Theor Appl Genet (1994) 89: 116–120

T. A. Schild · V. Wagner · H. Geldermann

Variants within the 5′-flanking regions of bovine milk protein genes: I. κ-casein-encoding gene

Received: 27 December 1993 / Accepted: 11 February 1994

Abstract In order to identify DNA variants within the 5′-flanking region of the bovine κ-casein (κCn)-encoding gene, this area of the gene from 13 cows belonging to seven breeds (Holstein Friesian, Brown Swiss, German Simmental, Jersey, Galloway, Scottish Highland and Ceylon Dwarf Zebu) was analysed. For each individual, about 1 kb of the 5′-flanking region including exon I was amplified by polymerase chain reaction (PCR). The biotinylated PCR product was immobilized on magnetic beads followed by direct bidirectional sequencing using an automated DNA sequencer. Fifteen DNA variants were identified, some of which are located within potential regulatory sites and possibly involved in the expression of the κ-casein encoding gene.

Key words κ-casein encoding gene · DNA regulatory sequences · Comparative DNA sequencing DNA variants · Bovine

Abbreviations *AP2* activator protein 2 · *bp* base pair(s) *GRE/RC* glucocorticoid response element/reverse complement · *HNF3* hepatocyte nuclear factor 3 *κCn* κ-casein · *MGF* mammary gland-specific nuclear factor · *nt* nucleotid(s) · *OCT1* octamer-binding site 1 *PA* polyacrylamide · *PCR* polymerase chain reaction *PMF* pregnancy-specific mammary nuclear factor *kb* kilobase(s) or 1000 bp

Introduction

Caseins are one of the best developed models in biochemical and genetic research, yet questions connected with the regulation and expression of these genes still remain open. In vitro studies have revealed that casein gene expression is regulated at both the transcriptional and the posttranscriptional levels by a complex interplay of different hormones (Rosen et al. 1989). All four casein genes are clustered within less than 200 kb of bovine chromosome 6 (Threadgill and Womack 1990; Ferretti et al. 1990). It is assumed that the three genes coding for the calcium-sensitive α_{S1}-, α_{S2}- and β-caseins have evolved from one ancestral gene that resulted from exon shuffling (Jones et al. 1985) by intergenic duplication (Yu-Lee et al. 1986; Bonsing and Mackinlay 1987). In contrast, the κ-casein encoding gene (*κCn*) evolved differently and consists of five exons distributed over a total length of approximately 13 kb (Alexander et al. 1988). Amino acid sequence similarities suggest that κCn is related to the fibrinogens (Jollès et al. 1986).

Four variants have been described for the coding region of *κCn* gene (Eigel et al. 1984; Miranda et al. 1993); their pylogeny is E<A>B>C. The variants are known to be associated with both the physico-chemical properties and the relative amount of milk proteins (Oloffs et al. 1992; Mao et al. 1992). Several studies have reported that cows carrying the *κCn* BB genotype produced milk with a significantly higher protein content (Gonyon et al. 1987; Aleandri et al. 1990; Bovenhuis et al. 1992). Other data have shown that differences in allelic protein expression in the milk of heterozygous *κCn* cows exist (Van Eenennaam and Medrano et al. 1991). The results indicate that the B allele is associated with the greater amount of total *κCn* present in the milk of AB cows.

The cause of this greater expression is not understood but may be due to differences in *cis*-acting sequences involved in the quantitative expression of *κCn* gene. Since DNA variants which alter gene expression might be located upstream from the 5′-flanking region, as a first step in an extended investigation we have sequenced, analysed and compared the 5′-flanking region of this milk protein encoding gene within highly divergent cows in order to analyse the DNA variants.

