Title	Bovine kappa-casein gene promoter haplotypes with potential implications for milk protein expression
Author(s)	Keating, A. F.; Davoren, P.; Smith, T. J.; Cairns, Michael T.
Publication Date	2007-05-25
Publication Information	Keating, AF,Davoren, P,Smith, TJ,Ross, RP,Cairns, MT (2007) 'Bovine kappa-casein gene promoter haplotypes with potential implications for milk protein expression'. Journal Of Dairy Science, 90 :4092-4099.
Publisher	Elsevier ScienceDirect
Link to publisher's version	http://dx.doi.org/10.3168/jds.2006-687
Item record	http://hdl.handle.net/10379/5813
DOI	http://dx.doi.org/10.3168/jds.2006-687

INTERPRETIVE SUMMARY

Bovine *Kappa-Casein* Promoter Haplotypes With Potential Implications For Milk Protein Expression by A.F. Keating, P. Davoren, T.J. Smith, R.P. Ross and M.T. Cairns.

Bovine kappa (κ)-casein protein comprises approximately 8% of total milk protein and is important in maintaining the structure of the casein protein complex. Different genetic haplotypes of κ-casein have been associated with varying levels of κ-casein protein in milk. This study investigated the regulatory region of the κ-casein gene and found two haplotypes in this area, which differed at three locations. A screen of animals representing dairy, dual-purpose and beef breeds found that one haplotype was more common in animals bred for dairy purposes. The transcriptional activity of each haplotype was investigated in a mammary cell model.

RUNNING HEAD: GENETIC VARIATION IN MILK PROTEIN EXPRESSING GENES

Bovine *Kappa-Casein* Promoter Haplotypes with Potential Implications For Milk Protein Expression.

A.F. Keating[*†], P. Davoren[*], T.J. Smith[‡], R.P. Ross[†] and M.T. Cairns.[*]

[*] National Diagnostics Centre, National University of Ireland, Galway, Ireland.

‡ National Centre for Biomedical Engineering Science, National University of Ireland, Galway.

[†] Teagasc, Dairy Products Research Centre, Moorepark, Fermoy, Co. Cork, Ireland.

[*] The first two named authors contributed equally to this paper and should be considered joint first authors.

Corresponding Author: Dr. Michael Cairns, National Diagnostics Centre, National University of Ireland, Galway. Telephone: 0035391492094, Fax no. 0035391586570. Email: michael.cairns@nuigalway.ie

Abstract

Genetic analysis of the *kappa-casein (CSN3)* promoter regions of 42 cattle representing nine different breeds revealed that two distinct haplotypes (A and B) exist at this locus, differing from each other by single base changes at positions –514 (T,G), -426 (T,C) and –384 (T,C), where haplotype A has bases T, T and T and haplotype B has bases G, C and C. The AA and AB haplotypes were found to occur at a higher frequency in the animals tested with 69.0% and 21.4% being homozygous and heterozygous respectively. The sequences that include these polymorphisms are potentially important in transcriptional regulation of the *κ-casein* gene as they contain putative sites for binding of a number of transcription factors. Linkage disequilibrium between the *κ-casein* promoter haplotype and either one of the two major κ-casein coding sequence haplotypes was not evident. The A allele is dominant in all groups (dairy, beef and dual purpose) with an allele frequency of 80% and is higher among high yielding dairy animals (88.9%) than among beef animals (75%). The AB haplotype is comparatively rare in the dairy cattle (11.1%) compared to both dairy and dual purpose animals. The BB haplotype, though rare overall (9.5%), is much higher in dual purpose animals (18.8%) than dairy (5.6%) animals. In contrast the B allele is much more representative of the κ–casein promoters from other ruminants.

(**Key words**: *Kappa*-casein, promoter haplotype, expression studies)

INTRODUCTION

The bovine *κ-casein* gene *(CSN3)* consists of five exons distributed over approximately 13kb of the bovine genome (Alexander et al., 1988; Martin et al., 2002). Eleven genetic haplotypes of *κ-casein* have been identified to date that differ in the coding region (Farrell et al., 2004). The two most common genetic haplotypes, designated A and B, differ in two positions at codons 136 (A-Thr: B-Ile) and 148 (A-Asp: B-Ala) of exon 4 (Mercier et al., 1973; Ron et al., 1994) and can be distinguished by an RFLP-PCR reaction (Nierop Groot et al., 1995). Milk from BB cows reportedly contains higher amounts of protein (Ng-Kwai-Hang et al., 1984; Ron et al., 1994; Bobe et al., 1999) and is associated with shorter rennet coagulation time, firmer curd, and higher cheese-yield than milk from AA animals (Marziali and Ng-Kwai-Hang, 1986; Schaar, 1984; Martin et al., 2002).

