

Development of a Single Nucleotide Polymorphism Genotyping Microarray Platform for the Identification of Bovine Milk Protein Genetic Polymorphisms

S. Chessa,*¹ F. Chiatti,* G. Ceriotti,* A. Caroli,† G. Consolandi,‡ G. Pagnacco,* and B. Castiglioni§

*Dipartimento di Scienze e Tecnologie Veterinarie per la Sicurezza Alimentare, Università degli Studi di Milano, Via Trentacoste 2, Milano 20134 Italy

†Dipartimento di Scienze Biomediche e Biotecnologie, Università degli Studi di Brescia, Viale Europa 11, Brescia 25123 Italy

‡Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, Via Fratelli Cervi 93, Segrate 20090 Italy

§Istituto di Biologia e Biotecnologia Agraria, Consiglio Nazionale delle Ricerche, Via Bassini 15, Milano 20133 Italy

ABSTRACT

The objective of this study was to develop and validate a fast method for typing the main mutations of bovine milk protein genes by using microarray technology. An approach based on the ligation detection reaction (LDR) and a universal array (UA) was used. Polymorphisms in both the coding and noncoding sequences of α_{S1} -casein, β -casein, κ -casein, and β -lactoglobulin genes were considered because of their well-known effects on milk composition and cheese production. A total of 22 polymorphic sites, corresponding to 21 different variants, were included in the diagnostic microarray. First, a multiplex PCR was developed to amplify all the DNA target sequences simultaneously. Second, the LDR-UA assay was implemented. The method was validated by analyzing 100 Italian Friesian DNA samples, which were also genotyped by conventional methods both at the protein level by means of milk isoelectrofocusing and at the molecular level using PCR-RFLP and PCR-single strand conformation polymorphism techniques. The genotypes obtained using the LDR-UA approach were in full agreement with those obtained by the conventional analyses. An important result of the LDR-UA assay was a more accurate genotyping of the different milk protein alleles than was found with conventional typing methods. At the κ -casein gene, in fact, 4 samples were heterozygous (3 reference samples and 1 validation sample) for an allele coding for Thr₁₃₆ and Ala₁₄₈. This variant, which can be considered as the wild type of the genus *Bos*, is not usually identifiable by the conventional typing methods used. The multiplex PCR-LDR-UA approach developed provides for an accurate, inexpensive, and high-throughput assay that does

not exhibit false positive or false negative signals, thus making it highly suitable for animal genotyping.

Key words: milk protein gene, cattle, microarray

INTRODUCTION

More than 95% of the proteins contained in ruminant milk are coded by 6 well-characterized structural genes (Martin et al., 2002), 2 (*LALBA* and *LGB* genes) for the main whey proteins, α -LA and β -LG, and 4 for the CN, α_{S1} -CN, α_{S2} -CN, β -CN, and κ -CN (*CSN1S1*, *CSN1S2*, *CSN2*, and *CSN3*, respectively). The 4 CN genes are tightly linked in a 250-kb cluster (Ferretti et al., 1990; Threadgill and Womack, 1990) on chromosome 6 (Hayes et al., 1993; Popescu et al., 1996). A recent revision of milk protein nomenclature considering only protein polymorphisms (Farrell et al., 2004) includes 8 α_{S1} -CN, 4 α_{S2} -CN, 12 β -CN, 11 κ -CN, 11 β -LG, and 3 α -LA variants within the genus *Bos*.

In addition to the effects of milk protein variants on milk composition and cheese-making properties (reviewed by Grosclaude, 1988; Di Stasio and Mariani, 2000; Martin et al., 2002), statistically significant associations with several milk production traits have also been identified for sites of polymorphism within noncoding regions in the CN complex (for a review see Martin et al., 2002). This is the case of a polymorphism within the short interspersed nucleotide element Bov-A2 in the second intron of *CSN3* described by Damiani et al. (2000b) and of the polymorphisms of the *CSN1S1* promoter described by Prinzenberg et al. (2003). Thus, the CN cluster polymorphism has to be considered as a whole complex in which expression sequence polymorphisms could help explain the productive implications of the different CN loci and the results obtained from the detection of CN gene effects on productive traits at the haplotype level (Braunschweig et al., 2000; Ikonen et al., 2001; Boettcher et al., 2004).

The importance of going deeper into the knowledge on milk protein polymorphisms in cattle is therefore

Received April 12, 2006.

Accepted August 1, 2006.

¹Corresponding author: stefania.chessa@unimi.it

evident. The availability of a fast method that allows the simultaneous typing of a great number of mutations affecting milk protein structure and composition could help researchers as well as breeding associations to identify the genetic milk protein variations.

Microarray technology offers the potential of opening new doors in the study of genome complexity, thanks to the extreme degree of parallelization (Ramsay, 1998). Arrays can be used to study DNA, with the primary aim being the identification and the genotyping of genetic polymorphisms. Hybridization and hybridization plus enzymatic processing are the 2 diverse approaches currently available to discriminate among different alleles when using microarray technology (Hacia, 1999; Kurg et al., 2000; Pastinen et al., 2000). As an example applied to animal characterization, an oligonucleotide microarray based on the arrayed primer extension technique was recently described, allowing genotyping in genes of interest for bovine milk protein biosynthesis (Kaminski et al., 2005).

A different approach, based on a ligation detection reaction (**LDR**) combined with a universal array (**UA**), could be considered. This technique allows for the unequivocal detection of several amplicons in a single experiment (Consolandi et al., 2004). The procedure is based on the discriminative properties of DNA ligase and requires the design of 2 adjacent oligonucleotide probes for each target sequence. One probe, named the "single nucleotide polymorphism (**SNP**)-specific discriminating" probe, is 5'-fluorescently labeled, whereas the other, named the "common" probe, contains a unique, artificial 3'-sequence (**cZipCode**) that is complementary to a "ZipCode" sequence within the UA. As first described by Gerry and coworkers (1999), the UA consists of selected oligonucleotides with similar hybridization characteristics and minimal cross-hybridization. This kind of array is termed "universal" because sequences complementary to the immobilized probes could be appended to any set of specific common probes during target-DNA processing. These ZipCodes or "tag" oligonucleotides can be used to anchor their complementary sequences to a precise array location.

Here we applied this approach to the detection of 22 polymorphic sites (corresponding to 21 different variants) in bovine milk protein genes. A multiplex PCR (**mPCR**) was also developed to amplify the DNA fragments containing these polymorphisms at the same time, with the aim of making a diagnostic genotyping assay available for the simultaneous detection of different polymorphisms in the milk protein genes.

MATERIALS AND METHODS

DNA Reference Samples

A total of 25 DNA samples belonging to different cattle breeds (Borgou, Brown Swiss, Carora, Holstein-

Friesian, Reggiana, Somba, and Sudanese Zebu Peul) were chosen from a DNA collection of animals previously typed for the milk protein polymorphisms and for the *CSN3* intron II Bov-A2 element and used as a reference to develop the microarray. They had been typed for the different variants as described in Table 1. The conventional analyses also include the typing method for the *CSN1S1* promoter polymorphisms developed in the present work, described hereafter. The samples were chosen on the basis of milk protein variability in Holstein-Friesian cattle. The following alleles were therefore considered: *CSN1S1*B*, *CSN1S1*C*, *CSN2*A¹*, *CSN2*A²*, *CSN2*A³*, *CSN2*B*, *CSN2*C*, *CSN2*I*, *CSN3*A*, *CSN3*B*, *CSN3*C*, *CSN3*E*, *CSN3*H*, *LGB*A*, *LGB*B*, each one resulting in a corresponding protein variant named by the same letter (Farrell et al., 2004). All these variants were identified in the Holstein-Friesian, except for *CSN3*C* and *CSN3*H*. The first one was found in other taurine breeds, including the Brown Swiss (Formaggioni et al., 1999). The latter was included in the microarray as a marker of *Bos indicus* introgression (Jann et al., 2004). Moreover, the polymorphism within the *CSN3* intron II Bov-A2 element (Damiani et al., 2000b) and the *CSN1S1* promoter mutations (Prinzenberg et al., 2003) were integrated into the diagnostic chip as indicators of the genetic variability in the noncoding regions of milk protein genes. Table 2 shows the genotypes of the DNA reference samples, which are discussed further in the Results section.

CSN1S1 Promoter Analysis

For the *CSN1S1* promoter region, a new PCR-single-strand conformation polymorphism (**SSCP**) protocol was developed as the conventional analysis for comparing the results obtained with the LDR-UA assay. A 712-bp fragment containing part of the bovine 5' region of the *CSN1S1* gene was PCR amplified using the following amplification protocol: 25 μ L of a reaction mixture containing 2 μ L of DNA solution (100 to 150 ng), 10 pmol of primer *CSN1S1* 5' region forward (5'-CCA-CACTAGTACACCCAAAATGA-3') and *CSN1S1* 5' region reverse (5'-GAATGAAAGATGAGACAGAGGAAA-3'), and 1 \times PCR Master Mix (Fermentas, Vilnius, Lithuania) with an initial denaturation step of 94°C for 5 min, followed by 35 cycles of 94°C for 60 s, 56°C for 60 s, 72°C for 60 s, and a final extension step of 72°C for 7 min using PTC-200 DNA (MJ Research Inc., Waltham, MA). Primers were designed on the basis of the bovine genomic *CSN1S1* sequence (GenBank Accession No. X59856) using Primer3 software, available online (Rozen and Skaletsky, 2000).