Communicated by E. J. Eisen

T. A. Schild · V. Wagner · H. Geldermann (✉)
Fachgebiet Tierzüchtung,
Institut für Tierhaltung und Tierzüchtung (470),
Universität Hohenheim, Garbenstr. 17,
D-70593 Stuttgart, Germany

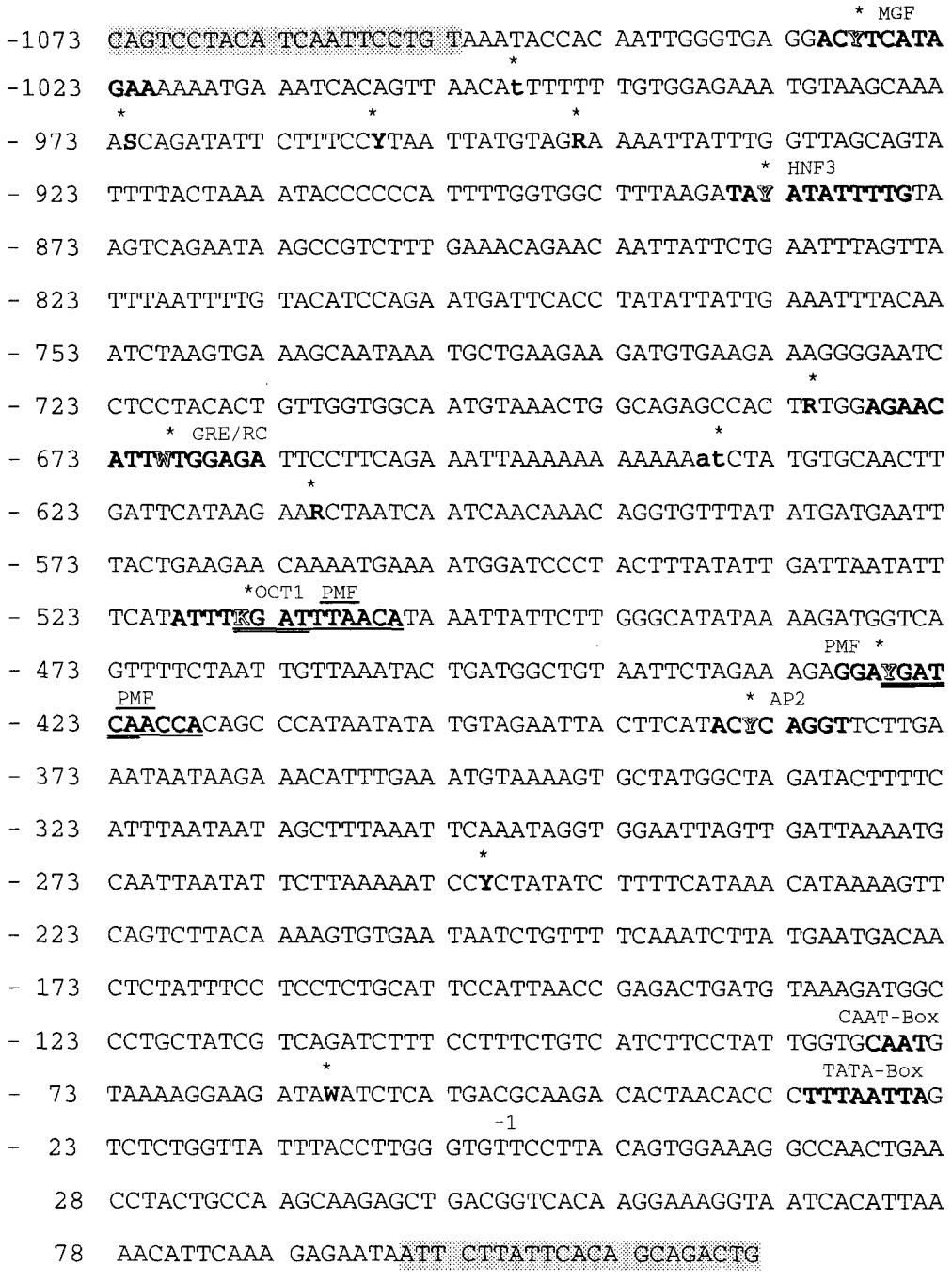

Fig. 1 Analysed region of the bovine κ-casein encoding gene. Variants are indicated by an *asterisk* (*capital letters* represent base substitutions, *small letters* base deletions); CAAT-box, TATA-box and potential protein binding sites hit by mutations are displayed by *shaded letters* (region shared by two factors are *doubly underlined);* PCR-primers are *hatched* (KCN-5b: 5′-biotin-CAGTCCTACATCAATTCCTGT-3′; KCN-3: 5′-CAGTCTGCTGTGAATAAGAAT). The 1.2 kb fragment of the promoter region including exon I was amplified by polymerase chain reaction (PCR) in 100 ml of a solution consisting of 0.25 m*M* of each of the dNTPs, 10 m*M* TRIS-HCl, pH 9.0, 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 10 pmol of each primer, 1 μg of genomic DNA and 2.5 units of *Taq*-Polymerase (Pharmacia Biotech, Freiburg, FRG) in a Biometra TRIO-Thermoblock (Biometra, Göttingen, FRG). Each cycle consisted of 1.5 min at 95 °C, 2 min at 54 °C and 4 min at 72 °C with the exception of 3 min denaturation in the first cycle and 7 min DNA chain extension in the last cycle

