

Available online at <u>http://www.ijabbr.com</u>

International journal of Advanced Biological and Biomedical Research

Volume 2, Issue 5, 2014: 1726-1733



Mapping of transcription factor binding Region of kappa casein (CSN3) gene in Iranian Bacterianus and Dromedaries camels

A.A. Khabiri^{*}, M. Tahmoorespur, M.R. Nassiri, M.H. Sekhavati

Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran.

ABSTRACT

 κ -casein is a glycosilated protein in mammalian milk that plays an essential role in the milk micelles. Control of κ -casein expression reflects this essential role, although an understanding of the mechanisms involved lags behind that of the other milk protein genes. Transcriptional regulation, a first mechanism for controlling the development of organisms, is carried out by transcription factors binding sites (TFBSs). We determined the 5' region for Iranian Dromedaries and bactrianus camel κ -casein gene. We also determine their regulatory TFBSs regions. This region contained 14 potential TFBSs The four of them C/EBP-α, Oct1, MGF/STAT5 and TPB have almost perfectly conserved.

Key words: κ-casein, Bioinformatics analysis, Camels, Transcription factors

INTRODUCTION

Kappa casein (CSN3) is the milk protein that determines the size and specific function of milk micelles and its cleavage by chymosin is responsible for milk coagulation. (Yahvaoui et al., 2003). Caseins are distinguished into four protein fractions, namely, alpha S1 (39-46% of total caseins), alpha S2 (8-11%), beta (25-35%), and kappa (8-20%) encoded by genes: CSN1S1, CSN2, CSN1S2 and CSN3 respectively (Farrell et al, 2004). Variation in a genes promoter may change its expression or alter the amino acid sequence and affect the functional properties of the protein. k-casein differs from other caseins in its solubility over a broad range of calciumion concentrations and contains a hydrophilic C-terminal region (Yahyaoui et al., 2003). the results on association studies involving only coding region variants are not always consistent (Heck et al, 2009) and this might be attributed to the presence of intragenic haplotypic combination of variants in the regulatory and noncoding regions (Prinzenberg et al, 2003) Moreover, casein gene expression is also known to be differentially regulated by hormones and most of the potential hormone receptor binding sites occur within the 5'-flanking region of casein genes (Prinzenberg et al,. 2003). Thus, mutations at these regulatory regions might also have enduring effect on milk protein gene regulation at transcriptional level (Prinzenberg et al., 2003 and Szymanowska et al., 2004) either individually or as interior intragenic haplotypes. For CSN1S1, mutations in the promoter region have been reported to influence the protein-coding efficiency of the relevant structural gene by changing the binding affinity towards their respective nuclear transcription factors (TFs) (Malewski et al., 1998 and M. Szymanowska et al., 2004) and can thus be considered as functional candidate for milk protein content.

Sequence variation within Alpha S1 5'-flanking region has been widely studied in several species like cattle involving mainly B. taurus (Prinzenberg et al, 2005 and Kuss et al, 2005) yak (Bai et al, 2010), buffalo (Chianese et al, 2009), goat (Ramunno et al, 2004), and sheep (Bhure et al, 2008). However, to date, very few studies are available on the non-coding region of kappa casein in camel. Further, due to evolutionary divergence, Iranian Bacterianus and Dromedaries are expected to have variations in the candidate genes related to dairy traits. Keeping in view the scanty information available Iranian Bacterianus and Dromedaries and other breed of camel, the present study was aimed to first: sequence the promoter of kappa casein; second: search for TFBSs region on Iranian Bacterianus and Dromedaries; third: identify homology in the regulatory domains for camel's kappa casein from different mammalian spaces.

MATERIALS AND MTHODS

Blood samples and DNA extraction

Blood samples were obtained from 5 Bacterian and 10 Dromedary camel randomly from Ardebil Camel Rearing Station and Mashhad Abattoir, respectively. Blood samples were immediately applied near the ice, was transported to the laboratory and stored until DNA isolation. DNA was extracted from 250 μ l of blood, using a commercial kit (Bioneer Cat .NO .K-3032, USA) following the manufacturer's protocol. Restriction products were analyzed by electrophoresis on 0.8% agarose gel and the purity of the obtained DNA was verified by Nano Drop ND-2000 spectrophotometer (Thermo, USA).