Table 1. Conventional molecular analyses used to genotype all the DNA samples considered in this study

DNA target	Method	Alleles	Authors
<i>CSN1S1</i> exon 17	PCR-SSCP ¹	<i>B, C</i>	Jann et al. (2002)
<i>CSN2</i> exon 6, 7	PCR-SSCP	<i>A¹, A², A³, B, C, I</i>	Barroso et al. (1999)
<i>CSN3</i> exon 4	PCR-SSCP	<i>A, B, C, E, H</i>	Prinzenberg et al. (1999)
<i>LGB</i> exon 3, 4	PCR-RFLP	<i>A, B</i>	Medrano and Aguilar-Cordova (1990)
<i>CSN1S1</i> promoter	PCR-SSCP	<i>1, 2, 3, 4</i>	Present work
<i>CSN3</i> intron II ²	PCR-RFLP	<i>A, B</i>	Damiani et al. (2000a)

¹PCR-SSCP = PCR-single-strand conformation polymorphism.

²Bov-A2 element.

For the SSCP analysis, 6 µL of PCR product was added to 8 µL of denaturation solution (0.05% of xylene-cyanol, 0.05% of bromophenol blue, 0.02 M EDTA in deionized formamide). After heat denaturation at 95°C for 8 min, the samples were immediately chilled on ice and then run overnight with a Penguin Dual Gel Water-Cooled Electrophoresis System (OWL Scientific Inc., Woburn, MA) on 10% acrylamide:bisacrylamide gels (33:1) with 0.5% glycerol in 0.5× Tris-borate-EDTA buffer at 280 V and 5°C for 16 h. Bands were visualized by silver staining (Bassam et al., 1991).

The method was applied to the 25 reference DNA samples. Among the samples showing different migration patterns, 8 were randomly chosen for sequencing. Primers used for sequencing were the same as those

used for the PCR-SSCP techniques. The PCR products were sequenced by Primm srl (Milano, Italy). The nucleotide sequences were analyzed with Bioedit software (Hall, 1999).

Validation Samples

A total of 100 milk samples of the Italian Friesian breed from 25 Northern Italy herdbook farms were collected for the validation analyses of the LDR-UA assay applied to animal genotyping. As a first step, all milk samples were typed by isoelectrofocusing (IEF; Erhardt et al., 1998). This analysis allows simultaneous screening of the main milk protein variants at the phenotype

Table 2. Breed and genotype of the 25 DNA reference samples

Breed ¹	<i>CSN1S1</i> exon 17	<i>CSN2</i> exon 6, 7	<i>CSN3</i> exon 4	<i>LGB</i> exon 3, 4	<i>CSN1S1</i> promoter	<i>CSN3</i> intron II
SO	BB	A ¹ A ¹	AB	AB	22	BB
SO	BB	A ¹ A ¹	BW ²	BB	11	BB
SO	BB	A ¹ A ¹	BW ²	BB	12	BB
HF	BB	A ¹ A ¹	EE	AB	22	AB
SO	BB	A ¹ A ²	AA	BB	13	BB
SO	BB	A ¹ A ²	AW ²	BB	11	BB
SO	BB	A ¹ A ²	AB	BB	11	BB
BS	BB	A ¹ A ²	AC	BB	12	AB
HF	BB	A ¹ A ²	AE	BB	23	AA
ZP	BB	A ¹ I	AH	BB	23	BB
SO	BB	A ² A ²	AA	AB	33	BB
SO	BB	A ² A ²	AB	AB	12	BB
BS	BB	A ² B	AA	AA	24	AB
BS	BB	A ² B	AB	BB	44	AB
CA	BB	A ² C	AB	AB	12	BB
ZP	BB	A ² I	AA	AB	23	BB
RE	BB	BB	AB	AB	44	BB
RE	BB	BB	AB	AB	44	BB
ZP	BB	BI	AH	BB	24	BB
BO	BC	A ¹ A ²	HH	AA	12	BB
HF	BC	A ² A ³	AA	AA	12	AB
RE	BC	A ² B	AB	AA	24	BB
RE	CC	A ² A ²	BB	AA	11	BB
ZP	CC	A ² A ²	BH	BB	12	BB
BO	CC	A ² A ²	HH	BB	11	BB

¹BO = Borgou; BS = Brown Swiss; CA = Carora; HF = Holstein-Friesian; RE = Reggiana; SO = Somba; ZP = Sudanese Zebu Peul.

²Allele W = ancestral allele detected by means of the ligation detection reaction-universal array assay.

Table 3. Primer sequences, melting temperature (T_m) values, and predicted sizes of PCR products for the simultaneous amplification of the target milk protein regions

Name	Sequence 5' → 3'	DNA target	T_m , °C	Amplicon size, bp
<i>CSN1S1</i> ex17F	TGGATGCCTATCCATCTGGT	<i>CSN1S1</i> exon 17	60.3	333
<i>CSN1S1</i> ex17R	CACTGCTCCACATGTTCTCTG	<i>CSN1S1</i> exon 17	60.3	
<i>CSN2</i> ex6F	CATCAATAAGGTAAAACCCCTCA	<i>CSN2</i> exon 6	59.3	275
<i>CSN2</i> ex6R	TTTGTCAAAGTTTTTATTTCTTGAC	<i>CSN2</i> exon 6	60.0	
<i>CSN2</i> ex7F	TCCAGGATGAACTCCAGGAT	<i>CSN2</i> exon 7	59.5	545
<i>CSN2</i> ex7R	CATCAGAAGTTAAACAGGCACAG	<i>CSN2</i> exon 7	59.0	
<i>CNS3</i> ex4F	TAGGTCACCTGCCAAATTC	<i>CSN3</i> exon 4	60.0	462
<i>CSN3</i> ex4R	ATTAGCCCATTTCGCCTTCT	<i>CSN3</i> exon 4	60.1	
<i>LGB</i> ex3F	TGGCTCCATCTGACTTCTCC	<i>LGB</i> exon 3	60.3	404
<i>LGB</i> ex3R	CACGGCAGTGTCTTCATCAC	<i>LGB</i> exon 3	60.3	
<i>LGB</i> ex4F	GAGTTGGGCTTCCAGAGTGA	<i>LGB</i> exon 4	60.4	494
<i>LGB</i> ex4R	CCGGTATATGACCACCCTCT	<i>LGB</i> exon 4	58.7	
Bov-A2F	AGCTACAAGCCCATGGATA	<i>CSN3</i> intron 2	59.5	423
Bov-A2R	TGCTGCTGCTGCTAAGTCAC	<i>CSN3</i> intron 2	60.5	
<i>CSN1S1</i> PromF	TTTGCAGGAAAAGATTAGACCA	<i>CSN1S1</i> promoter	60.1	518
<i>CSN1S1</i> PromR	GGTATGGCAGACTTTTGCT	<i>CSN1S1</i> promoter	60.3	

level, and it is a useful tool for checking milk protein gene expression.

A commercial kit (GFX Genomic Blood DNA Purification kit, Amersham Biosciences, Piscataway, NJ) was used for the DNA extraction directly from milk. The extraction was performed starting with 300 μ L of fresh milk to recover at least 5 ng/ μ L of genomic DNA at a final volume of 100 μ L.

To test the quality of the DNA extraction from milk, a PCR assay for the *CSN2* fragments was performed using the specific primers described in Table 3 and the amplification protocol described above for the *CSN1S1* promoter region PCR-SSCP method. Samples giving weak or no bands on a 2% agarose gel were amplified using a GenomiPhi DNA Amplification kit (Amersham Biosciences UK Limited, Buckinghamshire, UK) following the instructions provided by the manufacturer. All DNA validation samples were typed by the LDR-UA assay developed in the present work and by the conventional methods described in Table 1.

mPCR Setup

An mPCR was developed to amplify all the DNA target sequences simultaneously. The PCR primers were purchased from Primm srl. For the *CSN1S1* promoter region and exon 17 of *CSN1S1* (allele *B* and *C* discriminating fragment), 2 pairs of primers were designed on the basis of the bovine genomic sequence GenBank Accession No. X59856. For the *CSN2* polymorphic sites considered, the primers were designed on the basis of the bovine genomic sequence GenBank Accession No. X14711 to amplify parts of exon 6 and 7 of the gene. For the fourth exon of *CSN3* and the Bov-A2 element, primers were designed on the basis of the bovine geno-

mic sequence GenBank Accession No. AY380228. Finally, for the *LGB* polymorphic sites considered (exon 3 and exon 4), primers were designed on the basis of the bovine genomic sequence GenBank Accession No. X14710. Primers were designed using Primer3 software (Rozen and Skaletsky, 2000) to ensure similar thermodynamic behavior and to allow amplification of all the fragments in the same reaction. The primer list and the melting temperatures (T_m) are shown in Table 3.

The amplification reaction was performed in a 50- μ L mixture containing 15 μ L of DNA solution (at least 75 ng), 1 \times AccuPrime PCR Buffer II, 1 μ L of AccuPrime *Taq* DNA Polymerase (Invitrogen, Milano, Italy), 10 pmol of primers for *CSN1S1* exon 17, *LGB* exon 3, Bov-A2 element, and *CSN1S1* promoter, and 16 pmol of primers for *CSN2* exon 6 and exon 7, *CSN3* exon 4, and *LGB* exon 4. The following conditions were used for the PCR assay: an initial denaturation step of 94°C for 5 min was followed by 35 cycles of 94°C for 30 s, 56°C for 30 s, 68°C for 60 s, and a final extension step of 68°C for 7 min using a PTC-200 DNA engine thermal cycler (MJ Research Inc., Waltham, MA).

