrano. 1999. Differential expression of bovine β -lactoglobulin A and B promoter variants in transiently transfected HC11 cells. *J. Dairy Res.* 665:37-544
- Graml, R., G. Weiss, G. Buchberger, and F. Pirchner. 1989. Different rates of synthesis of whey protein and casein by alleles of the beta-lactoglobulin and alpha s1-casein locus in cattle. *Genet. Sel. Evol.* 21:547-554.
- Hayes, H. C and E. J. Petit. 1993. Mapping of the β -Lactoglobulin gene and of an immunoglobulin M-heavy chain-like sequence to homologous cattle, sheep and goat chromosomes. *Mamm. Genome*, 4:207-210
- Lum, S. L., P. Dovc, and J. F. Medrano. 1997. Polymorphism of bovine β -lactoglobulin promoter and differences in the binding affinity of activator protein 2 transcription factor. *J. Dairy Sci.* 80:1389-1397.
- Moioli, B., F. Pilla, and C. Tripaldi. 1998. Detection of milk protein genetic polymorphisms in order to improve dairy traits in sheep and goats: a review. *Small Ruminant Res.* 27:185-195.
- Ng Kwai-Hang, K. F. and F. Grosclaude. 1992. Genetic polymorphism of milk protein. In: P.F. Fox (Ed.). *Advanced Dairy Chemistry. Vol. 1 Proteins.* pp 405-455. Elsevier Science Publishers Ltd.
- Sánchez-Belda, A. 1986. Raza Canaria. In: *Catalogo de razas autoctonas españolas. Vol. 1. Especies ovina y caprina.* Ministerio de Agricultura, Pesca y Alimentacion Publ. pp 163-173. Madrid, Spain.
- Watson, C. J., K. E. Gordon, M. Robertson, and A. J. Clark. 1991. Interaction of DNA-binding proteins with a milk protein gene promoter in vitro: identification of a mammary gland-specific factor. *Nucleic Acids Res.* 19:6603-6610.

Key Words: Beta-Lactoglobulin, Promoter, Goat, Polymorphism

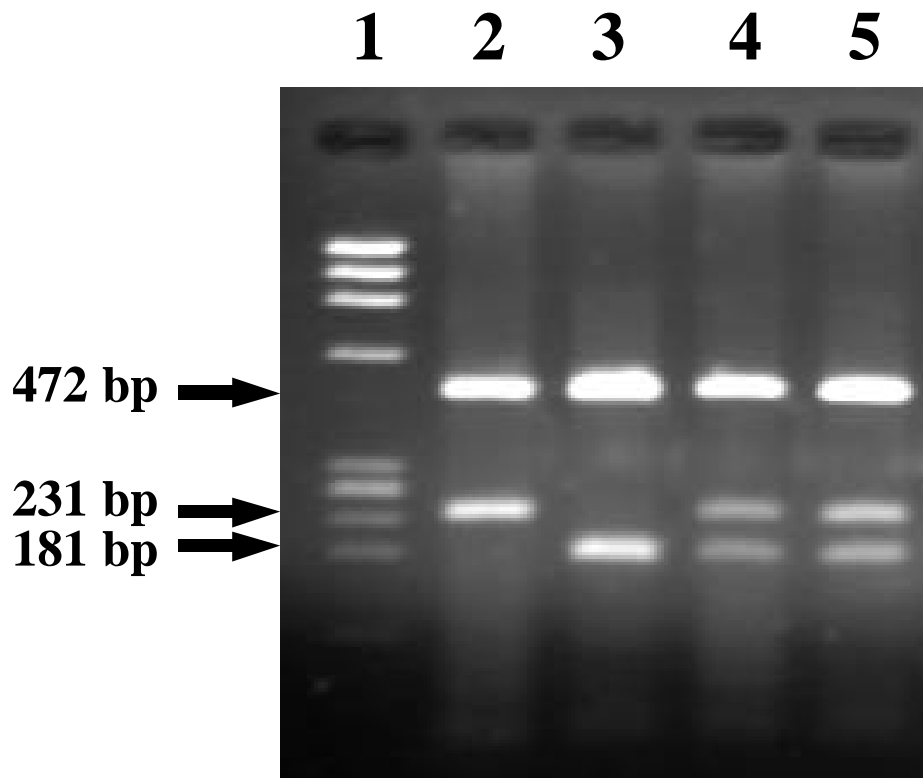


Figure 1: Genotyping by PCR-RFLP of the two variants found in the *b-Lactoglobulin* proximal promoter region. Lane 1: ϕ x174-*Hae*III marker (Promega); lane 2: -60T/-60T; lane 3: -60C/ -60C; lane 4: -60C/-60T, lane 5: -60C/-60T.

5. *“Rapid Communication: Partial nucleotide sequence of the goat stearoyl coenzyme A desaturase cDNA and gene structure”*

Yahyaoui, M.H., Sanchez, A. and Folch, J.M

Journal of Animal Science (2002) **80**:866-867

Rapid Communication: Partial nucleotide sequence of the goat stearyl
coenzyme A desaturase cDNA and gene structure

M. H. Yahyaoui¹, A. Sánchez, and J. M. Folch

Unitat Genètica i Millora, Departament de Ciència Animal i dels Aliments, Facultat de
Veterinària, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain

¹To whom correspondence should be addressed. Phone: (34) 935812087; fax: (34)

935812106; E-mail: Habib.Yahyaoui@uab.es.

Name of the Sequence. Stearoyl coenzyme A desaturase (SCD) cDNA.

Genus and Species. *Capra hircus*.

Origin of Clone. Total RNA was isolated from a mammary gland sample from a lactating Murciano-Granadina goat using Trizol reagent (Life Technologies, Rockville, MD) and following supplier's instructions. The first strand of the cDNA was synthesised using the M-MuLV reverse transcriptase (Life Technologies) and two different primers: an oligo-d(T)₃₀ and E5R (5'-CCACAGCTCCCAGGGAAACC-3') designed over the ovine SCD exon 5 (Ward et al., 1998). The cDNA was amplified by PCR using two sets of primers: E1F (5'-CAGCACAGCAGGTCGGGTCCG-3') and E5R for the amplification of the region comprising from exons 1 to 5, and E4F (5'-CTGATCCCCACAATTCCCGAC-3') with 3NCR (5'-GCTCTAGAGAGAAAGGGAGCATGCTGG-3') for the amplification from exon 4 to 6. The 3NCR primer was designed 22 bp upstream of the poly(A) tail of the sheep cDNA sequence. The cDNA bands of interest were isolated from an agarose gel and cloned using TOPO-TA cloning kit (Invitrogen, Carlsbad, CA). Several clones from different RT-PCR products were sequenced on both strands using an automated sequencer (ABI310, Applied Biosystems, Foster City, CA). For the determination of gene structure (intron sizes and exon-intron boundaries) introns were amplified using primers complementary to the flanking exons and genomic DNA from Murciano-Granadina goats. Preliminary position of exons was deduced from the sequence of the bovine *SCD* gene (J.F. Medrano, personal communication). The PCR was performed in a total volume of 50 µl with 100-250 ng of genomic DNA and using 2.5 units of Long Expand PCR System (Roche Molecular Diagnostics, Indianapolis, IN). The generated fragments from multiple independent PCR reactions of each region were purified with Concert kit (Life Technologies) and directly used in sequencing reactions.

Comparison with Related Sequences. Within the coding region, the nucleotide sequence is well conserved between the ruminant species with a similarity of 98 and 95% between the caprine *SCD* sequence and those of sheep and cow, respectively. The degree of similarity of the caprine sequence is lower with pig (89%), human (87%), and mouse *SCD-1* and *SCD-2* (82%) genes. Comparison of the protein sequences between species shows that the amino acid substitution are concentrated

in the amino terminal one-third of the protein. The eight histidine residues functionally essential on the catalytic properties of the enzyme (Shanklin et al., 1994) are spatially conserved.

EMBL/GenBank Accession Number. AF339909.

Gene Structure. The exon-intron junctions were determined by comparing the genomic sequence with the cDNA sequence and with the 5' and 3' splice consensus sequences (Shapiro and Senapathy, 1987). The structure of the caprine *SCD* gene is very similar to that reported of rodent and human. It spans approximately 12 Kb and consists of six exons and five introns (Figure 2). Exon sizes range from 131 bp (3rd exon) to up to 854 bp (exon 6) and the gene contains five introns varying in size from .5 to 3.7 Kb, approximately. The totality of the 3' UTR is located in exon 6. All exon-intron boundaries have the splice consensus sequences GT / AG.

Comments. *SCD* is an iron-containing enzyme involved in the biosynthesis of the monounsaturated fatty acids. It catalyses the Δ^9 -cis desaturation of palmitoyl (16:0) and stearoyl-CoA (18:0) to produce palmitoleoyl (16:1) and oleoyl-CoA (18:1), respectively. Two highly homologous genes (*SCD-1* and 2) in rodents (Ntambi et al., 1988; Kaestner et al., 1989; Mihara et al., 1990) and a single ovine (Ward et al., 1998) and human (Zhang et al., 1999) *SCD* genes have been cloned and characterised. In addition, partial sequences of bovine and porcine cDNAs have been reported (GenBank accession numbers: AF188710 and Z97186). Recently, a third specie of *SCD* gene has been identified in mouse skin (Parimoo et al., 1999). The *SCD* gene is expressed in several tissues including adipose, liver, brain, kidney, skin, and mammary gland of lactating animals. Differences in the regulation of gene expression at transcriptional and post-transcriptional (tandem polyadenylation sites generating alternative transcripts, mRNA stability) levels between species have been reported (Ntambi, 1999; Zhang et al., 1999). In sheep, the *SCD* is transcribed from a single gene and it is widely expressed in several tissues with a higher level of expression in adipose tissue, liver, and mammary gland of lactating animals (Ward et al., 1998). Unlike rodents and human, the ovine and the partial caprine 3'-untranslated regions of the mRNA are shorter and did not contain the motifs (AUUUA) characteristic of destabilisation of the mRNA sequence (Padget et al., 1986; Jackson 1993). This could be related to the differences on the lipid

metabolism in ruminants such as the nature of diet usually low in fat and the biohydrogenation activity of unsaturated fatty acids in the rumen. Further studies are necessary to evaluate the role of mRNA stability in the regulation of the ruminant *SCD* genes.

Literature Cited.

Jackson, R.J. 1993. Cytoplasmic regulation of mRNA function: the importance of the 3' untranslated region. *Cell* 74:9-14.

Kaestner, K.H., J.M., Ntambi, T.J Kelly Jr, and M.D. Lane 1989. Differentiation-induced gene expression: a second differentially expressed gene encoding stearoyl coenzyme A desaturase. *J. Biol. Chem.* 264:14755-14761.

Mihara, K. 1990. Structure and regulation of rat liver microsomal stearoyl coenzyme A desaturase gene. *J. Biochem.* 108:1022-29.

Ntambi, J.M. 1999. Regulation of stearoyl coenzyme A desaturase by polyunsaturated fatty acids and cholesterol. *J. Lipid Res.* 40:1549-58.

Ntambi, J.M., S.A Buhrow, K.H. Kaestner, R.J. Christy, E. Sibley, T.J. Kelly Jr, and M.D. Lane 1988. Differentiation-induced gene expression in 3T3-L1 preadipocytes: characterization of a differentially expressed gene encoding stearoyl coenzyme A desaturase. *J. Biol. Chem.* 263:17291-300.

Padgett, R.A., P.J. Grabowski, M. Konarska, S. Seiler, and P.A. Sharp 1986. Splicing of messenger RNA precursors. *Ann. Rev. Biochem.* 55:1119-50.

Parimoo, S., Y. Zheng,, K. Eilersten, L. Ge, S. Prouty, J. Sundberg, and K. Stenn 1999. Identification of a novel stearoyl coenzyme A desaturase gene and expression of the SCD gene family in mouse skin. *J. Inv. Derm. Symp. Proc.* 4:320-22.

Shanklin, J., E. Whittle, and B.G. Fox 1994. Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl coenzyme A desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. *Biochem. J.* 33:12787-94.

Shapiro, M.B. and P. Senapathy 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.*15:7155-74.

Ward, R.J., M.T. Travers, S.E., Richards, R.G. Vernon, A.M. Salter, P.J. Buttery, and M.C. Barber 1998. Stearoyl coenzyme A desaturase mRNA is transcribed from a single gene in the ovine genome. *Biochim. Biophys. Acta* 1391:145-56.

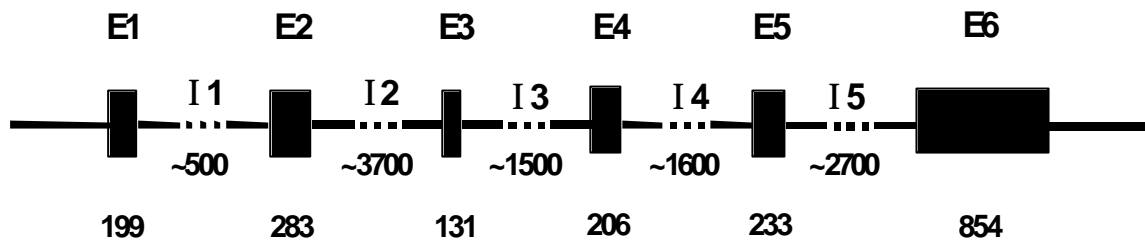
Zhang, L., L. Ge, S. Parimoo, K. Stenn, and S. Prouty 1999. Human stearyl coenzyme A desaturase: alternative transcripts generated from a single gene by usage of tandem polyadenylation sites. *Biochem. J.* 340:255-64.

Key words. Stearoyl coenzyme A desaturase, cDNA, goat.

Figure 1. Partial nucleotide sequence of the goat stearyl coenzyme A desaturase cDNA. The translation initiation (ATG) and stop (TGA) codons of the 1080 nucleotide open-reading frame are underlined. The corresponding protein sequence of 359 amino acids is denoted by one letter amino acid designation. The histidine residues spatially conserved in the family of membrane desaturases are indicated by bold script. Black triangles mark the exon-exon boundaries for exons 1 through 6.

GCGATCCAGCGCACAGCGTGCAGCGGAAGGTCCCGAGCACAGCGCCGCGGAT 52
 CGCCACGCATCGCCACGCAAAGCAGGCTCAGGAAGTACCTACACTCAGTTCCGACTGC 112
 CCCGAAGTCCGCTCCGCAGTCTGAGCCCCGAGAAAGTATCCCAGTGTCCGAGAGCCAG 172
ATGCCGGCCCACTTGCTGCAAGAGGAGATCTCTAGCTCCTACACGACCACCACCATC 232
 M P A H L L Q E E I S S S Y T T T T T I
 ACAGCACCTCCTTCCAAGGTCTGCAGAATGGAGGGGGCAAATTGGAGAAGACTCCCCTA 292
 T A P P S K V L Q N G G G K L E K T P L
 TACTTGGGAAGAAGACATCCGCCCTGAAATGAGAGATGACATCTATGACCCAACCTTACCAG 352
 Y L E E D I R P E M R D D I Y D P T Y Q
 GATAAGGAGGGCCCAAAGCCCAAGCTTGAGTATGTTTGGAGAAACATCATCCTCATGGGT 412
 D K E G P K P K L E Y V W R N I I L M G
 CTGTTACACTTGGGAGCCCTGTATGGGATCACATTGATCCCCACCTGCAAGATATACACC 472
 L L H L G A L Y G I T L I P T C K I Y T
 TTTCTCTGGGTGTTATTCTACTATATGATGAGTGCCCTGGGCATCACAGCAGGGGTCCAT 532
 F L W V L F Y Y M M S A L G I T A G V H
 CGCCTGTGGAGTCACCGAACCTACAAAGCTCGGCTGCCCTGCGGGTCTTCCTGATCATC 592
 R L W S H R T Y K A R L P L R V F L I I
 GCCAACACCATGGCGTTCAGAATGACGTTTTTGAATGGTCCCGAGATCACCGTGCCAC 652
 A N T M A F Q N D V F E W S R D H R A H
 CACAAGTTTTTCAGAAACGGATGCTGATCCCCACAATTCCCGACGTGGCTTTTTTCTTCTCT 712
 H K F S E T D A D P H N S R R G F F F S
 CACGTGGGTGGCTGCTTGTGCGCAAACACCCAGCTGTCAGAGAAAAGGGTGTACGCTA 772
 H V G W L L V R K H P A V R E K G A T L
 GATTTATCCGACCTAAGAGCTGAGAAGCTGGTGTATGTTCCAGAGGAGGTA[▼]CTACAAACCT 832
 D L S D L R A E K L V M F Q R R Y Y K P
 GGTGTCCTGCTGTTGTGCTTCATCCTGCCACACTCGTGCCGTGGTATCTATGGGGTGAA 892
 G V L L L C F I L P T L V P W Y L W G E
 ACGTTTCAAACAGCCTTTTTTTTTGCCACCCTTCTCCGGTACGCTGTTGTGCTCAATGCC 952
 T F Q N S L F F A T L L R Y A V V L N A
 ACCTGGCTGGTGAACAGCGCTGCCCATATGTATGGATATCGCCCTTACGACAAGACCATC 1012
 T W L V N S A A H M Y G Y R P Y D K T I
 AACCCCGAGAGAATATCCTGGTTTCCCTGGGAGCTGTGGGTGAGGGCTTCCACAACCTAC 1072
 N P R E N I L V S L G A V G E G F H N Y
 CACCACACCTTTCCTTATGACTACTCTGCCAGCGAGTACCGCTGGCACATCAACTTTACC 1132
 H H T F P Y D Y S A S E Y R W H I N F T
 ACATTCTTCATTGATTGCATGGCTGCCATCGGTCTGGCTTATGACCGGAAGAAAGTATCC 1192
 T F F I D C M A A I G L A Y D R K K V S
 AAGGCTGCCGCCTTGGCCAGGATGAAAAGAAGCTGGAGAGGAAAGCTGCAAGAGTGGCTGA 1252
 K A A A L A R M K R T G E E S C K S G
 GTTTCTGGTCCCTTGGATTCCTTTTTCCAAAAGCCAGCTGGGCAGAGGTTTAATGTTCTGT 1312
 TTATAAACTACTGAATAATGCTACCAGGATGCTAAAGATGACGTTAACCCATCACAGTAC 1372
 AGTATTCTTTAAAATTTCTTTTTTAAATTGAAAGCCAACAACCTCTGCCTTTATGATGCTA 1432
 AGCTCATGTTCTTATTTCTTCTTATCTTCTTTCTTCTTCTGTTCCATTATCTTTCCCT 1492
 TTGTTTTGTCCCTATCACCTTCCTTTCTTTCTCCTCATTGCCCCCAAGGCAAGCAGTT 1552
 GGTCAGTCACTGGTGGGTTTCCAGCTTCCAAAGCCTAGACGACCCTTCTGTAGTCTAAAA 1612
 CAAGTGGTCTTTGCCCGGCTGACCCTTTCTTGTAGCTGTCTGAGCTTTAAGGTGGATGG 1672
 CCCGAGCTAGATATATGACAGAATCTTCTGGGAAGGGCCCTGATGATCTTCGGCCAGGC 1732
 TTTTGCTAGATGAAATGGAAAAATAACTTTATTTTGGCACCAAAGTCAAAAAACAGGTAA 1792
 ATTGTCAGGGGAGAGAGTACAGCATGCATGATGTGATTGATAAATAGGGTGTAGTTGAAGTG 1852
 GGAAACAAGGCAGGAAGCTGCTGCTGTGATCAGACACCCCTGTCTGCCATCAC 1906

Figure 2. Structure of the caprine *SCD* gene. Exons are depicted schematically as boxes drawn approximately to scale, and introns by the connecting horizontal lines. Exon numbers and sizes (in bp) are indicated above and under the boxes, respectively. Estimated sizes of introns are also indicated.



6. *“Genetic mapping of the goat stearoyl coenzyme A desaturase gene to chromosome 26 using a biallelic polymorphism”*

Yahyaoui, M.H., Vaiman, D., Sanchez, A. and Folch, J.M.
Animal Genetics (Submitted)

Genetic mapping of the goat stearyl coenzyme A desaturase gene to
chromosome 26 using a biallelic polymorphism

M. H. Yahyaoui¹, D. Vaiman², A. Sánchez¹, and J. M. Folch¹

¹Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat
Autònoma de Barcelona, Bellaterra 08193, Spain

²Laboratoire de Génétique Biochimique et de Cytogénétique, INRA 78353, Jouy-en-Josas,
France

Source/description:

Polymorphism was detected within a 447-bp-long PCR amplification product of the goat *stearoyl coenzyme A desaturase (SCD)* gene, containing the complete exon 5 sequence (239 bp). Preliminary evidence of polymorphism was observed by comparison between the cDNA and genomic DNA nucleotide sequences¹ from different animals. The polymorphic site consists of a single nucleotide substitution (G to T) in exon 5, at position 931 of the cDNA sequence (Accession number: AF339909) with no amino acid change. The substitution results in the disappearance of a *RsaI* recognition site. Therefore, the *RsaI* digestion of the 447 bp PCR product from the G allele produces three fragments of 98, 111, and 238 bp, whereas the T allele generates only two fragments of 98 and 349 bp (Figure 1).

Primer sequences. Primers were located in the intron 4 (forward primer) and in the intron 5 (reverse primer) of the *SCD* gene. Forward primer: 5'-AGT GTA GAA GGG ACA GCC CAG C-3'; reverse primer: 5'-GTG GAA TGA CAC ATG GAG AGG G-3'.

PCR Conditions: The PCR reaction was performed in a 25- μ l final volume, containing 0.625 units of *Taq* DNA polymerase (Life Technologies) with 1xPCR buffer, 2 mM MgCl₂, 200 μ M of each dNTP, 0.4 μ M of each primer and approximately 100 ng of goat genomic DNA. Thermal cycling conditions were 95°C for 5 min, 10 cycles of 97°C for 15 s, 63°C for 1 min, and 72°C for 90 s, then 25 cycles of 95°C for 30 s, 63°C for 1 min, and 72°C for 90 s, followed by a final extension at 72°C for 5 min.

