

## Analysis of the Sheep MUC1 Gene: Structure of the Repetitive Region and Polymorphism

R. Rasero,<sup>\*1</sup> L. Bianchi,<sup>†</sup> E. Cauvin,<sup>\*</sup> S. Maione,<sup>\*</sup> S. Sartore,<sup>\*</sup> D. Soglia,<sup>\*</sup> and P. Sacchi<sup>\*</sup>

<sup>\*</sup>Dipartimento di Produzioni Animali, Epidemiologia ed Ecologia, University of Torino, Grugliasco 10095, Italy

<sup>†</sup>Dipartimento di Biologia Vegetale e Biotecnologie Agroambientali e Zootecniche, University of Perugia, Perugia 06121, Italy

### ABSTRACT

An investigation was undertaken with the aim of studying the repetitive region of the MUC1 gene and analyzing its polymorphisms in some Italian sheep breeds. Two primers previously used for the goat MUC1 gene analyses allowed for the amplification of 4 different alleles. The sequence analysis showed that the repetitive region of the sheep MUC1 gene is an array of 60-bp repeats, in accordance with the information reported in humans, cattle, and goats. The polypeptide sequence encoded by the consensus repeat was very similar to the corresponding sequences of goats and cattle. The average homology of all repeated units was 82%; when the repeats were compared with the derived consensus repeat, homology dropped to 78%. The repeats were not all perfectly conserved, but the sequence homology was nevertheless clearly sufficient to preserve the mechanism giving rise to the variable-number tandem-repeat polymorphism. In spite of their reduced sequence homology, the sheep repeats shared a high number of potential glycosylation sites. The conservation of the exact number and position of glycosylation sites did not seem to be very important for the purpose of functional integrity, but glycosylation appeared to be conserved as a bulk property. Analysis of the polymorphism in 6 Italian breeds showed that the sheep repetitive region seemed to be less variable and smaller in size than the repetitive region of the goat. The findings of this study suggest that ruminants can be a useful model to study the mechanisms by which the variation in the repeat number and the extracellular domain size can modulate the effectiveness of MUC1 as a cell-surface shield.

**Key words:** sheep, milk, MUC1, polymorphism

### INTRODUCTION

The MUC1 gene codes for the protein backbone of a mucin-type glycoprotein present on the apical surface

of the secreting mammary cell during lactation and on the milk fat globule membrane (MFGM; Mather, 2000). All mucins share some features: They contain large domains characterized by the presence of repetitive regions rich in sites for O-glycosylation, the so-called Ser-Thr-Pro-rich domains (Julenius et al., 2004). In particular, MUC1 contains a repetitive region showing, in humans, cattle, and goats, a 20-AA motif that is tandemly repeated a variable number of times, giving rise to a variable-number tandem-repeat (VNTR) polymorphism (Campana et al., 1992; Patton et al., 1995).

In humans, the MUC1 gene maps to 1q21 (Middleton-Price et al., 1988); in domestic ruminants, the MUC1 gene has been FISH-mapped to BTA3q13 (*Bos taurus*), BBU6q13 (*Bubalus bubalis*), OAR1p13 (*Ovis aries*), and CHI3q13 (*Capra hircus*; Perucatti et al., 2006).

The MUC1 gene is extensively studied as a target for the antigen-specific immune therapy of cancer (Gendler, 2001). Important roles of MUC1 in protecting epithelia and in preventing or reducing infectious diseases have been proposed (Patton, 1999; Lillehoj et al., 2001). Moreover, the complex structure of the carbohydrate side chains, heavily sialylated, suggests that it may bind a wide spectrum of pathogenic microorganisms (Liu et al., 2005). Recently, it has been demonstrated that bovine MUC1 is able to reduce infection by the neuraminidase-sensitive rotavirus RRV strain in vitro (Kvistgaard et al., 2004). The relationship between adherence of *Escherichia coli* and expression of MUC1 in the endometrium was studied in bitches at different stages of the estrus cycle: MUC1 gene expression markedly decreased in the canine endometrium at the early stage of diestrus, resulting in increased *E. coli* adherence. This could be one of the factors inducing pyometra (Ishiguro et al., 2006).