The DNA solution containing the PCR products was then purified by GFX PCR DNA and Gel Band Purification kits (Amersham Pharmacia Biotech Inc., Piscataway, NJ), eluted in 40 μ L of autoclaved water and quantified using a Bioanalyzer 2100, and a DNA 7500 kit for 100- to 7,500-bp dsDNA sizing and quantification (Agilent Technologies, Milano, Italy).

LDR Probe Design

According to the LDR procedure described by Conso-landi et al. (2004), we designed specific probes surrounding the polymorphic sites of interest listed in Ta-

Table 4. List of single nucleotide polymorphisms (SNP) considered in each locus to distinguish the different variants by means of the proposed ligation detection reaction-universal array assay

Locus	SNP position	Nucleotides involved	Alleles	Reference sequence ¹	AA involved
<i>CSN1S1</i>	17,807	A	<i>B</i>	X59856	Glu ₁₉₂
		G	<i>C</i>		Gly ₁₉₂
<i>CSN2</i>	6,690	G	<i>A¹-A²-A³-B-I</i>	X14711	Glu ₃₇
		A	<i>C</i>		Lys ₃₇
	8,101	C	<i>A²-A³-I</i>		Pro ₆₇
		A	<i>A¹-B-C</i>		His ₆₇
	8,178	A	<i>A¹-A²-A³-B-C</i>		Met ₉₃
		C	<i>I</i>		Leu ₉₃
	8,219	C	<i>A¹-A²-B-C-I</i>		His ₁₀₆
		A	<i>A³</i>		Gln ₁₀₆
8,267	C	<i>A¹-A²-A³-C-I</i>	Ser ₁₂₂		
	G	<i>B</i>	Arg ₁₂₂		
<i>CSN3</i>	12,951	G	<i>A-B-E-H</i>	AY380228	Arg ₉₇
		A	<i>C</i>		His ₉₇
	13,065	C	<i>A-B-C-E</i>		Thr ₁₃₅
		T	<i>H</i>		Ile ₁₃₅
	13,068	C	<i>A-C-E-H</i>		Thr ₁₃₆
		T	<i>B</i>		Ile ₁₃₆
	13,104	A	<i>A-C-E-H</i>		Asp ₁₄₈
		C	<i>B</i>		Ala ₁₄₈
	13,124	A	<i>A-B-C-H</i>		Ser ₁₅₅
		G	<i>E</i>		Gly ₁₅₅
13,165	A	<i>A-C-E-H</i>	Ala ₁₆₈		
	G	<i>B</i>	Ala ₁₆₈		
<i>LGB</i>	3,982	TGA	<i>A</i>	X14710	Asp ₆₄
		CGG	<i>B</i>		Gly ₆₄
	5,174	C	<i>A</i>		Asn ₈₈
		T	<i>B</i>		Asn ₈₈
	5,223	G ²	<i>A</i>		Val ₁₀₅
		T	<i>B</i>		Val ₁₀₅
5,261	GGT	<i>A</i>	Val ₁₁₈		
	GGC or TGC ³	<i>B</i>	Ala ₁₁₈		
<i>CSN1S1</i> promoter	289	A	<i>2-3-4</i>	AF549502	—
		C	<i>1</i>		—
	Del ⁴ 395–396	TT	<i>1</i>		—
			<i>2-3-4</i>		—
	Del ⁴ 397–400	GTTT	<i>3-4</i>		—
			<i>1-2</i>		—
	444	A	<i>2-3-4</i>		—
		G	<i>1</i>		—
484	G	<i>1-2-3</i>	—		
	C	<i>4</i>	—		
<i>CSN3</i> intron II ⁵	8,371	A	<i>A</i>	AY380228	—
		G	<i>B</i>		—

¹GenBank Accession Number. The table reports the sequence to which the SNP position refers.

²Reference sequence: X14712.

³Reference sequence: M27732.

⁴Del = Deletion.

⁵Bov-A2 element.

ble 4 to be used for allele identification. Two discriminating probes (allele-specific probes designed to have their 3' position placed just over the polymorphic site) and a common probe were selected for each SNP, except for the 2 deletions 395–396 and 397–400 of the *CSN1S1* promoter, for which 2 common probes were designed (Table 5).

The discriminating probes carried a Cy3 or a Cy5 fluorochrome at their 5' termini, whereas the common

probes carried a phosphate group in the same position. As a positive control, we designed a probe pair (*CSN2*-UNI) perfectly matching the *CSN2* sequence of all samples considered, and this was taken as an internal reference (**IR**) for the normalization process in genotyping calculations. The sequences of the selected probes are shown in Table 5. When a probe sequence contained an ambiguity code, this base was replaced with degenerate bases (Table 5) during oligonucleotide synthesis. Oligo-

Table 5. The discriminating and common probes for the single nucleotide polymorphism (SNP)-specific ligation detection reaction and their corresponding ZipCodes

SNP	5' mod	Discriminating probe	T _m ¹ , °C	Common probe	T _m ² , °C	ZipCode
CSN1S1-17807-A	Cy3	TCTGACATCCCTAAATCCCATTTGGCTCTGAGAACAGTGA	67	AAAGACTACTATGCCACTGTGGTAAATTTTAA ²	67	20
CSN1S1-17807-G	Cy5	TCTGACATCCCTAAATCCCATTTGGCTCTGAGAACAGTGG	68	ATGACTGCATATTG		
CSN2-6690-G	Cy3	TTTTAACAGAAAATTTGAGAAATTTTCAGAGTGAGG	69	AAAGCAGCAAAAACAGAGGTAATTTTGTTCACATA	68	38
CSN2-6690-A	Cy5	TTTTAACAGAAAATTTGAGAAATTTTCAGAGTGAGA	68			
CSN2-8101-C	Cy3	AGTCTCTACTATCCCTCCCTGGCCATCCG ³	67	TAAACAGCCTCCACAAAACATCCCTCCTCTTACTC	67	40
CSN2-8101-A	Cy5	AGTCTCTACTATCCCTCCCTGGCCATCCCA	67			
CSN2-8178-A	Cy3	GTGGTGTGGCCGCTTCCCTCAGCCTGAAGTAA	68	TGGGAGTCTCCAAAATGAAAGGAGGCTATGGCTCCTA	68	41
CSN2-8178-C	Cy5	GTGGTGTGGCCGCTTCCCTCAGCCTGAAGTAC	68			
CSN2-8219-C	Cy3	GTCTCCAAAATGAAAGGAGGCTATGGCTCCTTAAGCAC	68	AAIGAATAATGCCCTTCCCTAAATATCCAGTTGAGCCCTTTT	67	42
CSN2-8219-A	Cy5	GTCTCCAAAATGAAAGGAGGCTATGGCTCCTTAAGCAA	68			
CSN2-8267-C	Cy3	CCTTCCCTAAATATCCAGTTGAGCCCTTACTGAAAGC	67	CAGAGCCTGACTCTCCTACTGATGTTGAAAATCTGCACC	67	44
CSN2-8267-G	Cy5	CCTTCCCTAAATATCCAGTTGAGCCCTTACTGAAAGG	67			
CSN3-12951-G	Cy3	CCAGCCAACTACCATGGGCCAG	68	TCACCCACACCCACATTTATCATTTATGCG	69	1
CSN3-12951-A	Cy5	GCCAGCCAACTACCATGGCCAG	68			
CSN3-13065-C	Cy3	TAGTGGTCCAGCTACAAAGTACACCTAC	68	CACCGAAGCAGTAGAGAGCAGCTGT	67	2
CSN3-13065-T	Cy5	CTAGTGGTCCAGCTACAAAGTACACCTAT	68			
CSN3-13068-C	Cy3	GTGAGCCTACAAAGTACACCTACAC	67	CGAAGCAGTAGAGAGCAGCTGTAGCTA	68	3
CSN3-13068-T	Cy5	TGGTGGCCCTACAAAGTACACCTACAC	68			
CSN3-13104-A	Cy3	GTAGAGAGCAGCTGTAGCTACTCTAGAAAGA	69	TTCTCCAGAAATTTATGAGAGCCACC	68	4
CSN3-13104-C	Cy5	GAGAGAGCAGCTGTAGCTACTCTAGAAAGC	68			
CSN3-13124-A	Cy3	GCTTAGAAGATTTCCAGAAGTTAATTGAGA	68	GCCCCCTGAGATCAACACAGTCC	69	5
CSN3-13124-G	Cy5	CTTAGAAGATTTCCAGAAGTTAATTGAGG	68			
CSN3-13165-A	Cy3	ATCAACACAGTCCAAAGTACTTCAACTGCA	68	GTCTAAAACCTCTAAGGAGACATCAAAAGAGAC	69	6
CSN3-13165-G	Cy5	CAACACAGTCCAAAGTACTTCAACTGCG	68			
LGB-3982-TGA	Cy3	GTGGCTCAGAAAAGCAGTGTCTTTGAGGAGAAATGA	68	TGAGTGTGCTCAGAAAAGATCATTTGCAGAAAAAACCC ⁻²	67	15
LGB-3982-CGG	Cy5	TGGCTCAGAAAAGCAGTGTCTTTGAGGAGAAATGA	68	AAAGATCCC		
LGB-5174-C	Cy3	TCACTTTCTCCCGTCTGATCTCTTCCAGCCTTGAAC	68	GAGAACAAGTCTCTTGTGCTGGACACCCGACTACAAAA ⁻²	68	16
LGB-5174-T	Cy5	TCACTTTCTCCCGTCTGATCTCTTCCAGCCTTGAAT	67	AGTACC		
LGB-5223-T	Cy3	CCTGTGCTGGACACCGACTACAAAAGTACTGTCTCT	68	TCTGCATGGAGAACAGTGTGAGCCCGAGCAAAAAG	68	18
LGB-5223-G	Cy5	CTTGTGCTGGACACCGACTACAAAAGTACTGTCTCT	68			
LGB-5261-GGT	Cy3	AGAACAGTCTGAGCCCGAGCAAAAGCCTGGT	67	CTGCCAGTGCCTGGGTGGGTGCCAAC	67	19
LGB-5261-KGC	Cy5	AGAACAGTCTGAGCCCGAGCAAAAGCCTKGC ³	68			
Bov-A2-8371-A	Cy3	TAGACTGCAGTCCATGGGGTCACTAA	69	GAGTCAGACAGCAGCTGTGTAACTTTTCAAG	69	14
Bov-A2-8371-G	Cy5	GACTGCAGTCCATGGGGTCACTAG	68			
PromCSN1S1-289-A	Cy3	TTGAAGACCCCAATTTTCCCAAGAAATTTCA	68	TTTACAGGTATTGAAATTTTTCAAAAGGTACAAAAGGAAA	68	7
PromCSN1S1-289-C	Cy5	GAAGACCCCAATTTTCCCAAGAAATTTCC	69			
PromCSN1S1-Del 396-396	Cy5	CATAAATCTAGGGTGTGTTGGGTTTGGGTTTGTG	67	TTTGTAAATTTAGAACAAATGCCAATTTCCATTTCCCT	67	8
PromCSN1S1-395-396	Cy3	CCATAAATCTAGGGTGTGTTGGGTTTGGGTTTGTG	67	TTTGTAAATTTAGAACAAATGCCAATTTCCATTTCCCT	66	
PromCSN1S1-Del 397-400	Cy5	CTAGGTTTGTGTTGGGTTTGGGTTTGGGTTTGTG	67	TTAATTTAGAACAAATGCCAATTTCCATTTCCCTGTATA	67	9
PromCSN1S1-397-400	Cy3	ATAAATCTAGGGTGTGTTGGGTTTGGGTTTGGT	67	AATTTAGAACAAATGCCAATTTCCATTTCCCTGTATAAT	67	
PromCSN1S1-444-A	Cy3	CAATGCCAATTCATTTCTGTATAATGAGTCA	69	CTTCTTTGTGTGTAACCTCTCTTATAGAAATTTCTTTG	68	10
PromCSN1S1-444-G	Cy5	ATGCCAATTCATTTCTGTATAATGAGTCCG	68	GAAGTGAACAGAACAAATTTGATTTCCATTTCTGTAG	69	12
PromCSN1S1-484-G	Cy3	GTAAACTCTCTTGAATTTCTGGGAGAG	68	TTTCCAGGCCTTTCTGCTGTACCAGGAGCCCTGTACTC	68	35
PromCSN1S1-484-C	Cy5	GTAAACTCTCTTGAATTTCTGGGAGAG	68			
CSN2-UNI-Cy3	Cy3	GAAAAGCAGTGCCTATCCCAAGAGAGATATGCCCA	68			
CSN2-UNI-Cy5	Cy5	GAAAAGCAGTGCCTATCCCAAGAGAGATATGCCCA	68			