A number of studies have been carried out to analyze the promoter region of the *κ-casein* gene. Coding sequence associated allele-specific polymorphisms have been identified in the distal promoter regions at positions –2035 (G/T), -1651 (A/T) and –999 (-/T) (Robitaille et al., 2005). Fifteen single nucleotide changes were identified in thirteen cows from seven different breeds in a study by Schild et al. (1994) and a *Dde*I RFLP polymorphism within a 214bp fragment of the *κ-casein* promoter has also been identified (Kaminski, 1996). While these polymorphisms have been noted in the κ–casein gene promoter, it is unclear whether they are characteristic of a distinct haplotype or if they play a functional role by modulating gene expression.

In this study, polymorphism occurrence in the *κ-casein* gene promoter and the potential effect on transcriptional activity of the promoter were investigated. The

bovine breeds chosen were considered representative of the Irish herd and included dairy, dual-purpose and non-dairy (beef) breeds. Potential links between promoter polymorphisms and the two most common coding sequence genetic polymorphisms were also investigated.

MATERIALS AND METHODS

Blood sampling and DNA extraction

Approximately 15ml of blood was collected into heparinized tubes from the coccygeal vein of 42 cattle representing nine breeds. Breeds sampled were high genetic merit Holstein-Friesian, low genetic merit Holstein-Friesian (Kennedy et al., 2003), Irish-Friesian, Dutch-Friesian, Limousin, Montbeliarde, Charlois, Normande, Norwegian Red and Kerry. With the exception of the Limousin, Charlois, and Kerry animals, these animals were all part of the extended herd at Teagasc, Moorepark, Fermoy, Co. Cork. Limousin and Charlois samples were from Teagasc, Grange, Co. Meath, and Kerry samples were obtained from Duchas, Muckross House, Killarney, Co. Kerry. Blood samples were obtained for all animals and DNA extractions were carried out using the Gentra Capture Column™ (Gentra, UK) system for approximately 200µl of whole blood per animal. Blood was stored at –80^{o}C and DNA was stored at –20^{o}C until further use.

Primer design

Primers were designed as in Table 1 and synthesized by MWG Biotech.

Polymerase Chain Reaction (PCR)

PCR was carried out from a starting template of approximately 200ng of genomic DNA in a final volume of 50µl containing 1X *Taq* DNA polymerase buffer (Invitrogen), 1.5mM $MgCl_2$, 200µM dNTPs (Promega), 0.3µM each primer, and 1U

120 *Taq* polymerase (Invitrogen). The reaction was amplified for 35 cycles. An initial
121 incubation at 95^{o}C for 2min was followed by 35 cycles of 95^{o}C for 1min, Ta for 1min
122 and 72^{o}C for 1min. (Annealing was carried out at 2^{o}C below the melting temperature
123 (Tm) specific for each primer pair). Cycling was performed in a DNA Engine thermal
124 cycler (MJ Research). Two constructs were designed and the corresponding fragments
125 for both the A and B haplotypes were amplified from the genomic DNA of
126 homozygous animals by PCR using a proof-reading Taq polymerase. The primers
127 pAK20 and pAK21 incorporated restriction enzyme sites for *Pst*I and *Xho*I
128 respectively allowing cloning into the plasmid pCMVtkLUC. Six AK constructs
129 (AK1 to AK3 for each allele) were subsequently amplified from the two purified
130 plasmid templates and cloned into a luciferase reporter vector pGL3-Basic. Each
131 fragment used the same downstream primer (pAK21, *Xho*I site) and one of the three
132 upstream primers (pGLAK1-pGLAK3) incorporating a *Mlu*I restriction site. All
133 plasmids were sequenced on construction and again after a number of transfections.
134 Primers pAK06, pAK08, pAK20 – pAK24 were used to amplify κ-casein promoter
135 fragments for sequencing. Primers pAKp1; pAKp4 were used in PCR-RFLP to
136 genotype κ-casein coding sequence variants.

137 ***Restriction digests***

138 Restriction digests were carried out in a final volume of 20μl containing 10μl of PCR
139 product, 1X reaction buffer and 1U restriction enzyme. Restriction enzymes used for
140 screening purposes were *Bcl*I and *Hin*dIII and for cloning purposes were *Pst*I, *Mlu*I
141 and *Xho*I. Reactions were incubated at temperatures recommended by the
142 manufacturer (Roche Diagnostics).

143 ***Sequencing and Bioinformatics***

144 Sequencing of PCR products was carried out by MWG Biotech. The resulting
145 sequences were analysed using the Vector NTI® Suite of software (Invitrogen)
146 Alignment of sequences for all 42 animals was carried out, and potential polymorphic
147 sites identified. Examination of chromatogram sequence files to confirm homozygotic
148 and heterozygotic animals was also performed. Vector NTI® was also used for
149 promoter multi-species alignments: sequences aligned (with accession no.) were
150 bovine (M75887), yak (AF194988), goat (Z33882), horse (AY579426), mouse
151 (AJ309571), rabbit (AJ309572), sheep (L31372) and water buffalo (AJ628346).
152 MatInspector (Genomatix Software GmbH) was used to identify potential
153 transcription factor binding sites in the bovine promoter sequences (Rel. Professional
154 7.4.3 Jul 2006) (Cartharius et al., 2005).