The size of MUC1 may be an important factor in epithelia protection: The large size of MUC1 in human milk could explain why human MFGM is a more effective microbial inhibitor than the bovine homolog (Kvistgaard et al., 2004). In dairy species, the interest in the MUC1 VNTR-length polymorphism comes from the possible relationship between its protein variant size and the resistance to mastitis (Patton, 1999). The VNTR polymorphism was also found to be associated

Received May 16, 2006.

Accepted July 4, 2006.

<sup>1</sup>Corresponding author: roberto.rasero@unito.it

with yield, health, and reproductive traits in Holstein cows (Hens et al., 1995).

In previous investigations, the MUC1 gene repetitive region and its polymorphism in cattle and goats was analyzed (Rasero et al., 2002; Sacchi et al., 2004). However, information concerning the MUC1 gene in sheep is still lacking. The present investigation was undertaken with the aim of studying the repetitive region of the MUC1 gene and analyzing its polymorphism in some Italian sheep breeds.

## MATERIALS AND METHODS

### Sampling and DNA Extraction

Blood samples were collected from 206 individuals belonging to 7 Italian sheep breeds (38 Biellese, 38 Frabosana, 30 Langhe, 24 Savoiarda, 30 Sambucana, 20 Sopravissana, and 26 Tacola). Genomic DNA was obtained from 200  $\mu$ L of whole blood using the Nucleo Spin Blood Quick Pure method (Macherey-Nagel GmbH, Düren, Germany). Thirty individual milk samples from Langhe sheep were also collected.

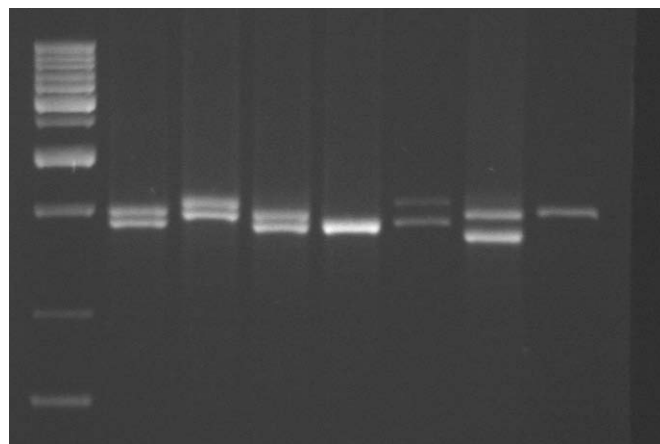
### PCR Amplification

Amplifications were carried out on a GeneAMP PCR System 2400 thermal cycler (Applied Biosystems, Foster City, CA) using the primer set PG4 and PG9, derived from the goat MUC1 gene sequences flanking the repetitive region (Sacchi et al., 2004). A GC-rich solution was added to improve the reaction.

The FastStart Taq DNA Polymerase (Roche Applied Science, Penzberg, Germany) was used under the following conditions: a denaturation step at 95°C for 10 min, 10 cycles at 95°C for 30 s, 62°C for 30 s, and 72°C for 2 min 30 s, followed by 30 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 2 min 30 s. A final extension step of 72°C for 7 min was carried out for all reactions. Amplified fragments were separated by agarose gel electrophoresis.

### Sequencing Strategy

Some PCR fragments were purified by the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) and cloned into the pDrive Cloning Vector using the Qiagen PCR Cloning Kit (Qiagen). The fragments were cycle sequenced using an ABI Prism 310 Genetic Analyzer (Applied Biosystems) by the dideoxy chain termination method with fluorescence dye terminators (Applied Biosystems). Sequencing on both strands was performed using the 2 vector primers as well as the PG4 and PG9 primers. The sequences obtained were compared and



**Figure 1.** Agarose gel electrophoresis of PG4-PG9 PCR products from individual sheep samples. Lane 1, 1-kb DNA ladder, from bottom 750 bp, 1 kb, 1.5 kb, 2 kb, 2.5 kb, 3 kb, 3.5 kb, and so on; lane 2, MUC1 2/3; lane 3, MUC1 3/4; lane 4, MUC1 2/3; lane 5, MUC1 2/2; lane 6, MUC1 2/4; lane 7, MUC1 1/3; lane 8, MUC1 3/3.

aligned by the Clustal W program (Thompson et al., 1994).

### Sequence Analysis

The repetitive region structure and the coding sequence were predicted by alignment with the bovine (GenBank AF399757) and the caprine (GenBank AY388993 and AY388994) sequences. The corresponding peptides were deduced by the GENSCAN program (Burge and Karlin, 1997). Prediction and structural characterization of O-glycosylation sites were performed using the approach suggested by Julenius et al. (2004; <http://www.cbs.dtu.dk/services/NetOGlyc/>).