PCR amplification

DNA regions (-1162 to +50 of CSN3 gene cluster) from the promoter region of CSN3 gene were amplified using the Personal CyclerTM thermo cycler (Biometra, Germany). The specific primers were designed by using Primer Premier 5, according to the available nucleotide sequences on the NCBI GenBank (EMBL ID HE863813) database and were synthesized commercially. The specific primers were as follows:

CSN3- F: 5'- TTGAAACTCTGCCATCTTTCTC-3'

CSN3- R: 5'- TGTGCCTGTCAGGTCTTGC-3'

PCR reactions were performed in a final volume of 25 μ l with the reaction mixture containing 2.5 μ l of 10X PCR buffer, 2 μ l MgCl2 (mM) and 2 μ l dNTPs (Mix), 1 μ l of the DNA solution (50 to 100 ng/ μ l), 1.5 μ l of mix primer (5 pmol/ μ l), 0.125 U/ μ l of *EX Taq* DNA polymerase (Takara, Japan), and some deionized water making up a final volume of 25 μ l. PCR program was performed with an initial denaturation step at 94°C for 10 min followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 62°C for 30 sec and extension at 72°C for 45 sec, and a final extension at 72°C for 10 min. The PCR correctness was assessed by electrophoresis of each sample (4 μ l) on 0.8 % agarose gel and was sequenced by specific designed primers (Macro Gen, South Korea).

Bioinformatics tools

Then, SNP discovery, homology searches, comparison among sequences, and multiple alignments were accomplished using Basic Local Alignment Search Tool (BLAST) and CLC Main workbench version 5.5.

Interspersed elements were found by these two Web Server: (<u>http://www.softberry.com/berry.phtml</u>), (<u>http://alggen.lsi.upc.edu/recerca/menu_recerca.html</u>).

RESULTS AND DISCUSSION

The entire sequences (1212 bp) of the Iranian Bactrianus and Dromedaries CSN3 gene plus the analysis of 5' flanking region was amplified. The accuracy of these fragments was visualized on using electrophoresis 0.8% agarose gel electrophoresis (Figure 1).

camelus dromedarius					camelus bactrianus								
C-	М	1	2	3	4	5	C-	Μ	1	2	3	4	5
	100bp							100bp					
	=		12	12bj)			=		12	12bp		
	=	-	-	-	-	-					-	-	-
								10					
	=												
	-												

Figure1: PCR products of promoter region of CSN3gene on agarose 0.8%

Sequences have been published on NCBI gene database or GenBank under accession numbers: KJ755350 and KJ755349 for Dromedarius and Bactrianus, respectively.

A homology and similar locations of most putative transcription binding sites were already reported among ovine, caprine, bovine, murine and rabbit κ -casein promoters (Gerencser et al., 2002), horse, donkey and zebra CSN3 5' flanking region (Hobor et al., 2008) C.dromedarius, C.bactrianus, E.asinus, B.taurus CSN3 5' flanking region (Pauciullo et al, 2013). In this work we performed a comparative analysis between two sequence that obtain in this work with two published sequences of the camels (EMBL ID: HE863813; AJ409280) and B.taurus (EMBL ID: AY380228).

Using manual and bioinformatics searches, 14 potential TFBSs were identified within about 1050 bp of the CSN3 gene promoter region. While, Pauciullo et al, 2013 report 25 putative consensus sequences for transcription factor binding sites in their study, that some of this region are similar with present work. The TATA-box location was performed by using analysis the available sequences were identified (figure1).