155 ***Generation of truncated promoter luciferase reporter constructs***

156 Three promoter fragments of decreasing length were amplified for each haplotype
157 using primers that incorporated restriction sites at their 5' ends. Promoter fragments
158 AK1, AK2 and AK3 were amplified with primers pGLAK1, pGLAK2 and pGLAK3
159 respectively paired with pAK21 (Table 1) from a common plasmid for each
160 haplotype. Fragment sizes were 505bp, 752bp and 1.3kb and were designated
161 pGLAK1, pGLAK2 and pGLAK3 respectively. Digestion was carried out to create
162 compatible ends for sub-cloning into the pGL3-Basic (Promega) promoter-less
163 expression vector. Ligation reactions were transformed into TAM Ultracomp
164 chemically competent *E. coli* cells (Active Motif Europe) as described by the
165 manufacturer. All constructs were verified by sequencing using the pGL3for and
166 pGL3rev sequencing primers (Table 1)

167 ***Cell culture***

168 Human mammary MCF7 cells (ATCC – HTB22) were cultured in Eagles Minimum
169 Essential Media (Invitrogen) containing 10% (v/v) foetal bovine serum (Invitrogen)
170 and 1% (v/v) non-essential amino acids (Invitrogen). Cells were routinely passaged
171 every 3-4 days by washing with phosphate buffered saline (PBS) and treating with
172 3ml EDTA-trypsin (Sigma) to remove adherent cells. Cells were maintained in a
173 humidified incubator at 37^{o}C and 5% CO_2. Cells were grown to 80% confluency and
174 transiently co-transfected with luciferase reporter vectors using Fugene 6 transfection
175 reagent (Roche Diagnostics). Four independent transfections were carried out. Six-
176 well transfection plates were seeded at a concentration of 3 x 10^5 cells per well and
177 incubated overnight at 37^{o}C and 5% CO_2. Experimental constructs (1μg) were co-
178 transfected with 100ng of pRL-SV40 plasmid (Promega) to control for transfection
179 efficiency. Forty-eight hours after transfection, media was removed from the wells
180 and 300μl Passive Lysis Buffer (Promega) added. Following incubation at room
181 temperature for 10min, samples were either assayed immediately as described below
182 or stored at -20^{o}C for later analysis.

183 ***Luciferase assay***

184 The Dual Luciferase Assay Kit (Promega) was used to measure both Renilla and
185 Firefly luciferase expression by the reporter vectors on a Tecan Spectrafluor Plus
186 luminometer using the Magellan software (Tecan). Mean Firefly luciferase activity
187 values were corrected for variations in transfection efficiency using the corresponding
188 mean Renilla luciferase figures.

189 ***Statistical analysis***

190 Linkage disequilibrium of the promoter haplotype and coding sequence haplotype was
191 analysed by χ2-test: the null hypothesis was that no association existed between
192 haplotypes. Statistical analysis of the reporter construct luciferase values was carried

193 out using either the ANOVA function or individual t-test function of SPSS for
194 Windows (Rel. 12.0.1).

195 **RESULTS**

196 ***Amplification and sequence analysis of the κ- casein gene promoter***

197 Nine bovine breeds were chosen that were considered representative of the Irish herd.
198 Amplification and sequencing of the *κ-casein* promoter region (2072bp) identified
199 three polymorphic sites at positions –514, -426 and –384 (T/G, T/C and T/C). The
200 three SNPs were found to be linked: animals were either homozygous or heterozygous
201 for all three of the base variations, creating three promoter haplotypes - AA, AB or
202 BB. The base change from T to C at –426 destroys a recognition site for the
203 restriction enzyme *Bcl*I, allowing a rapid screen for this base change and hence for the
204 κ-casein promoter haplotypes (Figure 1). All animals were also genotyped for the
205 SNPs present in exon 4 of the *κ-casein* gene to determine whether a relationship
206 exists between promoter haplotypes and the A and B major coding sequence
207 haplotypes. The two major coding region haplotypes were again identified by
208 restriction digestion: a 443bp fragment was amplified and digested with restriction
209 enzyme *Hin*dIII (Nierop Groot et al., 1995). Statistical analysis of these results using
210 χ^2-testing indicated that there was no apparent link between κ-casein promoter and
211 coding sequence haplotypes A and B (P = 0.89). These results, however, do not
212 preclude the possibility that the promoter haplotypes are linked to other *κ-casein*
213 coding sequence genetic haplotypes. The frequency of the BB coding region
214 haplotype is known to be low and this was confirmed here where a frequency of 4.7%
215 among all animals was found. The B allele frequency was 19.4% in dairy breeds,
216 37.5% in dual-purpose breeds and 31.3% in beef breeds. Although the A allele is the
217 dominant allele in all groups of animals (dairy: 71.4%; dual purpose: 62.5%; beef:

68.8%), the AA homozygote is only more frequent than the AB heterozygote in the dairy animals (66.7% against 62.5%) whereas in the non-dairy animals the AB heterozygote is nearly twice as frequent as the AA homozygote (62.5% against 33.3%). This probably reflects the hybrid vigour of the dual purpose animals compared to those specifically bred as high-yielding dairy animals.

