Journal of Animal & Plant Sciences, 31(5): 2021, Page: 1240-1251
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2021.5.0324

POLYMORPHISMS OF BETA CASEIN (*CSN2*) CDS AND INFERENCE OF ITS VARIANTS IN RIVER AND SWAMP WATER BUFFALO (*BUBALUS BUBALIS*)

X. Y. Fan[1, a], J. Xue[1, a], L. H. Qiu [1, a], R. P. Wang[2, *] and Y. W. Miao[1,*]

[1]Faculty of Animal Science and Technology, Yunnan Agricultural University, Kunming, Yunnan, China.
[2]Faculty of Animal Husbandry and Veterinary Medicine, Yunnan Vacational and Technical College of Agricultural, Kunming, Yunnan, China
[a]These authors contributed equally to this study.
*Corresponding author's email: 2415184531@qq.com (R. P. Wang), yongwangmiao1@126.com (Y. W. Miao)

ABSTRACT

β-casein (*β*-CN) has an important effect on surface properties of casein micelles and milk-clotting properties. However, current understanding of buffalo *CSN2* gene polymorphisms is not sufficient. In this study, the polymorphisms in the complete coding sequence (CDS) of buffalo *CSN2* were detected using PCR product direct sequencing. The CDS lengths of *CSN2* gene in river and swamp buffalo were the same, which was 675 nucleotides and encoded a peptide with 224 amino acid residues. A total of 15 single nucleotide polymorphisms (SNPs) were identified in two types of buffalo. Among them, c.72C>T, c.161A>G, c.167C>T, c.168A>G, c.236G>C, c.249C>G, c.350T>C and c.391G>T were non-synonymous, which leads to the changes of p.Gly10Ser, p.Gln39Arg, p.Thr41Met, p.Gly64Ala, p.Asn68Lys, p.Met102Thr and p.Val116Phe in the mature peptide of buffalo *β*-CN, respectively. However, it was predicted that these substitutions had no effect on the function of buffalo *β*-CN. Fourteen haplotypes were defined based on the SNPs found in buffalo, and accordingly, 4 protein variants and 6 synonymous variants were added in the genetic variants of buffalo *β*-CN, named variant B^2, B^3, B^4, C, C^1, C^2, C^3, D, E and F, respectively. The main variant in river buffalo was variant B, whereas in swamp buffalo was variant C. All the variants determined in buffalo did not exist in the animals of *Bos* genus. In addition, there were 11 amino acid differential sites of *β*-CN between buffalo and *Bos* genus, of which Thr at residue 41 was located at the phosphorylation site. Furthermore, the results revealed that both types of buffalo *β*-CN were "A^2-like", which indicates that the *β*-CN in buffalo milk is beneficial to human health.

Keywords: Buffalo, *CSN2*, polymorphisms, haplotype, genetic variants

Published online first January 21, 2021 Published Final August 07, 2021.

INTRODUCTION

Cattle casein genes are arranged in a cluster on BTA 6 including α_{S1}-casein, α_{S2}-casein, *β*-casein and *κ*-casein (Vinesh *et al*., 2013). About 80% of the protein consists of the caseins in cow's milk (Miluchová *et al*., 2014). As the second most abundant casein fraction (25-30%) (Kamiński *et al*., 2007; Mishra *et al*., 2009), *β*-CN is the most hydrophobic, and has important effect on surface properties of casein micelles and milk-clotting properties (Vinesh *et al*., 2013). In their natural state, *β*-CN binds with calcium citrate, phosphate and other inorganic ions in the formation of casein micelles, and calcium phosphate in micelle serves as a cementing agent (Choi *et al*., 2011; Vincent *et al*., 2014). The *β*-CN can also help to absorb minerals in the intestinal tract, like Fe and Zn (Bhattacharya *et al*., 2008).

Bos taurus CSN2 gene consists of 9 exons. Its CDS is 675 bp in length and encodes a peptide composed of 224 amino acid residues, of which the first 15 amino acid residues form a signal peptide (Caroli *et al*., 2009). The *β*-CN variants of *Bos* genus have been extensively studied for decades. To date, 15 protein variants (A^1, A^2, A^3, B, C, D, E, F, G, H^1, H^2, I, J, K and L) of *β*-CN are identified at the protein and DNA levels in *Bos* genus (Farrell *et al*., 2004; Gallinat *et al*., 2013). Among them, A^1 and A^2 are the most common forms in *Bos* genus, while variant B is less common (Farrell *et al*., 2004). The presence of His residue at position 67 makes variant A^1 induce the occurrence of the beta-casomorphin-7 (BCM-7) substance (Hanusová *et al*., 2010). BCM-7 has showed many immunological activities like chronic inflammatory responses and have been implicated in potential risk of insulin-dependent diabetes, atherosclerosis, human ischemic heart diseases and sudden infant death syndrome (Ganguly *et al*., 2013; Miluchová *et al*., 2014). Thus, the *β*-CN has potential effect on human health. Besides health promoting properties, the variant A^2 has been suggested to be related to higher protein and milk yield (Ganguly *et al*., 2013).