	20		40		60
A CTCCTA	ACTTTCTAAG	ACATGAAATA	CTTCTAAGAA	AAAGAGGACT	ACAACACAAT
.TCATG.	A	. T		A .	
	80		100		120
GCTGTTATTA	TTTCTTGTTT	GTGTTCTGGC	AGTCCT - CAT	TGATTCCTGT	AAATACCTCA
C					
	Α	A . C	A	CA	A
	140 1		GT-1 100	C	/ EBP-a 180
ATCTGGCCAA		GCTAAGGTGA	AGTCACAGTT	AACATTITT	TCCCCAGAGA
				· · · · · · · · · · · ·	-
GTG.		. AA AA	. A		GT G
Oct-1	200	AP1	220	Oct-1	240
	AAAGAAAATA		AATCATCTAA		TGGTTAGCTA
	••••••••••••••••••••••••••••••••••••••		'		
A	C . G	тст	T.GG	. A	AG
	260	YYI	280		300
TATTTTACCA	AAATATCTCC			ATATACTTTG	TAAGTTAGAA
		····			
. .	C.C	TG	. C A	T	C
	320		340		360
	T				F/STAT5
TGAGCTGTCT	TTGAAACAAA				TTGTATTTCC
				· · · · · · · · · · · · · · · · · · ·	<u></u>
.AC	G .				CA
CGAATGTTTC	380		400		420
COMMISTIC	ī	TTGAAATTTA	T		420 I
	ACCTGTATTA		AAAATCTAAA	TGCA	
	ī	ττςααάτττα	T		420 I
	ACCTGTATTA		ААААТСТААА	TGCA	420 1
AA	ACCTGTATTA		ААААТСТААА	TGCA	420 1
AA	ACCTGTATTA		AAAATCTAAA	TGCA	420 I AAATGCTGAA 480
AA	ACCTGTATTA		ААААТСТААА	TGCA	420 1
AA	ACCTGTATTA		AAAATCTAAA	TGCA	420 1
AA	ACCTGTATTA		AAAATCTAAA	TGCA	420 1 AAATGCTGAA 480 1
AA	ACCTGTATTA		AAAATCTAAA CG 460 1	TGCA	420 1 AAATGCTGAA 480 1
AA GAAGATGTGA	ACCTGTATTA		AAAATCTAAA CG 460 1 ACTGTTGGTG	TGCA	420 1 AAATGCTGAA 480 1
AA GAAGATGTGA	ACCTGTATTA		AAAATCTAAA CG 460 1	TGCA	420 1 AAATGCTGAA 480 1 CTGGCAGAGC 540
AA GAAGATGTGA	ACCTGTATTA		AAAATCTAAA CG 460 1 ACTGTTGGTG	TGCA	420 1 AAATGCTGAA 480 1 CTGGCAGAGC 540
AA GAAGATGTGA	ACCTGTATTA		AAAATCTAAA CG 460 1 ACTGTTGGTG	TGCA	420 1 AAATGCTGAA 480 1 CTGGCAGAGC 540
AA GAAGATGTGA	ACCTGTATTA		AAAATCTAAA CG 460 1 ACTGTTGGTG	TGCA	420 1 AAATGCTGAA 480 1 CTGGCAGAGC CTGGCAGAGC 540 1 CTGGCAGAGC
AA GAAGATGTGA	ACCTGTATTA		AAAATCTAAA CG CG 460 1 ACTGTTGGTG 520 1	TGCA	420 1 AAATGCTGAA 480 1 CTGGCAGAGC 540