Polymorphism and allele frequency:

Protocol for rapid genotyping of the polymorphic site was developed by PCR-RFLP procedure using the *RsaI* endonuclease (Roche): ten microliters of the PCR product was digested overnight at 37°C and separated by electrophoresis in a 2% agarose gel. A total of 99 genomic DNA samples from animals of different Spanish (Canaria n=27, Malagueña n=12, Payoya n=7, and Murciano-Granadina n=29) and French (Saanen n=29) breeds were analysed. The frequencies of the G allele were 0.92, 0.69, 1, 1, and 1 respectively, in the five breeds.

Chromosomal location and comparative mapping information:

Chromosomal location was obtained by genetic analysis of the genotypes of goat reference half-sib families². Amongst the twelve progenitors three appeared polymorphic and made it

possible to obtain 55 informative meioses. Two-point analysis was carried out using the CRI-MAP software and resulted in five significant lod scores with five markers of CHI26, INRA081 (?=0.09; Lod = 4.36), INRA172 (? = 0.13; Lod = 3.34), BM4505 (? = 0.11; Lod = 3.66), LSCV046 (? = 0.13; Lod = 4.18) et LSCV052 (? = 0.03; Lod = 5.46). Haplotype analysis was carried out by visual analysis of the recombinant chromosomes, suggesting the most probable order: Centromere – HEL11 - Bm1314 – INRA081 – INRA172 – LSCV052 – SCD – BM4505 – LSCV046 – Telomere (Figure 2). Compared with the use of the build option of CRIMAP which was used in the map construction of CHI26³, haplotype reconstitution of the chromosome results in an order much more similar to the one described in the sheep map of OAR22 (the homologue of CHI26). The cytogenetic mapping of LSCV046 to 26q13 and of LSCV052 to 26q21, together with the very close association of LSCV052 (originating from a BAC containing the gene *PAX2*) indicate that the two genes are close. These results are in accordance with those previously reported using FISH technique⁴. The human chromosome region is HSA10, homologous to segments of goat chromosomes 13, 28 and 26. In humans, *SCD* has been mapped to HSA 10q24.32, whereas *PAX2* is mapped to HSA 10q25.1. The interspecific homologous mapping of the two genes suggests a good conservation in this mammalian chromosome segment.

References

- 1 Yahyaoui, M.H. *et al.* (2002) *J. Anim. Sci.* **80**, 886-887.
- 2 Vaiman, D *et al.* (1996) *Genetics* **144**, 279-305.
- 3 Schibler L. *et al.* (1998) *Genome Res.* **8**, 901-15.
- 4 Bernard, L. *et al.* (2001) *Gene* **281**, 53-60.

Figure 1. Genotyping of the mutation detected in the *SCD* exon 5 by PCR-RFLP using *RsaI* endonuclease. Lane 1: homozygous for G, lanes 2 and 3: heterozygous for G/T, lane 4: homozygous for T.

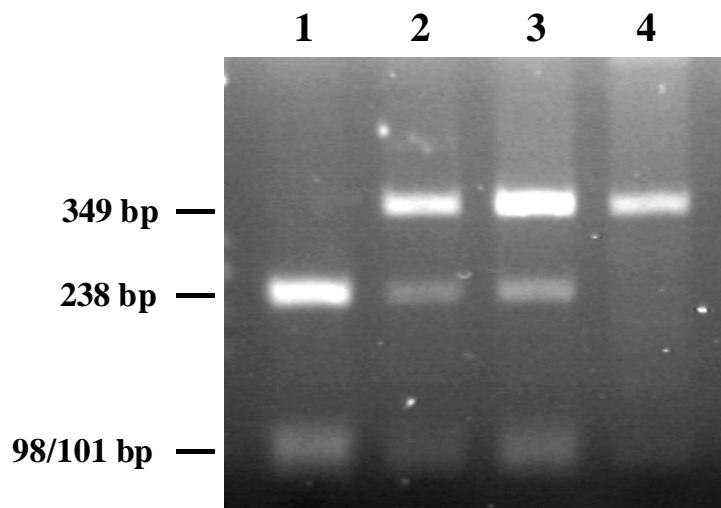
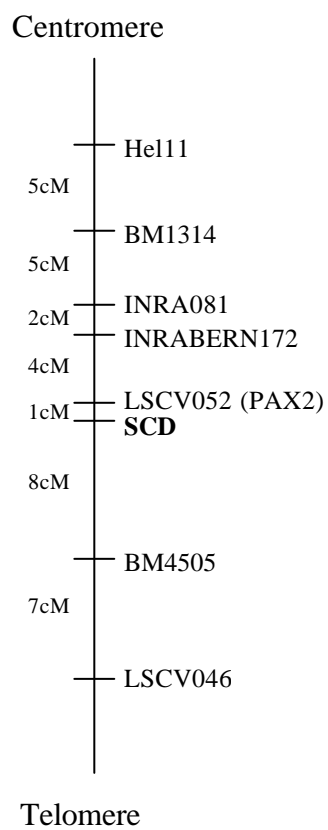


Figure 2. Partial genetic map of goat chromosome 26, encompassing *SCD*.



DISCUSSION

1. Mutation scanning

For the initial screening of mutations in the *k-casein*, *b-LG*, and *SCD* genes, the BESS T-Scan detection kit (Base Excision Sequence Scanning, Epicentre Technologies) was used as a low-cost method that was relatively rapid and simple. The method relies on the use of limiting amounts of dUTP during PCR reaction followed by enzyme treatment with uracil *N*-glycosylase (UNG) and endonuclease VI. UNG hydrolyzes the uracil-glycosidic bond (base excision) releasing uracil and creating abasic sites in the DNA. Endonuclease VI specifically cleaves the phosphodiester bond 3' to the abasic sites, generating a series of DNA fragments that produce a ladder pattern virtually identical to a thymine sequencing ladder (Hawkins and Hoffman, 1997). Polymorphisms are indicated by a change in the ladder pattern when compared with one of a known sequence. While all mutations involving thymine are theoretically detected by this method, a base change is undetectable when it only involves both cytosine and guanine (C to G or G to C). The method have been used for the identification of mutations in hereditary colorectal cancer (Brieger et al., 1999), virus isolates (Charrel et al., 1999; Hawkins and Hoffman, 1999), and human cestodes parasites typing (Yamashaki et al., 2002).

In this study, the BESS method has been coupled with capillary electrophoresis and fluorescent detection using an automated ABI 310 DNA sequencer (Applied Biosystems). The sense and antisense DNA strands were simultaneously screened by labeling forward and reverse primers with HEX (6-carboxy-hexachlorofluorescein) and FAM (6-phosphoramidite-fluorescein) fluorescent dyes, respectively. In order to determine optimal reaction conditions, goat and sheep DNA fragments that have been previously sequenced were generated by PCR and used following the original protocol. The amplified regions contain parts of proximal promoter and exon 1 of the *b-LG* gene (710 bp) and the major part of κ -casein exon 4 (459 bp).

Various amounts of dUTP (corresponding to final concentrations of 4, 8, 12, and 16 μM) were used in the PCR reaction. The best patterns in terms of signal intensity, persistence, and noise (background) were obtained with 16 and 4 μM of dUTP for κ -casein and β -LG, respectively. This is not unexpected given the nucleotide composition of the analyzed fragments. While the AT content of the κ -casein fragment is relatively high (55.2 %), the β -LG fragment contains only 36.6% of AT nucleotides; the *b-LG* gene is located in a GC-rich region of the genome (Bernardi et al., 1985). During PCR amplification, the dUTPs are

incorporated in a random manner similar to ddNTPs incorporation in Sanger sequencing, when any given dTTP is replaced by dUTP in a portion of DNA molecules. As a result, the signal intensity decreases with the fragments size, and depending on the nucleotide sequence, fragments larger than 300-350 bp are difficult to detect. The use of high amounts of dUTP in the PCR amplification of GC-rich fragments resulted in an excessive incorporation of uracil, giving rise to a high proportion of short fragments with weak signal intensities for longer bands. Thus, lowering dUTP quantities may limit signal decrease for such fragments.

In addition to dUTP concentration and after duplicating several BESS experiments, it was also apparent that PCR conditions heavily influenced the appearance of the final peak profile, and that some poor-quality templates would not yield satisfactory results. Modified BESS protocols with PCR product purification (from agarose gel) prior to excision and cleavage reactions have been reported (Brieger et al., 1999). Here, we have optimized PCR conditions (through primer design and use of hot-start PCR using *Taq* Gold DNA polymerase) to obtain a specific amplification (single band in agarose gel) rather than purifying PCR product, which is not indicated in the original protocol.

Regarding enzyme treatment, although AT-rich fragments require theoretically more PCR product for excision and cleavage reactions, neither amount of PCR product nor incubation time seem to affect significantly signal intensities. Figure 6 shows representative results of BESS T-scan analysis. All potential mutations were sequenced to confirm results. The two fluorescent dyes (Hex and Fam) produced similar signal peak intensities in both genes analyzed. The peak patterns represent fragments ending with dUTP and are virtually identical to the “T” lane of a conventional sequencing reaction. The appearance, disappearance, or change in intensity peaks indicates the presence of a mutation. As already mentioned, signal peak intensity decreases with fragment length but it depends also on base composition of the template. Signal intensity is more persistent in the case of β -LG (up to 600 bp) than for κ -casein, where T peaks longer than 350 bp were almost undistinguishable from background. This is probably due to the high amount of thymine content of κ -casein exon 4, which resulted in a large number of digestion products. The final peak correspond to the full-length PCR product that remains uncleaved and is a consequence of formulating the enzyme mix to generate longer excision-cleavage fragments.

The peak pattern should theoretically correspond to a “T” lane sequencing ladder, which represents the banding profile of dUTP incorporation during PCR reaction. However,

there was a lot of inconsistency between obtained peak profiles and corresponding sequences, and false positive were quite frequent. These false positive peaks were especially located close to fluorescent primer (< 50 bp) where signal intensities and background are very high. Small fragments are detected concurrently with PCR artifacts (primer-dimer, remaining labeled primers) resulting in a scrambled signal difficult to interpret. To overcome this problem, primers should be placed outside of the region of interest.

In addition to these false positives, subtraction and shifting peaks were obvious on the gel pattern (thymine in the sequence with no corresponding peak in the profile), especially in κ -casein. However, these “missing” peaks were not observed when analyzing β -LG and SCD fragments. Of the seven mutations found in κ -casein exon 4 fragment, five have been missed and only two have been correctly detected by BESS (Figure 7). These mutations are potentially detectable since they correspond to thymine substitutions on one of the DNA strands. Despite several attempts, no differences in the peak profiles for these mutations were observed by comparing control and mutated samples. This may be explained by the non-recognition of these cleavage sites by the BESS excision enzyme, whose activity depends on nucleotide composition and sequence context. The fact that missing peaks were not observed in β -LG and SCD peak profiles suggest that the BESS method works better on low AT-content templates. Contrary to the false positive peaks that can be overcome by primer design, the problem of the missing peaks is more difficult to surmount, and constitute a serious limitation of the method in the detection of mutations.

Overall, although BESS T-scan proved to be a relatively simple and rapid protocol that resulted in easy data analysis when coupled to fluorescent detection, it requires further optimizations for each individual PCR product to be screened. The method appears to be suitable only for initial scanning of short fragments (200-300 bp) with more effectiveness when the AT-content of the target is relatively low. The main advantage of the method is that allows approximate localization of the detected mutations.

Figure 6. (next page) BESS technique.

Typical aspect of pick pattern generated by BESS T-Scan method. The vertical axis represents fluorescence intensity (number of fragments at a given cleavage site), while the horizontal axis represents increasing molecular size. The final pick corresponds to the PCR product that remains uncleaved after enzyme treatment. The analyzed fragment is 459-bp long from κ -casein exon 4.

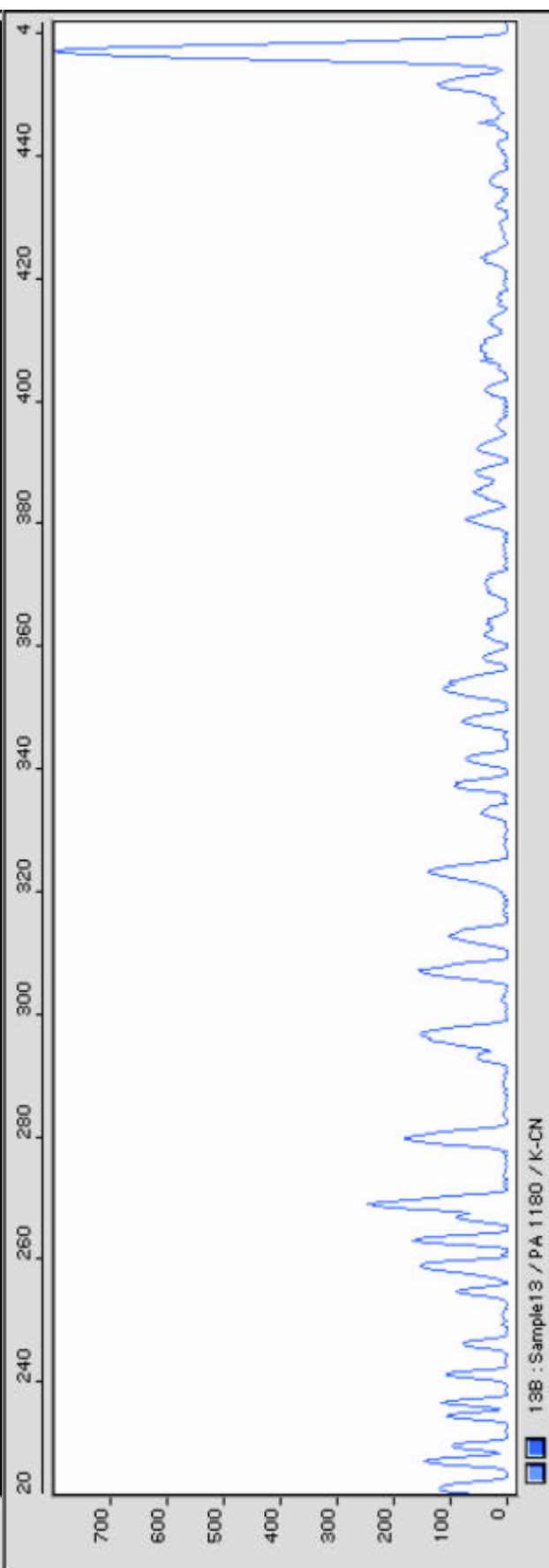
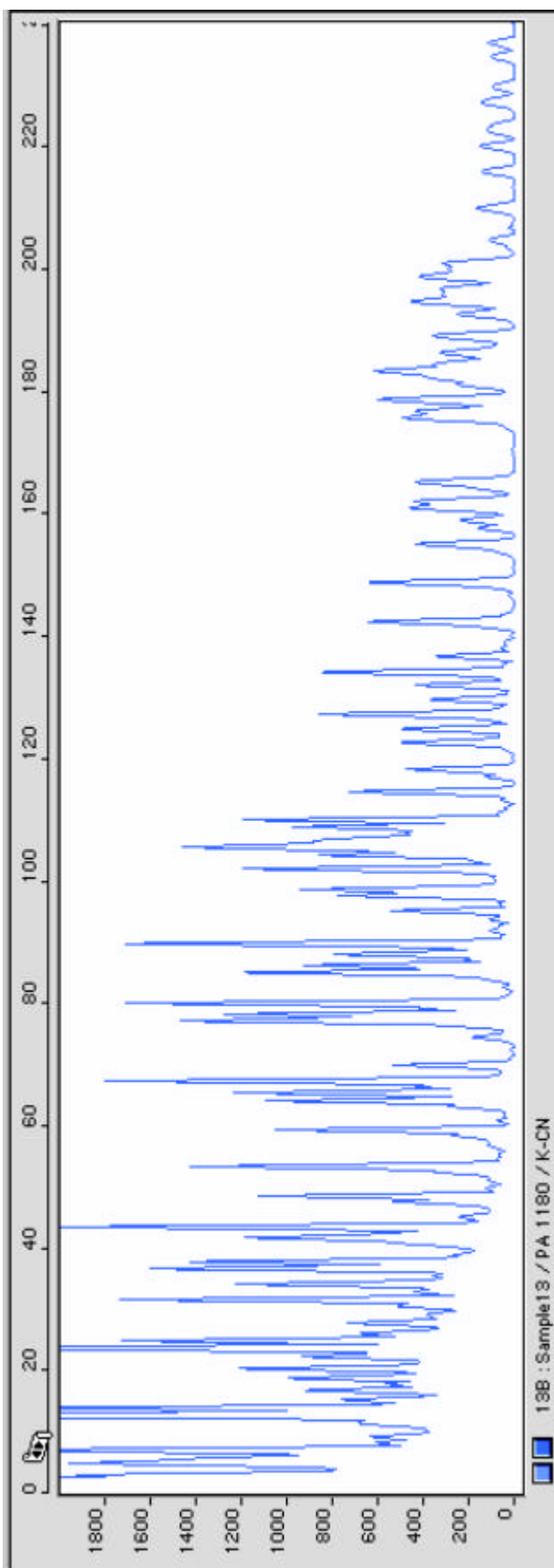
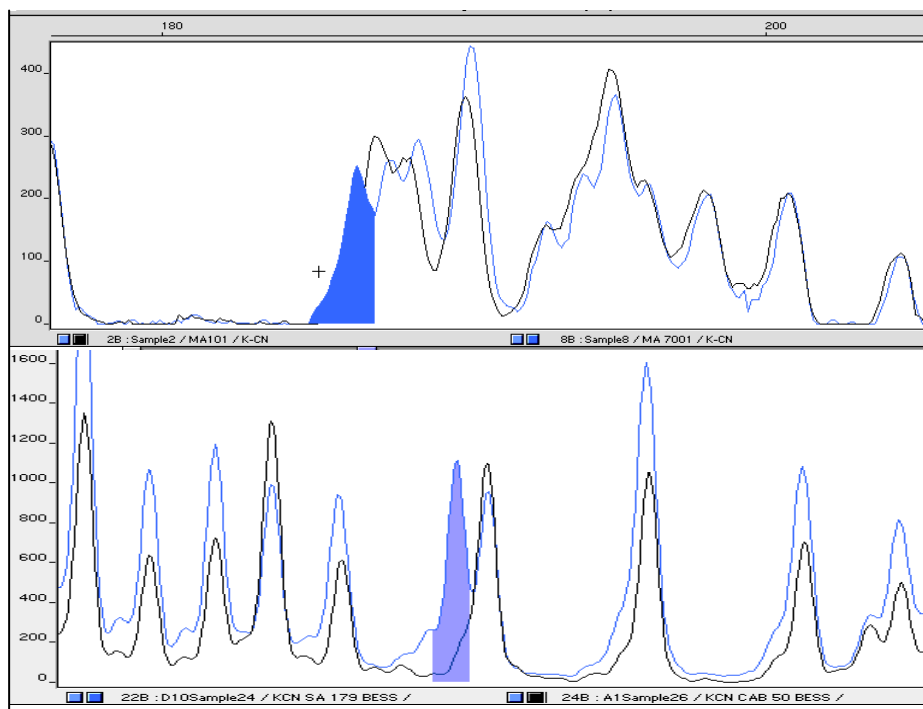
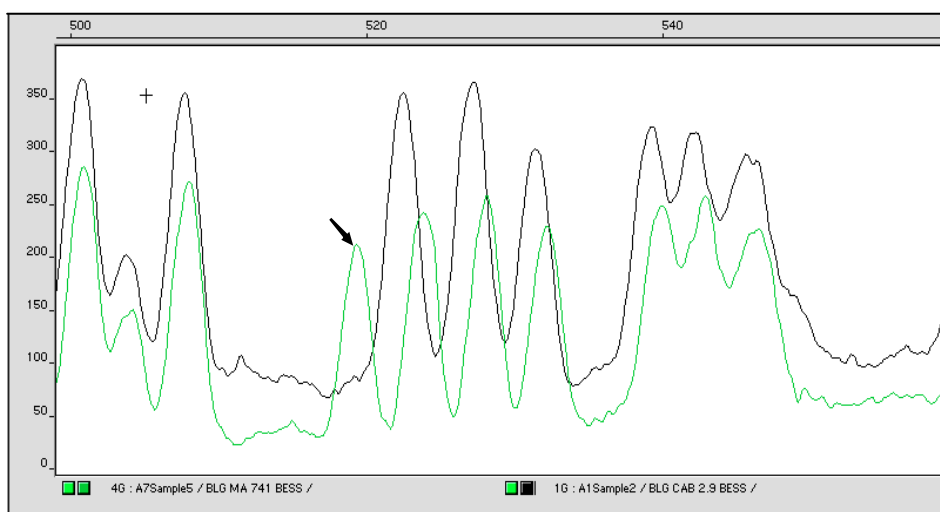


Figure 7. Mutations detected by BESS T-Scan method. **(a)** In kappa casein exon 4. Additional picks (corresponding to a thymine base in the sequence) are shown in mutated samples (in blue) compared to control (black). **(b)** In the proximal promoter region of the β -*LG* gene. The additional pick (arrow) corresponds to thymine in mutated sample (green) compared to control (black).



(a)



(b)

In addition to BESS T-scan, we also evaluated the enzyme mismatch cleavage (EMC) method, available as a commercial kit (Passport; Pharmacia Biotech). The EMC assay is based on the use of T4 endonuclease VII that cleaves DNA within 6 bp on the 3' side of the mismatches in double-stranded DNA (Solaro et al., 1993). These mismatches are created after hybridization of wild type and mutated DNA fragments. The formation of new cleavage products indicates the presence of mutation, and the size of the fragments indicates the approximate location of the mutation.

Fluorescently labeled κ -casein and β -LG fragments from goat and sheep (the same used in BESS) were amplified by PCR and hybridized (goat *versus* sheep for each fragment) following supplier recommendations. Sequence alignment revealed 10 mismatches between goat and sheep κ -casein fragments, and 29 mismatches in the case of the β -LG, from which one in the first exon and 28 in the proximal promoter region. The generated duplexes were then digested with T4 endonuclease VII, and the resulting fragments analyzed on an ABI310 automated sequencer.