¹T_m = melting temperature.²The nucleotide sequence continues below and is not another sequence.³The sequence contains a degenerate base: I = inosine, K = G or T.

nucleotide probes were purchased, HPLC-purified, and checked by matrix-assisted laser desorption ionization-mass spectrometry by Thermo-Hyaid GmbH (Ulm, Germany).

The specificity of the probe sequences was checked in the GenBank database using BLAST to verify the absence of homology with other sequences. A ZipCode sequence (Consolandi et al., 2004) was randomly assigned to each polymorphic site, and the ZipCode complementary sequence (cZipCode) was attached to the 3'-terminal position of the related common probe. All oligonucleotides had similar thermodynamic features and T_m values of 67 to 69°C.

The possibility of hairpin loop formation in the 24 common probes with cZipCode sequences was checked by using the Oligo Analyzer software available online (Integrated DNA Technologies Inc., 2006).

In addition, we designed a probe pair matching a synthetic template, 50 bp long, (5'-AGCCGCGAACAC CACGATCGACCGCGCGCGCAGCTGCAGCTTGC TCATG-3'), used as a ligation control, and a sequence complementary to ZipCode 66 as a control for the hybridization reaction (5'-GTTACCGCTGGTGCTGCCG CCGGTA-3').

UA Preparation

To prepare a UA, ZipCode sequences were chosen from among those described in the literature (Gerry et al., 1999; Chen et al., 2000). The 5' amino-modified ZipCodes, carrying an additional poly(dA)₁₀ tail at their 5' end, were dissolved in 100 mM phosphate buffer (pH 8.5) at a final concentration of 50 μ M and spotted 4 times onto Codelink slides (Amersham Biosciences Europe GmbH, Milano, Italy) by a pin contact arrayer (Microgrid II Compact, BioRobotics, Cambridge, UK). Printed slides were processed according to the manufacturer's protocol.

Batch quality control was performed by hybridization with 1 μ M 5' Cy3-labeled poly(dT)₁₀ in a solution containing 5 \times SSC and 0.1 mg/mL of salmon sperm DNA at room temperature for 1 h, then washing for 15 min in 2 \times SSC. The fluorescent signal was checked by laser scanning. By using 8 subarrays per slide, associated with a multichamber hybridization system, 8 samples were run in parallel on the same slide.

LDR and Hybridization onto UA

The LDR was performed in a final volume of 20 μ L, containing 20 mM Tris-HCl (pH 7.5), 20 mM KCl, 10 mM MgCl₂, 0.1% NP₄O, 0.01 mM ATP, 1 mM dithiothreitol, 2 pmol of each discriminating oligo, 2 pmol of each common probe, and 5 to 50 fmol of purified PCR

product. The reaction mixture underwent a denaturation step for 2 min at 94°C, and then 1 μ L of 4 U/ μ L of Pfu DNA ligase (Stratagene, La Jolla, CA) was added. The LDR was carried out for 30 cycles of 94°C for 30 s and 65°C for 4 min in a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA). The hybridization solution (65 μ L), consisting of the ligation reaction mix, 10 pmol each of the Cy3- and Cy5-labeled hybridization control probes, 5 \times SSC, and 0.1 mg/mL of salmon sperm DNA was heated at 94°C for 2 min and applied onto the UA using a multisampling hybridization chamber (Press-to-Seal silicone isolator, Schleicher & Schuell, Dassel, Germany). The hybridization reaction was carried out in the dark at 65°C for 2 h, in a temperature-controlled oven. After removal of the chamber, the array was washed in prewarmed 1 \times SSC-0.1% SDS for 15 min at 65°C.

Signal Detection and Statistical Data Analysis

Fluorescent signals were measured with a ScanArray Lite laser scanning system and ScanArray Express Microarray Analysis System 2.1 software (PerkinElmer Life Sciences, Boston, MA) using the green laser for Cy3 dye (λ_{ex} 543 nm/ λ_{em} 570 nm) and the red laser for Cy5 dye (λ_{ex} 633 nm/ λ_{em} 670 nm). Both the laser and the photomultiplier tube gain varied between fluorochromes and experiments. Spot analysis of 5- μ m-resolution TIFF images was carried out by using the ScanArray Express Microarray Analysis System software supplied with the scanner. The quantification method chosen was the "fixed circle." Spot intensities were calculated using the mean intensity option. Data analysis for each experiment was performed by using the following calculations:

- 1) The local background was subtracted from the intensity of each spot.
- 2) Fluorescent intensities (**FI**) of quadruplicate spots were averaged for each channel and the coefficient of variation was calculated.
- 3) The ratio between the average FI of the Cy3 and Cy5 channels of the IR (corresponding to the CSN2-UNI probe) was calculated ($FI_{Cy3}/FI_{Cy5} = R_{IR}$). This value was used for correcting the fluorophore imbalance (and other channel-related biases). Each average FI_{Cy3} for a given ZipCode was divided by R_{IR} to obtain a corrected fluorescent intensity (cFI_{Cy3}).

Genotype Determination

To define the genotype for each SNP, we calculated the allelic fraction (**AF**), which estimates the relative

amount of allele 1 (Cy3-labeled) with respect to allele 2 (Cy5-labeled) by using the following equation:

$$AF = cFI_{Cy3}/(cFI_{Cy3} + FI_{Cy5}),$$

as described by Liljedahl et al. (2003). The AF values of each SNP were plotted simultaneously to analyze their distribution as a function of the 3 possible genotypes (11, 12, and 22). A regression analysis of the AF values on the number of 2 alleles in the genotype was also performed. A specific program was developed in SAS software (SAS Institute, Inc., Cary, NC) to reconstruct the different alleles on the basis of the SNP analyzed (Table 4).