GCACTAACAC CTATA-BOXI ^{6D} 1.180 1.200 GCACTAACAC CCTTTAATTA GTCTCCAGTT ATTTACCTTG GGTCTTGAAT GA		560		580		600
C.		AAAAGGTTAA		AGTATACATT		ACCATGAACT 461
C						
TACTGAAGAA TAAAATGCAA ATGGCTTIGC TACTTIATAT TGATTAAAAT ATTTCATAT 521 TACTGAAGAA TAAAATGCAA ATGGCTTIGC TACTTIATAT TGATTAAAAT ATTTCATAT 521 C A. GC TAGTTAAAATGCAA ATGGCTTIGC TACTTIATAT TGATTAAAAT ATTTCATAT 521 TAGATTTAAC TTGTATGAAG TIGCTCTIGG GTATTAAGA AATGGTCTGT TITCTAATTI 581 S51 TAGATTTAAC TIGTATGAAG TIGCTCTIGG GTATTAAGA AATGGTCTGT TITCTAATTI 581 S51 TAGATTTAAC TIGTATGAAG TIGCTCTIGG GTATTAAGA AATGGTCGT TITCTAATTI 581 S51 T. T. CTG AGATGACAA ATGATGACGA AAGATGACCA ACCACAGCCC ATAATATATA 641 S51 C./EPBs TO TO TO GTTGTGGACCA GCTATTATCA TITTAACAAT ACCTTCCAAT GAAGAATGACA GTATAATAT 1000 TO TO GATTGGGACA GCTATTATCA TITTAAAAAT CCTICCAAT TCAAATAGGT GGAACTCGTT 760 TO TO TO GATTGAGAAT GCAATTAATT TITTAAAAAT ACCTICCAAT TCAAATAGGT GGAACTCGTT 760 TO TO TO GATTGAGAAT GCAATTAATT TITTAAAAT ACCTICCAAT TCAAATAGGT GGAACTCGTT 760 TO TO TO GATTGAGAAT GCAATTAATT TITTAAAAAT ACCTAAAATAAT CCTICCAAT TCAAATAGGT GGAACTCGTT 760 TO TO GC.A. G.T. A G. T. IC A T. G. T. T. A						
TACTGAAGAA TAAAATGCAA ATGCCTITGC TACTITATAT TGATTAAAAT ATTICATATS 521 TACTGAAGAA TAAAATGCAA ATGCCTITGC TACTITATAT TGATTAAAAT ATTICATATS 521 S21 C. A. A.CC. 645 700 727 TACTGAAGAA TAAAATGCAA ACC. 645 645 700 720 TAGATTAAC TIGTATGAAG TIGCTCTTGG GTATTTAAGA AATGGTCTGT TITCTAATT 581 700 71 71 A	c				-GG	
C. A. A. CC. 521 TAGATTTAAC TIGTATGAAG TIGCTCTIGG GTATTTAAGA AATGGTCTGT TITCTAATTS 581 581 581 T. A. A. A. G. C. A. A. G. A. G. C. A. A. C. A. G. G. C./EPBa 70 TCTGGAAGG AAGATGACCA ACCACAGCCC ATAATATA 641 641 641 641 C.A.A.A.TGATTGTAA TICTGGAAGG AAGATGACCA ACCACAGCCC ATAATATA 641 669 700 670 670 C.AG.T. A. C. A. G. 671 670 GATTGTGCACCA GCTATTATCA TITTACAATA ACCTTCCAAT TCCAAATAGGT GGAACTCGTT 760 760 770 770 GATTGTGCACCA GCTATTATCA TITTACAATA CCTTCCAAT TCCAAATAGGT GGAACTCGTT 770 770 770 770 GATTGAGAAT GCAATTAATT TITTACAATA CCTCACTATAT TCAAATAGGT GGAACTCGTT 770 770 770 770 GATTGAGAAT GCAATTAATT TITTACAATA CCTCAATATATAGGT GGAACA CTCTGTTTT 770 770 770 770 GATTGAGAAT GCAATTAATT TITTACAATA CCTCAATATATAGGT GGAAACACTCTCTTTT 770 770 770 GATTGAGAAT GCAATTAATAT TITTTACAATA CC A. <t< th=""><th></th><th>ĩ</th><th></th><th>1</th><th></th><th>Oct-1</th></t<>		ĩ		1		Oct-1
C. A. A.CC. 70 77 TAGATTTAAC TIGTATGAAG TIGCTCTTGG GTATTTAAGA AATGGTCTGT TITCTAATT 581 77 77 TAGATTTAAC TIGTATGAAG TIGCTCTTGG GTATTTAAGA AATGGTCTGT TITCTAATT 581 581 77 C. A. A. C. A. G 70 C./EPB-a 70 70 70 70 70 CTAAAAATCA ACATTGTAA TICTGGAAGG AAGATGACCA ACCACAGCCC ATAATATAT 641 641 T. T.CTG G.CC A. A.G. 670 70 GTAGGACCA GCTATTATCA TITTAACAAT ACCTTCCAAT 661 700 70 GTAGGACCA GCTATTATCA TITTAACAAT ACCTTCCAAT TAGATGAAAAAGGT GGAACTGGT 760 700 70 GATGGGACA GCAATTAATT TITTAAAAAC GCTCTATAAAAAAGGT GGAATGACCA TTTCCAAAAAAGGT GGAATGACCA 70 700 GCAACI """"""""""""""""""""""""""""""""""""						