In theory, the cleavage products should correspond to the size range of mismatches as expected from the sequence. Unfortunately, the enzyme appeared to have a high non-specific cleavage that gives rise to high intensity background peaks. The signal to noise ratio was not sufficient to enable the identification of peaks corresponding to the mutations (mismatches) from the non-specific peaks. In addition, the templates did not exhibit identical cleavage patterns when treated with the enzyme in repeated experiments and under identical conditions. A high background signal have been reported by several authors, indicating that the T4 endonuclease VII cleavage is dependent on the particular mismatch as well as the surrounding sequence context (Solaro et al., 1993; Pottmeyer and Kemper, 1992; Youil et al., 1995, 1996). Specifically, G-containing mismatches are less preferred than C-containing mismatches (Pottmeyer and Kemper, 1992). However, this should not be a problem here because the two generated heteroduplexes were analyzed (both primers were fluorescently labeled). Only C:G mismatches (three are found in β -LG heteroduplexes) are expected to be difficult to be detected.

It is possible that the κ -casein and β -LG fragments analyzed here contain the sequence motif(s) that is responsible for the nonspecific cleavage of the enzyme. It is also possible that the enzyme vial was of sub-optimal quality, or was contaminated in some

fashion. However, due to time constraints, this was no further investigated and the method was not used for mutation screening.

2. Kappa casein polymorphism

At the beginning of this work, no genetic variants for goat κ -casein had been characterized, although several works using different techniques (protein electrophoresis, chromatography, capillary electrophoresis) had reported polymorphism in the protein.

The full coding sequence of the caprine *k-casein* gene, which encompasses exons three (9 amino acids) and four (162 amino acids) was screened for polymorphisms. A 459 bp-long fragment of exon 4 which contain the major part of the coding sequence (141 amino acids out of a total of 171) was initially screened by BESS Tscan method, and potential polymorphic samples were sequenced. Subsequently, the remaining coding region (exon 3 and the lasting part of exon 4) was analyzed by sequencing.

Due to the confusion now present in the literature caused by the assignation of the same name to different variants, in particular as concerns the B variant, a new nomenclature for the different alleles was proposed during this work (Yahyaoui et al., 2003; In press). This nomenclature takes into account the chronological order report of the corresponding published references, and is used in this thesis to refer to the different variants.

As aforementioned, a total of seven polymorphic sites were detected in the 459 bp fragment of exon 4, from which two by BESS method (Figure 7) and five by direct sequencing (Yahyaoui et al., 2001). All were point mutations corresponding to base transitions, where T and G were substituted by C and A, respectively. It is well known that the frequencies of the different SNPs types are not equal, with transitions much more common than transversions, representing about 2/3 of all SNPs found in the human genome (Wang et al., 1998; Halaschuka et al., 1999; Dawson et al., 2001). The higher level of C/T substitution (including reverse complement G/A) probably reflects the deamination of 5-methylcytosine that occurs frequently at CpG dinucleotides (Duncan and Miller, 1980) and results in converting cytosine to thymine. This molecular mechanism must be a major cause of the great deficiency of CG dinucleotides observed in genomic sequences (Miller and Kwok, 2001).

Three polymorphisms are synonymous substitutions corresponding to amino acids residues tyrosine, leucine, and threonine, at positions 43, 56, and 131, respectively. The

other four mutations produce codon changes of valine to isoleucine (at positions 65 and 119 of the protein), alanine to valine (at position 156), and serine to proline (at position 159). These mutations result in three genetic variants designated A, B, and C (Yahyaoui et al., 2001). Variant A corresponds to amino acid sequence of the mature protein deduced from the caprine κ -casein cDNA (Coll et al., 1993), while the substitution of valine by isoleucine at position 119 constitutes the difference between variants A and B. These variants were first observed by Mercier et al. (1976a) when sequencing the caseinomacropeptide from the milk of an Alpine-Saanen goat. They found either valine or isoleucine at this position; however, they considered that isoleucine is the predominant residue because it was found in the sequence of bovine and ovine species. The κ -casein C variant differs from the A variant by four amino acid substitutions: isoleucine for valine at positions 65 and 119, valine for alanine at position 156, and proline for serine at position 159.

In addition to these seven mutations, we also detected a single nucleotide polymorphism at position 104 of exon 4 producing an amino acid substitution of glutamine to arginine (at position 44 of the protein). However, this mutation was found to be very rare in the Spanish and French breeds; from more than 150 animals analyzed, only one animal carried the mutation (in heterozygous way). The presence of this polymorphism have been later confirmed by Caroli et al., (2001) in Italian and German breeds where it was found with higher frequencies. This polymorphism results in D variant, the only one with arginine at position 44 of the protein, while all other variants present glutamine at this position. This amino acid substitution, located in the N-terminal region of the protein, leads to a change in the net charge of the protein. It most likely corresponds to the polymorphism reported by Di Luccia et al. (1990) in a local Italian goat breed, and by Law and Tziboula (1993) in the British Saanen breed and in a commercial herd in Greece, since it was postulated to be located in the N-terminal region with substitution of an acid residue for a neutral residue.

In collaboration with the F. Pilla group (University of Molise, Italy), different Italian goat breeds have been included in the κ -casein polymorphism analysis. An additional mutation in exon 4 (at position 242) was detected when sequencing the remaining coding region (exons 3 and 4) of *k-casein* gene from different Spanish, French, and Italian goat breeds. This mutation produce codon change of aspartic acid to glycine (at position 90 of the protein) and resulted in the E variant, which appears to be specific of the Italian Montefalcone goat (Angiolillo et al., 2002) and is not found in other breeds.

Finally, two other variants, κ -casein F and κ -casein G, were observed in Italian breeds and they result from the associations between these different mutations (Yahyaoui et al., 2003, In press). The F variant was also found in the Spanish wild goat *Capra pyrenaica sp. hispanica*.

Regarding exon 3, sequence analysis revealed a single nucleotide substitution (A to G) at position 27 (from the start of the exon) in Spanish and Italian breeds. This mutation is located at the third base of the codon (glutamine) and therefore does not produce any amino acidic change.

In total, six non synonymous mutations were detected within κ -casein exon 4 that result in seven genetic variants: A, B, C, D, E, F, and G (Yahyaoui et al., 2003, In press). Three mutations are located in the N-terminal region (at positions 44, 65, and 90) and the other three in the caseinomacropetide portion of the protein (Figure 8). None of these mutations affect glycosylation or phosphorylation site, the mutation involving serine residue at position 159 is not located within the phosphorylation recognition sequence of kinases as described by Mercier, (1981).

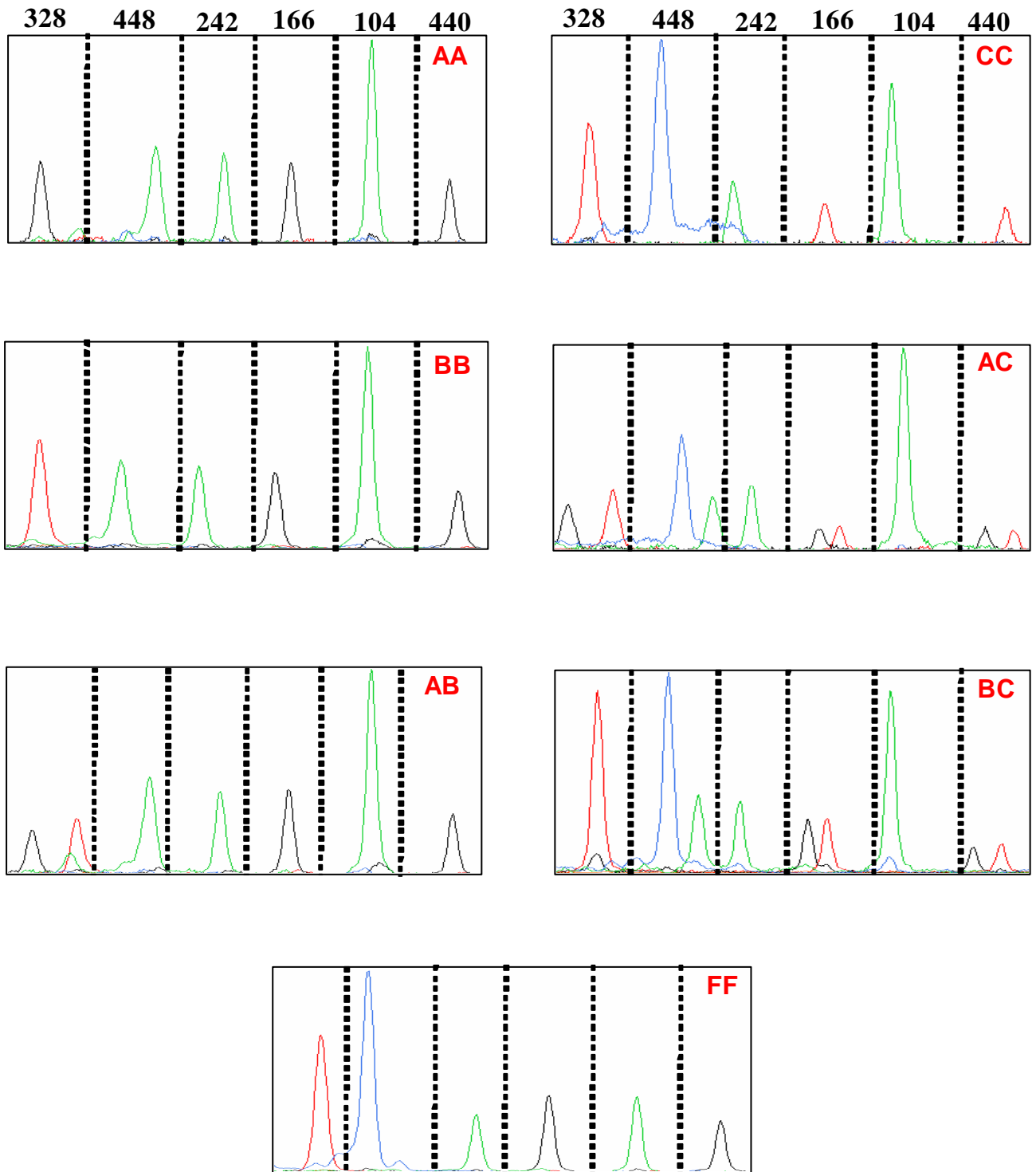
Given the high variability observed in the *k-casein* gene, the importance of developing molecular methods for a rapid and easy genotyping of all variants is evident. Using PCR-RFLP for the endonucleases *Alw44I*, *BseNI*, and *HaeIII*, we have reported the frequency of the C and E alleles in different European goat breeds (Yahyaoui et al., 2001; Angiolillo et al., 2002). However, these endonucleases do not discriminate between alleles A and B or between C and the new allele G. The PCR-RFLP analysis indicated that variants A and B are present at high frequencies in the analyzed breeds. We therefore designated allele-specific primers for differentiation of these two alleles by AS-PCR. Nevertheless, and due to the poor discrimination property of DNA polymerase, the method requires extensive optimization of PCR conditions for reproducible results. In addition, a duplicate reaction is required for each allele typing rendering the method labor-intensive and time consuming, and therefore not suitable for routine use.

The primer extension analysis (PEA) allows genotyping of known SNP alleles by using appropriate primers and fluorescent detection. This technique uses DNA polymerase to extend an oligonucleotide primer immediately adjacent to a known mutation on an amplified target molecule, in the presence of fluorescently labelled dideoxy-nucleotides which act as a terminal base. We have applied this method to κ -casein typing using the

Figure 8. Multiple alignment of the predicted amino acid sequences of κ -casein protein (including signal peptide). Sequences were obtained from GeneBank (accession numbers are indicated in boldface) and aligned using the Multialin software. The position of the chymosin bond is indicated by the vertical bar. Positions of the six polymorphic positions in goat κ -casein are indicated in yellow. Conserved amino acids are indicated by asterisk. Caprine and bovine κ -casein sequences correspond to A and B variants, respectively.

	1				50	
camel	.MKSFFLVVT	ILALTLPFLG	AEVQNQEPT	CFEKVERLLN	EKTVKYFPIQ	
pig	MMKSSFLIVP	ILALTLPFLG	AEEQNQEKL	RCESDKRLFN	EKVKYIPIY	
human	.MKSFLLVFN	ALALTLPFLA	VEVQNQKQPA	CHENDERPFY	QKTAPYVPMY	
goat	MMKSFFLVVT	ILALTLPFLG	AQEQNQEPI	CCEKDERFFD	DKIAKYIPIQ	
sheep	MMKSFFLVVT	ILALTLPFLG	AQEQNQEPI	CCEKDERFFD	DKIAKYIPIQ	
cow	MMKSFFLVVT	ILALTLPFLG	AQEQNQEPI	RCEKDERFFS	DKIAKYIPIQ	
rabbit	MMKHFLVFN	ILAVTLPFLA	ADIQNQEPT	CRENEERLFH	QVTAPYIPVH	
mouse	MMRNFIVVMN	ILALTLPFLA	AEIQNPDSNC	RGEKNDIVYD	EQRVLYTPVR	
rat	MMRNFIVVMN	ILALTLPFLA	AEVQNPDNSC	R.EKNEVVYD	VQRVLYTPVS	
		** *****	**	*	* *	
	51				100	
camel	FVQSRYPYSG	INYYQHRLAV	PIN.NQFIPY	PNYAKPVAIR	LHAQIPQCQA	
pig	YMLNRFPSYG	F.FYQHRSV	SPN.RQFIPY	PYYARPVVAG	PHAQKPQWQD	
human	YVPNSYPYSG	TNLYQRRPAI	AIN.NPYVPR	TYANPAVVR	PHAQIPQRQY	
goat	YVLSRYPYSG	LNYYQRPVA	LIN.NQFLPY	PYYAKPVAVR	SPAQTLQWQV	
sheep	YVLSRYPYSG	LNYYQRPVA	LIN.NQFLPY	PYYAKPVAVR	SPAQTLQWQV	
cow	YVLSRYPYSG	LNYYQKPVA	LIN.NQFLPY	PYYAKPAVR	SPAQILQWQV	
rabbit	YVMNRYPQYE	PSYYLRRQAV	PTL.NPFMLN	PYYVKPIVFK	PNVQVPHWQI	
mouse	SVLN.FNQYE	PNYYHYRPSL	PATASPYMY	PLVVRLLLLR	SPAPIKQWS	
rat	SVLN.RNHYE	PIYYHYRTSV	PV..SPYAYF	PVGLKLLLLR	SPAQILKQWP	
		* *		*	*	
	101				150	
camel	LPNI.....	..DPPTVERR	PRPRPSFLAI	PPKKTQDKTV	NPAINTVATV	
pig	QPNV.....	..YPPTVARR	PRPHASFLAI	PPKKNQDKTA	I PAINSIATV	
human	LPNS.....	..HPPTVVRR	PNLHPSFLAI	PPKKIQDKII	I PTINTIATV	
goat	LPNTVPAKSC	QDQPTTLARH	PHPHLSFMAI	PPKKDQDKTE	Y PAINTIASA	
sheep	LPNAVPAKSC	QDQPTAMARH	PHPHLSFMAI	PPKKDQDKTE	I PAINTIASA	
cow	LSNTVPAKSC	QAQPTTMARH	PHPHLSFMAI	PPKKNQDKTE	I PTINTIASG	
rabbit	LPNI.....	..HQPKVGRH	SHPF..FMAI	LPNKMQDKAV	TPTTNTIAAV	
mouse	MPNF.....	..PQSAGVPY	AIPNPSFLAM	PTNENQDNTA	I PTIDPITPI	
rat	MPNF.....	..PQPVGVPH	PIPNPSFLAI	PTNEKHDNTA	I PASNTIAPI	
	*		** *		*	
	151				199	
camel	EPPVIPT...	...AEPVNT	VVIAEASSEF	ITTSTPETTT	VQITSTEI.	(Y10082)
pig	EPTIVPATEP	IVNAEPIVNA	VVTPEASSEF	LITSAPETTT	VQVTSPVV.	(X51977)
human	EPTPTPA...	...TEPTVDS	VVTPEAFSES	IITSTPETTT	VAVTPPTA.	(M73628)
goat	EPTVHSTP..	..TTEAIVNT	VDNPEASSE.	SIASASETNT	AQVTSTEV.	(X60763)
sheep	EPTVHSTP..	..TTEAVVNA	VDNPEASSE.	SIASAPETNT	AQVTSTEV.	(X51822)
cow	EPT..STP..	..TIEAVEST	VATLEASPE.	VTESPPEINT	VQVTSTAV.	(M36641)
rabbit	EPTPIPT...	...TEPVVST	EVIAEASPEL	IIS..PETTT	EATAASAAA	(Z18243)
mouse	VSTPVPT...	...MESIVNT	VANPEAST..	VSINTPETTT	VPVSSTAA.	(M10114)
rat	VSTPVST...	...TESVNT	VANTEAST..	VPISTPETAT	VPVTSPAA.	(K02598)
		*	**	** *		

Figure 9. Genotyping of kappa casein variants by primer extension analysis. The six mutations (positions indicated below) were analyzed simultaneously. Genotypes are indicated in red.



Multiplex SnaiPshot ddNTP kit (Applied Biosystems), and developed a set of primers that allows to distinguish unambiguously the different goat κ -casein alleles. Allele discrimination is based on both primer size and color assigned to the fluorescent ddNTPs incorporated during extension reaction (Figure 9). The primers were optimized to be used simultaneously in a single reaction, thus reducing cost, effort, and time of genotyping. Furthermore, and because an automated instrument was used, the method is suitable for large screening and genotyping in selection schemes and industrial purposes.

Although the primers were optimized to be used simultaneously in a single reaction, alleles A, C, D, and E can be individually differentiated from others by single genotyping using corresponding primers. Allele E can also be distinguished by PCR-RFLP with *HaeIII* endonuclease, being this allele the only one with guanine (Gly) at position 242. Using the same technique, the *BseNI* allows discrimination of alleles A, B, E, and F from others, whereas *Alw44I* distinguishes only A, B, and E alleles. Consequently, the F allele can be differentiated combining these two endonucleases.

Allele frequencies of the different κ -casein variants in several goat breeds are given in Table 6. Kappa casein A and B are the most frequent variants in the majority of analyzed breeds; κ -casein B is the predominant variant, with the exception of the Canaria breed where allele A is the most frequent (0.6 vs. 0.4). The A variant, present in all other breeds, is not found in the Italian Teramana and Montefalcone breeds.

Apart from E allele, which is specifically present at high frequency (0.41) in the Italian Montefalcone breed, other variants occur at relatively low frequencies, less than 15 %. The C variant is particularly present in Saanen (13 %), whereas the F alleles is found only in the Italian breeds Teramana, Girgentana, and Sarda. Paradoxically, these three breeds are in danger of extinction (FAO; <http://fao.org/dad-is>), and despite their small number, they present a high variability at the κ -casein locus. This may be explained by the fact they consist of a localized and heterogeneous populations that remained without specific selection pressure, and probably having distinct origin compared to other European goat breeds. The G variant is comparatively more frequent than F allele, and is also found in Maltese and Cilentena Nera, a two long established breeds with large populations. As already mentioned, the D variant is not found in Spanish and French breeds, and its occurrence in Italian breeds is low. This variant is also present in Cashemere goat (3.5 %).

Table 6. Frequencies of the kappa casein variants in different goat breeds.

Breeds	N° animals	Variants						
		A	B	C	D	E	F	G
<u>Spanish breeds</u>								
Murciano-Granadina	30	0.37	0.63	-	-	-	-	-
Malagueña	11	0.45	0.55	-	-	-	-	-
Canaria	30	0.58	0.42	-	-	-	-	-
<u>Italian breeds</u>								
Teramana	28	-	0.70	0.02	0.10	-	0.14	0.04
Montefalcone	17	-	0.59	-	-	0.41	-	-
Girgentana	19	0.34	0.45	-	0.05	-	0.05	0.11
Sarda	19	0.31	0.61	-	-	-	0.08	-
Maltese	29	0.09	0.88	-	0.015	-	-	0.015
Garganica	45	0.08	0.92	-	-	-	-	-
Cilentena Nera	32	0.2	0.78	-	-	-	-	0.02
<u>French breeds</u>								
Alpine	28	0.34	0.66	-	-	-	-	-
Saanen	28	0.39	0.48	0.13	-	-	-	-
<u>Other breeds</u>								
Cashemere	15	0.035	0.93	-	0.035	-	-	-
<u>Wild goat</u>	23	-	0.02	-	-	-	0.98	-

In addition to allele D, variants F, G, and E are not found in the analyzed Spanish and French breeds. As regard the Spanish wild goat *Capra pyrenaica*, only B and F alleles are present. However, in contrast to the low frequencies of the F allele observed in goat breeds, this variant is by far the most frequent allele found in this specie; from 23 animals genotyped, 22 are homozygous FF and only one heterozygous BF.

From the 28 expected genotypes, only 18 were observed in the analyzed breeds. All populations were in Hardy-Weinberg equilibrium at κ -casein locus, with the exception of Alpine, Maltese, Cilentena Nera, and Garganica where a deficiency of heterozygous genotypes was observed.

When considering the distribution of milk protein variants among different breeds, it is of interest to know the relative appearance in time of different alleles, which may provide clues about the origin and relationships between the different populations. Identification of the original variant is generally based on two different criteria: sequence data and frequency (Ng –Kwai Hang and Grosclaude, 1992). When compared to the protein (or nucleotide) sequence of related species, the original variant should normally give the closest sequence alignment. In addition, the corresponding allele is expected to have usually the highest frequency over a large number of populations.

Among the seven genetic variants characterized, only the B variant was found in all studied populations and is predominant over other alleles, except in Spanish Canaria breed. It could therefore be considered as the original variant on the basis of frequency criterion. However, when considering comparative sequence data, variant F is identical at position 159 to the κ -casein of other species (sheep, cattle, pig, mouse, rat, rabbit, camel, and human; Figure 8), all having proline, while the B variant presents serine at this position. In addition, the F variant is identical to the sequence of the κ -casein exon 4 obtained from the Spanish wild goat. The similarity of sequences clearly indicates that this variant is likely the original type.