RESULTS

The objective of the study was to use microarray technology to develop and validate a fast method for genotyping the main mutations of bovine milk proteins in both the coding and noncoding sequences. To validate the method, the samples were analyzed by conventional DNA approaches.

Milk was chosen for the DNA extraction because it is particularly suitable for the routine genetic analyses that could be performed by means of the LDR-UA assay. In fact, milk is more suitable than blood or hair in dairy cattle, where a screening of milk protein loci in the female population could be carried out with the DNA chip developed. Considering that the DNA extracted from milk for validation analyses is sometimes qualitatively poor (20% of the samples), giving low or no product in the PCR quality test, a method was found to use this DNA. In fact, the whole genome amplification procedure described and applied to these samples gave an amount of DNA quantitatively and qualitatively sufficient to perform mPCR amplification (data not shown).

Development of the *CSN1S1* Promoter SSCP Analysis

For the *CSN1S1* promoter region, an SSCP analysis was developed as a fast and reliable alternative to the microarray analysis for typing this region and for testing a number of samples sufficient to find the references samples for all the alleles included in the microarray analysis. The SSCP analysis developed revealed 4 different allelic patterns. The sequencing confirmed that they corresponded to the 4 alleles described by Prinzenberg et al. (2003) and were named 1, 2, 3, and 4 according to the authors and the sequence analysis. The migration patterns of alleles 2 and 3 were quite

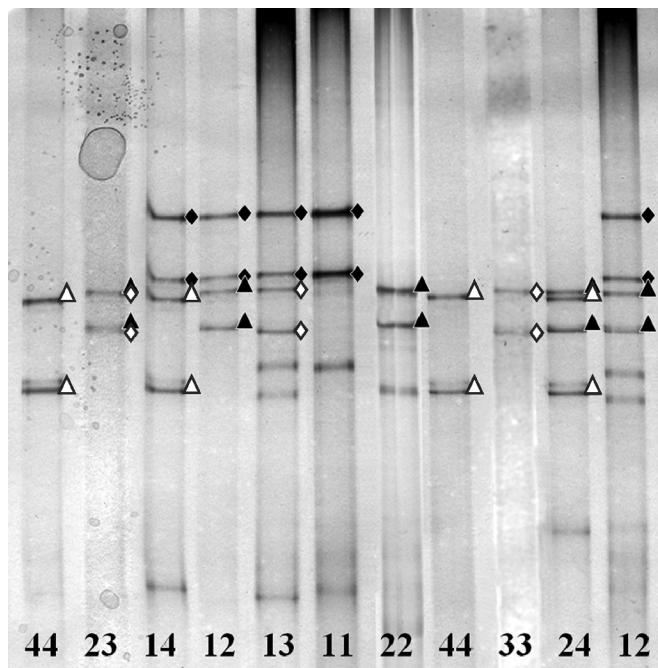


Figure 1. Polymerase chain reaction-single strand conformation polymorphism typing of the *CSN1S1* promoter region for the simultaneous detection of alleles 1, 2, 3, and 4. Genotypes are indicated for each sample. Two bands for each allele are also shown (solid diamond: allele 1; solid triangle: allele 2; open diamond: allele 3; open triangle: allele 4). In sample 23, the overlapping of the slower bands and the small difference between the faster bands of the 2 alleles is shown.

similar (Figure 1), so in some cases it was necessary to sequence the samples to be sure of the genotype.

Development of the mPCR

The mPCR described here permitted the simultaneous amplification of 8 different DNA fragments: a 275-bp fragment of the *CSN2* sixth exon, a 333-bp fragment of the *CSN1S1* 17th exon, a 404-bp fragment of the *LGB* third exon, a 423-bp fragment of the second intron of *CSN3* containing the Bov-A2 element, a 462-bp fragment of the *CSN3* fourth exon, a 494-bp fragment of the *LGB* fourth exon, a 518-bp fragment of the *CSN1S1* promoter region, and a 545-bp fragment of the *CSN2* seventh exon. The amount of PCR product varied among experiments and fragments but was always between 5 and 50 fmol, and was therefore sufficient for the subsequent analyses. Examples of the results obtained are shown in Figure 2.

LDR, Hybridization onto the UA, and Genotyping

Some LDR-UA results are shown in Figure 3. With regard to the efficiency and the behavior of all the probes, the SNP determinations performed by the

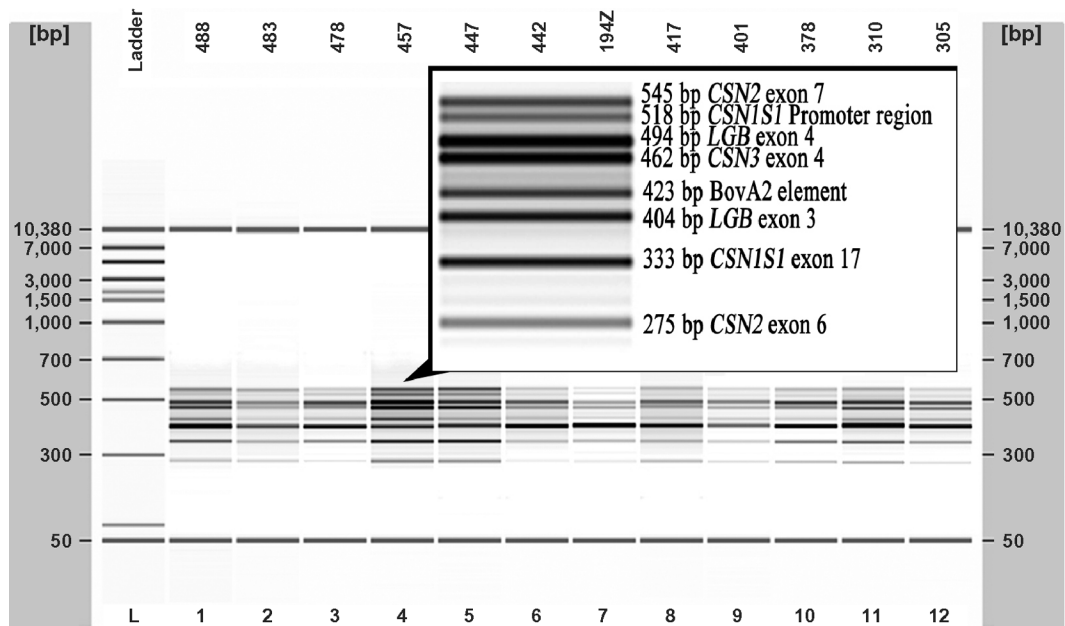


Figure 2. Example of multiplex PCR amplification of DNA from 12 different animals. The separation of each DNA fragment was visualized using a Bioanalyzer 2100 (Agilent Technologies, Milano, Italy).

mPCR-LDR-UA approach were in agreement with genotyping carried out by protein IEF, conventional DNA molecular methods, and direct sequencing, except for some detection limits of IEF and PCR-SSCP, summarized in Table 6, showing the allele frequencies in the validation samples.

For *CSN3*, the mPCR-LDR-UA approach allowed us to detect, in both the reference samples (Table 2) and in the validation samples (Table 6), the occurrence of an allele (here named *W*) coding for Thr₁₃₆ and Ala₁₄₈. This allele was not detectable by either IEF or PCR-SSCP. In addition to *CSN3*W* and the polymorphisms in the noncoding sequences, IEF cannot detect *CSN2*I* comigrating with *CSN2*A*², and *CSN3*H* comigrating with *CSN3*A*. The latter allele was not detected in the Italian Holstein samples analyzed.

For the *CSN1S1* promoter, the sometimes ambiguous discrimination of alleles 2 and 3 by PCR-SSCP ideally had to be confirmed by sequencing. This typing difficulty was overcome by using the DNA chip developed in this study.

The differences in performance of the entire process consisting of mPCR-LDR-UA were normalized by adjustments of the fluorescent signal intensity of each spot relative to the signal obtained from the IC, corresponding to a region of the *CSN2* gene and associated with ZipCode 35. Because of our probe design strategy, this region was heterozygous in all the samples analyzed.

Afterward, the AF values of each SNP analyzed were plotted simultaneously as a function of the 3 possible genotypes, as shown in Figure 4. We obtained a very well-defined distribution corresponding to the 3 possible genotypes: AF ranging from 0.87 to 1.00 for the homozygotes for allele 1, from 0.01 to 0.21 for the homozygotes for allele 2, and 0.36 to 0.65 for heterozygotes. All the typed SNP fell within one of the 3 distinct non-overlapping areas, defining the 3 possible genotypes unequivocally. The regression analysis was highly significant ($P < 0.0001$, $R^2 = 0.998$), indicating the particular robustness of the genotype classification of the basis of the AF values obtained.