C. A. C. 60 70 72 TAGATTTAAC ITGTATGAAG TTGCTCTIGG GTATTTAAGA AATGGTCTGT TTTCTAATT 561 581 581 581 T. C. A. A. C. 581 T. C. A. A. C. 581 T. C. A. A. C. A. G. CIANAATCA ATGATTGTAA TTCTGGAAGG AAGATGACCA ACCACAGCCC ATAATATA 641 641 641 641 T. T. C.TG GC A. A. G. 760 GTAGATACA ATGATTGTAA TTCTGGAAGG AAGATGACCA ACCACAGCCC ATAATATAT 641 G. 641 641 T. T.C.TG GC A. A. G. T. 641 GTGTGTGACCA GCTATTATCA TTTTACAAAT GCCTTCCAAT TTAAATAGA GGAACTCCTT 760 760 GTGTGTGGACCA GCTATTATCA TTTTAACAAT ACCTCAAT 776 776 776 GCAA. T. GCAATATAT TTTTAACAAT GCATCCAAT TT 776						
TAGATTTAAC TIGTATGAAG TIGCTCTTGG GTATTTAAGA AATGGTCTGT TITCTAATT 561				• • • • • • • • • • •		
Sett 551 .T. A. A. A. C. A. A. G. A.		680		700		
T. A. A. A. A. S79 C./EPB.a T0 T0 T0 T0 T0 C1AAAAATCA_ATGATTGTAA TICTGGAAGG AGATGACCA ACCACAGCCC ATAATATATA 641 T.<.T.CTG	TAGATTTAAC	TTGTATGAAG	TTGCTCTTGG			
.T A AAT C. A. A. G A G. 700 C/EPE _a CTAAAAAICA ATGATIGTAA TTCTGGAAGG AAGATGACCA ACCACAGCCC ATATATATA 641 						515000000000000000000000000000000000000
C/EPEx 740 740 740 740 CTAAAAATCA_ATGATTGTAA TTCTGGAAGG AAGATGACCA ACCACAGCCC ATAATATATA 641 641 641						
C / EPB-a CTAAAAATCA ATGATTGTAA TTCTGGAAGG AAGATGACCA ACCACAGCCC ATAATATATA 641 	.T		AT	. C A A .	GA	G 700
		ĩ		760 I		780
GATTGAGAAT GCAATTAATT TITTAAAAAT ACCTTCCAAT TCAAATAGGT GGAACTCGT 760 GTTGTGACCA GCTATTATCA TITTAACAAT ACCTTCCAAT TCAAATAGGT GGAACTCGT 760 	CTAAAAATCA	ATGATTGTAA				
TT.CTG						
B00 F00 F						639
GTTGTGACCA GCTATTATCA TTTTAACAAT ACCTTCCAAT TCAAATAGGT GGAACTCGTT 760 GATTGAGAAT GCAATTAATT TTTCAAAACT CCTCTATAT TTTTCATAAA ACTAAAAT 820 GATTGAGAAT GCAATTAATT TTTTAAAAACT CCTCTATATT TTTTCATAAA ACTAAAAAT 820 GATTGAGAAT GCAATTAATT TTTTAAAAACT CCTCTATATT TTTTCATAAA ACTAAAAAT 820 GATTGAGAAT GCAATTAATT TTTTAAAACT CCTCTATATT TTTTCATAAA ACTAAAAAAT 820 GATTGAGAAT GCAATTAATT TTTTAAAACT CCTCTATATT TTTTCATAAA ACTAAAAAAT 820 GCATTCATTA AAAGGTGGA TAGTCTCTTT TCAAGCTTTA TAAATGACAA CTCTGTTTCC 880 SCARGI TTTA AAAAGGTGGA TAGTCTCTTT TCAAGCTTTA TAAATGACAA CTCTGTTTCC 880 MAAAGGTGGA TAGTCTCTTT TCAAGCTTTA TAAATGACAA CTCTGTTTCC 880 SCARG.CT.AC. GT.A. A. A. A. A. TCCTATGCAC TTCACTAACC ACAACTGACA CAAAGGATGAC TCTGCTATCG TCAAATCTTT 940 1,40 1,060 TCCTATGCAC TTCACTAACC ACAACTGACA CAAAGGATGAC TCTGCTATCG TCAAATCTTT 940	TT.CTC	GGC	AA.	. G T		G 760
		860		880)	900
A. C.A. G.T. A. C.T. TC A. T. G. TA. T. A. 758 GATTGAGAAT GCAATTAATT TITTAAAAACT CCTCTATATT TTTTCATAAA A. A. TG CA. T. A. TA. C. A. C. C. CATAATT TTTTCATAAA CArG1 900 CATTGAGAAT GCAATTAATT CArG1 900 CAGCT.AC. A. A. CAG.CT.AC. 900 1,040 1,040 1,040 1,040 1,040 1,040 1,040 1,040 1,040 1,040	GTTGTGACCA	GCTATTATCA	TTTTAACAAT	ACCTTCCAAT	TCAAATAGGT	GGAACTCGTT 760
A. A. C. T. T.C. A						
.C.AG.T. AC.T.TC AT. .G. TA TA	•••	ΑΑ				
GATTGAGAAT GCAATTAATT TTTTAAAAACT CCTCTATATT TTTTCATAAA ACATAAAAA 820 818						
GATTGAGAAT GCAATTAATT TITTAAAAACT CCTCTATATT TITTCATAAA ACATAAAAAT 820 820		920		940		
A.A.TG CA.T.A.TA. C.A.G. A.A.TG CA.T.A.TA. C.A.G. CATT CATT TPB ¹⁰⁰⁰ 1020 GCATTCTTTA AAAAGGTGGA TAGTCTCTTT TCAAGCTTTA TAAATGACAA CTCTGTTTCC 880 1020 1020 1020 GCATTCTTTA AAAAGGTGGA TAGTCTCTTT TCAAGCTTTA TAAATGACAA CTCTGTTTCC 880 880 880 880 CAGCT.AC. GT.A. A. GCAATGACAA CTCACTATA AAAGGTGAA A. 878 CAG.CT.AC. GT.A. A. GCAATGACAA CTCACTAACC ACAACTGACA CAAAGATGAC TCTGCTATCG TCAAATCTTT 940 1000 TCCTATGCAC TTCACTAACC ACAACTGACA CAAAGATGAC TCTGCTATCG TCAAATCTTT 940 1000 1000 938	GATTGAGAAT	GCAATTAATT	TTTTAAAACT	CCTCTATATT	TTTTCATAAA	ACATAAAAAT 820