The genetic haplotypes for the *κ-casein* gene promoter for all forty two animals screened are listed in Table 2. The frequencies of each haplotype of the *κ-casein* promoter were as follows: AA (69%), AB (21.5%) and BB (9.5%). The BB haplotype was therefore quite rare in comparison to the AA haplotype. In fact the A allele frequency (80%) was again dominant over the B allele frequency (20%). When the breed differences were examined the frequency of the A allele in dairy animals was 88.9% compared with 71.9% in the dual-purpose breeds and 75% in the beef breeds. It was interesting to note that while the allele frequency was similar for the beef and dual purpose animals, the beef animals showed the greater heterozygosity (50% as compared to 18.8%).

Functional analysis of the κ-casein promoter alleles

Two constructs were designed and the corresponding fragments for both the A and B haplotypes were amplified from the genomic DNA of homozygous animals by PCR using a proof-reading Taq polymerase. Six AK constructs (AK1 to AK3 for each allele) were subsequently amplified from the two purified plasmid templates and cloned into the luciferase reporter vector pGL3-Basic. All plasmids were sequenced on construction and again after a number of transfections.

MCF7 cells were transfected (using Fugene 6) with purified plasmid DNA preparations (endotoxin-free) of the six constructs together with a Renilla transfection efficiency control. Forty eight hours after transfection luciferase activity was measured using the Dual Luciferase Assay Kit. Activities were background corrected and adjusted for transfection efficiency. Four independent transfections were carried out and all values were presented as a percentage of the AK1-A value (Figure 2). The AK3 promoter fragment (1.3kb), irrespective of haplotype, generally resulted in the highest luciferase expression of the three constructs. Statistically both the AK3-A and AK3-B alleles were significantly higher than their AK1 counterparts ($P<0.05$). In most instances (7 of 8) luciferase expression levels under the control of the AK3 promoter fragment were also higher than their corresponding AK2 allele in the same experiment but statistical confirmation was lacking.

There were no significant differences detected between the two haplotypes in MCF7 cells but both AK3-B and AK1-B were consistently higher (7 of 8 within experiment comparisons) than their AK3-A and AK1-A counterparts. The closest significant difference was between the AK1-B allele and the AK1-A allele ($P = 0.10$).

DISCUSSION

Two haplotypes of the *κ-casein* promoter, differing in three nucleotide positions, have been identified. These nucleotide changes are present in regions of proposed importance for the regulation of the κ-casein gene expression. One of these base changes destroyed a recognition site for the restriction enzyme *Bcl*I allowing a rapid screen by a PCR/RFLP method. Different frequency patterns for the A allele in dairy (88.8%), dual-purpose (71.9%) and beef (75%) breeds were revealed. Animals were also screened for known genetic haplotypes in exon 4 of the coding sequence of the *κ-casein* gene, however statistical analysis of both promoter and coding sequence haplotypes suggested that no linkage disequilibrium existed between the promoter and coding regions.

Putative transcription factor binding sites were identified in the *κ-casein* promoter sequence from both the literature and a search of transcription factor databases. The three polymorphic sites identified were in regions predicted to be of importance in the regulation of milk protein gene expression (Figure 1). Functional studies of the κ-casein promoter have been carried out previously in mouse cell lines (Adachi et al., 1996; Rijnkels et al., 1995) and have provided evidence of pregnancy-specific and lactation-specific transcription factor binding sites in a segment of the promoter that includes the -384 and -426 polymorphic sites investigated in this study (Adachi et al., 1996). AK3 expression (approximately 1.3kb of promoter) was found to be higher in MCF7 cells than AK1 expression (approximately 500bp of promoter) irrespective of haplotype. This would be consistent with the reported presence of two crucial MGF/STAT5 binding sites highly conserved in six species in the additional stretch of promoter (Gerencser et al., 2002). Similar experiments were carried out in triplicate in

CHO cells (data not shown): although there were no significant differences between the different sized fragments of each allele, AK1 tended to be stronger than AK3 (opposite that of MCF7 cells) and AK1-A was significantly greater than AK2-B. The tissue-specific nature of the activation factor in the longer AK3 fragment is therefore borne out by the absence of activation in CHO-K1 cells. It is also in agreement with recent findings that a 2040bp κ–casein promoter fragment was stronger than a 925bp promoter fragment (Debeljak et al., 2005).