DISCUSSION

In the work described herein, we applied the LDR combined with a UA platform for the simultaneous identification of 22 single base modifications situated in 8 different regions of milk protein genes. For SNP genotyping, many microarray techniques are available (Hacia, 1999; Kurg et al., 2000; Pastinen et al., 2000). Using the LDR-UA assay, we overcame one of the major limitations of DNA microarray approaches based on differential hybridization with allele-specific oligonucleotide probes (Hacia, 1999). Optimal hybridization conditions are difficult to determine for large sets of different probes, which need to be hybridized on a DNA chip at the same time. Other array-based SNP genotyp-

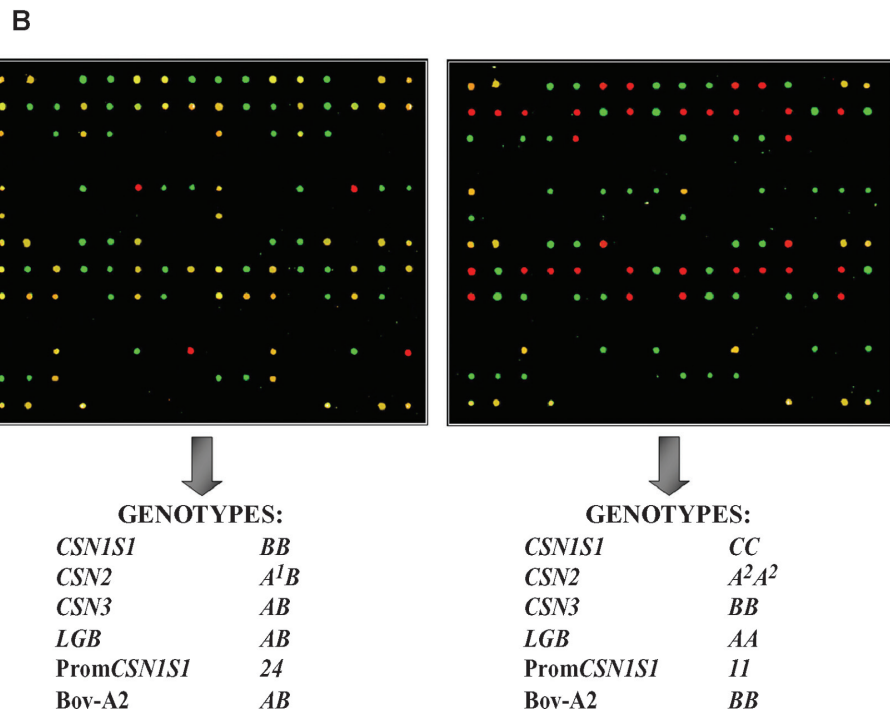
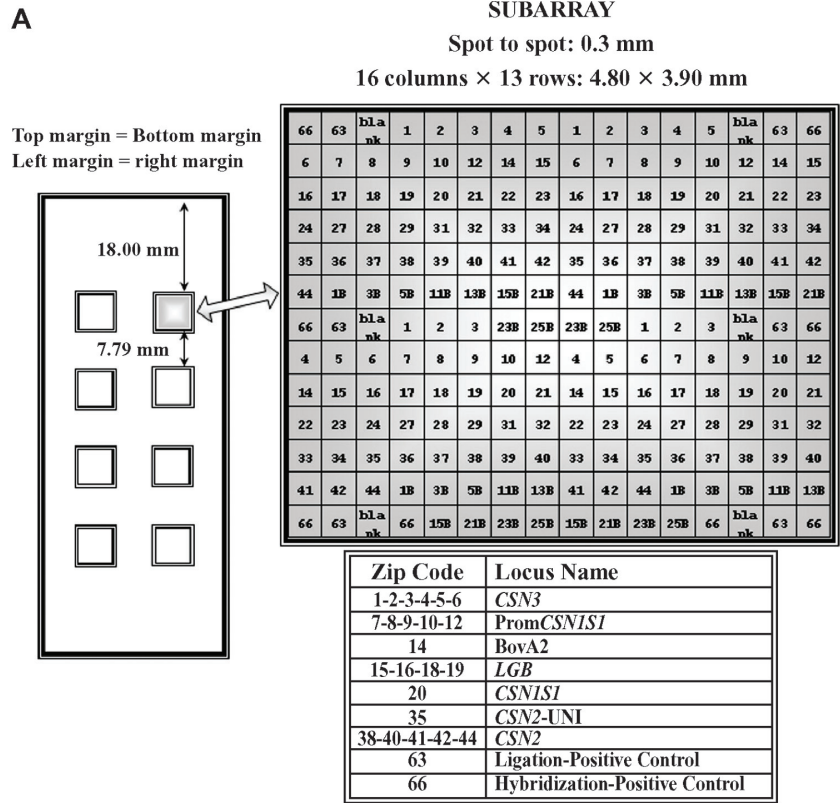


Figure 3. Some results of the ligation detection reaction-universal array (LDR-UA) approach. (A) Universal array deposition scheme. Among all ZipCodes spotted, 23 corresponded to the single nucleotide polymorphisms selected. ZipCodes 63 and 66 corresponded to the ligation and hybridization control, respectively. (B) Examples of LDR-UA results: On the left is a validation sample; on the right is a reference sample. The composite images were achieved by overlapping Cy3 fluorescent signals (green palette) and Cy5 fluorescent signals (red palette). Green and red spots represent the 2 homozygous conditions. Yellow spots are the combination of Cy3 and Cy5 dyes, corresponding to heterozygous sites.

Table 6. Allele frequencies of the sample of Italian Friesian cattle (n = 100) obtained from the genotyping by ligation detection reaction-universal array assay (LDR-UA), isoelectrofocusing (IEF), and conventional molecular methods (CM)

DNA target	Alleles	LDR-UA	IEF	CM
<i>CSN1S1</i> exon 17	<i>B</i>	0.995	0.995	0.995
	<i>C</i>	0.005	0.005	0.005
<i>CSN2</i> exon 6, 7	<i>A</i> ¹	0.310	0.310	0.310
	<i>A</i> ²	0.635	0.645 ¹	0.635
	<i>A</i> ³	0.005	0.005	0.005
	<i>B</i>	0.040	0.040	0.040
	<i>C</i>	0.000	0.000	0.000
	<i>I</i>	0.010	ND	0.010
<i>CSN3</i> exon 4	<i>A</i>	0.710	0.710	0.710
	<i>B</i>	0.185	0.190 ²	0.190 ²
	<i>C</i>	0.000	0.000	0.000
	<i>E</i>	0.100	0.100	0.100
	<i>H</i>	0.000	ND	0.000
	<i>W</i>	0.005	ND	ND
<i>LGB</i> exon 3, 4	<i>A</i>	0.470	0.470	0.470
	<i>B</i>	0.530	0.530	0.530
<i>CSN1S1</i> promoter	<i>1</i>	0.070	—	0.070
	<i>2</i>	0.695	—	0.695 ³
	<i>3</i>	0.185	—	0.185 ³
	<i>4</i>	0.050	—	0.050
<i>CSN3</i> intron II ⁴	<i>A</i>	0.695	—	0.695
	<i>B</i>	0.305	—	0.305

¹*CSN2***A*² and *CSN2***I* are not differentiable (ND) by IEF.

²The ancestral *CSN3* allele (*W*) is not differentiable from *CSN3***B* by PCR-single-strand conformation polymorphism (PCR-SSCP) and IEF. A dash (—) indicates nonevaluable by IEF.

³Alleles 2 and 3 of the *CSN1S1* promoter are not clearly separated by PCR-SSCP and require much expertise to be discriminated.

⁴Bov-A2 element.

ing methods exist, such as single base primer extension on the microarray (Pastinen et al., 2000) and arrayed primer extension technologies (Kurg et al., 2000; Kaminski et al., 2005), but optimal probes must be designed for each SNP and multiplexing may be difficult or impossible. In the UA-based approach, the optimization of hybridization conditions for each probe set is not required. During the hybridization reaction, the cZipCode sequences appended to the common probes hybridize with the complementary ZipCodes of the UA. Inasmuch as the ZipCodes were already designed for this purpose, it is unnecessary to set new hybridization conditions, and new probe pairs can be added to the array without further optimization, thus reducing costs and setup time.

Our approach provides for an accurate, inexpensive, and high-throughput assay that does not exhibit false positive or false negative signals, thus making it highly suitable for animal genotyping. All the LDR probes selected were designed to have high theoretical T_m to perform the ligation reaction at 65°C, which prevented problems caused by secondary structures of the target DNA. In the LDR, the specificity of the hybridization

probes and the selectivity of the ligation reaction are combined to increase discrimination power. Furthermore, by means of the LDR it is possible to target several PCR amplicons at the same time in a single ligation reaction. The development of an mPCR reaction for the simultaneous amplification of 8 different fragments allows for increasing throughput and lowers time and costs. The good quality of the mPCR products, in terms of length and concentration, was taken as a guarantee of successful LDR.

The whole procedure, from DNA extraction to signal analysis, required almost 10 h for the typing of 24 samples. Because the number of animals assayed depends on how many arrays can be examined, we increased this number by means of “arrays of arrays.” The use of 8 subarrays per slide format associated with a multi-chamber hybridization system allowed the simultaneous genotyping of 8 different samples at once, either homozygous or heterozygous for each SNP analyzed, at minimized reagent costs (Figure 3A).

As pioneered by Gerry et al. (1999) and Favis et al. (2000), our platform also allowed insertion and deletion polymorphisms to be genotyped. For the *CSN1S1* promoter, this approach was applied by designing 2 common probes, differing at their 3'-nucleotide, to detect 2 deletions at the *CSN1S1* promoter. An innovative result obtained in the present work was the successful use of 2 different fluorochromes for the deletion genotyping.

To test the feasibility of using LDR-UA as a genotyping assay to discriminate milk protein genes, a set of 100 animals was successfully analyzed. An important result of our LDR-UA genotyping was a more accurate genotyping of the different milk protein alleles. In the validation samples, new findings were obtained from the analysis of *LGB* and *CSN3*. For *LGB*, the G nucleotide at gene position 5,223 (Table 4) was never detected in the samples, suggesting that this synonymous SNP may be a sequencing mistake or a rare mutation not affecting the *LGB* typing.