A.A.TG CA.T.A.TA.CA. CCCATAGT.938 CARCI 980 TPB ¹⁰⁰ 1020 GCATTCTTTA AAAAGGTGGA TAGTCTCTTT TCAAGCTTA TAAATGACAA CTCTGTTTCC 880					.	
A.A.TG CA.T.A.TACA. C. CATAGT. 938 CARCI 900 TPB ¹⁰⁰⁰ 1020 GCATTCTTTA AAAAGGTGGA TAGTCTCTTT TCAAGCTTTA TAAATGACAA CTCTGTTTCC 880 880 8878 GCATCCTTTA AAAAGGTGGA TAGTCTCTTT TCAAGCTTTA TAAATGACAA CTCTGTTTCC 880 876 GCATTCTTTA AAAAGGTGGA TAGTCTCTTT TCAAGCTTTA TAAATGACAA CTCTGTTTCC 880 876 GCATCCTTTA CGTA. AGA.GATCGA. 878 CAG.CT.ACGTA. AGA.GATCGA. 998 1.040 1.060 1.080 1 1.040 1.060 1.080 1 1.040 1.060 1.080 1 1.040 1.060 1.080 1 1.040 1.060 1.080 1 1.100 1.120 1.140 1 1.120 1.140 1.140 1 1 1.120 1.140 1 1.120 1.140 1.140 1 1.120 1.140 1.140 1 1.120 1.140 1.140 1 1.120 1.140 1.140 1	• • • • • • • • • • •	•••••				
GCATTCTTTA AAAAGGTGGA TAGTCTCTTT TCAAGCTTTA TAAATGACAA CTCTGTTTCC 880	A.A.TG	CA.T.A.TA.				
GCATTCTTTA AAAAGGTGGA TAGTCTCTTT TCAAGCTTTA TAAATGACAA CTCTGTTTCC 880	C1			TDD ^{1,00}	0	1,020
			TAGTCTCTTT			CTCTGTTTCC 880
CAG.CT.AC. GT.A. AG. ATCG. GATCG. ATCG. ATGG. ATGG. <t< td=""><td> </td><td>· · · · · · · · · · · · · · ·</td><td></td><td>· · · · · · · · · · ·</td><td></td><td></td></t<>		· · · · · · · · · · · · · · ·		· · · · · · · · · · ·		
CAG.CT.AC. GT.A. AG. ATCG. A998 1,040 1,060 1,060 1,060 1 TCCTATGCAC TTCACTAACC ACAACTGACA CAAAGATGAC TCTGCTATCG TCAAATCTT						
1,040 1,060 1,060 1,080 TCCTATGCAC TTCACTAACC ACAACTGACA CAAAGATGAC TCTGCTATCG TCAAATCTTT 940						
TCCTATGCAC TTCACTAACC ACAACTGACA CAAAGATGAC TCTGCTATCG TCAAATCTT 940 940 938 938 938 938 938 CCTT C.T GAG TG TG.C G.C. G.G. 1058 1,100 1,120 1,140 1.140 1.140 1.140 1.140 CCTTTTTGTC ATCTTCCTAT TGGGTGTAAT GTAGGAGGAA GATAAATCAC ATGATGCAAA 1000 998						
940 938 938 938 938 938 938 938 938		1				1
938	TECTATGEAC					
1,100 1,120 1,140 CCTTTTTGTC ATCTTCCTAT TGGGTGTAAT GTAGGAGGGAA GATAAATCAC ATGATGCAAA 1000						
CCTTTTTGTC ATCTTCCTAT TGGGTGTAAT GTAGGAGGAA GATAAATCAC ATGATGCAAA 1000 998	CT		GAGTG			
1000 998		1		1		1
998	CCTTTTTGTC	ATCTTCCTAT	TGGGTGTAAT			
C. C.						
GCACTAACAC CC TATA-BOX1 ⁽⁶⁾ 1.180 1.200 GCACTAACAC CC TTTAATTA GTCTCCAGTT ATTTACCTTG GGCTCTGAAT GA 1052 A I I I I						
IAIA-BUXI GCACTAACAC CCTTTAATTA GTCTCCAGTT ATTTACCTTG GGTCTTGAAT GA 1052 1046 C. G. 1046 C. G.	C		.			
1045 A 1052 C. dromedarius this study 1045 >99% C. Bactrianus this study		TATA-BOX160		1,180	0	1,200
A	GCACTAACAC	CCTTTAATTA	GTCTCCAGTT	ATTTACCTTG		
A						
AGCCT. ACAGTGGAAA 1177 1052 C. dromedarius this study 1045 >99% C. Bactrianus this study		1000 CH295 C205 C305 ACC 3				
1052 C. dromedarius this study 1045 >99% C. Bactrianus this study	A					
1045 >99% C. Bactrianus this study						
1043 >99% C. dromedarius (AJ409280) GGCC 1181 >65% B. Taurus (AY380228)	···· 1045	>99% C. Bactrianus >99% C. dromedari >99% C. dromedari	this study us (HE 863813) us (AJ409280)			