### SDS-PAGE Analysis of MUC1

Fat globules were obtained from individual Langhe milk samples by centrifugation and the MFGM were purified by repeated ethanol-ether extractions (Sacchi et al., 1995). All samples were analyzed by SDS-PAGE on 4 to 15% gradient Ready Gel Precast gels (Bio-Rad, Hercules, CA). The MUC1 fractions were identified using protocols by Sacchi et al. (1995).

## RESULTS AND DISCUSSION

### PCR Amplification

Analysis of the 206 sheep samples allowed the overall identification of 4 different fragments of approximately 1,500 bp on agarose gel electrophoresis (Figure 1). Each individual sheep showed 1 or 2 bands, in accordance with the polymorphic model previously deduced for the

	1	10	20	30	40	50	60																																								
<i>Ovis aries</i>	A	G	C	C	C	G	G	C	T	G	C	C	T	T	C	C	A	A	C	C	A	G	C	A	G	C	C	C	C	G																	
<i>Capra hircus</i>	A	G	C	C	C	G	G	C	T	G	C	C	T	T	C	C	A	A	C	C	A	G	C	A	G	C	C	C	C	G																	
<i>Bos taurus</i>	A	G	C	C	A	G	C	T	G	C	T	T	C	C	A	G	C	C	A	G	C	A	G	C	C	T	C	G	A	C	T	C	C	A	A	C	C	A	G	C	A	G	C	C	C	C	A
	*****																																														

**Figure 2.** Alignment of sheep, goat, and bovine MUC1 repeat consensus sequences.

human, bovine, and caprine MUC1 genes. Therefore, we called these fragments alleles 1, 2, 3, and 4, in descending order of electrophoretic mobility.

### Cloning and Sequence Analysis

The following pairs of nonoverlapping sequences were obtained: GenBank DQ389165S1 and GenBank DQ389165S2 for allele 1, GenBank DQ389166S1 and GenBank DQ389166S2 for allele 2, GenBank DQ389167S1 and GenBank DQ389167S2 for allele 3, and GenBank DQ389168S1 and GenBank DQ389168S2 for allele 4. It was not possible to obtain 2 overlapping sequences from the PG4 and PG9 primers because of the length of the fragments. Also, it was not possible to choose more internal primers because of the presence of a repetitive structure.

The alleles contained a repetitive region and 2 unique-sequence flanking regions. The sequences of the 2 flanking regions were identical in the 4 sequenced fragments. The 5' end sequence showed 90 and 95% homology with the portion identified upstream from repetitive regions in cattle and goats, respectively. The 3' end showed homology equal to 91 and 99% with the portion identified downstream from the repetitive region in cattle and goats, respectively.

Analysis of the repetitive region showed an array of 60-bp units tandemly repeated as in humans, cattle, and goats. The average homology of all repeated units with the derived consensus sequence was 87%. The repeats were not all perfectly conserved, but the sequence homology was clearly sufficient to preserve the mechanism giving rise to the VNTR polymorphism.

A certain degree of homology is necessary for the repeats to recognize each other during the generation of alleles of different length. In fact, the most polymorphic MUC1 genes are found in species showing the greatest conservation of repeats (Olsen, 1999). In mice, no MUC1 gene polymorphism was found and a homology of 75% was observed, whereas the most polymorphic MUC1 gene was found in humans (30 different alleles among 69 random individuals), where repeats shared 97 to 100% homology (Gendler et al., 1988; Spicer et al., 1991).

Ruminants hold a middle-ranking position. In goats, the repeats are highly conserved (95% homology), and

15 different alleles were observed in 178 animals belonging to 6 Italian breeds (Sacchi et al., 2004), whereas only 4 alleles were found in the present investigation in 206 individual sheep belonging to 7 breeds. The sheep repeat consensus sequence differs from the sequences of goats and cattle by 2 and 4 nucleotides, respectively, so the degree of homology between sheep and goats is 97% and that between sheep and cattle is 93% (Figure 2).