The reasons of the divergence in the conclusions of the frequency and of the sequence conservation criteria are unclear. Because the κ -casein plays a critical role in several important physiological processes, it seems unlikely that this gene would be completely free of the selective pressure. By analyzing the pattern of nucleotide substitution in κ -casein sequences (377 bp from exon 4 and 407 bp from intron 4) for distantly related species in *bovidae* family, Ward et al. (1997) showed that positive selection has accelerated

their divergence at the amino acid level. Generally, the level of polymorphism is shown to be exceptionally elevated in proteins for which positive selection has been demonstrated (Hughes and Nei, 1989; Metz and Palumbi, 1996). The low frequency of variant F could have come through selection and it was very early supplanted by type B. In this regard, it is important to notice that the chemical and nutritional properties of the amino acid residues serine and proline (in F and B variants, respectively) are very different.

It has been suggested that the occurrence of certain casein variants in bovine specie is correlated with selection of improvement in production traits (Lin et al., 1986; Jann et al., 2002). As a consequence, selection pressure is expected to lead to increasing fixation of alleles linked to production traits. This would also explain the relatively elevated occurrence of C allele (13 %) shown in French Saanen goat comparatively to other goat breeds; French Saanen is a highly milk selected breed. However, it seems more plausible that genetic drift and founder effects have played a determinant role in the allelic distribution of κ -casein variants, and the prevalence of the B variant may therefore have occurred through these effects. Nevertheless, the distribution of κ -casein variants remains to be established. Until further developments by typing more goat populations either wild type or goat breeds, especially non-European breeds, the F variant could be considered as the ancestral type. The different alleles were most likely derived from this original type by successive mutations following two distinct trunks: A, B, E which occur, when found, at high frequencies, and C, D, G which are observed at low frequencies.

In bovine specie, comparative sequence data revealed that variant A is identical with κ -caseins of other species at position 136 (threonine). On the contrary, at position 148 (alanine), it is variant B which is identical with the same proteins. Because variant A is predominant in the majority of the analyzed breeds, it is considered as the original type of κ -casein.

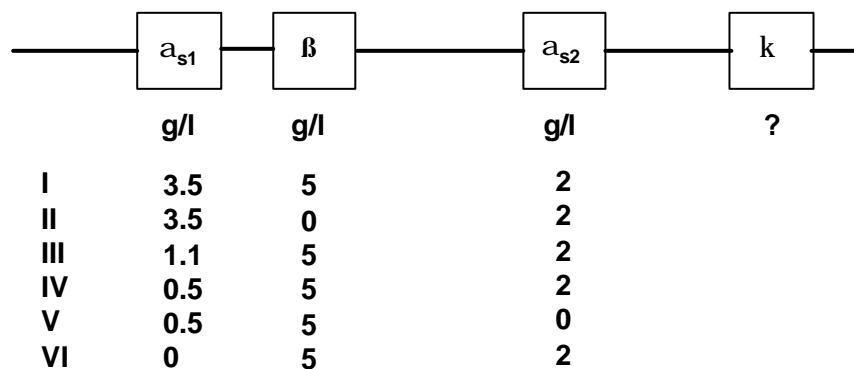
With the characterization of κ -casein variants, the four casein genes are now shown to be highly polymorphic in goat specie, with at least 12 alleles for *as1-casein*, 4 alleles for *as2-casein* and 3 alleles for *b-casein* genes (revised in Moioli et al., 1998; Barroso et al., 1999). Alleles of the three calcium-sensitive genes have been associated with strong differences in their level of expression resulting in different amounts of the corresponding protein in the milk: four levels of α s1-casein (3.5, 1.1, 0.45, and 0 g/l), two of α s2-casein (2.0 and 0 g/l), and two of β -casein (5.0 and 0 g/l). The alleles associated with a null

amount of protein are characterized by the presence of nonsense mutations in the case of α s2- and β -casein (Persuy et al., 1999; Ramunno et al., 2001) whereas both a large deletion and large insertion are responsible for the absence of α s1-casein in goat milk (Martin et al., 1999).

In this regard, the high level of genetic variability observed at the casein loci and its association with milk traits allows to obtain groups of animals producing types of milk which could be more suitable for specific technologies of transformation or for specific needs of human nutrition (Rando et al., 2000). For example, milk obtained from goats homozygote for the α s2-casein null allele would be more suitable for the production of “humanized” milk, since human milk is characterized by the absence of α s2-casein (Martin et al., 1996). In addition, the allergenicity of milks may vary according to the proportion of constitutive proteins (Bevilacqua et al., 2001). On the other hand, the high quantitative polymorphism observed in goat casein locus constitutes a valuable tool to investigate casein micelle organization (Martin et al., 1999; Chanut et al., 1999)

Although 16 different quantitative haplotypes resulting from the different levels of expression of the calcium-sensitive genes are theoretically possible, only six were observed (Rando et al., 2000; Figure 10), as a result of the very low frequency of some alleles and the linkage disequilibrium. Since κ -casein is linked to other calcium-sensitive genes, it is necessary to take into account κ -casein variants when considering haplotype studies in order to include milk protein variants in breeding programmes. In fact, its characterized polymorphism may result in further haplotypes at the casein cluster, which have to be determined and evaluated for their physiological role.

Figure 10. Quantitative haplotypes observed at the goat calcium-sensitive *loci*. (from Rando et al., 2000).



In this context, several studies have demonstrated the effects of particular genetic variants of milk proteins on the composition and physico-chemical properties of milk (revised in Grosclaude, 1988; Ng Kwai Hang, 1998). As regards κ -casein, it has been consistently shown that the bovine B variant is associated with a higher proportions of fat, higher total casein content, and a higher ratio of caseins to whey proteins (Ng Kwai Hang, 1998; Di Stasio and Mariani, 2000). The BB genotype has also been associated with superior milk manufacturing properties (heat stability, coagulation time, curd firmness) resulting in higher cheese yield (Schaar, 1984; Marziali and Ng-Kwai Hang, 1986). The bovine κ -casein genotype is already listed in some artificial insemination catalogues and selection for B variant is practiced in several countries (Ng Kwai Hang, 1998).

On the other hand, significant differences have been found in the content of variants A and B of κ -casein in milks of heterozygous animals (Van Ennenman and Medrano, 1991). This differential expression has been suggested to be related to allele-specific mutations within potential regulatory sites in the 5' flanking region of the gene (Shild et al., 1994). However, no such polymorphisms have been found in the proximal promoter region of A, B, and E alleles (Debeljak et al., 2000). In addition, electromobility shift assays (EMSA) and DNaseI footprinting analyses showed no differences in DNA-protein binding for the different allele promoters.

In caprine specie, the κ -casein polymorphism detected by Diaz-Carillo (1993) in milk samples from Murciano-Granadina goat breed by protein electrophoresis appeared to be associated with significant effects on milk yield and total casein content (Angulo et al., 1994). In particular, significant associations were found with α s1-casein content, whereas no significant effect was shown as regard κ -casein. However, only a small number of animals were included in the study, and the genotypes were not uniformly distributed (three animals homozygous for thr polymorphism). This polymorphism has not been characterized, and it most probably correspond to the B (or A) variant described here, since only these two variants have been found in this breed (Table 6).

The influence of the κ -casein variants characterized in this work on the quality and technological properties of milk still remain to be evaluated in further studies. The primer extension method reported here should be useful for rapid genotyping of large number of animals.

3. Beta-lactoglobulin polymorphism

Initiation of transcription requires the specific interaction of multiple transcriptional nuclear factors with *cis*-acting DNA sequences in gene promoters and enhancers. Most of the regulatory sequences are located in the proximal regions of milk protein gene promoters, within 200-250 bp adjacent to the initiation of the transcription site. Since mutations in 5'-flanking regions of genes may be located in transcription factor-binding sites, they can affect the efficiency of the transcription of a relevant structural gene, and therefore influence its expression and the physiological traits of animals. The *b-LG* gene promoter has been well characterized in sheep (Watson et al., 1991), and the recognition motifs for several transcription factors are conserved in bovine (Graml et al., 1989; Lum et al., 1997) and caprine (Folch et al., 1994) genes. Study of the ovine promoter region functionality (by 5' shortened constructs) in transgenic mice indicated that the first 410 bp upstream from the transcription unit was sufficient for high and tissue-specific expression, in particular, the region -406 to -149 bp appears to be essential, and contains the basic hormone response elements (Whitelaw et al., 1992).

Here, we have studied polymorphism within the proximal promoter region of the caprine *β-LG* gene. The analyzed fragment contains 588 bp of the proximal promoter and part of exon 1 (122 bp). Initial screening with BESS method revealed the presence of two potential mutations from which only one was confirmed by sequence analysis (Figure 10). The second detected mutation was considered to be an artifact since it was located close to fluorescent primer where signal and background intensities are very high. The polymorphic site consists of single nucleotide substitution (C to T) at position -60 upstream to transcription initiation site. This SNP affected a *SmaI* target site which is present in the previously reported *β-LG* sequence (Folch et al., 1994) but not in the new allele, allowing rapid genotyping by PCR-RFLP using this endonuclease.

Beta-lactoglobulin was the first milk protein in which polymorphism was evidenced by electrophoresis of bovine milk samples (Aschaffenburg and Drewry, 1955). In this specie, a dozen of genetic variants have been detected (Table 4), but only the most frequent alleles, A and B, were shown to be associated with differences in milk yield and composition. Milk from animals with genotype AA presents a higher content of *β-LG* protein than milk from animals BB, with an average content of 2.58 mg/ml for allele A compared to 1.7 mg/ml for allele B (Ng-Kwai Hang and Grosclaude, 1992). Due to a such

considerable difference in β -LG content associated with the two types, the structural *b-LG* gene (A compared to B) is generally considered as a gene with major effect for the β -LG amount in the milk (Grosclaude, 1988). In fact, the different level of the protein synthesis is also evident in milk of heterozygous AB animals where there is on average 1.5 times more β -LG A than β -LG B (Graml et al., 1989; Lum et al., 1997). Measurements of mRNA levels by allele-specific hybridization indicated a relative proportion of approximately 60 % of β -LG A and 40 % of β -LG B (Ford et al., 1993).

Several studies have been focused on the origin of this different transcriptional activity of β -LG A and B variants (revised in Martin et al., 2002). Nine polymorphic sites associated with allelic variants have been found in the 733 bp of the proximal promoter region (Lum et al., 1997), from which four are located in the recognition motifs of different transcription factors (MPBF, milk protein binding factor; NF-1, nuclear factor 1; and AP-2, activator protein 2). Transient transfection experiments of HC11 cells of β -LG A and B promoters have demonstrated that these nine allele-specific mutations are responsible for the differential expression of A and B alleles (Folch et al., 1999).

In addition to this differential expression, quantitative effects of these common variants on milk composition and cheese-making properties have been reported (revised in Ng-Kwai Hang, 1998; Di Stasio and Mariani, 2000). Allele B of β -LG is associated with high casein and fat content, whereas AA milk was shown to contain more whey and total proteins than those of the other genotypes. The reduced synthesis of β -LG B allele is compensated for by an increased synthesis of other major milk proteins, e.g. α -lactalbumin and caseins (Hill, 1993). As a consequence, milk from animals with BB genotype is characterized by a higher casein number (ratio of caseins to whey proteins) resulting in higher cheese yield.

The β -LG protein was considered to be monomorphic in the goat specie, although protein variants with faster electrophoretic mobility have been reported by Russian authors (Stupniskii and Il'Chenko, 1967; Macha, 1970). As far as coding region is concerned, no genetic variants are referred in *b-LG* gene. Two polymorphic sites (a single nucleotide substitution and a 10-bp insertion) have been identified and characterized in Spanish and French breeds, and localized in the 3'-non coding region (exon 7) of the gene (Pena et al., 2000). These polymorphisms are also observed in Italian breeds (Pappalardo et al., 2001). The polymorphism reported in this work is not located in binding sites of any transcription

factor, and probably would not affect the efficiency of the transcription. Allele frequencies for the analyzed breeds (Murciano-Granadina, Payoya, Malagueña, Canaria, and Saanen) were similar with the exception of Canaria breed, in which allele –60T was not found. This difference may be explained by the different origin of this breed (Sanchez-Belda, 1986). This may also explain the prevalence of κ -casein variant A in this breed in contrast to all other breeds analyzed where allele B is predominant. The four populations were in Hardy-Weinberg equilibrium at β -LG locus for this position.

The goat β -LG gene has been cytogenetically mapped in the chromosome 11q28 (Hayes et al., 1993), allowing the use of this novel polymorphism as a marker for this position in genome mapping.

4. Stearoyl Coenzyme A Desaturase

Stearoyl coenzyme A desaturase (SCD) catalyses the Δ^9 -cis desaturation of palmitoyl- and stearoyl-CoA to produce palmitoleoyl- and oleoyl-CoA, respectively (Enoch et al., 1976). Palmitoleic acid and oleic acid are the major constituents of the membrane phospholipids, and in ruminants, oleic acid is the main unsaturated fatty acid found in adipose tissue and milk triglycerides.

At the beginning of this thesis, no information was available regarding goat SCD mRNA sequence and the structural organization of the gene. Thus, as a first step toward studying the polymorphism of the *SCD* gene that may affect gene expression or enzymatic activity of SCD, we characterized the caprine cDNA and genomic organization of the gene (Yahyaoui et al., 2002). The first strand of cDNA was synthesized from total RNA isolated from a mammary gland sample from a lactating Murciano-Granadina goat. Then, two overlapping fragments of 1.1 and 1.6 Kb were generated by PCR using primers designated over the 5'-untranslated region and exon 4 (forward primers), and exon 5 and the 3'-untranslated region (reverse primers) of the ovine SCD cDNA (Ward et al., 1998). These fragments were then cloned and sequenced along both directions.

The obtained sequence encompasses the entire protein coding region (1077 bp encoding for 359 amino acids) and partial sequences of the 5'- and 3'-untranslated regions. It differs from the sequence reported by Bernard et al. (2001) at a number of locations: three nucleotides in the coding region (exons 2, 3, and 6) resulting in one variant amino acid; three positions in the 3'-untranslated region and other three in the 5'-untranslated

region. The SCD mRNA size is approximately 5 Kb in ruminant species, as estimated by northern-blot analysis (Bernard et al., 2001).

A GenBank database search using Blast software (Altschul et al., 1990) revealed that the obtained SCD cDNA sequence was highly homologous with those of sheep and cattle, whereas the similarity is lower with those of pig, man, and mouse SCDs (SCD1, 2, and 3). Within the coding region, the percentage of nucleotide similarity between the caprine SCD sequence and those of sheep and cow, are of 98 and 95%, respectively, whereas it is lower with pig (89%), human (87%), and mouse SCD-1, 2 and 3 (82%).

The partial 3'-UTR sequence is contained in exon 6 and is highly homologous to that of sheep SCD (\approx 98%) but is more divergent with the other species compared. The structural AU motifs (e.g. AUUUA) characteristic of mRNA destabilization sequences reported in mouse (Sessler et al., 1996), rat (Mihara, 1990), and human (Zhang et al., 1999) SCD mRNA are also found in the caprine 3'-UTR (Bernard et al., 2001). These AU elements have been shown to play active roles in the selective degradation of several mRNAs in response to various factors (Vakalopoulou et al., 1991). Studies on the *SCD* genes regulation in mouse adipocytes and yeast indicated that the effect of the polyunsaturated fatty acids on these genes could be at the level of mRNA stability (Sessler et al., 1996; Gonzalez et al., 1996). It has been proposed that these motifs could be implicated in polyunsaturated fatty acids action on SCD1 and SCD2 mRNA in mouse adipocytes (Ntambi et al., 1999). However, the exact pathways and mechanisms of this regulation are not yet known, and no binding or mediator proteins were characterized.

Comparison of the protein sequences between species shows that the amino acid substitutions are concentrated in the amino terminal one-third of the protein. The amino acid identity of the caprine SCD compared to those of ovine, bovine, porcine, and human proteins, is 98, 94, 87, and 87%, respectively. Such a high level of sequence identity of the open reading frame is a feature observed in the *SCD* gene family (Zhang et al., 2001). The eight histidine residues functionally essential on the catalytic properties of the enzyme and found in the desaturases and hydroxylases of several species, including mammals (Shanklin et al., 1994), are conserved in the caprine sequence. These histidine residues (which are spatially conserved) act as iron-centre for the electron transfer and the insertion of the double bond into fatty acids (Shanklin et al., 1994).

The structure of the caprine *SCD* gene is very similar to that reported in human (Zhang et al., 1999) and rodents (Ntambi et al., 1988; Kaestener et al., 1989; Mihara et al., 1990). It spans approximately 15 Kb and consists of six exons and five introns. All exon-intron boundaries follow the general the splice-donor and acceptor-consensus sequences (GT/AG).

As above mentioned, three nucleotide differences in the coding region were observed when comparing our cDNA sequence (GenBank accession n°: AF339909) to that reported by Bernard and co-workers (AF325499). Two differences are synonymous mutations in exons 2 and 3 corresponding to amino acid residues threonine and methionine at positions 15 and 109 of the protein, respectively. The third nucleotide difference (exon 6) results in one amino acid variant: alanine to valine at position 344 in the C-terminal region of the protein. In order to examine the occurrence of this polymorphism, a total of 40 animals (from Murciano-Granadina, Canaria, and Saanen) were genotyped by primer extension analysis. However, the alanine variant was not found in the analyzed breeds. It is possible that this variant is present at higher frequencies in other breeds, as observed for some κ -casein variants.

On the other hand, we have also screened exon 5 for polymorphisms by BESS method. A total of fifty genomic DNA samples from animals of different Spanish (Canaria, Malagueña, Payoya, and Murciano-Granadina) and French (Saanen) breeds were analyzed but no additional mutations were detected. The exon five was chosen because mutations in this exon have been reported in bovine specie, which results in two variants differing by an amino acid substitution (alanine to valine) designated A and B (Medrano et al., 1998). These variants are found in Holstein, Jersey, and Brown Swiss breeds.

In addition to these polymorphisms, other three single nucleotide substitutions were observed by comparison of the nucleotide sequence data from the cDNA and genomic DNA (amplified by PCR for the determination of gene structure). Two SNPs affect exon 5 at positions 904 (C to T) and 931 (G to T) of the cDNA. The latter mutation results in the disappearance of the *RsaI* recognition site, allowing rapid genotyping by PCR-RFLP. These nucleotide substitutions are located at the third base of the codon (asparagine-244 and arginine-253) and therefore do not produce any amino acidic change. The third SNP is situated in the 3'-untranslated region. In this region, a trinucleotide (TGT) deletion polymorphism in the 3'-UTR sequence has been also reported in Alpine goat by Bernard et

al., (2001). Thus, a total of five polymorphisms were identified in the 3'-UTR indicating a high level of variation, which could affect the stability of the messenger.

The SNP in exon 5 at position 931 (G to T) was used as RFLP marker for chromosome localization of the *SCD* gene. The polymorphism was not found in Canaria, Payoya, and Malagueña breeds, however, the number of animals analyzed from these two ultimate breeds is limited (Yahyaoui et al., submitted manuscript). The frequency of G allele is 0.08 and 0.31 in the Murciano-Granadina and the French Saanen breeds, respectively.

Mapping of the goat *SCD* gene have been realized by D. Vaiman in the INRA (Jouy-en-Josas, France). Chromosomal location was obtained by genetic analysis of the genotypes of goat reference half-sib families (Vaiman et al., 1996). Amongst the twelve progenitors three appeared polymorphic and made it possible to obtain 55 informative meioses. Two-point analysis was carried out using the CRI-MAP software and resulted in five significant lod scores with five markers of CHI26, INRA081 ($\theta = 0.09$; Lod = 4.36), INRA172 ($\theta = 0.13$; Lod = 3.34), BM4505 ($\theta = 0.11$; Lod = 3.66), LSCV046 ($\theta = 0.13$; Lod = 4.18) and LSCV052 ($\theta = 0.03$; Lod = 5.46). Haplotype analysis was carried out by visual analysis of the recombinant chromosomes, suggesting the most probable order: Centromere – HEL11 – Bm1314 – INRA081 – INRA172 – LSCV052 – *SCD* – BM4505 – LSCV046 – Telomere. Compared with the use of the build option of CRIMAP which was used in the map construction of CHI26 (Schibler et al., 1998), haplotype reconstitution of the chromosome results in an order much more similar to the one described in the sheep map of OAR22 (the homologue of CHI26). The cytogenetic mapping of LSCV046 to 26q13 and of LSCV052 to 26q21, together with the very close association of LSCV052 (originating from a BAC containing the gene *PAX2*) indicate that the two genes are close. These results are in accordance with those previously reported using FISH technique, assigning *SCD* gene to CHI 26q21 (Bernard et al., 2001). The human chromosome region is HSA10, homologous to segments of goat chromosomes 13, 28 and 26. In humans, *SCD* has been mapped to HSA 10q24.32, whereas *PAX2* is mapped to HSA 10q25.1. The interspecific homologous mapping of the two genes suggests a good conservation in this mammalian chromosome segment.

The *SCD* gene has been mapped to bovine chromosome 26q21 (Campbell et al., 2001) and the presence of QTLs affecting milk fat percentage has been reported especially

on this chromosome (Zhang et al., 1998). In mouse, the cluster of *SCD* genes has been assigned to mouse chromosome 19 and it is overlapped by an identified QTL which is associated with elevated plasma levels of free fatty acids and triglycerides (Tabor et al., 1998).

As aforementioned, the SCD activity results from the transcription of three related genes in rodents (mouse and rat). On the contrary, the SCD mRNA is transcribed from a single gene in sheep (Ward et al., 1998) and in human (Zhang et al., 1999), whereas no information is available concerning goat specie.