Figure1. Alignment between the nucleotide sequences of the promoter region dromedarius and bactrianus CSN3 gene with the published sequences of the camels (EMBL ID: AJ409280, HE863813) and cattle (EMBL ID: AY380228) sequences. The boxes represent show TATA-box.

The initial and final regions (approximately 200, 900-1050 bp) upstream of the transcription start site was found to be the least conserved within the entire promoter. The adjacent upstream region of

approximately 800 bp contained form 14 TFBSs that found, 4 putative transcription factor binding sites almost perfectly conserved: three Octamer bind protein (Oct-1), one Mammary Gland Factor/STAT5 (MGF/STAT5), two enhancer bind protein (C/EBP-a), and two Tata Binding Protein (TBP). The MGF/STAT5 and (C/EBP- α) sites were reported to be very well conserved in more than 9 species (Pauciullo et al., 2013; Gerencser et al., 2002; Hobor et al., 2008; Malewski, 1998). Here confirm them also in camels increasing the probability that these factors are necessary for the regulation of the gene expression (Pauciullo et al., 2013). It interacts directly with STAT 5, reported as a crucial functional activator of casein gene expression (Vanselow et al., 2006). Oct-1 is not known to be a strong transcriptional activator by itself, but in conjunction with other co-activators. For instance, Oct-1 and STAT5 are both involved in the hormonal induction of casein gene expression (Zhao et al. 2002). They bind DNA cooperatively and they interact in the stabilization of the transcription complex. However, Oct-1 can also interact with the TBP components (Zwilling et al., 1994). Yin Yang 1 (YY1) activator protein 1 (AP1) were conserved in equids (Pauciullo et al., 2013; Hobor et al., 2008). Nuclear factor 1 (NF1) conserved among primates and equids (Pauciullo et al., 2013; Hobor et al., 2008). This site found in the camel promoter were identied as common motifs in 28 milk protein gene (Akos et al, 2001). For GT1 and CArG1 are not available reports prior to this study.

The table1 shows their location, and sequences of this TFBSs.

Factor	Iranian dromedarius and bactrianus camel
GT_1	-1000: AAGGTGAAGTCACAGTTA
C / EBP-α	-978: ATT TTTTTTCCCC AGA
	-422: TCTAAAAATCAATG
Oct-1	-963: AAT GTATACA AAA
	-923: TAA GCAAATTAT TTG
	-493: AAT ATTTCATATT TAG
AP1	-946: AAT ATTCTTC TCT
	-604: GCA ATTGATTC ATA
MGF / STAT5	-796: TTA ATTTTGTATT TCC
NHF-1	-507: TAT TATTGATTAA AAT
TPB	-195: ATA AAACATAAAA AAT
	-147: CAAGCTTTATAAATGA
CArG1	-180: ATT CTTTAAAAG GTG
YY1	-888: AGC CCATA ATA

Table1. Occurrence of putative transcriptions factor binding sites in the 5' region of Iranian dromedaries and bactrianus camel κ -casein. Positions are relative to the TATA boxes

CONCLUSION

We performed a genome-wide search transcription factor binding sites in Iranian dromedaries and bactrianus camel and identified a total of 14 TFBSs. The distribution of these regulatory motifs along promoter region sequences, which allowed us to accurately detect DNA regulatory

sequences that influence gene expression. Our work greatly expanded our knowledge of regulatory sites in camel and is a valuable step toward building a genome-wide regulatory network of camel.

REFERENCES

Yahyaoui M.H., Angiolillo A., Pilla F., Sanchez A. and Folch, J.M. Characterization and genotyping of the caprine j-casein variants. *Journal of Dairy Science*. 86, 2715–20, 2003.