Materials and methods

Selection of animals

About 2000 cows, selected according to extremely high or low milk protein yields out of eight dairy breeds, were further analysed for their milk protein variants by the isoelectric focusing of milk samples using the method of Seibert et al. (1985). Consequently, 300 animals with different and rare genotype combinations were chosen. From these, 8 animals belonging to four breeds (Jersey, German Simmental, German Friesian, German Brown Swiss) were selected according to their casein and whey protein contents and used for DNA analysis. In addition, five cows from three additional breeds (Galloway, Scottish Highland and Ceylon Dwarf Zebu) were considered in order to obtain a highly divergent group of animals.

Direct sequencing of PCR amplified DNA fragments

After 35 cycles of PCR amplification, 90 μl of the mixture was added to 0.4 mg of Dynabeads M280 streptavidin (Dynal, Hamburg, FRG). Both DNA strands were prepared for sequencing reactions following the protocol of Hultman et al. (1991) with slight modifications.

Immobilized PCR products were sequenced using the AutoRead Sequencing Kit (Pharmacia Biotech, Freiburg, FRG) and four fluorescent primers (KCN1: 5′-fluorescein-GCTGTGAATAAGAATTATTC-3′, KCN2: 5′-fluorescein-TGAAAACAGATTATTCACAC-3′, KCN3: 5′-fluorescein-ATGCCCAAGAATAATTTATG-3′, KCN4: 5′-fluorescein-TAGGTGAATCATTCTGGATG-3′) on an A.L.F. DNA sequencer (Pharmacia Biotech, Freiburg, FRG). Cycle Sequencing reactions were performed using the Ready Reaction Dye Deoxy Terminator Cycle Sequencing Kit (Applied Biosystems,