The 4 allele sequences discussed previously were used to deduce the corresponding peptides. It was possible to obtain 39 individual peptide repeats showing an average homology of 82%; when the repeats were compared with the derived consensus repeat, homology dropped to 78% (Figure 3). Overall, only 3 residues were 100% conserved, specifically the Ser at positions 5 and 17, and the Arg at position 19. On the other hand, the alignment of 38 repeats from 3 different bovine alleles showed that as many as 12 residues were perfectly conserved (Rasero et al., 2002).

O-Glycosylation sites were predicted for the 4 alleles and were validated by alignment with the complete bovine MUC1 primary structure (GenBank L41543 and GenBank AF399757). Sheep predicted peptides were shorter than the complete native mucin backbone, and the analysis of partial sequences could produce unreliable results, particularly at the ends of fragments. This comparative approach seemed to be quite accurate owing to the highly conserved MUC1 core structure in the nonrepetitive flanking regions. All Ser-Thr residues are recognized as potential glycosylation sites.

In spite of their reduced sequence homology, the sheep repeats shared a high number of potential glycosylation sites, because the presence of Ser, Thr, and Pro were highly conserved across the repeats from the different alleles. The function of mucin-type glycosylation is to protect the protein from cleavage, to make it bind more water, and to be stiffer and extended. The conservation of the exact number and position of glycosylation sites does not seem to be very important for the purpose of the functional integrity of the mucin, but the level of glycosylation would be conserved as a bulk overall property (Julenius et al., 2004). This does not exclude the possibility that individual sites may be highly conserved, as evidenced by the Ser found at positions 5 and 17 of the repeat unit (Figure 3).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
<i>Ovis aries</i>	Ser	Pro	Ala	Ala	Ser	Pro	Gly	His	Asp	Gly	Ala	Ser	Ser	Pro	Thr	Ser	Ser	Pro	Ala	Pro
<i>Capra hircus</i>	Ser	Pro	Ala	Ala	Ser	Pro	Gly	His	Asp	Gly	Ala	Ser	Thr	Pro	Thr	Ser	Ser	Pro	Ala	Pro
<i>Bos taurus</i>	Ser	Pro	Ala	Ala	Ser	Pro	Gly	His	Asp	Gly	Ala	Ser	Thr	Pro	Thr	Ser	Ser	Pro	Ala	Pro

**Figure 3.** Alignment of sheep, goat, and bovine polypeptide consensus sequences.

The alignment of the deduced consensus repeats of the different ruminant species showed that the sheep consensus sequence was very similar to the homologs of cattle and goats, because they shared 95% of their AA, that is, 19 of 20 residues (Figure 3). The observed substitution involved the presence of Thr rather than Ser in position 13 but only in 51% of the repeats. This substitution did not change the functional significance of the residue.

### Polymorphism Analysis

On the basis of the sequence analyses, the differences observed in individual sheep were in accordance with the VNTR model assessed in other species. A size of approximately 1,500 bp, corresponding to 21 repeats, was estimated for allele 3. The differences between alleles could not be directly computed as the number of repeats. In fact, the repetitive region was not completely sequenced because of its size. Thus, an estimate of allelic differences was attempted using a heterozygous reference sample from cattle with alleles differing for 3 repeats (Figure 4). Sheep alleles 1, 2, 3, and 4 differ from each other in 1 repeat unit. The proposed PCR protocol was efficient for the analysis of sheep MUC1 polymorphisms, allowing a single repeat unit difference between alleles to be detected.

The number of alleles, observed heterozygosity, and allele frequencies for each breed, analyzed by the PCR protocol developed, are reported in Table 1. All sheep breeds showed a low degree of polymorphism, with a mean heterozygosity of 0.26. Allele 3 was by far the predominant allele among all breeds. Allele 1 was rare and was found only in one heterozygous Tacola sheep.