Farrell Jr H. M., Jimenez-Flores R and Bleck G. T. Nomenclature of the proteins of cows' milk sixth revision," Journal of Dairy Science, vol. 87, no. 6, pp. 1641–1674, 2004.

Heck J. M. L., Schennink A., van Valenberg H. J. F., Effects of milk protein variants on the protein composition of bovine milk, *Journal of Dairy Science*, vol. 92, no. 3, pp. 1192–1202, 2009.

Prinzenberg E. M., Weimann C., Brandt H., Polymorphism of the bovine CSN1S1 promoter: linkage mapping, intragenic haplotypes and effects on milk production traits. *Journal of Dairy Science*, vol. 86, no. 8, pp. 2696–2705, 2003.

Szymanowska M., Strzalkowska N., Siadkowska E., Krzyzewski J., Ryniewicz Z and Zwierzchowski L. 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. *Animal Science Papers and Reports*, vol. 21, pp. 97–108, 2003.

Malewski T. Computer analysis of distribution of putative cisand trans- regulatory elements in milk protein gene promoters. *BioSystems*, vol. 45, no. 1, pp. 29–44, 1998.

Prinzenberg E. M., Brandt H., Bennewitz J., Kalm E and G. Erhardt, Allele frequencies for SNPs in the α s1-casein gene (CSN1S1) 5' α s1 flanking region in European cattle and association with economic traits in German Holstein. *Livestock Production Science*, vol. 98, no. 1-2, pp. 155–160, 2005.

Kuss A. W., Gogol J., Bartenschlager. H and H. Geldermann. Polymorphic binding site in bovine CSN1S1 shows quantitative differences in protein binding associated with milk protein expression. *Journal of Dairy Science*, vol. 88, no. 6, pp. 2246–2252, 2005.

Bai W. L., Yin R. H., Dou Q. L. A single nucleotide polymorphism and sequence analysis of CSN1S1 gene promoter region in Chinese BOS Grunniens (YAK). *Animal Biotechnology*, vol. 21, no. 1, pp. 36–41, 2010.

Chianese L., Quarto M., Pizzolongo F. Occurrence of genetic polymorphism at the αs1-casein locus in Mediterranean water buffalo milk. *International Dairy Journal*, vol. 19, no. 4, pp. 181–189, 2009.

Ramunno L., Cosenza G., Rando A. the goat α S1-casein gene: gene structure and promoter analysis. *Gene*, vol. 334, no. 1-2, pp. 105–111, 2004.

Bhure S. K. and B. Sharma. the PCR amplification, sequencing and computer-aided analysis of ovine α S1-casein gene promoter, *Journal of Biotechnology*, vol. 7, no. 4, pp. 478–481, 2008.

Pauciullo A., Shuiep E., Cosenza G., Ramunno L. and Erhardt G. Molecular characterization and genetic variability at κ-casein gene (CSN3) in camels. *Gene*, 513(1), 22-30, 2013.

Gerencsér, A., Barta, E., Boa, S., Kastanis, P., Bosze, Z., Whitelaw, C.B. Comparative

analysis on the structural features of the 5' flanking region of k-casein genes from six different species. *Genet.* Sel. Evol. 34, 117–128, 2002.

Hobor, S., Kunej, T., Dovc, P. Polymorphisms in the kappa casein (CSN3) gene in horse

and comparative analysis of its promoter and coding region. Anim. Genet. 39, 520-530, 2008.

Malewski, T. Computer analysis of distribution of putative cis- and trans regulatory

elements in milk protein gene promoters. Biosystems 45, 29-44, 1998.

Vanselow A. DNA-remethylation around a STAT5-binding enhancer in the alphaS1-casein promoter is associated with abrupt shutdown of alphaS1-casein synthesis during acute mastitis. *Endocrinol.* 37, 463–477, 2006.

Zhao, F.Q., Adachi, K., Oka, K., 2002. Involvement of Oct-1 in transcriptional regulation

of β -casein gene expression in mouse mammary gland. Biochim. Biophys. Acta 1577, 27–37.

Zwilling, S., Annweiler, A., Wirth, T., 1994. The POU domains of the Oct-1 and Oct-2

transcription factors mediate specific interaction with TBP. Nucleic Acids Res. 22, 1655–1662.

Akos Gerencser, Endre Barta, Simon Boa, Petros Kastanis, Zsuzsanna Bosze, C. Bruce and A.Whitelaw. 2001. Comparative analysis on the stuctural features of the 5' flanking region of k-casein genes from six different species. Genet.Sel Evol., 34: 117-128.