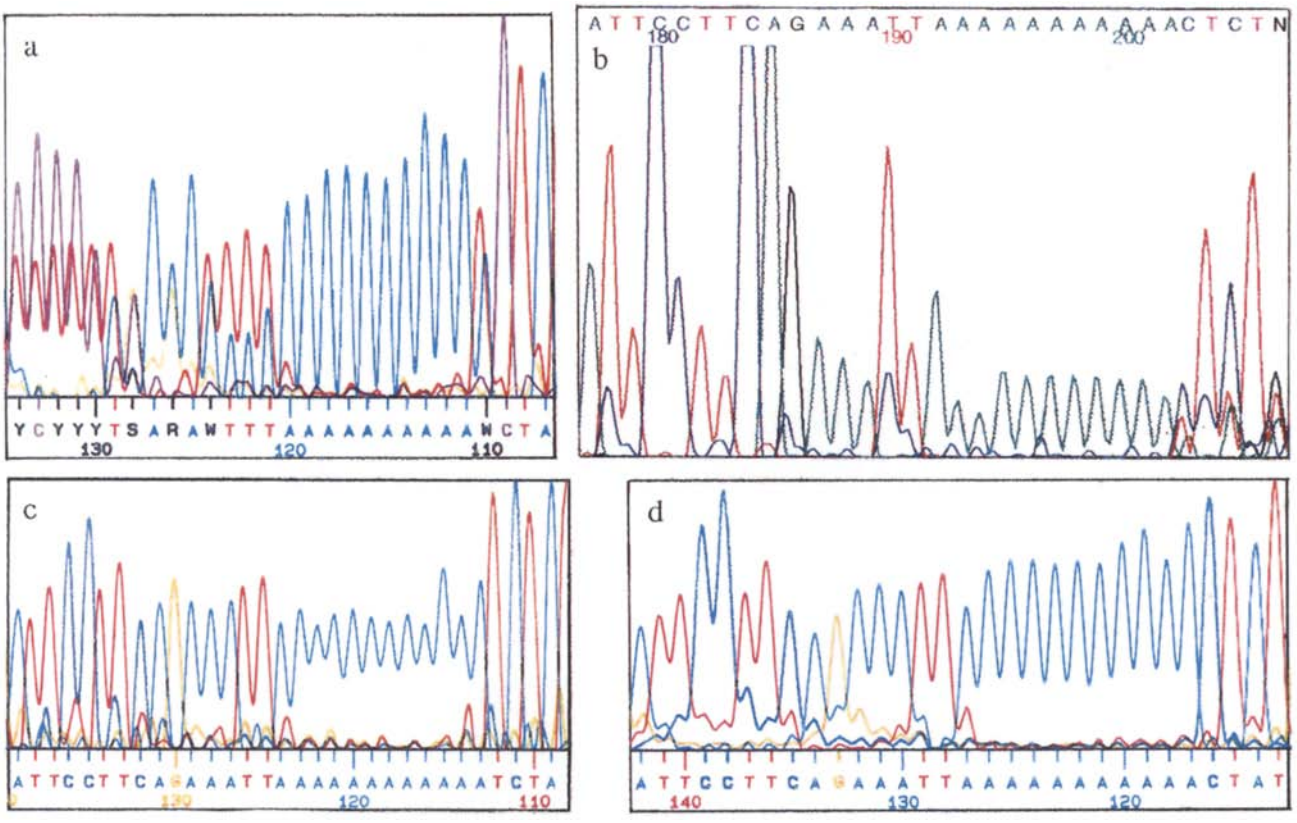

Fig. 2a–d Verification of a heterozygous AT deletion at position –639/–638 by bidirectional sequencing. Solid phase sequencing was performed using T7 polymerase and dye primers with the A. L. F. DNA sequencer (Pharmacia, Freiburg, FRG) on 6% PA gels, reverse strand sequencing by dye deoxy terminator cycle sequencing using *Taq* DNA polymerase on an ABI 373A DNA sequencer (Applied Biosystems, Weiterstadt, FRG) on 7% PA gels. **a** Heterozygous animal, solid phase, **b** same animal (Galloway 1) reverse strand, **c** homozygous animal (Scottish Highland 1) without deletion, solid phase, **d** homozygous animal (Galloway 2) with AT deletion, solid phase

Weiterstadt, FRG) and two unlabeled primers (KCN5: 5′-AGTCCTACATCAATTCCTG-3′, KCN6: 5′-GCCGTCTTTGAAACAGAAC-3′) on an ABI 373A DNA sequencer (Applied Biosystems, Weiterstadt, FRG).

In order to test the ability to identify nucleotide substitutions in the heterozygous state by solid phase sequencing, three independent sequencing reactions in both directions were performed for three individuals on an A. L. F. DNA Sequencer. T7 polymerase sequencing in conjunction with dye primers produced a very even signal height. Thus, heterozygous substitutions were identified as two half-height peaks coinciding at the same position. Identification of the heterozygous mutations was reproducable in all three control reactions.

Results and discussion

Identification of polymorphic sites

After sequencing of the 5′-flanking region of the *κCn* gene of 13 cows, a total of 15 DNA-variants was identified by the alignment of sequencing results. The locations of these variants within the promoter region (+115/–1073) is shown in Fig. 1. Furthermore, in comparison to the 5′-flanking region of the bovine *κCn* gene, as determined by Groenen et al. (EMBL data base, accession no. M75887) there were 12 additional bases (Fig. 1, positions –1052: A; –1048: A; –1043: A; –1022/–1021: AA; –1014: A; –999/998: TT; –867: A; –744: A; –694: G; –588: T) and four differences in DNA primary structure (Fig. 1: positions –978/–977; –972/–971; –929/–928: GC instead of CG; –559/–558: AT instead of TC) in all of the animals analysed. The high number of variants identified does not necessarily represent all mutations that exist for cattle within the analysed region, although cows from a very heterogenous group of animals were considered. The confirmation of evidence for breed specific variants, their breed distribution and the detection of further variants requires the analysis of a larger group of animals.