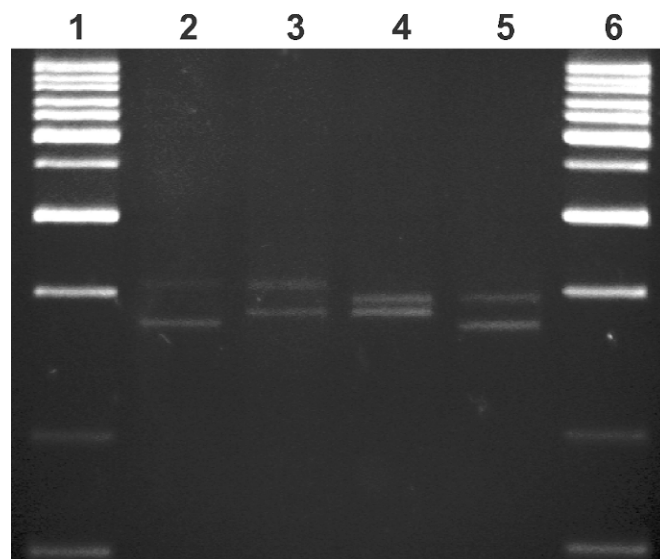
In Italian goat breeds, the same primer set allowed for the identification of 15 different alleles ranging from 1,500 to 3,000 bp, with a mean heterozygosity value of 0.8 across 6 breeds (Sacchi et al., 2004). Therefore, the sheep repetitive region seemed to be, on average, less variable and smaller in size than that of the goat. In this respect, sheep mucin was more similar to bovine mucin, which showed only 5 alleles, although their frequencies were more evenly distributed (Sacchi et al., 1995).

### SDS-PAGE Analysis of MUC1

Small individual differences in mobility were observed in the SDS-PAGE analysis. The resulting electrophoretic pattern did not allow us to differentiate the protein variants identified at the DNA level. A single repeat unit difference, corresponding to 20 AA residues, appeared too weak and was not clearly detectable. In addition, variation in SDS-PAGE mobility of the mucin bands may not depend exclusively on the diversity of the polypeptide size, but also on the variable number and structure of the carbohydrate side chains (Patton and Patton, 1990).

### CONCLUSIONS

The sheep MUC1 gene contained a repetitive region of 60 bp units. The motif was tandemly repeated for a variable number of times in individual molecules, giv-



**Figure 4.** Agarose gel electrophoresis of PG4–PG9 PCR products from individual sheep samples. Lanes 1 and 6, 1-kb DNA ladder, from bottom 750 bp, 1 kb, 1.5 kb, 2 kb, 2.5 kb, 3 kb, 3.5 kb, and so on; lane 2, reference sample from a MUC1 heterozygous bovine showing alleles differing for 3 repeats; lane 3, MUC1 heterozygous sheep showing alleles differing for 2 repeats (MUC1 2/4); lane 4, sheep MUC1 heterozygous 2/3; lane 5, sheep MUC1 heterozygous 1/3.

**Table 1.** Observed heterozygosity (H) and allele frequencies

Sheep breed	H	Allele frequency			
		1	2	3	4
Biellese	0.263	0.00	0.07	0.80	0.13
Frabosana	0.132	0.00	0.06	0.91	0.03
Langhe	0.233	0.00	0.12	0.88	0.00
Savoiarda	0.167	0.00	0.08	0.92	0.00
Sambucana	0.300	0.00	0.28	0.72	0.00
Sopravvissana	0.550	0.00	0.30	0.67	0.03
Tacola	0.308	0.02	0.08	0.85	0.06

ing rise to a length polymorphism much like what has been observed in the homologous genes present in other species. In spite of the high phylogenetic relationship among domestic ruminant species, bovine, caprine, and ovine MUC1 mucins showed some differences. Goat MUC1 was hypervariable, showing highly conserved repeats. Bovine mucin was less variable in spite of a high degree of conservation among units, whereas sheep MUC1 showed a lower degree of polymorphism, probably because of the reduced homology among units. Nevertheless, the glycosylation properties were well conserved as a bulk. These findings suggest that ruminants may be a useful model to study the mechanisms by which the variation in the repeat number and the extracellular domain size can modulate the effectiveness of MUC1 as a cell-surface shield.

### ACKNOWLEDGMENTS

This work was supported by Ministero dell'Università e della Ricerca, PRIN 2005, Prot. 2005075887-004.