Three linked polymorphisms have been identified in the promoter of the bovine κ–casein gene. If these SNPs occur within transcription factor binding sites they may alter the expression of the *cis*-related transcript and ultimately could have effects on milk yield or milk fat/protein composition. All three polymorphisms were noted in a previous study (Schild et al., 1994) but there was no suggestion that these were linked in any of the breeds they investigated. The promoter constructs were all functional in MCF7 cells (and CHO cells), but although the promoter haplotypes did not reveal a significant difference between promoter haplotypes there was some indication that expression from the haplotypes differed and that haplotype B was expressed more strongly than haplotype A in MCF7 cells.

At positions 1 (-514) and 2 (-426) there are reported (Schild et al., 1994) potential sites for binding of the transcription factor PMF ($\underline{T}GATN_{1\text{-}2}A\mathbf{T}CA$) although the sequences differ from the consensus with an A or a C instead of the T (marked in bold). In both positions the B haplotype with a G (position 1) or a C (position 2) substitution at the (underlined) T position is likely to be less favourable for PMF binding than the A haplotype (Figure 1). Surprisingly the A haplotype is unique at

311 position 1 (T) amongst the published κ–casein promoter sequences (goat, sheep,
312 horse, mouse, water buffalo and rabbit) which without exception have the B
313 haplotype-like G at this position. Of the published sequences there is no consensus at
314 position 2 though all the ruminants in common with the B haplotype share the C at
315 this position – the T is only seen in the A haplotype and the mouse sequence. It should
316 be stated that although the literature on κ–casein suggests that the polymorphisms at
317 positions 1 and 2 overlap with sites for the binding of transcription factor PMF the
318 evidence for this factor binding site comes from studies of the beta casein gene in the
319 non-ruminant mouse (Lee and Oka, 1992). Information on the sequence variations
320 that bind this transcription factor is very scant, nor is it included in the major
321 transcription factor databases.
322
323 At position 3 there is a reported potential AP-2 site (CCCCAGGC) as originally
324 described by Mitchell et al. (1987) and mentioned in relation to kappa-casein by
325 Schild et al. (1994); however, only the inner six nucleotides of the eight match (see
326 figure 1). In this case haplotype B (ACCCAGGT) is expected to bind more tightly
327 than haplotype A (ACTCAGGT). Alignment of the sequences across species suggests
328 that the C found in the B haplotype is the more common nucleotide with only the
329 horse and the yak (like Bos taurus in the Bovidae family) sharing the T with haplotype
330 A. Although CREB, OCT1 and CEBP surround this site, MatInspector does not
331 identify a transcription factor binding site that might be affected by the
332 polymorphism.
333
334 It appears therefore that haplotype A, even though it is much more common in the
335 breeds analysed in this study, is quite unique across species and that haplotype B is

more consistent with other species, especially the ruminants. Indeed haplotype B (G/C/C) is shared by goat, sheep and water buffalo.

Analysis using the MatInspector program (Cartharius et al., 2005) suggests that the major transcription factor binding sites present in the regions around the three polymorphic sites, before selecting for factors known only to be expressed in mammary tissue, are HNF6 and OCTB (position 1), HAML and HOXF, (position 2) and CREB, HOXF, HOMF, PARF, CEBP and OCTB (position 3). Haplotype A also presents a number of unique transcription factor binding sites not identified in haplotype B at one of the polymorphic positions (GFI1, HOXC, HOXF, EVI1) whereas haplotype B has only one unique site not identified in haplotype A (SRF). The identification of a novel site only reflects that the polymorphism has made the sequence more homologous to the consensus matrix for that site where one version (absent) falls below a threshold score (0.75/Optimized) and the other version is above it (present). Of these transcription factor binding sites, analysis suggests (Malewski & Zwierzchowski, 2002) that the most relevant to the expression of milk proteins are HOXF, OCTB, CEBP and EVI1.

Of course computer-aided analysis can only suggest potential transcription factor binding sites. Possibly the most relevant is HOXF which has potential sites in 2 of the 3 positions but has a more optimal site (at position 1) in haplotype A. Similarly haplotype A might favour binding of EVI1. In addition it has been suggested that OCT1 (ATTTG**C**AT) may also bind to position 1 (Schild et al., 1994) although the C (bold) is replaced by G in both bovine haplotypes A and B: haplotype B may bind OCT1 preferentially as the G (underlined) in haplotype A is replaced by a T.