5. Concluding remarks and future prospects

The advances in biotechnology techniques, primarily being developed for the human health industry, are increasingly used in the field of animal production. In particular, the use of the tools of molecular biology to identify and manage natural variation in milk composition contributes to accelerate the breeding process, allowing better control over the production and composition of milk. For example, milk protein variants in goat (α s1-casein) and in cattle (β -LG and κ -casein) are already being included in selection schemes (Manfredi et al., 1995; Ng-Kwai Hang, 1998). Another example is the recent identification of a specific genetic polymorphism of a gene involved in milk fat synthesis (DGAT; diacylglycerol acyl-transferase) that appears to modify milk composition (Grisart et al., 2002).

The biological significance (e.g. milk composition) of the κ -casein variants described in this thesis still remain to be evaluated in further studies. Due to the tight linkage of casein *loci*, it could be more appropriate to study the effect of the casein *haplotypes* rather than single *loci* taken separately. For this purpose, the method reported here should be useful for genotyping casein variants in which single nucleotide polymorphisms are involved.

Additional investigations should also be done in order to better understand the mechanisms controlling the expression of the *k-casein* gene. In this context, several studies have attempted to use κ -casein regulatory sequences to drive transgene expression in mice. However, both ruminant (bovine and caprine) and rabbit genomic clones were not or were poorly expressed under their own promoter in transgenic mice (Ninomiya et al., 1994; Rijnkels et al., 1995; Persuy et al., 1995; Baranyi et al., 1995). In contrast, the *k-casein*

gene is correctly expressed under the control of the β -casein regulatory sequences in transgenic mice (Persuy et al., 1995; Gutierrez et al., 1996) and transgenic cattle (Brophy et al., 2003). From the fact that repetitive SINE and LINE elements have been identified in the 5'-flanking region of the ruminants and human *k-casein* genes (Coll et al., 1995; Gerencser et al., 2002), it can be speculated that the presence of these elements could alter the correct expression of the gene. Another possibility is that common regulatory elements, most likely in the 5' proximal region of the cluster, might be involved in the expression of the entire casein gene locus. Such elements, known as locus control region (LCR), have been well described for human β -globin gene cluster (Grosveld et al., 1987; revised in Li et al., 2002).

Overall, the caprine specie has been subjected to relatively reduced levels of artificial selection comparatively to the bovine and ovine species. This may be related to the high level of genetic variability observed in caprine casein genes, with several alleles characterized by strong differences in their level of expression. In this regard, some breeds or populations may possess unique alleles or combinations of alleles that are absent in other selected breeds, and may be useful as reservoirs of genetic diversity for commercial domestic species. This emphasizes the importance of the preservation of such high genetic variability in these populations through programs of genetic resource conservation.

In addition to proteins, lipid fraction constitutes the second important component in dairy products. Ruminant animals have relatively high ratios of saturated: monounsaturated fatty acids in their lipids, which are consumed by humans as meat and milk products. An important selection objective is to increase the proportion of unsaturated fatty acids to make healthier products. The *SCD* gene constitutes a prime candidate gene in this purpose. The significance of the polymorphisms in the goat *SCD* gene reported here in relation to the fatty acid composition of milk (and to messenger stability for those located in the 3' untranslated region) remain to be evaluated. Strategies to manipulate the expression of the stearoyl-CoA desaturase gene could result in alteration of the fatty acid composition of tissues and milk of ruminants. In this regard, further investigations of the expression patterns and promoter analysis of the ruminant *SCD* genes will be necessary.

CONCLUSIONS

CONCLUSIONS

- 1) Although BESS T-scan proved to be a relatively simple and rapid protocol that resulted in easy data analysis when coupled to fluorescent detection, it requires further optimizations for each individual PCR product to be screened. The method appears to be suitable only for initial scanning of short fragments (200-300 bp) with more effectiveness when the AT-content of the target is relatively low. The main advantage of the method is that allows approximate localization of the detected mutations.
- 2) A total of ten mutations were detected in the kappa casein coding region by using BESS method and sequencing. Four are synonymous mutations (3 in exon 4 and one in exon 3) whereas other six produce amino acid changes. All these non-synonymous mutations, located in exon 4, are single nucleotide transitions. The association between the different mutations (*haplotypes*) resulted in seven genetic variants, designated κ -casein A, B, C, D, E, F, and G.
- 3) A procedure for rapid and simultaneous genotyping for all κ -casein variants was developed. The method is based on primer extension analysis coupled with capillary electrophoresis and fluorescent detection.
- 4) Kappa casein A and B are the most common variants found in several Spanish, French, and Italian breeds. Variant B is predominant in all these breeds, with the exception of the Canaria breed, where variant A is prevalent. The F variant is predominant in the Spanish wild type goat *Capra. pyrenaica sp. hispanica*.
- 5) Comparative sequence data suggest that the F variant is most likely the original type of kappa casein in caprine specie. Other variants were derived from this allele by successive mutations following two different trunks.
- 6) A single nucleotide polymorphism was detected by BESS method in the proximal promoter region of the *b-LG* gene. However, this mutation is not located in the sequence recognition of

any known transcription factor. The polymorphism is found in Murciano-Granadina, Malagueña, Payoya, and Saanen, but it is absent in the Canaria breed.

7) The goat *SCD* cDNA and gene were isolated and characterized. The open reading frame contains 1077 nucleotides encoding for a mature protein of 359 amino acids, with a high level of similarity with ovine (98%) and bovine (95%) homologues. The structural organization of the caprine *SCD* gene is similar to those of rodents and human. It spans approximately 15 Kb and consists of six exons and five introns.

8) Several single nucleotide polymorphisms were detected in the coding region (exons 5 and 6) and in the 3'-untranslated region. A PCR-RFLP protocol for genotyping the SNP at position 931 of the cDNA was developed. The polymorphism was used to map the caprine *SCD* gene by linkage analysis to CHI 26q13-26q21

REFERENCES

REFERENCES

- Addeo, F., Mauriello, R., Moio, L., Laezza, P., Chianese, L. and Di Luccia, A. 1992. Ovine casein variant identification using electrophoretic, immunochemical and chromatographic techniques. *Milchwissenschaft* **47**: 283-287.
- Addeo, F., Soulier, S., Pelissier, J.P., Chobert, J.M., Mercier, J.C. and Ribadeau-Dumas, B. 1978. Preparation and fractionation of goat κ -casein: analysis of the glycan and peptide components. *Journal of Dairy Research* **45**: 191-196.
- Ahmadian, A. and Lundeberg, J. 2002. A brief history of genetic variation analysis. *BioTechniques* **32**: 1122-1137.
- Akey, J.M., Zhang, K., Xiong, M., Doris, P. and Jin, L. 2001. The effect that genotyping errors have on the robustness of common linkage-disequilibrium measures. *American Journal of Human Genetics* **68**: 1447-1456.
- Alais, C. and Jolles, P. 1961. Comparative study of CMP obtained after rennin digestion of the caseins of the milk of cow, sheep, and goat. II. study of non-peptidic part. *Biochimica et Biophysica Acta* **51**: 309-314.
- Alexander L.J., Stewart, A.F., Mackinlay, A.G., Kapelinskaya, T.V., Tkach, T.M. and Gorodetsky, S.I. 1988. Isolation and characterization of the bovine kappa casein gene. *European Journal of Biochemistry* **178**: 395-401.
- Alexander, L.J., Hayes, G., Pearse, M.J., Beattie, C.W., Stewart, A.F., Willis, L.M. and Mackinlay, A.G. 1989. Complete sequence of the bovine β -lactoglobulin cDNA. *Nucleic Acids Research* **17**: 16739.
- Ali, S. and Clark, A.J. 1988. Characterization of the gene encoding ovine beta-lactoglobulin. *Journal of Molecular Biology* **199**: 415-426.
- Alonso, L., Fontecha, J., Lozada, L., Fraga, M.J. and Juarez, M. 1999. Fatty acid composition of caprine milk: Major, branched-chain, and trans fatty acids. *Journal of Dairy Science* **82**: 878-884.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.
- Alves, A.M. and Carr, E.J. 1988. Dot blot detection of point mutations with adjacently hybridizing synthetic oligonucleotides probes. *Nucleic Acids Research* **16**: 8723.
- Angiolillo, A., Yahyaoui, M.H., Sanchez, A., Pilla, F. and Folch, J.M. 2002. Characterization of a new genetic variant in the caprine κ -casein gene. *Journal of Dairy Science* **85**:2679-2680.
- Angulo, C., Diaz-Carillo, E., Muñoz, A., Alonso, A., Jiménez, I. And Serradilla, J.M. 1994. Effect of electrophoretic goat's kappa casein polymorphism on milk yield and main components yield. *V World Congress on Genetics Applied to Livestock Production Guelfh* **19**: 333-336.
- Aschaffenburg, R. 1961. Inherited casein variants in cow's milk. *Nature* **192**:431-432.
- Aschaffenburg, R. and Drewry J. 1955. Occurrence of different β -lactoglobulins in cow's milk. *Nature* **176**: 218-219.
- Aschaffenburg, R. and Drewry J. 1957. Genetics of the β -lactoglobulins of cow's milk. *Nature* **180**: 376-378.
- Barany, F. 1991. Genetic disease detection and DNA amplification used cloned thermostable ligase. *Proceedings of the National Academy of Sciences of USA* **88**: 189-193.

- Baranyi, M., Aszodi, A., Devinoy, E., Fontaine, M.L., Houdebine, L.M. and Bosze, Z. 1996. Structure of the rabbit κ -casein encoding gene: expression of the cloned gene in the mammary gland of transgenic mice. *Gene* **174**: 27-34.
- Barber, M.C., Clegg, R.A., Travers, M.T. and Vernon, R.G. 1997. Lipid metabolism in the lactating mammary gland. *Biochimica et Biophysica Acta* **1347**: 101-126.
- Barber, M.C., Ward, R.J., Richards, S.E., Salter, A.M., Buttery, P.J., Vernon, R.G. and Travers, M.T. 2000. Ovine adipose tissue monounsaturated fat content is correlated to depot-specific expression of the stearoyl-CoA desaturase gene. *Journal of Animal Science* **78**: 62-68.
- Baron, H., Fung, S., Aydin, A., Bahring, S., Luft, FC. And Schuster, H. 1996. Oligonucleotide ligation assay (OLA) for the diagnosis of familial hypercholesterolemia. *Nature Biotechnology* **14**: 1279-1282.
- Barroso, A., Dunner, S. and Cañon J. 1999. Polimorfismo genético de las lactoproteínas de los rumiantes domésticos: revisión. *ITEA* **95A**: 143-179.
- Barroso, A., Dunner, S. and Cañon, J. 1998. Technical note: detection of bovine kappa casein variants A, B, C, and E by means of PCR-SSCP. *Journal of Animal Science* **76**: 1535-1538.
- Bates, G. and Lehrach, H. 1994. Microsatellite instability in human solid tumors. *Cancer* **82**: 1808-1820.
- Bell, K. 1962. One-dimensional starch gel electrophoresis of bovine skim milk. *Nature* **195**: 705-706.
- Bernard, L., Leroux, C., Hayes, H., Gauthier, M., Chilliard, Y. and Martin, P. 2001. Characterization of the caprine stearoyl-CoA desaturase gene and its mRNA showing an unusually long 3'-UTR sequence arising from a single exon. *Gene* **281**: 53-61.
- Bernardi, G., Olofsson, B., Filipski, G., Zerial, M., Salinas, J., Cuny, G., Meunier Rotival, M. and Rodier, F. 1985. The mosaic genome of warm-blooded vertebrates. *Science* **228**: 953-958.
- Bevilacqua, C., Martin, P., Candalh, C., Fauquant, J., Piot, M., Roucayrol, A.M., Pilla, F. and Heyman, M. 2001. Goats' milk of defective alpha(s1)-casein genotype decreases intestinal and systemic sensitization to beta-lactoglobulin in guinea pigs. *Journal of Dairy Research* **68**: 217-227.
- Bingham, E.W. 1979. Role of mammary casein kinase in the phosphorylation of milk proteins. *Journal of Dairy Research* **46**: 181-185.
- Blumberg, B.S. and Tombs, M.P. 1958. Possible polymorphism of bovine α -lactalbumin. *Nature* **181**: 683-684.
- Bonnet, G., Tyagi, S., Libchaber, A. and Kramer, F.R. 1999. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proceedings of the National Academy of Sciences of USA* **96**: 6171-6176.
- Bonsing, J. and Mackinlay, A.G. 1987. Recent studies on nucleotide sequences encoding the caseins. *Journal of Dairy Research* **54**: 447-461.
- Botstein, D., White, R.L., Skolnick, M., and Davis, R.M. 1980. Construction of a genetic linkage map in man using restriction length polymorphism. *American Journal of Human Genetics* **32**: 314-331.
- Bray, M.S., Boerwinkle, E. and Doris, P.A. 2001. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry. *Human Mutation* **17**: 296-304.

- Brew, K., Castellino, F.G., Vanaman, T.C. and Hill, R.L. 1970. The complete amino acid sequence of bovine α -lactalbumin. *Journal of Biological Chemistry* **245**: 4570-4582.
- Brieger, A., Trojan, J., Raedle, J., Roth, K. and Zeuzem, S. 1999. Identification of germline mutations in hereditary nonpolyposis colorectal cancer using base excision sequence scanning analysis. *Journal of Clinical Chemistry* **45**: 1564-1567.
- Brody, E.P. 2000. Biological activities of bovine glycomacropeptide. *British Journal of Nutrition* **84**: Supplement 1: S39-S46.
- Brookes, A.J. 1999. The essence of SNPs. *Gene* **234**: 177-186.
- Brophy, B., Smolenski, G., Wheeler, T., Wells, D., L'Huillier, P. and Laible, G. 2003. Cloned transgenic cattle produce milk with higher levels of β -casein and κ -casein. *Nature Biotechnology* **21**: 157-162.
- Campbell, E.M.G., Gallagher, D.S, Davis, K.S., Taylor, J.F. and Smith, S.B. 2001. Rapid communication: mapping of the bovine stearoyl coenzyme A desaturase (*SCD*) gene to BTA26. *Journal of Animal Science* **79**: 1954-1955.
- Caroli, A., Jann, O., Budelli, E., Bolla, P., Jager, S. and Erhardt, G. 2001. Genetic polymorphism of goat κ -casein (*CSN3*) in different breeds and characterization at DNA level. *Animal Genetics* **32**: 226-230.
- Chadwick, R.B., Conrad, M.P., McGinnis, M.D., Johnston-Dow, L., Spurgeon, S.L. and Kronick, M. 1996. Heterozygote and mutation detection by direct automated fluorescent DNA sequencing using mutant *Taq* DNA polymerase. *Biotechniques* **20**: 676-683.
- Chanat, E., Martin, P. and Ollivier-Bousquet, M. 1999. Alpha(S1)-casein is required for the efficient transport of beta- and kappa-casein from the endoplasmic reticulum to the Golgi apparatus of mammary epithelial cells. *Journal of Cell Science* **112**: 3399-3412.
- Charrel, R.N., Levy, N., Tesh, R.B. and Chandler, L.J. 1999. Use of base excision sequence scanning for detection of genetic variation in St Louis encephalitis virus isolates. *Journal of Clinical Microbiology* **37**: 1935-1940.
- Cheng, J. Kasuga, T., Mitchelson, K.R., Lightly, E.R.T., Watson, N.D., Martin, W.J. and Atkinson, D. 1994. Polymerase chain reaction heteroduplex polymorphism analysis by entangled solution capillary electrophoresis. *Journal of Chromatography* **677**: 169-177.
- Chilliard, Y., Ferlay, A., Mansbridge, R.M and Doreau, M. 2000. Ruminant milk fat plasticity: nutritional control of saturated, polyunsaturated, *trans* and conjugated fatty acids. *Annales de Zootechnie* **49**: 181-205.
- Chiofalo, L. and Micari, P. 1987. Attuali conoscenze sulle varianti delle proteine del latte nelle popolazione ovine allevate in sicilia. Osservazioni sperimentali. *Scienza e Tecnica Lattiero-Casearia* **38**: 104-114.
- Cho, R.J, Mindrinos, M., Richards, D.R, Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T.L., Stammers, M., Federspiel, N., Theologis, A., Yang, W.H., Hubbell, E., Au, M., Chung, E.Y., Lashkari, D., Lemieux, B., Dean, C., Lipshutz, R.J., Ausubel, F.M., Davis, R.W., and Oefner, P.J. 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genetics* **23**: 203-207.
- Christie, W.W. 1983. The composition and structure of milk lipids. In: *Developments in dairy chemistry*. Vol. II P.F. Fox Ed. Elsevier Applied Sciences, London.

- Chung, M., Ha, S., Jeong, S., Bok, J., Cho, K., Baik, M. and Choi, Y. 2000. Cloning and characterization of bovine stearoyl-CoA desaturase 1 cDNA from adipose tissues. *Biosciences Biotechnology Biochemistry* **64**: 1526-1530.
- Clare, D.A. and Swaisgood, H.E. 2000. Bioactive milk peptides: a prospectus. *Journal of Dairy Science* **83**: 1187-1195.
- Clegg, R.A., Barber, M.C., Pooley, L., Ernens, I., Larondelle, Y. and Travers, M.T. 2001. Milk fat synthesis and secretion: molecular and cellular aspects. *Livestock Production Science* **70**: 3-14.
- Cohen, J.B. and Levinson, A.D. 1988. A point mutation in the last intron responsible for increased expression and transforming activity of the c-Has-ras oncogene. *Nature* **334**: 119-124.
- Cohen, P., Miyazaki, M., Socci, N.D., Hagge-Greenberg, A., Liedtke, W., Soukas, A.A., Sharma, R., Hudgins, L.C., Ntambi, J.M. and Friedman, J.M. 2002. Role for stearoyl-CoA desaturase in leptin-mediated weight loss. *Science* **297**: 240-243.
- Coll, A. 1995b. Caracterizació del gen de la kappa caseina de cabra per la seva utilització com transgen. *PhD Tesis*. Faculty of Veterinary, Universidad Autonoma de Barcelona, Spain.
- Coll, A., Folch, J. M., and Sanchez, A. 1993. Nucleotide sequence of the goat kappa casein cDNA. *Journal of Animal Science* **71**: 2833.
- Coll, A., Folch, J. M., and Sanchez, A. 1995a. Structural features of the 5' flanking region of the caprine kappa casein gene. *Journal of Dairy Science* **78**: 973-977.
- Cook, H.W. and McMaster, C.R. 2002. Fatty acid desaturation and chain elongation in eukaryotes. In: *Biochemistry of lipids, lipoproteins and membranes*. D.E. Vance and J.E. Vance Ed. (4th Ed.). Elsevier Science B.V. pp 181-204.
- Cotton, R.G. 1993. Current methods of mutation detection. *Mutation Research* **285**: 125-144.
- Cotton, R.G., Rodrigues, N.R. and Campbell, R.D. 1988. Reactivity of cytosine in single base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proceedings of the National Academy of Sciences of USA* **85**: 4397-4401.
- Dagleish, D.G. 1992. Bovine milk properties and the manufacturing quality of milk. *Livestock Production Science* **35**: 75-93.
- Dallerac, R., Labour, C., Jallon, J.M., Knipple, D.C., Reolofs, W.L. and Wicker-Thomas, C. 2000. A Δ^9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of USA* **97**: 9449-9454.
- Dawson, E., Chen, Y., Hunt, S., Slink, L.J., Hunt, A., Rice, K., Livingston, S., Bumpstead, S., Bruskiwich, R., Sham, P., Ganske, R., Adams, M., Kawasaki, K., Shimizu, N., Minoshima, S., Roe, B., Bentley, D. And Dunham, I. 2001. A SNP resource for human chromosome 22: extracting dense clusters of SNPs from the genomic sequence. *Genome Research* **11**: 170-178.
- De Vries, J.E., Vork, M.M., Roemen, T.H., deJong, Y.F., Cleutjens, J.P., Vander Vusse, G.J. and Bilsen, M. 1997. Saturated but not mono-unsaturated fatty acids induce apoptotic cell death in neonatal rat ventricular myocytes. *Journal of Lipid Research* **38**: 1384-1394.
- Debeljak, M., Susnik, S., Marinsek-Logar, R., Medrano, J.F. and Dovc, P. 2000. Allelic differences in bovine kappa casein gene which may regulate gene expression. *Pflugers Archives* **439**: R4-R6.

- Del Tito, B.J., Poff III, H.E., Novotny, M.A., Cartledge, D.M., Walker II, R.I., Earl, C.D. and Bailey, A.L. 1998. Automated fluorescent analysis procedure for enzymatic mutation detection. *Clinical Chemistry* **44**: 731-739.
- Demeyer, D. and Doreau, M. 1999. Targets and procedures for altering ruminant meat and milk lipids. *Proceedings of the Nutrition Society* **58**:593-607.
- Di Gregorio, P., Rando, A., Ramunno, L., Masina, P. and Pieragostini, E. 1991. DNA polymorphism at the casein loci in sheep. *Animal Genetics* **22**: 21-30.
- Di Gregorio, P., Rando, A., Ramunno, L., Masina, P. and Pieragostini, E. 1989. DNA polimorfismi nelle regioni del DNA che contengono i geni delle caseine di pecora e capra. *XXIV Simposio Internazionale Zootecnica* pp 275-282.
- Di Luccia, A., Mauriello, R., Chianese, L., Moio, L. and Addeo, F. 1990. Kappa casein polymorphism in caprine milk. *Scienza e Tecnica Lattiero-Casearia* **41**: 305-314.
- Di Stasio, L. and Mariani, P. 2000. The role of protein polymorphism in the genetic improvement of milk production. *Zootecnica Nutrizione Animali* **26**: 69-90.
- Diaz-Carillo, E. 1993. Análisis de contenidos y producción total de las fracciones caseínicas en la leche de tres razas de cabras españolas. Factores de variación y relaciones con otros componentes lacteos. *PhD Thesis*. University of Cordoba, Faculty of Sciences, Spain.
- Dils, R.R. 1986 Comparative aspects of milk fat synthesis. *Journal of Dairy Science* **69**: 904-910.
- Duncan, B.K. and Miller, J.H. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature* **287**: 560-561.
- Duriez, B., Sobrier, M.L., Duquesnoy, B., Tixier-Boichard, M., Decuypere, E., Coquerelle, G., Zeman, M., Goossens, M. and Amselem, S. 1993. A naturally occurring growth hormone receptor mutation: in vivo and in vitro evidence for the functional importance of the WS motif common to all members of the cytokine superfamily. *Molecular Endocrinology* **7**: 806-814.
- Dziuba, J. and Minkiewicz, P. 1996. Influence of glycosylation on micelle-stabilizing ability and biological properties of C-terminal fragments of cow's κ -casein. *International Dairy Journal* **6**: 1017-1044.
- Echols, H. and Goodman, M.F. 1991. Fidelity mechanisms in DNA replication. *Annual Reviews of Biochemistry* **60**: 477-511.
- Edlund, A., Johansson, T., Leidvik, B. and Hansson, L. 1996. Structure of the human kappa casein gene. *Gene* **174**: 65-69.
- Eigel, W.M., Butter, J.E., Ernstrom, C.A., Farrell, H.M., Harwalkar, V.R., Jenness, R. and Whitney, R.M. 1984. Nomenclature of proteins of cow's milk: fifth revision. *Journal of Dairy Science* **67**: 1599-1631.
- Enoch, M.G., Catala, A. and Strittmatter P. 1976. Mechanism of rat liver microsomal stearoyl coenzyme A desaturase: studies of the substrate specificity, enzyme-substrate interactions, and the function of lipid. *Journal of Biological Chemistry* **251**: 5095-5103.
- Erhardt, G. 1989. Evidence for a third allele at the beta lactoglobulin locus of sheep milk and its occurrence in different breeds. *Animal Genetics* **20**: 197-204.
- Faham, M. and Cox, D.R. 1995. A novel *in vivo* method to detect DNA sequence variation. *Genome Research* **5**: 474-482.
- Faham, M., Baharloo, S., Tomitaka, S., DeYoung, J. And Freimer, B. 2001. Mismatch repair detection (MRD): high-throughput scanning for DNA variations. *Human Molecular Genetics* **10**: 1657-1664.