Water buffalo (*bubalus bubalis*) as dairy, meat, and draught animals, have economic importance in the countries of tropical and subtropical areas (Michelizzi *et al*., 2010). Domestic buffalo are divided into river and

swamp buffalo. The former is mainly used for milk production, which produces about 2000 kg milk per year, while the latter is mainly used for draught and produces 500-600 kg milk per year. At present, more than 5% of the milk produced around the world is supplied by water buffalo (Michelizzi *et al*., 2010). In contrast to cattle, the polymorphisms of *CSN2* gene in buffalo have been less extensively investigated. So far, buffalo *CSN2* gene has been mapped on BBU7, and 4 protein variants (A, A^1, B and B^1) of *β*-CN in buffalo are identified at DNA level (Vinesh *et al*., 2013). For swamp buffalo, the polymorphisms in the *CSN2* gene have not been reported. The DNA sequencing has been successfully applied for identifying three new bovine *β*-CN variants (Gallinat *et al*., 2013). The objective of this study was to detect the polymorphisms of *CSN2* gene in two types of buffalo, to infer the possible *β*-CN variants in two types of buffalo, and to elucidate the differences of *β*-CN variants between buffalo and *Bos* genus.

MATERIALS AND METHODS

Sample and DNA sequences information: In accordance with the Guide for Animal Care and Use of Experimental Animals, all procedures for sample collection were performed and approved by the Institutional Animal Care and Use Committee of Yunnan Agricultural University.

A total of 126 mammary gland samples were obtained by percutaneous biopsy in this study, including 54 river buffalo which sampled from Binglangjiang buffalo (a river buffalo breed distributed in the west of Yunnan province of China), and 72 swamp buffalo (36 Dehong, 6 Yuanjiang, 6 Enshi, 6 Eshan, 3 Guizhou, 3 Yanjin, 6 Yongxiu and 6 Fu'an buffalo). Percutaneous biopsy was performed using the previously reported method (Bionaz *et al*., 2007). All samples without direct kinship were collected at random. The samples were transported by cryopreservation and then stored in a refrigerator at -80 °C in the laboratory. The river buffalo samples were obtained from Binglangjiang buffalo breeding farm in Tengchong, Yunnan Province, China, and the swamp buffalo samples were all collected from its main distributing areas in China where they are located. The *CSN2* sequences of buffalo, cattle, yak, zebu and bison published in NCBI database (https://www.ncbi.nlm.nih.gov/) were also included here for comparison (Table S1).

Total RNA extraction, first-strand cDNA synthesis and DNA isolation: The extraction process of RNA from the mammary gland via the RNAiso Plus kit (TaKaRa, Dalian, China) was described previously (Song *et al*., 2016). Their qualities and quantities were detected using agarose gel electrophoresis and the NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific, Waltham, USA). The cDNA was synthesized from 2 μg RNA via M-MLV reverse transcriptase (Takara) and oligo $(dT)_{18}$ primer. Genomic DNA from the mammary gland was isolated according to a standard Proteinase-K and phenol/chloroform method (Sambrock and Russell, 2001). Then, the cDNA and the DNA were diluted to 50 ng/μL using the sterile ddH_2O, and stored at -80 °C.

PCR and sequencing: The primers for amplifying CDS were designed based on the *CSN2* sequence of buffalo (Accession no. FN424088) (Table 1). PCR was executed in a final 25 μL reaction mixture containing 2.5 μL of 10×Taq DNA polymerase buffer (Mg^{2+} Plus), 0.4 μL of 10 mM each primer, 2.0 μL of 25 mM dNTPs (TaKaRa), 0.3 μL of 5 U/μL Taq DNA polymerase (TaKaRa), 2.0 μL of cDNA template, and 17.4 μL of sterile water. The mixture was denatured for 5 min at 95 °C, and PCR was run for 34 cycles at 95 °C for 40 sec (denaturation), 57.1 °C for 40 sec (annealing), 72 °C for 45 sec (extension), and a final extension for 5 min at 72 °C. In order to confirm the authenticity of the SNPs found at the cDNA level, we selected the corresponding samples to detect the SNPs at the DNA level using corresponding exon primers (which were designed based on the sequence with accession no. NC_037551) (Table 1). PCR products were detected using 2% agarose gel with EtBr staining. Then, the PCR products were sequenced in both directions using Sanger method.