CONCLUSIONS

In summary, haplotype A is the more common form seen in the bovine breeds studied and may potentially be favoured from the standpoint of stronger potential transcription factor binding. However, haplotype B is much more consistent with other known κ–casein genes and the results suggest that, at least in one mammary cell line, gene expression is stronger from this promoter form. It could be argued that the expression pattern may be different in bovine mammary cells, that the effects may have been more pronounced upon hormonal induction, that gene expression might not reflect protein expression, or that stronger expression of κ–casein may not be beneficial to overall milk characteristics, but this initial study does suggest that under these particular *in vitro* conditions expression is favoured by one allele over the other and that the haplotype should be investigated further, with a larger number of animals, to determine whether there is an association of haplotype with milk quality characteristics. The promoter A and B haplotypes do not appear to be associated with the coding region haplotypes A and B, but this does not exclude the possibility that they are linked to another factor that is key to milk quality.

ACKNOWLEDGEMENTS

This work was supported by Enterprise Ireland and by a Teagasc Walsh fellowship to AFK. Thanks to Benoit Houeix and Emma Holian (NUI, Galway) for assistance with the statistical analysis.

382 **REFERENCES**

383 Adachi, T., J.Y. Ahn, K. Yamamoto, N. Aoki, R. Nakamura, T. Matsuda. 1996.
384 Characterization of the bovine kappa-casein gene promoter. Biosci Biotechnol
385 Biochem. 60:1937-1940.

386

387 Alexander, L.J., A.F. Stewart, A.G. Mackinley, T.V Kapelinskaya, T.M. Tkach, and
388 S.I. Gorodetsky. 1988. Isolation and characterisation of the bovine κ-casein gene. Eur.
389 J. Biochem. 178: 395-401.

390

391 Bobe, G., D.C. Beitz, A.E. Freeman, and G.L. Lindberg. 1999. Effect of milk protein
392 genotypes on milk protein composition and its genetic parameter estimates. J. Dairy
393 Sci. 82: 2797-2804.

394

395 Cartharius, K., K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M.
396 Frisch, M. Bayerlein, and T. Werner. 2005. MatInspector and beyond: promoter
397 analysis based on transcription factor binding sites. Bioinformatics. 21: 2933-2942

398

399 Debeljak, M., P. Frajman, T. Lenasi, M. Narat, A. Baldi, and P. Dovc. 2005.
400 Functional analysis of the bovine beta- and kappa-casein gene promoters using
401 homologous mammary gland derived cell line. Arch.Tierz., Dummerstorf 4: 334-345.

402

403 Farrell, H.M. Jr., R. Jimenez-Florez, G.T. Bleck, E.M. Brown, J.E. Butler, L.K.
404 Creamer, C.L. Hicks, C.M. Hollar, K.F. Ng-Kwai-Hang, and H.E. Swaisgood. 2004.

405 Nomenclature of the proteins of cows' milk – sixth revision. J. Dairy Sci. 87: 1641-
406 1674.
407
408 Gerencser, A., E. Barta, S. Boa, P. Kastanis, Z. Bosze, and C.B. Whitelaw. 2002.
409 Comparative analysis on the structural features of the 5' flanking region of kappa-
410 casein genes from six different species. Genet. Sel. Evol. 34: 117-28.
411
412 Kaminski, S. 1996. *Dde*I RFLP at the 5′ region of bovine κ-casein gene. J. Appl.
413 Genet. 37: 173-178.
414
415 Kennedy, J., P. Dillon, K. O'Sullivan, F. Buckley, and M. Rath. 2003. Effect of
416 genetic merit and concentrate supplementation on grass intake and milk production
417 with Holstein-Friesian dairy. J. Dairy Sci. 86: 610-621.
418
419 Lee, C.S., and T. Oka. 1992. Progesterone regulation of a pregnancy-specific
420 transcription repressor to beta-casein gene promoter in mouse mammary gland.
421 Endocrinology. 131: 2257-2262
422
423 Malewski, T., and L. Zwierzchowski. 2002. Genes expressed in cattle mammary
424 gland – computational analysis of 5'-upstream sequences in search for factors
425 conferring a tissue- and stage-specific transcription. Animal Science Papers and
426 Reports. 20: 5-20
427

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.

Marziali, A.S., and K.F. Ng-Kwai-Hang. 1986. Effects of milk composition and genetic polymorphism on coagulating properties of milk. J. Dairy Sci. 69: 1793.

Mercier, J.C., G. Brignon, and B. Ribadeau-Dumas. 1973. Primary structure of bovine κ-casein B. Complete sequence. Eur. J. Biochem. 35: 222-235.

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

Ng-Kwai-Hang, K.F., J.F. Hayes, J.E. Moxley, and H.G. Monardes. 1984. Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. J. Dairy Sci. 67: 835-840.

Nierop Groot, M.N., H. O'Hara, R.J. FitzGerald, and R.P. Ross. 1995. Evaluation of allele-specific primers for rapid identification of bovine κ-casein genotypes by the polymerase chain reaction. Irish J. Ag. Food. Res. 34: 165-173.

Robitaille, G., M. Britten, J. Morisset, and D. Petitclerc. 2005. Polymorphism in the bovine κ-casein (CSN3) gene and the 5' flanking region: sequence analysis of CSN3 A and B alleles. Anim. Genet. 36: 184-185.