- Fischer, S.G. and Lerman, L.S. 1983. DNA fragments differing by single base-pair substitution are separated in denaturing gradient gel: correspondence with melting theory. *Proceedings of the National Academy of Sciences of USA* **80**: 1579-1583.
- Flower, D.R. 1996. The lipocalin protein family: structure and function. *Biochemistry Journal* **318**: 1-14.
- Folch, J.M., Coll, A. and Sanchez, A. 1993 Rapid communication: cloning and sequencing of the cDNA encoding goat beta lactoglobulin. *Journal of Animal Science* **71**: 2832.
- Folch, J.M., Coll, A. and Sanchez, A. 1994. Complete sequence of the caprine beta lactoglobulin gene. *Journal of Dairy Science* **77** 3493-3497.
- Folch, J.M., Coll, A., Hayes, H.C. and Sanchez, A. 1996. Characterization of a caprine β -lactoglobulin pseudogene, identification and chromosomal localization by *in situ* hybridization in goat, sheep, and cow. *Gene* **177**: 87-91.
- Folch, J.M., Dovic, P. and Medrano, J.F. 1999. Differential expression of β -lactoglobulin A and B promoter variants in transiently transfected HC11 cells. *Journal of Dairy Research* **66**: 537-544.
- Fox, P.F. 1989. The milk protein system. In: *Developments in Dairy Chemistry*. Vol. 4. Functional Milk Proteins. (Ed. P.F: Fox). London, Elsevier Science Publishers, pp. 1-53.
- Fujiwara, Y., Miwa, M., Nogami, M., Okumura, K., Nobori, T., Suzuki, T. and Ueda, M. 1997. Genomic organization and chromosomal localization of the human casein gene family. *Human Genetics* **99**: 368-373.
- Furet, J.P., Mercier, J.C., Soulier, S., Gaye, P., Hue-Delahaie, D. and Vilotte, J.L. 1990. Nucleotide sequence of the ovine kappa casein cDNA. *Nucleic Acids Research* **17**: 5286.
- Garg, M.L., Snoswell, A.M. and Sabine, J.R. 1986. Influences of dietary cholesterol on desaturase enzymes of rat liver microsomes. *Progress in Lipid Research* **25**: 639-644.
- Gaye, P., Mue-Delahaie, D., Mercier, J.C., Soulier, S., Vilotte, J.L. and Furet, J.P. 1986. Ovine beta lactoglobulin messenger RNA: nucleotide sequence and mRNA levels during functional differentiation of the mammary gland. *Biochimie* **68**: 1097-1107.
- Gerencser, A., Barta, A., Boa, S., Kastanis, P., Bosze, Z. and Whitelow, C.B. 2002. Comparative analysis on the structural features of the 5' flanking regions of k-casein genes from six different species. *Genetics Selection Evolution* **34**: 117-128.
- Gibson, J.P. 1991. The potential for genetic change in milk fat composition. *Journal of Dairy Science* **74**: 3258-3266.
- Ginger, M.R. and Grigor, M.R. 1999. Comparative aspects of milk caseins. *Comparative Biochemistry and Physiology Part B* **124**: 133-145.
- Godovac-Zimmerman, J., Krause, I., Barany, M., Fisher-Fruhloz, S., Juszczak, J., Erhardt, G., Buchberger, J. and Klostermeyer, H. 1996. Isolation and rapid sequence characterization of two novel beta lactoglobulins I and J. *Protein Chemistry* **15**: 743-750.
- Godovac-Zimmermann, J., Conti, A., Sheil, M. and Napolitano, L. 1990. Covalent structure of the minor monomeric beta lactoglobulin II component from donkey milk. *Biological Chemistry Hoppe Seyler* **371**: 871-879.
- Gonzalez, C.I. and Martin, C.E. 1996. Fatty acid responsive control of mRNA stability: Unsaturated fatty acid-induced degradation of the saccharomyces OLE1 transcript. *Journal of Biological Chemistry* **271**: 25801-25809.

- Gonzalzo, M.L. and Jones, P.A. 1997. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Research* **18**: 3739-3744.
- Gorodetskii, S.I., and Kaledin, A.S. 1987. Nucleotide sequence analysis of cow kappa casein cDNA. *Genetika* **23**: 596-604.
- Graml, R.G., Weiss, G., Buchberger, J. and Pirchner, F. 1989. Different rates of synthesis of whey protein and casein by alleles of the β -lactoglobulin and α s1-casein locus in cattle. *Genetics Selection Evolution* **21**: 547-553.
- Greenwood, A.D. and Burke, D.T. 1996. Single nucleotide primer extension: quantitative range, variability, and multiplex analysis. *Genome Research* **6**: 336-348.
- Griffin, T.J., Hall, J.G., Prudent, J.R. and Smith, L.M. 1999. Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry. *Proceedings of the National Academy of Sciences of USA* **96**: 6301-6306.
- Grobet, L., Poncelet, D., Royo, L.J., Brouwers, B., Pirottin, D., Michaux, C., Ménéssier, F., Zanotti, M., Dunner, S. and Georges, M. 1998. Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle. *Mammalian Genome* **9**: 210-213.
- Grosclaude, F. 1988. Le polymorphisme génétique des principales lactoprotéines bovines. Relation avec la quantité, la composition, et les aptitudes fromagères du lait. *INRA Productions Animales* **1**: 5-17.
- Grosclaude, F., Mahé, M.F., Mercier, J.C. and Ribadeau-Dumas, B. 1972. Localisation des substitutions des acides aminés différenciant les variants A et B de la caséine kappa bovine. *Annales de Génétique et de Sélection Animale* **4**: 515-521.
- Grosclaude, F., Ricordeau, G., Martin, P., Remeuf, F., Vassal, L. and Bouillon, J. 1994. Du gène au fromage: le polymorphisme de la caséine α s1 caprine, ses effets, son évolution. *INRA Productions Animales* **7**: 3-19.
- Gross, E., Arnold, N., Goette, J., Schwartz-Boeger, U. and Kiechle, M. 1999. A comparison of BRCA1 mutation analysis by direct sequencing, SSCP and DHPLC. *Human Genetics* **105**: 72-78.
- Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. 1987. Position-independent, high level expression of the human beta-globin gene in transgenic mice. *Cell* **51**: 975-985.
- Grummer, R.R. 1991. Effect of feed on the composition of milk fat. *Journal of Dairy Science* **74**: 3244-3257.
- Gutierrez, A., Meade, H.M., Ditullio, P., Pollock, D., Harvey, M., Jimenez-Flores, M., Anderson, G.B., Murray, J.D. and Medrano, J.F. 1996. Expression of a bovine κ -casein DNA in the mammary gland of transgenic mice utilizing a genomic milk protein gene as an expression cassette. *Transgenic Research* **5**: 271-279.
- Gyorfy, Z., Benko, S., Kusz, E., Maresca, B., Vigh, L. and Duda, E. 1997. Highly increased TNF sensitivity of tumor cells expressing the yeast delta 9-desaturase gene. *Biochemical and Biophysical Research Communications* **241**: 465-470.
- Hall, J.G., Eis, P.S., Law, S.M., Reynaldo, L.P., Prudent, J.R., Marshall, D.J., Allawi, H.T., A.L., Mast, Dahlberg, J.E., Kwiatkowski, R.W., de Arruda, M., Neri, B.P. and Lyamichev, V.I. 2000. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proceedings of the National Academy of Sciences of USA* **97**: 8272-8277.

- Halliday, J.A., Bell, K., and Shaw, D.C. 1991. The complete amino acid sequence of feline beta lactoglobulin II and a partial revision of the equine beta lactoglobulin II sequence. *Biochimica et Biophysica Acta* **1077**: 25-30.
- Halushka, M.K., Fan, J.B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R. and Chakravarti, A. 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature Genetics* **22**: 239-247.
- Hambling, S.G., McAlpine, A.S. and Sawyer, L. 1992. β -lactoglobulin. In: *Advanced Dairy Chemistry*. Vol.1 Proteins. (Ed. P.F. Fox). London, Elsevier Applied Science, pp. 141-190.
- Hancock, J. M. 1996. Simple sequences in a "minimal" genome. *Nature Genetics* **14**: 14-15.
- Harris, S., McClenaghlan, M., Simons, J.P., Ali, S. and Clark, A.J. 1990. Gene expression in the mammary gland. *Journal of Reproduction and Fertility* **88**: 707-715.
- Hawkins, G.A. and Hoffman, L.M. 1997. Base excision sequence scanning. A new method for rapid scanning and mutation detection. *Nature Biotechnology* **15**: 803-804.
- Hawkins, G.A. and Hoffman, L.M. 1999. Rapid DNA mutation identification and fingerprinting using base excision sequence scanning. *Electrophoresis* **20**: 1171-1176.
- Hayes, H. and Petit, E.J. 1993. Mapping of the β -lactoglobulin gene and of an immunoglobulin M heavy chain-like sequence to homologous cattle, sheep, and goat chromosomes. *Mammalian Genome* **4**: 207-210.
- Hayes, H., Petit, E., Bouniol, C. and Popescu, P. Localization of of the α 2-casein gene to the homologous cattle, sheep and goat chromosomes 4 by in situ hybridization. *Cytogenetics and Cell Genetics* **64**: 282-285.
- Holland, P.M., Abramson, R.D., Watson, R. and Gelfand, D.H. 1991. Detection of specific polymerase chain product by utilizing the 5'→3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proceedings of the National Academy of Sciences of USA* **88**: 7276-7280.
- Horisberger, M., and Vonlanthen, M. 1980. Localization of glycosylated kappa casein in bovine casein micelles by lectin-labelled gold granules. *Journal of Dairy Research* **47**: 185-191.
- Horne, D.S. Ethanol stability. In: *Advanced Dairy Chemistry*. Vol. 1. Proteins. (Ed. P.F. Fox). London, Elsevier Applied Science, pp. 657-690.
- Hovig, E., Smith-Sorensen, B., Brogger, A. and Borresen, A.L. 1991. Constant denaturant gel electrophoresis, a modification of denaturing gradient gel electrophoresis, in mutation detection. *Mutation Research* **262**: 63-71.
- Howell, W.M., Jobs, M., Gyllensten, U. and Brookes, A.J. 1999. Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nature Biotechnology* **17**: 87-88.
- Hsu, T.M., Law, S.M., Duan, S., Neri, B.P. and Kwok, P.Y. 2001. Genotyping single-nucleotide polymorphisms by the invader assay with dual-color fluorescence polarization detection. *Clinical Chemistry* **47** (8): 1373-1377
- Hu, F.B., Manson, J.E. and Willet, W.C. 2001. Types of dietary fat and risk of coronary heart disease: a critical review. *Journal of the American College of Nutrition* **20**: 5-19.
- Huang, N., Cogburn, L.A., Agarwal, S.A., Marks, H.L. and Burnside, G. 1993. Over expression of a truncated growth hormone receptor in the sex-linked dwarf chicken: evidence for a splice mutation. *Molecular Endocrinology* **7**: 1391-1398.

- Hughes, A.L. and Nei, M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the USA* **86**: 958-962.
- Ideta, R., Seki, T. and Adachi, K. 1995. Sequence analysis and characterization of FAR-17c, and androgen-dependent gene in the flank organs of hamsters. *Journal of Dermatology Science* **9**: 94-102.
- Jackson, R.J. 1993. Cytoplasmic regulation of mRNA function: the importance of the 3' untranslated region. *Cell* **74**: 9-14.
- Jann, O., Prinzenberg, E., Brandt, H., Williams, J.L., Ajmone-Marsan, P., Zaragoza, P., Ozbeyaz, C. and Erhardt, G. 2002. Intragenic haplotypes at the bovine CSN1S1 locus. *Archives Tierzucht* **45**: 13-21.
- Jaubert, A. and Martin, P. 1992. Reverse phase HPLC analysis of goat caseins. Identification of as1 and as2 genetic variants. *Lait* **72**: 235-247.
- Jeffreys, A.J., Wilson V. and Thein S.L. 1985. Individual-specific fingerprint of human DNA. *Nature* **316**: 76-79.
- Jensen, R.G. 2002. The composition of bovine milk lipids: January 1995 to December 2000. *Journal of Dairy Science* **85**: 295-350.
- Jolles, J., Alais, C. and Jolles, P. 1968. The tryptic peptide with rennin-sensitive linkage of cow's kappa casein. *Biochimica Biophysica Acta* **168**: 591-593.
- Jolles, P. and Fiat, A. 1979. The carbohydrate portions of milk glycol-proteins. *Journal of Dairy Research* **46**: 187-191.
- Jolles, P. and Henschen, A. 1982. Comparison between the clotting blood and milk. *Trends in Biochemical Sciences* **7**: 325-328.
- Jolles, P., Loucheux-Lefebvre, M.H. and Henschen, A. 1978. Structural relatedness of κ -casein and fibrinogen γ -chain. *Journal of Molecular Evolution* **11**: 271-277.
- Jones, H.B., Standridge, K.M., Claycombe, K.J., Smith, P.J. and Mostaid-Moussa, N. 1998. Glucose induces expression of stearoyl coenzyme A desaturase 1 in 3T3-L1 adipocytes. *Biochemical Journal* **335** 405-408.
- Jones, W.K., Yu-Lee, L.Y., Clift, S.M., Brown, T.L. and Rosen, J.M. 1985. The rat casein multigene family. Fine structure and evolution of the beta-casein gene. *Journal of Biological Chemistry* **260**: 7042-7050.
- Jump, B.D. and Clarke, S.D. 1999. Regulation of gene expression by dietary fat. *Annual Review of Nutrition* **19**: 63-90.
- Jurinke, C., van den Boom, D., Jacob, A., Tang, K., Worl, R. and Koster, H. 1996. Analysis of ligase chain reaction products via matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Analytical Biochemistry* **237**: 174-181.
- Kaestener, K.H., Ntambi, J.M., Kelly, T.J. and Lane, D. 1989. Differentiation-induced gene expression in 3T3-L1 pre-adipocytes. A second differentially expressed gene encoding stearoyl coenzyme A desaturase. *Journal of Biological Chemistry* **264**: 14755-14761.
- Kambadur, R., Sharma, M., Smith, T.P.L. and Bass, J.J. 1997. Mutations in myostatin (GDF8) in double-musled Belgian blue and piedmontese cattle. *Genome Research* **7**: 910-915.
- Khanna, M., Park, P., Zirvi, M., Cao, W., Picon, A., Day, J., Paty, P. and Barany, F. 1999. Multiplex PCR/LDR for detection of K-ras mutations in primary colon tumours. *Oncogene* **18**: 27-38.

- Kim, Y.C. and Ntambi, J.M. 1999. Regulation of stearoyl coenzyme A desaturase genes: role in cellular metabolism and pre adipocyte differentiation. *Biochemical and Biophysical Research Communications* **266**: 1-4.
- King, J.B.W. 1969. The distribution of sheep beta lactoglobulin. *Animal Production* **11**: 53-57.
- Kirk, B.W., Feinsod, M., Favis, R., Kliman, R.M. and Barany, F. 2002. Single nucleotide polymorphism seeking long term association with complex disease. *Nucleic Acids Research* **30**: 3295-3311.
- Korkko, J., Annunen, S., Pihlajamaa, T., Prockop, D.J. and Ala-kokko, L. 1998. Conformation sensitive gel electrophoresis: comparison with denaturing gradient gel electrophoresis and nucleotide sequencing. *Proceedings of the National Academy of Sciences of USA* **95**:1681-5.
- Krook, A., Stratton, I.M. and O'Rahilly, S. 1992. Rapid and simultaneous detection of multiple mutations by pooled and multiplex single nucleotide primer extension: application to the study of insulin-responsive glucose transporter and insulin receptor mutations in non-insulin-dependent diabetes. *Human Molecular Genetics* **1**: 391-395.
- Kuppaswami, M.N., Hoffmann, J.W., Kasper, C.K., Spitzer, S.G., Groce, S.L. and Bajaj, S.P. 1991. Single primer extension to detect genetic diseases: experimental application to hemophilia B (factor IX) and cystic fibrosis genes. *Proceedings of the National Academy of Sciences of USA* **88**: 1143-1147.
- Kwok, P.Y. 2001. Methods for genotyping single nucleotide polymorphisms. *Annual Reviews of Genomics and Human Genetics* **2**: 235-258.
- Kwok, S., Kellogg, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J.J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Research* **18**: 999-1005.
- Landau, J.M., Sekowski, A. and Hamm, M.W. 1997. Dietary cholesterol and the activity of stearoyl coenzyme A desaturase in rats: evidence of an indirect regulatory effect. *Biochimica Biophysica Acta* **1345**: 349-357.
- Landegren, U., Kaiser, R., Sanders, J. and Hood, L. 1988. A ligase-mediated gene detection technique. *Science* **241**: 1077-1080.
- Lander E.S. et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.
- Lathrop, M., Nakamura, Y., O'Connell, P., Leppert, M., Woodward, S., Lalouel, J.M. and White, R. 1988. A mapped set of genetic markers for human chromosome 9. *Genomics* **3**: 361-366.
- Law, A.J.R. and Tziboula, A. 1993. Fractionation of caprine kappa casein and examination of polymorphism by FPLC. *Milchwissenschaft* **48**: 68-71.
- Lee, L.J., Kivak, K.J., Mullah, B., Graham, R.J., Vinayak, R.S. and Wodenberg, T.M. 1999. Seven-color homogeneous detection of six PCR products. *Biotechniques* **27**: 342-379.
- Lefevre, P., Diot, C., Legrand, P. and Douaire, M. 1999. Hormonal regulation of stearoyl coenzyme A desaturase activity and gene expression in primary cultures of chicken hepatocytes. *Archives of Biochemistry and Biophysics* **368**: 329-337.
- Leveziel, H., Metenier, L., Guerin, G., Cullen, P., Provot, C., Bertaud, M. and Mercier, J.C. 1991. Restriction fragment length polymorphism of ovine casein genes: close linkage between the α s1, α s2, β and κ -casein loci. *Animal Genetics* **22**: 1-10.

- Leveziel, H., Rodellar, C., Leroux, C., Pepin, L., Grohs, C., Vaiman, D., Mahe, M.F., Martin, P. and Grosclaude, F. 1994. A microsatellite within the bovine k-casein gene reveals a polymorphism correlating strongly with polymorphisms previously described at the protein as well as the DNA level. *Animal Genetics* **25**: 223-228.
- Li, J., Butler, J.M., Tan, Y., Lin, H., Royer, S., Ohler, L., Shaler, T.A., Hunter, J.M., Pollart, D.J., Monforte, J.A. and Becker, C.H. 1999. Single nucleotide polymorphism determination using primer extension and time of flight mass spectrometry. *Electrophoresis* **20**: 1258-1265.
- Li, Q., Peterson, R.K., Fang, X. and Stamatoyannopoulos, G. 2002. Locus control regions. *Blood* **100**: 3077-3086.
- Li, W.H. and Sadler, L.A. 1991. Low nucleotide diversity in man. *Genetics* **129**: 513-523.
- Lien, S. and Rogne, S. 1993. Bovine casein haplotypes: number, frequency and applicability as genetic markers. *Animal Genetics* **24**: 373-376.
- Lin, C.Y., Mcallister, A.J., Ng-Kwai Hang, K.F. and Hayes, J.F. 1986. Effects of milk protein loci on first lactation production in dairy cattle. *Journal of Dairy Science* **69**: 704-712.
- Lindersson, M., Lunden, A. and Andersson, L. 1995. Genotyping bovine milk protein using allele discrimination by primer length and automated DNA sizing technology. *Animal Genetics* **26**: 67-72.
- Lishanki, A., Ostrander, E.A. and Rine, J. 1994. Mutation detection by mismatch binding protein, MutS, in amplified DNA: application of the cystic fibrosis gene. *Proceedings of the National Academy of Sciences of USA* **91**: 2674-2678.
- Liu, Q. and Sommer, S.S. 1995. Restriction endonuclease fingerprinting (REF): a sensitive method for screening mutations in long, contiguous segments of DNA. *Biotechniques* **18**: 470-477.
- Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C. and Ward, D.C. 1998. Mutation detection and single-molecule counting using rolling-circle amplification. *Nature Genetics* **19**: 225-232.
- Lodes, A., Krause, I., Buchberger, J., Aumann, J. and Kloster-Meyer, H. 1996. The influence of genetic variants of milk proteins on the compositional and technological properties of milk: 1. Casein micelle size and the content of non glycosylated k-casein. *Milchwissenschaft* **51**: 368-373.
- Loeb, L.A. and Preston, B.D. 1986. Mutagenesis by apurinic/apyrimidic sites. *Annual Reviews of Genetics* **20**: 201-230.
- Lombardo, A.J. and Brown, G.B. 1997. A quantitative and specific method for measuring transcript levels of highly homologous genes. *Nucleic Acids Research* **24**: 4812-4816.
- Los, D.A. and Murata, N. 1998. Structure and expression of fatty acid desaturases. *Biochimica et Biophysica Acta* **1394**: 3-15.
- Lum, S.L., Dovic, P. and Medrano, J.F. 1997. Polymorphism of bovine β -lactoglobulin promoter and differences in the binding affinity of activator protein-2 transcription factor. *Journal of Dairy Science* **80**: 1389-1397.
- Lyamichev, V., Mast, A.L., Hall, J.G., Prudent, J.R., Kaiser, M.W., Takova, T., Kwiatkowski, R.W., Sander, T.J., deArruda, M., Arco, D.A., Neri, B.P. and Brow, M.A. 1999. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnology* **17**: 292-296.
- Macha, J. 1970. Protein polymorphism in goat's milk. *Zivocisna Vyroba* **15**: 801-805.