Identification of heterozygous substitutions or deletions

Within the amplified region of the *κCn* gene, two heterozygous deletions were identified. As Fig. 2 shows, the variant at position –638/–639 is due to an AT deletion. Thus, in the heterozygous state the sequences following the poly (A) stretch are unreadable because the signals from both alleles no longer coincide in most positions. By bidirectional solid phase sequencing the reverse picture was observed when sequencing both strands, thus verifying the heterozygous state. Out of the 13 animals analysed 6 were

Table 1 Binding sites of potential regulatory proteins within the 5′-flanking region of the bovine κ-casein encoding gene and their locations. Computer analysis was performed using the HUSAR programme package supplied by the German Cancer Research Center, Heidelberg, FRG. The recognition sites listed below only enclose those that are hit by variants

Factor[a]	Consensus sequence[b]	Sequence within the κ-casein encoding gene	Position	Homology [%]
MGF	ANTTCTTGGNA (1)	ACYTCATAGAA	−1031	82
HNF3	TATTGAYTTWG (2)	TAYATATTTTG	−886	82
GR/RC	AGRACANNNTGTACC (3)	AGAACATTWTGGAGA	−683	87
OCT1	ATTTGCAT (4)	ATTTKGAT	−519	88
PMF	TGAT$(N)_{1-2}$ATCA (5)	KGATTTAACA	−515	90
		GGAYGATCA	−420	89
		YGATCAACCA	−417	90
AP2	CCCCAGGC (6)	ACYCAGGT	−387	75

[a] See abbreviations for explanations
[b] (1) Schmitt-Ney et al. 1991; (2) Raymondjean et al. 1991; (3) Beato 1989; (4) Groenen et al. 1992; (5) Lee and Oka 1992; (6) Mitchell et al. 1987

heterozygous at this position. Furthermore, in the zebu individual sequenced only one A within the poly(A) stretch was found to be deleted, perhaps indicating a breed-specific variant (data not shown).

Further steps in analysing different rates in gene expression associated with allelic variation

The association of protein expression with *κCn* genotypes may be explained by the existence of polymorphisms in the regulatory regions of genes. Investigations on the influences of genetic polymorphisms in the 5′-flanking region of the human cytochrome P450IIE1 gene have demonstrated that mutations can alter the binding of *trans*-acting factors and change a gene's transcriptional regulation (Hayashi et al. 1991). Figure 1 shows that some of the variants identified for the 5′-flanking region of *κCn* gene are located within potential recognition sites of DNA binding proteins. Quantitative differences in gene expression may be the result. In vitro studies have demonstrated that mutations within the binding site of the mammary gland-specific nuclear factor (MGF) strongly affect transcriptional activity in vitro (Schmitt-Ney et al. 1991). The C/T transition at position −1029 is located within a potential MGF binding site of 82% homology to the sequence ANTTCTTGGNA, which might represent part of a MGF recognition site (see Table 1). Further investigation is required to verify the influence of these allelic variants in *κCn* gene regulation. Therefore, we are currently carrying out gene expression studies with various constructs of a luciferase vector, including the promoter region, with different sets of variants in order to identify those variants within the 5′-flanking region that influence gene expression. Moreover, the relationship between the mutations altering a potential *cis*-acting motif and the milk protein expression of the relevant cows is being analysed in a larger number of animals (Ehrmann et al., in preparation).

Acknowledgements We gratefully acknowledge the technical support of Pharmacia Biotechnology, Freiburg, FRG, and Applied Biosystems, Weiterstadt, FRG. This study was supported by a grant from the Deutsche Forschungsgemeinschaft.

References

Aleandri R, Buttazoni LG, Schneider JC, Caroli A, Davoli R (1990) The effects of milk protein polymorphisms on milk components and cheese-producing ability. J Dairy Sci 73:241–255

Alexander LJ, Stewart AF, Mackinlay AG, Kapelinskaya TV, Tkach TM, Gorodetsky SI (1988) Isolation and characterization of the bovine κ-casein gene. Eur J Biochem 178:395–401

Beato M (1989) Gene regulation by steroid hormones. Cell 56:335–344

Bonsing J, Mackinlay AG (1987) Recent studies on nucleotide sequences encoding the caseins. J Dairy Res 54:447–461

Bovenhuis H, Van Arendonk JAM, Korver S (1992) Associations between milk protein polymorphisms and milk production traits. J Dairy Sci 75:2549–2559