Table 1. Primer information for PCR and polymorphism identification.

Amplified region	Primer sequence (5' to 3')	Products length (bp)	Annealing temperature (°C)	Extension time (s)	Purpose
CDS	F: GCTCCTCCTTCACTTCTTGTCC R: AAAGTTGCCATATTTCCAGTCACA	793	57.1	45	SNP detection
Exon III	F: TATACCAATGAAAGTTAC R: TTGCCTCTGAATGAACAC	363	43.1	30	SNP confirmation
Exon VI	F: CAAAGGAATCTATTACAC R: TACCATTCTTAGACTGAA	257	43.9	30	SNP confirmation
Exon VII	F: CCAAACCAAGTGGAAGATT R: CATCAGAAGTTAAACAGCACAG	868	54.3	55	SNP confirmation

Data analysis: The sequences obtained in this study were checked and edited using Lasergene software package (DNAstar Inc.). And open reading frame (ORF) was determined exploiting program ORF Finder (http://www. ncbi. nlm. nih. gov/gorf/) and translated into amino acid sequence. Physicochemical characteristics of isoelectric point, theoretical molecular weight, hydropathy, O-glycosylation site, phosphorylation site and signal peptide were predicted by ComputepI/Mw tool (http://web.expasy.org/compute_pi/), ProtParam tool (http://web.expasy.org/protparam/), NetOGlyc 4.0 Server (http://www.cbs.dtu.dk/services/NetOGlyc/), NetPhos 3.1 Server (http://www.cbs.dtu.dk/services/NetPhos/) and SignalP software (http://www. cbs. dtu. dk/services/ SignalP/), respectively. The conserved domain of buffalo *β*-CN was analyzed using the Conserved Domain Architecture Retrieval Tool in BLAST at the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST). The single nucleotide polymorphisms (SNPs) were checked and outputted by Seqman (DNAstar Inc.) and Mega 7 (Kumar *et al*., 2016). The estimate of genotype and allele frequencies, and the Hardy-Weinberg balance test was conducted for each SNP using PopGen 1.32 (Yeh *et al*., 1999). Based on Bayesian model algorithm, the haplotypes with high confidence were deduced by PHASE, and the number of iterations is ≥ 100 (Stephens *et al*., 2001). All the inferred haplotypes were verified by clonal sequencing, and the process was described previously (Yuan *et al*., 2015). The effect of amino acid substitutions on the protein function was evaluated via program PANTHER (http://www.pantherdb.org/; Mi *et al*., 2017).

To investigate the probable phylogenetic relationship between the haplotypes of buffalo *CSN2* gene, median-joining network was constructed by the Network 5 (www.fluxus-engineering.com). The weights of transitions and transversions were 5 and 10, respectively. Furthermore, the phylogenetic tree was constructed by Mega 7 based on the haplotype sequences of *CSN2* gene. The maximum likelihood (ML) tree was constructed using the Kimura 2-parameter model after model selection tests. Robustness of the nodes was evaluated by the bootstrap method after 10000 replications.