- Manfredi, E., Ricordeau, G., Barbieri, M.E., Amigues, Y. and Bibé, B. 1995. Génotype caseines as1 et sélection des boucs sur descendance dans les races Alpine et Saanen. *Genetics Selection Evolution* **27**: 451-458.
- Martin, P., Brignon, C., Furet, J.P. and Leroux, C. 1996. The gene encoding as1-casein is expressed in human mammary epithelial cells during lactation. *Lait* **76**: 526-537.
- Martin, P., Ollivier-Bousquet, M. and Grosclaude, F. 1999. Genetic polymorphism of caseins: a tool to investigate casein micelle organization. *International Dairy Journal* **9**: 163-171.
- Martin, P., Szymanowska, M., Zwierzchowski, L. and Leroux, C. 2002. The impact of genetic polymorphisms on the protein composition of ruminant milks. *Reproduction Nutrition Development* **42**: 433-459.
- Marziali, A.S. and Ng-Kwai Hang, K.F. 1986. Effect of milk composition and genetic polymorphism on cheese composition. *Journal of Dairy Science* **69**: 1793-1799.
- Mashal, R.D., Koontz, J. and Sklar, J. 1995. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nature Genetics* **9**: 177-183.
- Mather, I.H. and Keenan, T.W. 1998. Origin and secretion of milk lipids. *Journal of Mammary Gland Biology and Neoplasia* **3**: 259-273.
- Medrano, J.F. and Aguilar-Cordova, E. 1990. Genotyping of bovine kappa-casein loci following DNA sequence amplification. *Biotechnology* **8**: 144-146.
- Medrano, J.F. and Aguilar-Cordova, E. 1990. PCR amplification of bovine β -lactoglobulin genomic sequences and identification of genetic variants by RFLP analysis. *Animal Biotechnology* **8**: 144-146.
- Medrano, J.F., Johnson, A., DePeters, E.J., and Islas, A. 1998. Genetic modification of the composition of milk fat: identification of polymorphisms within the bovine stearyl coenzyme A desaturase gene. *Life Sciences Conference Slovenia*, September 19-24 1998.
- Mei, R., Galipeau, P.C., Prass, C., Berno, A., Ghandour, G., Patil, N., Wolff, R.K., Chee, M.S., Reid, B.J. and Lockhart, D.J. 2000. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Research* **10**: 1126-1137.
- Mephram, T.B., Gaye, P., Martin, P. and Mercier, J.C. 1992. Biosynthesis of milk protein. In: *Advanced Dairy Chemistry*. Vol. 1. Proteins. (Ed. P.F: Fox). London, Elsevier Applied Science, pp. 491-543.
- Mercier, J.C. 1981. Phosphorylation of caseins. Present evidence for an amino triplet code post-translationally recognized by specific kinases. *Biochimie* **63**: 1-17.
- Mercier, J.C., Addeo, F. and Pelissier, J.P. 1976b. Structure primaire de la caseinomacropéptide caprine. *Biochimie* **58**: 1303-1310.
- Mercier, J.C., Chobert, J.M. and Addeo, F. 1976a. Comparative analysis of the amino acid sequences of the caseinomacropéptide from seven species. *FEBS Letters* **72**: 208-214.
- Metz, E.C. and Palumbi, S.R. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein binding. *Molecular Biology Evolution* **13**: 397-406.
- Miesel, H. 1997. Biochemical properties of bioactive peptides derived from milk proteins: potential nutraceuticals for food and pharmaceutical applications. *Livestock Production Science* **50**: 125-138.
- Mihara, K. 1990. Structure and regulation of rat liver microsomal stearyl coenzyme A desaturase gene. *Journal of Biochemistry* **108**: 1022-1029.

- Miller, R.D. and Kwok, P.Y. 2001. The birth and death of human SNPs: new experimental evidence and implications for human history and medicine. *Human Molecular Genetics* **10**: 2195-2198.
- Miller, W.C. and Ntambi, J.M. 1996. Peroxisome proliferators induce mouse liver stearyl coenzyme A desaturase 1 gene expression. *Proceedings of the National Academy of Sciences of USA* **93**: 9443-9448.
- Miranda, G., Anglade, P., Mahé, M.F. and Erhardt, G. 1993. Biochemical characterization of the bovine genetic kappa casein C and E variants. *Animal Genetics* **24**: 27-31.
- Monaco, H.L., Zanotti, G., Spadon, P., Bolognesi, M., Sawyer, L. and Eliopoulos, E.E. 1987. Crystal structure of the trigonal form of bovine β -lactoglobulin and of its complex with retinol at 2.5 Å resolution. *Journal of Molecular Biology* **197**: 695-706.
- Monti, J.C., Mermoud, A.F. and Jollès, P. 1989. Anti-bovine β -lactoglobulin antibodies react with a human lactoferrin fragment and bovine β -lactoglobulin present in human milk. *Experienta* **45**: 178-180.
- Moore, J.H. and Christie, W.W. 1979. Lipid metabolism in the mammary gland of ruminant animals. *Progress in Lipid Research* **17**: 347-395.
- Moore, J.H. and Christie, W.W. 1981. Lipid metabolism in the mammary gland of ruminant animals. In: *Lipid metabolism in ruminants*. W.W. Christie Ed. Pergamon Press, Oxford UK.
- Moreno, F.J., Recio, I., Olano, A. and Lopez-Fandino, R. 2001. Heterogeneity of caprine kappa casein macropeptide. *Journal of Dairy Research* **68**: 197-208.
- Morgan, F. and Gaborit, P. 2001. The typical flavor of goat milk products: technological aspects. *International Journal of Dairy Technology* **54**: 38-40.
- Morini, D., Losi, G., Catagnetti, G.B., Benevelli, M., Resmini, P. and Volonterio, G. 1975. L'influenza delle varianti genetiche della k-caseina sulla dimensione delle micelle caseiniche. *Scienza e Tecnica Lattiero-Casearia* **26**: 437-444.
- Moses, R.E. 2001. DNA damage processing defects and disease. *Annual Reviews of Genomics and Human Genetics* **2**: 41-68.
- Mullally, M.M., Meisel, H. and FitzGerald, R.J. 1997. Angiotensin-I-converting enzyme inhibitory activities of gastric and pancreatic proteinase digests of whey proteins. *International Dairy Journal* **7**: 299-303.
- Mullis, K.B. and Faloona, F.A. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed reaction. *Methods in Enzymology* **155**: 335-350.
- Murphy, M., Uden, P., Palmquist, D.L. and Wiktorsson, H. 1987. Rumen and total digestibilities in lactating cows fed diets containing full-fat rapeseed. *Journal of Dairy Science* **70**: 1572-1578.
- Myers, R.M., Fischer, S.G., Lerman, L.S. and Maniatis, T. 1985. Nearly all single base substitutions in DNA fragments joined to GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Research* **13**: 3131-3145.
- Myers, R.M., Larin, Z. and Maniatis, T. 1985a. Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**: 1242-1246.
- Nash, K.A. and Inderlied, C.B. 1996. Rapid detection of mutations associated with macrolide resistance in *Mycobacterium avium* complex. *Antimicrobial Agents and Chemotherapy* **40**: 1748-1750.
- Neelin, J.M. 1964. Variants of κ -casein revealed by improved starch gel electrophoresis. *Journal of Dairy Science* **47**: 506-509.

- Neville, M.C. and Picciano, M.F. 1997. Regulation of milk lipid secretion and composition. *Annual Review of Nutrition* **17**: 159-184.
- Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.G., Summers, C. et al. 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Research* **17**: 2503-2516.
- Ney, D.M. 1991 Potential of enhancing the nutritional properties of milk fat. *Journal of Dairy Science* **74** 4002-4012.
- Ng-Kwai Hang, K.F. 1998. Genetic polymorphism of milk proteins: relationships with production traits, milk composition and technological properties. *Canadian Journal of Animal Science* **78** (supplement): 131-147.
- Ng-Kwai Hang, K.F. and Grosclaude, F. 1992. Genetic polymorphism of milk proteins. In: *Advanced Dairy Chemistry*. Vol. 1. Proteins. (Ed. P.F. Fox). London, Elsevier Applied Science, pp. 405-455.
- Nickerson, D.A., Taylor, S.L., Weiss, K.M., Clark, A.G., Hutchinson, R.G., Stengard, J., Salomaa, V., Vertiainen, E., Boerwinkle, E. and Sing, F. 1998. DNA sequence diversity in a 9.7 Kb region of the human lipoprotein lipase gene. *Nature Genetics* **19**: 233-240.
- Nilsson, M., Krejci, K., Koch, J., Kwiatkowski, M., Gustavsson, P. and Ladegren, U. 1997. Padlock probes reveal single-nucleotide differences, parent of origin and in situ distribution of centromeric sequences in human chromosomes 13 and 21. *Nature Genetics* **16**:252-255.
- Ninomiya, T., Hirabayashi, M., Sagara, J. and Yuki, A. 1994. Functions of milk protein gene 5' flanking regions on human growth hormone gene. *Molecular Reproduction Development* **37**: 276-283.
- Nollau, P. and Wagener, C. Methods for detection of point mutations: performance and quality assessment. *Clinical Chemistry* **43**(7): 1114-1128.
- Ntambi, J. M., Buhrow, S.A, Kaestener, K.H., Christie, R.J., Sibley, E., Kelly, T.J. and Lane, D. 1988. Differentiation-induced gene expression in 3T3-L1 preadipocytes. Characterization of a differentially expressed gene encoding stearoyl coenzyme A desaturase. *Journal of Biological Chemistry* **263**: 17291-17300.
- Ntambi, J.M. 1992. Dietary regulation of stearoyl-CoA desaturase 1 gene expression in mouse liver. *Journal of Biological Chemistry* **267**: 10925-10930.
- Ntambi, J.M. 1995. The regulation of stearoyl-CoA desaturase. *Progress in Lipid Research* **34**: 139-150.
- Ntambi, J.M. 1999. Regulation of stearoyl coenzyme A desaturase by polyunsaturated fatty acids and cholesterol. A review. *Journal of Lipid Research* **40**: 1549-1558.
- Ntambi, J.M., Miyazaki, M., Store, J.P., Lan, H., Kendzioriski, C.M., Yandell, B.S., Song, Y., Cohen, P., Friedman, J.M. and Attie, A.D. 2002. Loss of stearoyl-CoA desaturase-1 function protects mice against adiposity. *Proceedings of the National Academy of Sciences of USA* **99**: 11482-11486.
- Nyren, P., Pettersson, B. and Uhlen, M. 1993. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Analytical Biochemistry* **208**: 171-175.
- O'Donovan, M.C., Oefener, P.J., Roberts, S.C., Austin, J., Hoogendoorn, B., Guy, C., Speight, G., Upadhyaya, M., Sommer, S.S. and McGuffin, P. 1998. Blind analysis of denaturing high performance liquid chromatography as a tool for mutation detection. *Genomics* **52**: 44-49.

- Orita, M., Iwahana, H., Kanazawa, H., Hatashi, K. and Sehiya, T. 1989. Detection of polymorphism of human DNA by gel electrophoresis as single-strand conformation polymorphism. *Proceedings of the National Academy of Sciences of USA* **86**: 2766-2770.
- Ouwehand, A.C., Salminen, S.J., Skurnik, M. and Conway, P.L. 1997. Inhibition of pathogen adhesion by β -lactoglobulin. *International Dairy Journal* **7**: 685-692.
- Palmquist, D.L., Beaulieu, A.A. and Barbano DM. 1993. Feed and animal factors influencing milk fat composition. *Journal of Dairy Science* **76**: 1753-1771.
- Papiz, M.Z., Sawyer, L., Eliopoulos, E.E., North, A.C.T., Findlay, J.B.C., Sivaprosadarao, R., Jones, T.A., Newcomer, M.E. and Kraulis, P.J. 1986. The structure of β -lactoglobulin and its similarity to plasma and retinol-binding protein. *Nature* **324**: 383-385.
- Pappalardo, M., Gallo, D., Cosenza, G., Pastore, N., Senese, C., Rubino, R., Marletta, D. and Ramuno, L. 2001. Molecular analysis of the goat β -lactoglobulin gene: preliminary results. *Proceedings of the ASPA XIV Congress, Firenze, 12-15 June, 2001*.
- Passey, R.G. and MacKinlay, A.G. 1995. Characterization of a second, apparently inactive copy of the bovine β -lactoglobulin gene. *European Journal of Biochemistry* **233**: 736-743.
- Pena, N.P., Sanchez, A., Coll, A. and Folch, J.M. 1999. Isolation, sequencing, and relative quantitation by fluorescent-ratio PCR of feline β -lactoglobulin I, II, and III cDNAs. *Mammalian Genome* **10**: 560-564.
- Peréz, M.D. and Calvo, M. 1995. Interaction of β -lactoglobulin with retinol and fatty acids and its role as a possible biological function for this protein: a review. *Journal of Dairy Science* **78** 978-988.
- Perez, M.D., Puyol, P., Ena, J.M. and Calvo, M. 1993. Comparison of the ability to bind lipids of beta lactoglobulin and serum albumin of milk from ruminant and non ruminant species. *Journal of Dairy Research* **60**: 55-63.
- Persuy, M.A., Legrain, S., Printz, C., Stinnakre, M.G., Lepourry, L., Brignon, G. and Mercier, J.C. 1995. High level, stage-, and mammary-tissue-specific expression of a caprine kappa-casein-gene-encoding minigene driven by a beta-casein promoter in transgenic mice. *Gene* **165**: 291-295.
- Persuy, M.A., Printz, C., Medrano, J.F. and Mercier, J.C. 1999. A single nucleotide deletion resulting in a premature stop codon is associated with marked reduction of transcripts from a goat β -casein null allele. *Animal Genetics* **30**: 444-451.
- Pinder S.J., Perry, B.N., Skidmore, C.J. and Savva, D. 1991. Analysis of polymorphism in the bovine casein genes by use of the polymerase chain reaction. *Animal Genetics* **22**: 11-20.
- Ponce-Castenada, M.V., Lopez-Casillas, F. and Kim, K. 1991. Acetyl coenzyme A carboxylase messenger ribonucleic acid metabolism in liver, adipose tissue, and mammary gland during pregnancy and lactation. *Journal of Dairy Science* **74**: 4013-4021.
- Pottmeyer, S. and Kemper, B. 1992. Endonuclease VII resolves cruciform DNA with nick and counter-nick and its activity is directed by local nucleotide sequence. *Journal of Molecular Biology* **223**: 607-615.
- Preaux, G., Braunitzer, G., Schrank, B. and Strangl, A. 1979. The amino acid sequence of goat β -lactoglobulin. *Hoppe-Seyler's Physiology and Chemistry* **360**: 1595-1604.

- Prinzenberger, E.M., Hiendleder S., Ikonen, T. and Erhardt, G. 1996. Molecular genetic characterization of new bovine genetic kappa casein alleles F and G and genotyping by PCR-RFLP. *Animal Genetics* **27**: 347-349.
- Prinzenberger, E.M., Krause, I. and Erhardt, G. 1999. SSCP analysis at the bovine *CSN3* locus discriminates six alleles corresponding to known protein variants and three new DNA polymorphisms. *Animal Biotechnology* **10**: 49-62.
- Puyol, P., Perez, M.D., Ena, J.M. and Calvo, M. 1991. Interaction of bovine beta lactoglobulin and other bovine and human whey proteins with retinol and fatty acids. *Agricultural Biological Chemistry* **55**: 2515-2520.
- Ramunno, L., Longobardi, E., Pappalardo, M., Rando, A., Di Gregorio, P., Cosenza, G., Mariani, P., Patore, N. and Masina, P. 2001. An allele associated with a non-detectable amount of α s2 casein in goat milk. *Animal Genetics* **32**: 19-26.
- Rasmussen, L.K., Hojrup, P. and Petersen, T.E. 1994. Disulphide arrangement in bovine caseins: localization of intrachain disulphide bridges in monomers of κ - and α s2-casein from bovine milk. *Journal of Dairy Research* **61**: 485-493.
- Recio, I., Molina, E., Ramos, M. and de Frutos, M. 1995. Quantitative analysis of major whey proteins by capillary electrophoresis using uncoated capillaries. *Electrophoresis* **16** 654-658.
- Recio, I., Perz-Rodriguez, M.L., Amigo, L. and Ramos, M. 1997. Study of the polymorphism of caprine milk casein by capillary electrophoresis. *Journal of Dairy Research* **64**: 515-523.
- Reddy, M., Kella, N.K.D. and Kinsella, J.E. 1988. Structural and conformational basis of the resistance of beta lactoglobulin to peptic and chymotryptic digestion. *Journal of Agricultural Food Chemistry* **36**: 737-742.
- Reddy, P.S. and Housman, D.E. 1997. The complex pathology of trinucleotide repeats. *Current Opinion in Cell Biology* **9**: 364-372.
- Reiss, J., Krawczak, M., Schloesser, M., Wagner, M. and Cooper, D. 1990. The effect of replication errors on the mismatch analysis of amplified DNA. *Nucleic Acids Research* **18**: 973-978.
- Rempel, W.E., Lu, M., El-Kandelgy, S., Kennedy, C.F., Irvin, L.R., Mickelson, J.R. and Louis, C.F. 1993. Relative accuracy of the halothane challenge test and a molecular genetic test in detecting the gene for porcine stress syndrome. *Journal of Animal Science* **71**: 1395-1399.
- Richards, R.L. and Sutherland, G.R. 1992. Dynamic mutations: A new class of mutations causing human disease. *Cell* **70**: 709-712.
- Richardson, B.C., Creamer, L.K. and Pearce K.M. 1973. Comparative micelle structure. I. The isolation and chemical characterization of caprine κ -casein. *Biochimica Biophysica Acta* **310**: 111-117.
- Ridley, M. 1991. A survey of science: The edge of ignorance. *The Economist* 1-22
- Rijnkels, M., Kooiman, P.M., deBoer, H.A. and Pieper, F.R. 1997a. Organization of the bovine casein gene locus. *Mammalian Genome* **8**: 148-152.
- Rijnkels, M., Kooiman, P.M., Krimpenfort, P.J., de Boer, H.A., and Pieper, F.R. 1995. Expression analysis of the individual bovine β -, α s2- and κ -casein genes in transgenic mice. *Biochemical Journal* **311**: 929-937.
- Rijnkels, M., Wheeler, D.A., deBoer, H.A. and Pieper, F.R. 1997b. Structure and expression of the mouse casein gene locus. *Mammalian Genome* **8**: 9-15.

- Robitaille, G., Ng-Kwai-Hang, K.F. and Monardes, H.G. 1991. Variation of the N-acetylneuraminic acid content of bovine κ -casein. *Journal of Dairy Research* **58**: 107-114.
- Rollema, H.S. 1992. Casein association and micelle formation. In: *Advanced Dairy Chemistry*. Vol. 1. Proteins. (Ed. P.F: Fox). London, Elsevier Applied Science, pp. 111-140.
- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. and Nyren, P. 1996. Real time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry* **242**: 84-89.
- Ross, P., Hall, L., Smirnov, I and Haff, L. 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature Biotechnology* **16**: 1347-1351.
- Russo, V., Davoli, R., Dall'olio, S. and Tedesci, M. 1986. Ricerche sul polimorfismo del latte caprino. *Zootecnica i Nutrizione Animale* **12**: 55-62.
- Saleeba, J.A., Ramus, S.J. and Cotton, R.G. 1992. Complete mutation detection using unlabeled chemical cleavage. *Human Mutation* **1**: 63-69.
- Sanchez-Belda, A. 1986. Raza Canaria. In: Catalogo de razas autóctonas españolas. Vol. 1. Especies ovina y caprina. Ministerio de Agricultura, Pesca y Alimentación Publ. pp 163-173, Madrid, Spain.
- Sarkar, G., Yoon, H.S. and Sommer, S.S. 1992. Dideoxy fingerprinting (ddE): a rapid and efficient screen for the presence of mutations. *Genomics* **13**: 441-443.
- Sawyer, L. 1982. The secondary structure of milk proteins and their biological function. *Journal of Dairy Science* **76**: 3062-3078.
- Sawyer, L. and Kontopidis, G. 2000. The core lipocalin, bovine β -lactoglobulin. *Biochimica et Biophysica Acta* **1482**: 136-148.
- Schaar, J. 1984. Effects of kappa casein genetic variants and lactation number on the renneting properties of individual milks. *Journal of Dairy Research* **51**: 397-403.
- Schibler, L., Vaiman, D., Oustry, A., Giraud-Delville, C. and Cribiu, E.P. 1998. Comparative gene mapping: a fine-scale survey of chromosome rearrangements between ruminants and humans. *Genome Research* **8**: 901-915.
- Schielben, S., Erhardt, G. and Senft, B. 1991. Genotyping of bovine kappa casein A, B, C, and E following DNA sequence amplification and direct sequencing of kappa casein E product. *Animal Genetics* **22**: 333-342.
- Schlee, P., Krause, I. and Rottmann, O. 1993. Genotyping of bovine beta lactoglobulin alleles A and B using the PCR. *Archives Tierzucht* **36**: 519-523.
- Schmidt, D.G. 1964. Starch gel electrophoresis of κ -casein. *Biochimica Biophysica Acta* **90**: 411-414
- Scholz, R.B., Milde Langosch, K., Jung, R. et al. 1993. Rapid screening for Tp53 mutations by temperature gradient gel electrophoresis: a comparison with SSCP. *Human Molecular Genetics* **2**: 2155-2158.
- Sessler, A.M., Kaur, N., Palta, P.J. and Ntambi, J.M. 1996. Regulation of stearyl-CoA desaturase-1 mRNA stability by polyunsaturated fatty acids in 3T3-L1 adipocytes. *Journal of Biological Chemistry* **271**: 29854-29858.
- Shanklin, J., Whitle, E. and Fox, B.G. 1994. Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, Stearyl-CoA desaturase and are conserved in alkane hydroxylase and xylene monooxygenase. *Biochemistry* **33**: 12787-12794.