Eigel WN, Butler JE, Ernstrom CA, Farrell HM, Harwalkar VR, Jenness R, Whitney RMcL (1984) Nomenclature of proteins of cow's milk: fifth revision. J Dairy Sci 67:1599–1631

Ferretti L, Leone P, Sgaramella V (1990) Long-range restriction analysis of the bovine casein genes. Nucleic Acids Res 18:6829–6833

Gonyon DS, Mather RE, Hines HC, Haenlein GFW, Arave CW, Gaunt SN (1987) Associations of bovine blood and milk polymorphisms with lactation traits: Holsteins. J Dairy Sci 70:2585–2598

Groenen MAM, Dijkhof RJM, Spira C, van der Poel JJ (1991) *B. taurus* κ-casein gene, exon 1. EMBL data base, accession no. M75887

Groenen MAM, Dijkhof RJM, van der Poel JJ, van Diggelen R, Verstege E (1992) Multiple octamer binding sites in the promoter region of the bovine α_{S2}-casein gene. Nucleic Acids Res 20:4311–4318

Hayashi SI, Watanabe J, Kawajiri K (1991) Genetic polymorphisms in the 5′-flanking region change transcriptional regulation of the human cytochrome P450IIE1 gene. J Biochem 110:559–565

Hultman T, Bergh S, Moks T, Uhlén M (1991) Bidirectional solid-phase sequencing of *in vitro*-amplified plasmid DNA. BioTechniques 10:84–93

Jollès P, Lévy-Toledano S, Fiat A-M, Soria C, Gillessen D, Thomaidis A, Dunn FW, and Caen JP (1986) Analogy between fibrinogen and casein. Eur J Biochem 158:379–382

Jones WK, Yu-Lee L-Y, Clift SM, Brown TL, Rosen JM (1985) The rat casein multigene family. J Biol Chem 260:7042–7050

Lee CS, Oka T (1992) A pregnancy-specific mammary nuclear factor involved in the repression of the mouse β-casein gene transcription by progesterone. J Biol Chem 267:5797–5801

Mao IL, Buttazzoni LG, Aleandri R (1992) Effects of polymorphic milk protein genes on milk yield and composition traits in Holstein cattle. Acta Agric Scand Sect. A Animal Sci 42:1–7

Miranda G, Anglade P, Mahé MF, Erhardt G (1993) Biochemical characterization of the bovine genetic κ-casein C and E variants. Anim Genet 24:27–31

Mitchell PJ, Wang C, Tjian R (1987) Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. Cell 50:847–861

Oloffs K, Schulte-Coerne H, Pabst K, Gravert HO (1992) Die Bedeutung der Proteinvarianten für genetische Unterschiede in der Käsereitauglichkeit der Milch. Zuechtungskunde 64:20–26

Raymondjean M, Pichard A-L, Gregori C, Ginot F, Kahn A (1991) Interplay of an original combination of factors: C/EBP, NFY, HNF3, and HNF1 in the rat aldolase B gene promoter. Nucleic Acids Res 19:6145–6153

Rosen JM, Poyet P, Goodmann H, Lee K-F (1989) Mechanisms by which prolactin and glucocorticoids regulate casein gene expression. Biochem Soc Symp 55:115–123

Schmitt-Ney M, Doppler W, Ball RK, Groner B (1991) β-casein gene promoter activity is regulated by the hormone-mediated relief of transcriptional repression and a mammary gland-specific nuclear factor. Mol Cell Biol 11:3745–3755

Seibert B, Erhardt G, Senft B (1985) Procedure for simultaneous phenotyping of genetic variants in cow's milk by isoelectric focusing. Anim Blood Groups Biochem Genet 16:183–191

Threadgill DW, Womack JE (1990) Genomic analysis of the major bovine milk protein genes. Nucleic Acids Res 18:6935–6942

Van Eenennaam AL, Medrano JF (1991) Differences in allelic protein expression in the milk of heterozygous κ-casein cows. J Dairy Sci 74:1491–1496

Yu-Lee L-Y, Richter-Mann L, Couch CH, Stewart AF, Makinlay AG, Rosen JM (1986) Evolution of the casein multigene family: conserved sequences in the 5′-flanking and exon regions. Nucleic Acids Res 14:1883–1901