### REFERENCES

- Burge, C., and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268:78–94.
- Campana, W. M., R. W. Josephson, and S. Patton. 1992. Presence and genetic polymorphism of an epithelial mucin in milk of the goat (*Capra hircus*). *Comp. Biochem. Physiol.* 103B:261–266.
- Gendler, S. 2001. MUC1, the renaissance molecule. *J. Mam. Gland Biol. Neoplasia* 6:339–353.
- Gendler, S., J. Taylor-Papadimitriou, T. Duhig, J. Rothbard, and J. Burchell. 1988. A highly immunogenic region of a human polymorphic epithelial mucin expressed by carcinomas is made up of tandem repeats. *J. Biol. Chem.* 263:12820–12823.
- Hens, J. R., J. W. Rogers, M. L. Huott, and S. Patton. 1995. Association of the epithelial mucin, PAS-1, with yield, health, and reproductive traits in Holstein dairy cows. *J. Dairy Sci.* 78:2473–2480.
- Ishiguro, K., E. Baba, R. Torii, H. Tamada, N. Kawate, S. Hatoya, V. Wijewardana, D. Kumagai, K. Sugiura, T. Sawada, and T. Inaba. 2006. Reduction of mucin-1 expression associated with increased *Escherichia coli* adherence in the canine uterus in the early stage of dioestrus. *Vet. J.* doi:10.1016/j.tvjl.2005.11.009
- Julenius, K., A. Molgaard, R. Gupta, and S. Brunak. 2004. Prediction, conservation analysis and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* 15:153–164.
- Kvistgaard, A. S., L. T. Pallesen, C. F. Arias, S. Lopez, T. E. Petersen, C. W. Heegaard, and J. T. Rasmussen. 2004. Inhibitory effects of human and bovine milk constituents on rotavirus infections. *J. Dairy Sci.* 87:4088–4096.
- Lillehoj, E. P., S. W. Hyun, B. T. Kim, X. G. Zhang, and D. I. Lee. 2001. Muc1 mucins on the cell surface are adhesion sites for *Pseudomonas aeruginosa*. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 280:L181–L187.
- Liu, C., A. K. Erickson, and D. R. Henning. 2005. Distribution and carbohydrate structures of high molecular weight glycoproteins, MUC1 and MUCX, in bovine milk. *J. Dairy Sci.* 88:4288–4294.
- Mather, I. H. 2000. A review and proposed nomenclature for major proteins of milk-fat globule membrane. *J. Dairy Sci.* 83:203–247.
- Middleton-Price, H., S. Gendler, and S. Malcolm. 1988. Close linkage of PUM and SPTA within chromosome band 1q21. *Ann. Hum. Genet.* 52:273–278.
- Olsen, K. M. 1999. Minisatellite variation in a single-copy nuclear gene: Phylogenetic assessment of repeat length homoplasy and mutational mechanism. *Mol. Biol. Evol.* 16:1406–1409.
- Patton, S. 1999. Some practical implications of the milk mucins. *J. Dairy Sci.* 82:1115–1117.
- Patton, S., S. J. Gendler, and A. P. Spicer. 1995. The epithelial mucin, MUC1, of milk, mammary gland and other tissues. *Biochim. Biophys. Acta* 1241:407–424.
- Patton, S., and R. S. Patton. 1990. Genetic polymorphism of PAS-I, the mucin-like glycoprotein of bovine milk fat globule membrane. *J. Dairy Sci.* 73:3567–3574.
- Perucatti, A., S. Floriot, G. P. Di Meo, D. Soglia, R. Rullo, S. Maione, D. Incarnato, A. Eggen, P. Sacchi, R. Rasero, and L. Iannuzzi. 2006. Comparative FISH mapping of mucin 1, transmembrane (*MUC1*) among cattle, river buffalo, sheep and goat chromosomes: Comparison between bovine chromosome 3 and human chromosome 1. *Cytogen. Genome Res.* 112:103–105.
- Rasero, R., P. Sacchi, S. Rosati, E. Cauvin, and S. Maione. 2002. Molecular analysis of the length polymorphic MUC1 gene in cattle. *J. Anim. Breed. Genet.* 119:342–349.
- Sacchi, P., A. Caroli, E. Cauvin, S. Maione, S. Sartore, D. Soglia, and R. Rasero. 2004. Analysis of the Muc1 gene and its polymorphism in *Capra hircus*. *J. Dairy Sci.* 87:3017–3021.
- Sacchi, P., E. Macchi, R. Rasero, and P. Fiandra. 1995. Two new variants of the bovine PAS-1 glycoprotein. *Anim. Genet.* 26:197–198.
- Spicer, A. P., G. Parry, S. Patton, and S. J. Gendler. 1991. Molecular cloning and analysis of the mouse homologue of the tumour-associated mucin, reveals conservation of potential O-glycosylation sites, transmembrane, and cytoplasmic domains and a loss of minisatellite-like polymorphism. *J. Biol. Chem.* 266:15099–15109.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.