- Sheffield, V.C., Beck, J.C., Kwiktek, A.E., Sandstrom, D.W. and Stone, E.M. 1993. The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* **16**: 325-332.
- Sheffield, V.C., Cox, D.R., Lerman, L.S. and Myers, R.M. 1989. Attachment of a 40-base pair GC-rich sequence (GC-clamp) to genomic DNA fragments by PCR results in improved detection of single base changes. *Proceedings of the National Academy of Sciences of USA* **86**: 232-236.
- Shild, T.A., Wagner, V. and Geldermann, H. 1994. Variants within the 5'-flanking regions of bovine milk protein genes: I. κ -casein gene. *Theoretical and Applied Genetics* **89**: 116-120.
- Shuster, D.E., Kehrl Jr, M.E., Ackermann, M.R. and Gilbert, R.O. 1992. Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proceedings of the National Academy of Sciences of USA* **89**: 9225-9229.
- Singer-Sam, J., LeBon, J.M., Dai, A. and Riggs A.D. 1992. A sensitive quantitative assay for measurement of allele-specific transcripts differing by a single nucleotide. *PCR Methods Applications* **1**: 160-163.
- Smith, J. and Modrich, P. 1996. Mutation detection with MutH, MutL and MutS mismatch repair proteins. *Proceedings of the National Academy of Sciences of USA* **82**: 4374-4379.
- Smith, S. 1994. The animal fatty synthase: one gene, one polypeptide, seven enzymes. *FASEB Journal* **8**: 1248-1259.
- Sokolov, B.P. 1990. Primer extension technique for the detection of single nucleotides in genomic DNA. *Nucleic Acids Research* **18**:3671.
- Solaro, P.C., Birkenkamp, K., Pfeiffer, P. and Kemper, B, 1993. Endonuclease VII of phage T4 triggers mismatch correction in vitro. *Journal of Molecular Biology* **230**: 868-877.
- Soulier, S., Ribadeau-Dumas, B. and Deanmur, R. 1974. Purification of κ -casein from sheep. *European Journal of Biochemistry* **50**: 445-452.
- Sourdioux, M., Douaire, M. and Delabrosse, Y. 1996. DNA polymorphisms of lipogenesis genes and analysis of linkage with fatness in turkeys. *Poultry Science* **5**: 1018-1026.
- Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**: 503-517.
- Stasiuk, S.X., Summers, E.L. and Demmer, J. 2000. Cloning of marsupial kappa-casein cDNA from the brushtail possum (*Trichosurus vulpecula*). *Reproduction Fertility and Development* **12**: 215-222.
- Stukey J.E., McDonough, V.M. and Martin, C.E. 1989. Isolation and characterization of OLE1, a gene affecting fatty acid desaturation from *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **264**:16537-16544.
- Stupiniskii, R.M. and Il'Chenko, M.D. 1967. Electrophoresis of goat's milk proteins. *Fiziol. Biokim. Sel'khoz Zhivot Respub. Mezhdv Temat. Nauch.* **5**: 62-65.
- Swaisgood, H.E. 1992. Chemistry of the caseins. In: *Advanced Dairy Chemistry*. Vol.1 Proteins. (Ed. P.F. Fox). London, Elsevier Applied Science, pp. 63-110.
- Syvanen, A.C. 1999. From gels to chips: "Minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Human Mutation* **13**: 1-10.
- Syvanen, A.C. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics* **2**: 930-942.

- Syvanen, A.C., Aalto-Setälä, K., Harju, L., Kontula, K. and Söderlund, H. 1990. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* **8**: 684-692.
- Tabor, D.E., Kim, J.B., Spiegelman, B.M. and Edwards, P.A. 1999. Identification of conserved cis-elements and transcription factors required for sterol-regulated transcription of stearoyl-CoA desaturase 1 and 2. *Journal of Biological Chemistry* **274**: 20603-20610.
- Tabor, D.E., Xia, Y.R., Mehrabian, M., Edwards, P.A. and Lusic, A.J. 1998. A cluster of stearoyl CoA desaturase genes, *Scd1* and *Scd2*, on mouse chromosome 19. *Mammalian Genome* **9**: 341-342.
- Takeuchi, M., Tsuda, E., Yoshikawa, M., Sasaki, R. and Chiba, H. 1984. Post-translational modification of κ -casein in the Golgi apparatus of mid-lactating mammary gland from rat and cow. *Agricultural Biological Chemistry* **48**: 2789-2797.
- Tapp, I., Malmberg, L., Rennel, E., Wick, M. and Syvanen, A.C. 2000. Homogeneous scoring of single nucleotide polymorphisms: comparison of the 5' nuclease TaqMan assay and molecular beacons probes. *Biotechniques* **28**: 732-738.
- Tee, M.K., Moran, C. and Nicholas, F.W. 1992. Temperature gradient gel electrophoresis: detection of a single base substitution in the cattle β -lactoglobulin gene. *Animal Genetics* **23**: 431-435.
- Thiede M.A., Ozols, J. and Stritmatter, P. 1986. Construction and sequence of cDNA for rat liver stearoyl coenzyme A desaturase. *Journal of Biological Chemistry* **261**: 13230-13235.
- Thompson, M.D., Dave, R.J. and Nakhashi H.L. 1985. Molecular cloning of mouse mammary gland κ -casein: comparison with rat κ -casein and rat and human γ -fibrinogen. *DNA* **4**: 263-271.
- Thompson, M.P., Kiddy, C.A., Pepper, L. and Zittle, C.A. 1962. Variation in the α s-casein fraction of individual cow's milk. *Nature* **195**:1001-1002.
- Threadgill, D.W. and Womack, J.E. 1990. Genomic analysis of the major bovine milk protein genes. *Nucleic Acids Research* **18**: 6935-6942.
- Tindall, K.R. and Kunkel, T.A. 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemical Journal* **27**: 6008-6013.
- Tocher, D.R., Leaver, M.J. and Hodgson, P.A. 1998. Recent advances in the biochemistry and molecular biology of fatty acyl desaturases. *Progress in Lipid Research* **37**: 73-111.
- Tyagi, S and Kramer, F.R. 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnology* **14**: 303-308.
- Tyagi, S., Bratu, D.P. and Kramer, F.R. 1998. Multicolor molecular beacons for allele discrimination. *Nature Biotechnology* **16**: 49-53.
- Underhill, P.A., Jin, L.A., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L. and Oefener, P.J. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high performance liquid chromatography. *Genomics* **7**: 996-1005.
- Vaiman, D., Schibler, L., Bourgeois, F., Oustry, A., Amigues, Y. and Cribiu, E.P. 1996. A genetic linkage map of the male goat genome. *Genetics* **144**: 279-305.
- Vakalopoulou, E., Schaack, J. and Shenk, T. 1991. A 32-kilodalton protein binds to AU-rich elements domains in the 3'-untranslated regions of rapidly degraded mRNAs. *Molecular Cell Biology* **11**: 3355-3364.

- Van Eenennaam, A.L. and Medrano, J.F. 1991. Differences in allelic protein expression in the milk of heterozygous kappa casein cows. *Journal of Dairy Science* **74**: 1491-1496.
- Venter, C. et al. 2001. The sequence of the human genome. *Science* **291**: 1304-1351.
- Vernon, R.G. 1980. Lipid metabolism in the adipose tissue of ruminant animals. *Progress in Lipid Research* **19**: 23-36.
- Vernon, R.G., Barber, M.C. and Finley, E. 1991. Modulation of the activity of acetyl-CoA carboxylase and other lipogenic enzymes by growth hormone, insulin and dexamethasone in sheep adipose tissue and relationships to adaptations to lactation. *Biochemical Journal* **274**: 543-548.
- Vernon, R.G., Barber, M.C. and Travers, M.T. 1999. Present and future studies on lipogenesis in animal and human subjects. *Proceedings of the Nutrition Society* **58**: 541-549.
- Wagner, R., Debbie P. and Radman, M. Mutation detection using immobilized mismatch binding protein (MutS). *Nucleic Acids Research* **23**: 3944-3948.
- Walstra, P. 1990. On the stability of casein micelles. *Journal of Dairy Science* **73**: 1965-1979.
- Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J. and Lander, E.S. 1998. Large scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. *Science* **280**: 1077-1082.
- Ward, R. J., Travers, T.M., Richards, E.S., Vernon, R.G., Salter, M.A, Buttery, P.J. and Barber, M.C. 1998. Stearoyl coenzyme A desaturase mRNA is transcribed from a single gene in the ovine genome. *Biochimica et Biophysica Acta* **1391**: 145-156.
- Ward, T. J., Honeucutt, R.L. and Derr, J.N. 1997. Nucleotide sequence evolution at the κ -casein locus: evidence for positive selection within the family bovidae. *Genetics* **147**:1863-1872.
- Wartell, R.M., Hosseini, S.H. and Moran Jr, C.P. 1990. Detecting base pair substitution in DNA fragments by temperature gradient gel electrophoresis. *Nucleic Acids Research* **18**: 2699-2705.
- Waters, K.M., Miller, C.W. and Ntambi, J.M. 1994. Insulin and dietary fructose induce stearoyl-CoA desaturase 1 gene expression of diabetic mice. *Journal of Biological Chemistry* **269**: 27773-27777.
- Waters, K.M., Miller, C.W. and Ntambi, J.M. 1997. Localization of a polyunsaturated fatty acid response region in stearoyl-CoA desaturase 1. *Biochimica Biophysica Acta* **1349**: 33-42.
- Watson, C.J., Gordon, K.E., Robertson, M. and Clark, A.J. 1991. Interaction of DNA-binding proteins with a milk protein gene promoter in vitro: identification of a mammary gland-specific factor. *Nucleic Acids Research* **19**: 6603-6610.
- Watts, J.L. and Browse, J. 2000. A palmitoyl-CoA-specific Δ^9 desaturase from *Caenorhabditis elegans*. *Biochemical and Biophysical Research Communications* **272**: 263-269.
- Wenz, H.M., Baumhueter, S., Ramachandra, S. and Wordwood, M. 1999. A rapid automated SSCP multiplex capillary electrophoresis protocol that detects the two common mutations implicated in hereditary hemochromatosis. *Human Genetics* **104**: 29-35.

- Whitelaw, C.B.A., Harris, S., McClenaghlan, M., Simons, J.P. and Clark, A.J. 1992. Position-independent expression of the ovine β -lactoglobulin gene in transgenic mice. *Biochemical Journal* **286**: 31-39.
- Wilkins Stevens, P., Hall, J.G., Lyamichev, V., Neri, B.P., Lu, M., Wang, L., Smith, L.M. and Kelso, D.M. 2000. Analysis of single nucleotide polymorphisms with solid phase invasive cleavage reactions. *Nucleic Acids Research* **29**: e77.
- Wilson, J.J., Zembayashi, M., Lunt, D.K. and Smith, B.S. 1993. Rapid communication: a *TaqI* RFLP in the stearoyl coenzyme A desaturase gene in DNA from purebred Japanese Black cattle. *Journal of Animal Science* **71**: 2575.
- Winter, E.F., Yamamoto, C., Almoguera, C. and Perucho, M. 1985. A method to detect and characterize point mutations in transcribed genes: amplification and overexpression of the mutant *c-Ki-ras* allele in human tumor cells. *Proceedings of the National Academy of Sciences of USA* **82**: 7575-7579.
- Wong, D.W.S., Camirand, W.M. and Pavlath, A.E. 1996. Structure and functionalities of milk proteins. *Critical Reviews in Food Science and Nutrition* **36**: 807-844.
- Wooster, R., Cleton-Jansen, A.M., Collins, N., Mangion, J., Cornelis, R.S., Cooper, C.S., Gusterson, B.A., Ponder B.A., von Deimling, A., Wiestler, O.D. et al. 1994. Instability of short tandem repeats (microsatellites) in human cancer. *Nature Genetics* **6**: 152-156.
- Woychik, J.H. 1964. Polymorphism in κ -casein cow's milk. *Biochemical and Biophysical Research Communications* **16**: 267-271.
- Wu, D.Y., Ugozzoli, L., Pal, B.K. and Wallace, R.B. 1989. Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proceedings of the National Academy of Sciences of USA* **86**: 2757-2760.
- Xu, J.F., Yang, Q.P., Chen, J.Y., Van Baalen, M.R and Hsu, I.C. 1996. Determining the site and nature of DNA mutations with the cloned MutY mismatch repair enzyme. *Carcinogenesis* **17**: 321-326.
- Yahyaoui, M.H., Coll, A., Sanchez, A. and Folch, J.M. 2001. Genetic polymorphism of the caprine kappa casein gene. *Journal of Dairy Research* **68**: 209-216.
- Yahyaoui, M.H., Sanchez, A. and Folch, J.M. 2002. Rapid communication: Partial nucleotide sequence of the goat stearoyl coenzyme A desaturase cDNA and gene structure. *Journal of Animal Science* **80**: 866-867.
- Yamashaki, H., Nakao, M., Sako, Y., Nakaya, K., Sato, M.O., Mamuti, W., Okamoto, M. and Ito, A. 2002. DNA differential diagnosis of human taenid cestodes by base excision sequence scanning thymine-base reader analysis with mitochondrial genes. *Journal of Clinical Microbiology* **40**: 3818-3821.
- Yan, H., Kinzler, K.W. and Vogelstein, B. 2000. Genetic testing, present and future. *Science* **289**: 1890-1892.
- Youil, R., Kemper, B. and Cotton, R.G.H. 1996. Detection of 81 of 81 known mouse β -globin promoter mutations with T4 endonuclease VII. The EMC method. *Genomics* **32**: 431-435.
- Youil, R., Kemper, B.W. and Cotton R.G. 1995. Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proceedings of the National Academy of Sciences of USA* **92**: 87-91.
- Yvon, M., Van Hille, I. and Pellisier, J.P. 1984. In vivo milk digestion in the calf abomasums. II. Milk and whey proteolysis. *Reproduction Nutrition Development* **24**: 835-.

- Zhang, L., Ge, L., Tran, T. Stenn, K. and Prouty, S.M. 2001. Isolation and characterization of the human stearoyl-CoA desaturase gene promoter: requirement of a conserved CCAAT cis-element. *Biochemical Journal* **357**: 183-193.
- Zhang, L., Parimoo, S., Stenn, K. and Prouty, S.M. 1999. Human stearoyl coenzyme A desaturase: alternative transcripts generated from a single gene by usage of tandem polyadenylation site. *Biochemical Journal* **340**: 255-264.
- Zheng, Y., Prouty, S.M., Harmon, A., Sundberg, J.P., Stenn, K.S. and Parimoo, S. 2001. Scd3-A novel gene of the stearoyl-CoA desaturase family with restricted expression in skin. *Genomics* **72**: 182-191.
- Zittle, C.A. and Custer, J.H. 1966. Identification of the κ -casein among the components of whole goat casein. *Journal of Dairy Science* **49**:788-791.

ANNEXES

Partial nucleotide sequence of the goat SCD cDNA

The translation initiation (ATG) and stop (TGA) codons of the 1080 nucleotide open-reading frame are underlined. The corresponding protein sequence of 359 amino acids is denoted by one letter amino acid designation. The histidine residues spatially conserved in the family of membrane desaturases are indicated by bold script. Black triangles mark the exon-exon boundaries for exons 1 through 6. The sequence has been submitted to GenBank under the accession number AF339909

```

CGGATCCAGCGCACAGCGTGCAGCGGAAGGTCCCGAGCACAGCGCCGCGGAT 52
CGCCACGCATCGCCACGCAAAGCAGGCTCAGGAAGTACACTCAGTTCCGACTGC 112
CCCGAACTCCGCTCCGCAGTCTGAGCCCCGAGAAAGTGATCCCAGTGTCCGAGAGCCAG 172
ATGCCGGCCACTTGCTGCAAGAGGAGAT▼TCTAGCTCCTACACGACCACCACCACCATC 232
M P A H L L Q E E I S S S Y T T T T T I
ACAGCACCTCCTTCCAAGTCTGCAGAATGGAGGGGGCAAATTGGAGAAGACTCCCCTA 292
T A P P S K V L Q N G G G K L E K T P L
TACTTGGAAGAAGACATCCGCCCTGAAATGAGAGATGACATCTATGACCCAACCTTACCAG 352
Y L E E D I R P E M R D D I Y D P T Y Q
GATAAGGAGGGCCCAAAGCCCAAGCTTGAGTATGTTTGGAGAAACATCATCCTCATGGGT 412
D K E G P K P K L E Y V W R N I I L M G
CTGTTACACTTGGGAGCCCTGTATGGGATCACATTGATCCCCACCTGCAAGATATACACC 472
L L H L G A L Y G I T L I P T C K I Y T
TTTCTCTGGGTGTTATTCTACTATATGATGAGTGCCCTGGGCATCACAGCAGGGGTCCAT 532
F L W V L F Y Y M M S A L G I T A G V H
CGCCTGTGGAGTCACCGAACCTACAAAGCTCGGCTGCCCTGCGGGTCTTCCTGATCATC 592
R L W S H R T Y K A R L P L R V F L I I
GCCAACACCATGGCGTTCAGAA▼TGACGTTTTTTGAATGGTCCCGAGATCACCGTGCCAC 652
A N T M A F Q N D V F E W S R D H R A H
CACAAGTTTTTCAGAAACGGATGCTGATCCCCACAATTCCCGACGTGGCTTTTTTCTTCTCT 712
H K F S E T D A D P H N S R R G F F F S
CACGTGGGTGGCTGCTTGTGCGCAAACACCCAGCTGTCAGAGAAAAGGGTGCTACGCTA 772
H V G W L L V R K H P A V R E K G A T L
GATTTATCCGACCTAAGAGCTGAGAAGCTGGTGTATGTTCCAGAGGAGGTACTACAAACCT 832
D L S D L R A E K L V M F Q R R Y Y K P
GGTGTCTGTGCTGTTGTGCTTCATCCTGCCACACTCGTGCCGTGGTATCTATGGGGTGAA 892
G V L L L C F I L P T L V P W Y L W G E
ACGTTTCAAACAGCCTTTTTTTTTTGCCACCCTTCTCCGGTACGCTGTTGTGCTCAATGCC 952
T F Q N S L F F A T L L R Y A V V L N A
ACCTGGCTGGTGAACAGCGCTGCCCATATGTATGGATATCGCCCTTACGACAAGACCATC 1012
T W L V N S A A H M Y G Y R P Y D K T I
AACCCCGAGAGAATATCCTGGTTTTCCCTGGGAGCTGTGGGTGAGGGCTTCCACAACCTAC 1072
N P R E N I L V S L G A V G E G F H N Y
CACCACACCTTTCCTTATGACTACTCTGCCAGCGAGTACCGCTGGCACATCAACTTTACC 1132
H H T F P Y D Y S A S E Y R W H I N F T
ACATTCTTCATTGATTGCATGGCTGCCATCGGTCTGGCTTATGACCGGAAGAAAGTATCC 1192
T F F I D C M A A I G L A Y D R K K V S
AAGGCTGCCGCCTTGGCCAGGATGAAAAGAACTGGAGAGGAAAGCTGCAAGAGTGGCTGA 1252
K A A A L A R M K R T G E E S C K S G

```

GTTTCTGGTCCCTTGGATTCCTTTTCCAAAAGCCAGCTGGGCAGAGGTTTAATGTTCTGT 1312
TTATAAACTACTGAATAATGCTACCAGGATGCTAAAGATGACGTTAACCCATCACAGTAC 1372
AGTATTCTTTAAAATTTTCTTTTTTAAATTGAAAGCCAACAACCTCTGCCTTTATGATGCTA 1432
AGCTCATGTTCTTATTTCTTCTTATCTTCTTTCTCTTCTGTTCCCATTATCTTTCCCT 1492
TTGTTTTGTCCCTATCACCTTCCTTTCTCTTTCTCCTCATTGCCCCCAAGGCAAGCAGTT 1552
GGTCAGTCACTGGTGGGTTTCCAGCTTCCAAAGCCTAGACGACCCTTCTGTAGTCTAAAA 1612
CAAGTGGTCTTTGCCCCGGCTGACCCTTTCCTTGAGCTGTCTGAGCTTTAAGGTGGATGG 1672
CCCGAGCTAGATATATGACAGAATCTTCTGGGAAGGGCCCTGATGATCTTCGGCCCAGGC 1732
TTTTGCTAGATGAAATGGAAAAATAACTTTATTTTGGCACCAAACCTGCAAAAACAGGTAA 1792
ATTGTCAGGGGAGAGAGTCAGCATGCATGATGTGATTGATAAATAGGGTGAGTTGAAGTG 1852
GGAAACAAGGCAGGAAGCTGCTGCTGTGATCAGACACCCTGTCTGCCCATCAC 1906

Multiple alignment of the seven caprine kappa casein variants (exon 4)

The nucleotide sequences of the different variants were deposited to GenBank under the accession numbers AF485339,

AF485340, AF485341, AY090465, AF486523, AY090466, and AY090467