Ron, M., O. Yoffe, E. Ezra, J.F. Medrano, and J.I. Weller. 1994. Determination of effects of milk protein genotypes on production traits of Israeli Holsteins. J. Dairy Sci. 77: 1106-1113.

Schaar, J. 1984. Effects of *kappa*-casein genetic variants and lactation number on the renneting properties of individual milks. J.Dairy. Res. 51: 397.

Schild, T. A., V. Wagner, and H. Geldermann. 1994. Variants within the 5′ flanking regions of bovine milk protein genes. I. κ-casein-encoding gene. Theor. Appl. Genet. 89: 116-120.

FIGURES

Table 1. Primers used in this study

Primer	Sequence	Tm
pAK06	CGGAATTCTGGGCATATAAAAGATGG	69.2
pAK08	AACTGCAGCTAAAAGCCAAAAGGAAGAAAAT	69.2
pAK20	CGCGCTGCAGTATTTGATATGGGGAATGTTCAGGCTCA	72.7
pAK21	CGCGCTCGAGTCCACTGTAAGGAACACCCAGGTAAAT	73.8
pAK22	CGCGCTGCAGCTCGTGTTAGCAGTTTGGAAATTCT	71.8
pAK23	CGCGCTCGAGTGGAATGCAGAGGAGGAAATAGAGTTGT	73.8
pAKp1	TGAGCAGGTATCCTAGTTATGGACT	62.8
pAKp4	GTCTTCTTTGATGTCTCCTTAGAG	59.9
pGLAK1	GCGCACGCGTGCATATAAAAGATGGTCAG	68.1
pGLAK2	GCGCACGCGTGAAGAAAGGGGAATCCTCC	72.3
pGLAK3	GCGCACGCGTCTCGTGTTAGCAGTTTGGAAATTC	71.9
pGL3for	CTAGCAAAATAGGCTGTCCC	57.3
pGL3rev	TATGTTTTTGGCGTCTTCC	52.4

1. Restriction site sequences: CTGCAG=*Pst*I, CTCGAG=*Xho*I, ACGCGT=*Mlu*I,

Table 2. *κ-casein* coding sequence and promoter haplotypes

Dairy	CDS haplotype	Promoter haplotype	Dual purpose	CDS haplotype	Promoter haplotype	Beef	CDS haplotype	Promoter haplotype
HF	AB	AA	N	AB	AA	C	AA	AB
HF	AA	AB	N	BB	AA	C	AB	AA
HF	AA	AA	N	AB	AA	L	AA	AA
HF	AB	AA	N	AB	AA	L	AB	AB
HF	BB	AA	M	AA	BB	L	AB	AA
HF	AA	AA	M	AB	AA	L	AB	AB
HF	AA	AA	M	AB	AA	L	AB	AA
HF	AA	AA	M	AA	BB	L	AA	AB
IF	AA	AA	K	AB	BB			
IF	AA	AA	K	AB	AA			
IF	AB	AA	K	AB	AB			
IF	AB	BB	K	AB	AB			
NR	AA	AB	K	AB	AA			
NR	AA	AA	K	AA	AA			
DF	AA	AA	K	AA	AB			
DF	AA	AA	K	AA	AA			
DF	AA	AA						
DF	AB	AA						
Allele frequency								
A	80.6	88.8		62.5	71.9		68.8	75.0
B	19.4	11.1		37.5	28.1		31.3	25.0
Haplotype frequency								
AA	66.7	83.3		31.3	62.5		37.5	50
AB	27.8	11.1		62.5	18.8		62.5	50
BB	5.6	5.6		6.3	18.8		0	0

1. Breed Abbreviations. HF = Holstein Friesian, IF = Irish Friesian, NR = Norwegian Red, DF = Dutch Friesian, N = Normande, M = Montbeliarde, K = Kerry, C = Charlois, L = Limousin.

Figure 1. *κ-casein* sequence haplotypes and RFLP screen

A

```
                                          BclI
                  -514                  -426                  -384
Haplotype A:  CATATTT TGATTTAACA- GAGGATGATCAAC-AATTACTTCATACTCAGGT
Haplotype B:     ATTT GGATTTAACA- GAGGACGATCAAC-AATTACTTCATACCCAGGT

Potential transcription factors
PMF, PMF, AP-2        TGATnnATCA    TGATnATCA              CCCCAGGC
GF1, EVI1, CREB         GATT          ganGATG      tnncTTCA
HOXC                 t TGATttaa
HOXF                   tGATTnaa                  aATTAnt
SRF1           CAnAnnn gg
OCT1               ATTTGCAT
```