The 2 main *CSN3* alleles in *Bos taurus* and *B. indicus* are *CSN3***A* (Thr₁₃₆-Asp₁₄₈) and *CSN3***B* (Ile₁₃₆-Ala₁₄₈; Farrell et al., 2004). The LDR-UA genotyping allows the identification of both SNP associated with these main κ -CN variants, thus providing additional information on the intragenic *CSN3* haplotype, which could not be evaluated by the other genotyping methods used. In fact, at *CSN3* 4 samples were found to be heterozygous for an allele coding for Thr₁₃₆ and Ala₁₄₈. This variant can be considered the wild type of the genus *Bos*, as first suggested by Addeo and Mercier (1977), who found the Italian water buffalo (*Bubalus arnee*) monomorphic for this intragenic haplotype, although until now, it has not been detected in *B. taurus* and *B. indicus*. In the most recent nomenclature of milk proteins, a variant

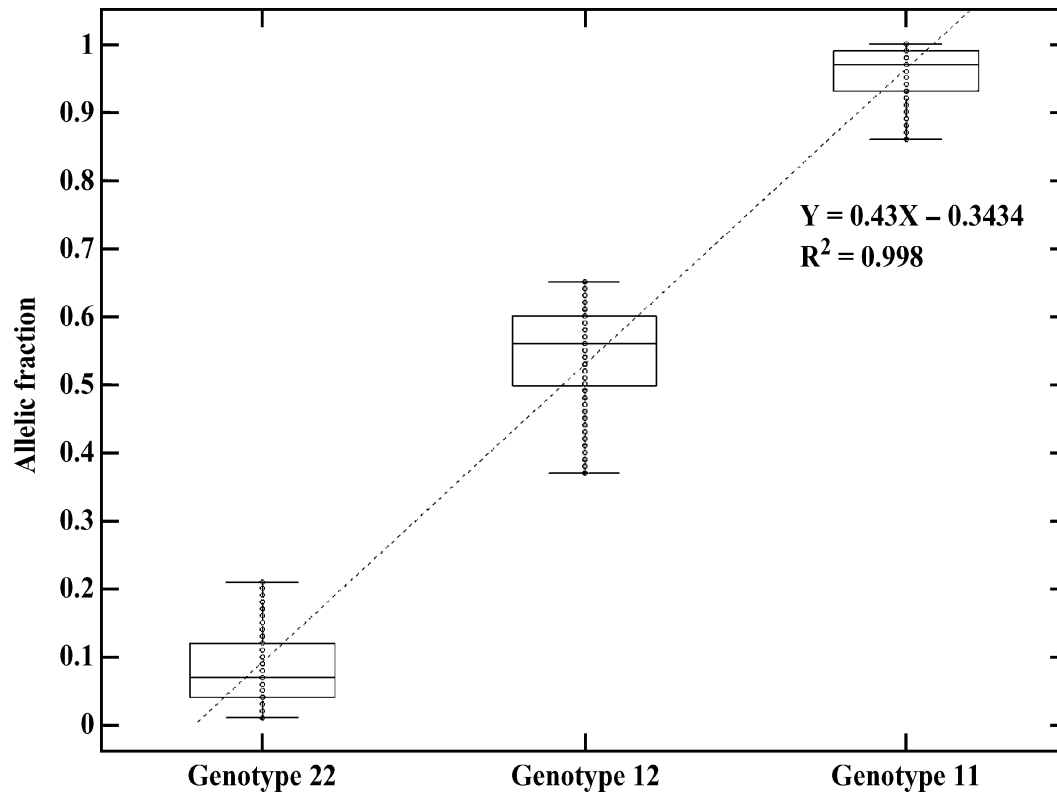


Figure 4. Boxplot of the allelic fraction (AF) values obtained for the different single nucleotide polymorphisms by means of the ligation detection reaction-universal array assay. Three areas were well defined as a function of the combination of the 2 alleles in the genotype: genotype 11, genotype 12, and genotype 22. Boxes represent the median (solid black horizontal line) and interquartile range. Whiskers correspond to the AF extreme values (minimum or maximum). Values were interpolated by a linear regression of the AF values on the number of alleles of one type in the genotype.

identified in the yak (*Bos grunniensis*) coding for Thr₁₃₆ and Ala₁₄₈ was renamed as G² (Farrell et al., 2004). This nomenclature includes milk protein variants from different *Bos* species, thus generating possible confusion from the phylogenetic point of view. The ancestral variant could usefully be defined as *W* (wild). Among the 4 samples in our work carrying this allele, 2 of the 2 *BW* were found in the Somba breed within the reference samples and the third was found in the Italian Friesian validation samples, whereas for the *AW*, one was found again within the Somba reference samples. Thus, the ancestral allele Thr₁₃₆-Ala₁₄₈ was found in the Italian Friesian sample analyzed at a low frequency (1/200 = 0.005), but its occurrence should be taken into account to have a more complete picture of the genetic variation at κ -CN as well as at the entire CN complex.

This finding in the Italian Friesian sample indicates the potential of our microarray-based assay to detect intragenic haplotypes not identifiable using conventional molecular techniques. This potential is of great interest for genetic analyses. The informative content of a locus is increased by the discovery of such haplotypes,

providing additional elements for biodiversity studies and phylogenetic analyses.

In conclusion, we demonstrated the great potential of the mPCR-LDR-UA method for rapid and reliable identification of milk protein polymorphisms in cattle. The method is currently under Italian Patent (Castiglioni et al., 2006). The next step in our mPCR-LDR-UA approach applied to bovine milk protein polymorphisms will be to include further mutations in the microarray that are important for other breeds. For Italian Brown typing, for example, the *CSN1S1*G* allele needs to be considered for its important expression effect (Rando et al., 1998), as does the *CSN1S1*F* allele, which was found in the same breed within the haplotype *CNS1S1*F-CSN2*A²-CSN3*B* (Boettcher et al., 2004). Other important noncoding milk protein polymorphisms could also be added to the chip, such as *LALBA* (Bleck and Bremel, 1993), *LGB* (Wagner et al., 1994; Ehrmann et al., 1997; Lum et al., 1997; Kaminski and Zabolwicz, 2000; Kuss et al., 2003), and CN genes (Schild and Geldermann, 1996; Szymanowska et al., 2003; Szymanowska, et al., 2004). As described in Con-

solandi et al. (2004), up to 54 different discriminating probes can be used in such a multiplex ligation reaction without affecting the efficiency of the method. The full potential of microarrays implying a large set of analysis to be performed in parallel could be exploited in detecting several SNP at once.

This technology offers the power of a microarray format, facilitating the genotyping of different genomic regions, which could be investigated for phylogenetic and association studies, for the evaluation of genetic distances among cattle breeds and the domestication history of bovine species, and for dairy cattle identification and parentage control. In selection, the chip could also be used to analyze CN haplotypes instead of single loci, which would help to exploit the effects of the entire CN cluster on milk yield and component traits. Interrogation of known SNP, compared with the most popular DNA markers (microsatellites), is attractive because they are genetically stable and amenable to high throughput with automated technology. They are considered a realistic alternative in livestock identification, kinship analysis, and linkage studies. The mPCR-LDR-UA technology could become a powerful tool for their detection and exploitation in animal breeding.

ACKNOWLEDGMENT

We thank PRIN 2003 (2003078003_004) for partial financial support.