B

```
                                                                MGF

Bovine     CATATTTTGATTTAACA---GAGGATGATCAAC---CTTCATACTCAGGTTCTTGAAA
Yak        -------------------------------------CTTCATACTCAGGTTCTTGAAA
Goat       CATATTTGGATTTAACA---GAGGACGACCAAC---CTTCATACCCAGGTTCTTGAAA
Horse      CCTATTTGGATTTGACT---GAAGAATACCAAC---CTTGCTATTCAGGTTCTTAAAC
Mouse      CATGCATGAATTTAACT---AACAATGACCAA----CTTGAAGACATAGTTCTTCAAG
Rabbit     TATATAAGAATTTAACA---AAAGAAAACCAAT---CTTAACATCTAAGTTCTTCAAC
Sheep      CATATTTGGATTTAACA---GAGGACGACCAAC---CTTCATACCCAGGCTCTTGAAG
W.buffalo  CATATTTGGATTTAACA---GAGGACGACCAAC---CTTCATACCCAGGTTCTTGAAA
                  -514                  -426                  -384
```

C

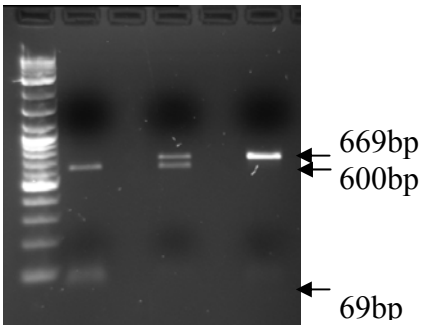

Figure 2. *κ-casein* promoter activity in MCF7 cells

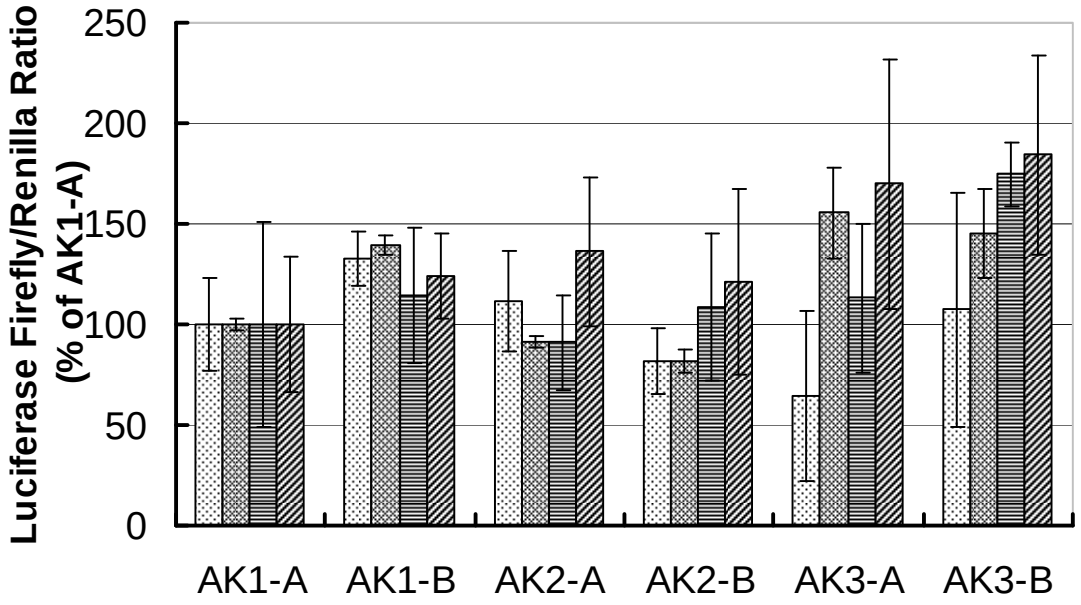

FIGURE LEGENDS

Table 1: Primers used in this study.

Table 2. Bovine breeds and *κ-casein* coding sequence and promoter haplotypes.

Figure 1. Promoter haplotypes of *κ-casein* in relation to potential transcription factor binding sites

A. Haplotypes (A & B) in the bovine *κ-casein* promoter sequence are aligned across the three linked polymorphic sites (SNPs highlighted). Transcription factor binding sites potentially affected by the polymorphism are given below the sequence and the nucleotide position is indicated about sequence. The *Bcl*I site used for the RFLP screen (see C) is marked. B. Alignment across the three linked polymorphic sites of *κ-casein* promoter sequences from other species (for accessions nos. see Materials and Methods). The neighbouring MGF binding site is also marked in bold.

Figure 2. *κ-casein* promoter activity in MCF7 cells

Plasmids AK1 (505bp promoter), AK2 (752bp promoter) and AK3 (1.3kb promoter) of both haplotypes (A & B) carrying the Firefly luciferase reporter under the control of the various *κ-casein* promoter fragments were transfected into MCF7 cells together with a Renilla luciferase control vector. Luciferase values were expressed as a ratio of Firefly to Renilla and corrected to the AK1-A value (100%). Results are from 4 independent experiments. Error bars show standard deviations for triplicates assayed in each experiment.