REFERENCES

- Addeo, F., and J. C. Mercier. 1977. Primary structure of the casein macropeptide of κ -casein of buffalo. *Biochimie* 59:375–379.
- Barroso, A., S. Dunner, and J. Canon. 1999. A multiplex PCR-SSCP test to genotype bovine β -casein alleles A1, A2, A3, β , and C. *Anim. Genet.* 30:322–323.
- Bassam, B. J., G. Caetano-Anollés, and P. M. Gresshoff. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* 196:80–83.
- Bleck, G. T., and R. D. Bremel. 1993. Sequence and single base polymorphisms of the bovine α -lactalbumin 5' flanking region. *Gene* 126:213–218.
- Boettcher, P. J., A. Caroli, A. Stella, S. Chessa, E. Budelli, F. Canavesi, S. Ghiroldi, and G. Pagnacco. 2004. Effects of casein haplotypes on production traits in Italian Holstein and Brown Swiss Cattle. *J. Dairy Sci.* 87:4311–4317.
- Braunschweig, M., C. Hagger, G. Stranzinger, and Z. Puhán. 2000. Associations between casein haplotypes and milk production traits of Swiss Brown cattle. *J. Dairy Sci.* 83:1387–1395.
- Castiglioni, B., S. Chessa, and G. Pagnacco. 2006. Metodo per la determinazione dei polimorfismi lattoproteici nei bovini. Italian Pat. no. MI2006A000529.
- Chen, J., M. A. Iannone, M. S. Li, J. D. Taylor, P. Rivers, A. J. Nelsen, K. A. Slentz-Kesler, A. Roses, and M. P. Weiner. 2000. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res.* 10:549–557.
- Consolandi, C., A. Frosini, C. Pera, G. B. Ferrara, R. Bordoni, B. Castiglioni, E. Rizzi, A. Mezzelani, L. R. Bernardi, G. De Bellis, and C. Battaglia. 2004. Polymorphism analysis within the HLA-A locus by universal oligonucleotide array. *Hum. Mutat.* 24:428–434.
- Damiani, G., E. Budelli, S. Florio, A. Caroli, and G. Pagnacco. 2000a. Polymorphism of κ -casein SINE Bov-A2 and CYP21-hydroxylase in some bovine breeds. *Zoot. Nutriz. Anim.* 3:145–148.
- Damiani, G., S. Florio, E. Budelli, P. Bolla, and A. Caroli. 2000b. Single nucleotide polymorphisms (SNPs) within Bov-A2 SINE in the second intron of bovine and buffalo κ -casein (CSN3) gene. *Anim. Genet.* 31:277–279.
- Di Stasio, L., and P. Mariani. 2000. The role of protein polymorphism in the genetic improvement of milk production. *Zoot. Nutr. Anim.* 26:69–90.
- Ehrmann, S., H. Bartenschlager, and H. Geldermann. 1997. Polymorphism in the 5' flanking region of the bovine β -lactoglobulin-encoding gene and its association with β -lactoglobulin in the milk. *J. Anim. Breed. Genet.* 114:49–53.
- Erhardt, G., J. Juszczyk, L. Panicke, and H. Krick-Saleck. 1998. Genetic polymorphism of milk proteins in Polish Red Cattle: A new genetic variant of β -lactoglobulin. *J. Anim. Breed. Genet.* 115:63–71.
- Farrell, H. M., Jr., R. Jimenez-Flores, G. T. Bleck, E. M. Brown, J. E. Butler, L. K. Creamer, C. L. Hicks, C. M. Hollar, F. Ng-Kwai-Hang, and H. E. Swaisgood. 2004. Nomenclature of the proteins of cows' milk—Sixth edition. *J. Dairy Sci.* 87:1641–1674.
- Favis, R., J. P. Day, N. P. Gerry, C. Phelan, S. Narod, and F. Barany. 2000. Universal DNA array detection of small insertions and deletions in *BRCA1* and *BRCA2*. *Nat. Biotechnol.* 18:561–564.
- Ferretti, L., P. Leone, and V. Sgaramella. 1990. Long range restriction analysis of the bovine casein genes. *Nucleic Acids Res.* 18:6829–6833.
- Formaggioni, P., A. Summer, M. Malacarne, and P. Mariani. 1999. Milk protein polymorphism: Detection and diffusion of the genetic variants in *Bos* genus. *Ann. Fac. Med. Vet. Un. Parma* 127–165.
- Gerry, P. N., N. E. Witowsky, J. Day, R. P. Hammer, G. Barany, and F. Barany. 1999. Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.* 292:251–262.
- Grosclaude, F. 1988. Le polymorphisme génétique des principales lactoprotéines bovines. Relations avec la quantité, la composition et les aptitudes fromagères du lait. *INRA Prod. Anim.* 1:5–17.
- Hacia, J. G. 1999. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat. Genet. (Suppl. 21):*42–47.
- Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nuc. Acids. Symp. Ser.* 41:95–98.
- Hayes, H., E. Petit, C. Bouniol, and P. Popescu. 1993. Localisation of the α_{S2} -casein gene (CASA2) to the homoeologous cattle, sheep, and goat chromosomes 4 by in situ hybridization. *Cytogenet. Cell Genet.* 64:281–285.
- Ikonen, T., H. Bovenhuis, M. Ojala, O. Ruottinen, and M. Georges. 2001. Associations between casein haplotypes and first lactation milk production traits in Finnish Ayrshire cows. *J. Dairy Sci.* 84:507–514.
- Integrated DNA Technologies. 2006. Oligo Analyzer software. Integrated DNA Technologies Inc., 1710 Commercial Park, Coralville, IA 52241. <http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx> Accessed Oct. 20, 2006.
- Jann, O., E. Ibeagha-Awemu, C. Özbeyaz, P. Zaragoza, J. L. Williams, P. Ajmone-Marsan, J. A. Lenstra, K. Moazami-Goudarzi, and G. Erhardt. 2004. Geographic distribution of haplotype diversity at the bovine casein locus. *Genet. Sel. Evol.* 36:243–257.
- Jann, O., E.-M. Prinzenberg, H. Brandt, J. L. Williams, P. Ajmone-Marsan, P. Zaragoza, C. Özbeyaz, and G. Erhardt. 2002. Intra-genic haplotypes at the bovine *CSN1S1* locus. *Arch. Tierzucht.* 45:13–21.
- Kaminski, S., A. Ahman, A. Rusc, E. Wojcik, and T. Malewski. 2005. MilkProtChip—A microarray of SNPs in candidate genes associated with milk protein biosynthesis—Development and validation. *J. Appl. Genet.* 46:45–58.
- Kaminski, S., and T. Zabolewicz. 2000. Associations between bovine β -lactoglobulin polymorphism within coding and regulatory sequences and milk performance traits. *J. Appl. Genet.* 41:91–99.

- Kurg, A., N. Tonisson, I. Georgiou, J. Shumaker, J. Tollett, and A. Metspalu. 2000. Arrayed primer extension: Solid-phase 3-color DNA resequencing and mutation detection technology. *Genet. Test.* 4:1–7.
- Kuss, A. W., J. Gogol, and H. Geldermann. 2003. Associations of a polymorphic AP-2 binding site in the 5'-flanking region of the bovine β -lactoglobulin gene with milk proteins. *J. Dairy Sci.* 86:2213–2218.
- Liljedahl, U., J. Karlsson, H. Melhus, L. Kurland, M. Lindersson, T. Kahan, F. Nystrom, L. Lind, and A. C. Syvanen. 2003. A microarray minisequencing system for pharmacogenetic profiling of anti-hypertensive drug response. *Pharmacogenetics* 13:7–17.
- Lum, L. S., P. Dovic, and J. F. Medrano. 1997. Polymorphisms of bovine β -lactoglobulin promoter and differences in the binding affinity of activator protein-2 transcription factor. *J. Dairy Sci.* 80:1389–1397.
- Martin, P., M. Szymanowska, L. Zwierzchowski, and C. Leroux. 2002. The impact of genetic polymorphisms on the protein composition of ruminant milks. *Reprod. Nutr. Dev.* 42:433–459.
- Medrano, J. F., and E. Aguilar-Cordova. 1990. Polymerase chain reaction amplification of bovine β -lactoglobulin genomic sequences and identification of genetic variants by RFLP analysis. *Anim. Biotechnol.* 1:73–77.
- Pastinen, T., M. Raitio, K. Lindroos, P. Tainola, L. Peltonen, and A. C. Syvanen. 2000. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.* 10:1031–1042.
- Popescu, C. P., S. Long, P. Riggs, J. Womack, S. Schmutz, R. Fries, and D. S. Gallagher. 1996. Standardization of cattle karyotype nomenclature: Report of the committee for the standardization of the cattle karyotype. *Cytogenet. Cell Genet.* 74:259–261.
- Prinzenberg, E.-M., I. Krause, and G. Erhardt. 1999. SSCP analysis at the bovine CSN3 locus discriminates six alleles corresponding to known protein variants (A, B, C, E, F, G) and 3 new DNA polymorphisms (H, I, AI). *Anim. Biotechnol.* 10:49–62.
- Prinzenberg, E. M., C. Weimann, H. Brandt, J. Bennewitz, E. Kalm, M. Schwerin, and G. Erhardt. 2003. Polymorphism of the bovine *CSN1S1* promoter: Linkage mapping, intragenic haplotypes, and effect on milk production traits. *J. Dairy Sci.* 86:2696–2705.
- Ramsay, G. 1998. DNA chips: State of the art. *Nat. Biotechnol.* 16:40–44.
- Rando, A., P. Di Gregorio, L. Ramunno, P. Mariani, A. Fiorella, C. Senese, D. Marletta, and P. Masina. 1998. Characterization of the *CSN1AG* allele of the bovine α_{S1} -casein locus by the insertion of a relict of a long interspersed element. *J. Dairy Sci.* 81:1735–1742.
- Rozen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. Pages 365–386 in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. S. Krawetz and S. Misener, ed. Humana Press, Totowa, NJ. Whitehead Institute for Biomedical Research, <http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi> Accessed Oct. 20, 2006.
- Schild, T. A., and H. Geldermann. 1996. Variants within the 5'-flanking regions of bovine milk-protein encoding genes. III. Genes encoding the Ca-sensitive casein α_{S1} , α_{S2} and β . *Theor. Appl. Genet.* 93:887–893.
- Szymanowska, M., T. Malewski, and L. Zwierzchowski. 2004. Transcription factor binding to variable nucleotide sequences in 5' flanking regions of bovine casein genes. *Int. Dairy J.* 14:103–115.
- Szymanowska, M., N. Strzalkowska, E. Siadkowska, J. Krzyzewski, Z. Ryniewicz, and L. Zwierzchowski. 2003. Effects of polymorphism at 5'-noncoding regions (promoters) of α_{S1} - and α_{S2} -casein genes on selected milk production traits in Polish Black-and-White cows. *Anim. Sci. Pap. Rep.* 21:97–108.
- Threadgill, D. W., and J. E. Womack. 1990. Genomic analysis of the major bovine milk protein genes. *Nucleic Acids Res.* 18:6935–6942.
- Wagner, V. A., T. A. Schild, and H. Geldermann. 1994. DNA variants within the 5'-flanking region of milk-protein-encoding genes. II. The β -lactoglobulin-encoding gene. *Theor. Appl. Genet.* 89:121–126.