RESULTS

PCR and sequence analysis: Being consistent with expectations, the PCR products of 793 bp were obtained. The products were subjected to bi-directional sequencing. The prediction showed that the obtained sequence contained a CDS of 675 bp. The CDS was compared by alignment with the *CSN2* sequences of *Bos* genus published in NCBI database, and the results displayed that the consistency of this CDS with that of the *CSN2* gene in cattle (XM_010806178), zebu (XM_019962870), yak (XM_005902037) and bison (XM_010852144) was 98.37 %, 98.67 %, 98.37 % and 98.52 %, respectively. As a result, the obtained CDS was determined to be the buffalo *CSN2* gene. The whole CDS of buffalo *CSN2* gene was 675 nucleotides in length (Fig 1) and the CDS length and structure of buffalo *CSN2* gene were the same as those of the species in *Bos* genus. Mean nucleotide composition of buffalo *CSN2* CDS consists of 24.00% A, 20.15% G, 25.04% T, and 30.81% C. Buffalo *CSN2* was predicted to encode a protein composed of 224 amino acids (AA) with a 15 AA signal peptide at N-terminus and a mature peptide of 209 AA. The mature peptide contained a casein domain (Fig. 1), which belonged to casein superfamily. The AA composition and basic molecular characteristics of buffalo *β*-CN variant C were slightly different from those of cattle *β*-CN variant A[1] (Table 2). The predicted results showed that there were 8 O-glycosylation sites and 17 phosphorylation sites in the mature peptide, which were different from that of cattle *β*-CN in number and location (Table 3)..

```
  1 ATGAAGGTCCTCATCCTTGCCTGCCTGGTGGCTCTGGCCCTTGCAAGAGAGCTGGAAGAACTCAATGTACCCGGTGAGATTGTGGAAAGC  90
  1 M  K  V  L  I  L  A  C  L  V  A  L  A  L  A  R  E  L  E  E  L  N  V  P  G  E  I  V  E  S  30
 91 CTTTCAAGCAGTGAGGAATCTATTACACACATCAATAAGAAAATTGAGAAGTTTCAGAGTGAGGAACAGCAGCAAACAGAGGATGAACTC 180
 31  L  S  S  S  E  E  S  I  T  H  I  N  K  K  I  E  K  F  Q  S  E  E  Q  Q  Q  T  E  D  E  L  60
181 CAGGATAAAATCCACCCCTTTGCCCAGACACAGTCTCTAGTCTATCCCTTCCCTGGGCCCATCCCTAACAGCCTCCCACAAAACATCCCG 270
 61  Q  D  K  I  H  P  F  A  Q  T  Q  S  L  V  Y  P  F  P  G  P  I  P  N  S  L  P  Q  N  I  P  90
271 CCTCTTACTCAAACCCCTGTGGTGGTGCCGCCTTTCCTTCAGCCTGAAATAATGGGAGTCTCCAAAGTGAAGGAGGCTATGGCTCCTAAG 360
 91  P  L  T  Q  T  P  V  V  V  P  P  F  L  Q  P  E  I  M  G  V  S  K  V  K  E  A  M  A  P  K  120
361 CACAAAGAAATGCCCTTCCCTAAATATCCAGTTGAGCCCTTTACTGAAAGCCAGAGCCTGACTCTCACTGATGTTGAAAATCTGCACCTT 450
121  H  K  E  M  P  F  P  K  Y  P  V  E  P  F  T  E  S  Q  S  L  T  L  T  D  V  E  N  L  H  L  150
451 CCTCTGCCTCTGCTCCAGTCTTGGATGCACCAGCCTCCCCAGCCTCTTCCTCCAACTGTCATGTTTCCCCCTCAGTCCGTGCTGTCCCTT 540
151  P  L  P  L  L  Q  S  W  M  H  Q  P  P  Q  P  L  P  P  T  V  M  F  P  P  Q  S  V  L  S  L  180
541 TCTCAGTCCAAAGTTCTGCCTGTTCCCCAGAAAGCAGTGCCCTATCCCCAGAGAGATATGCCCATTCAGGCCTTTCTGCTGTATCAGGAG 630
181  S  Q  S  K  V  L  P  V  P  Q  K  A  V  P  Y  P  Q  R  D  M  P  I  Q  A  F  L  L  Y  Q  E  210
631 CCTGTACTTGGTCCTGTCCGGGGACCCTTCCCTATTATTGTCTAA                                              675
211  P  V  L  G  P  V  R  G  P  F  P  I  I  V  *                                              224
```

Fig 1. Nucleotide sequence of buffalo *CSN2* (B1) and its deduced amino acid sequence. The predicted protein sequence is shown immediately above the nucleotide sequence. The signal peptide is shaded. The casein domain is boxed. The stop codon is indicated by an asterisk (*).

Table 2. Physicochemical characteristics of *β*-CN mature peptide for buffalo and cattle.

Basic physical and chemical properties	Buffalo C (QHB80269)	Cattle A[1] (X14711)
Formula	$C_{1080}H_{1689}N_{265}O_{310}S_6$	$C_{1081}H_{1692}N_{270}O_{310}S_6$
Number of amino acids	209	209
Molecular weight	23.54 KD	23.62 KD
Isoelectric point (pI)	5.01	5.24
Strongly acidic amino acid (D, E)	23	23
Strongly basic amino acid (K, R)	14	15
Polar amino acid (N, C, Q, S, T, Y)	54	54
Hydrophobic amino acid (A, I, L, F, W, V)	66	66
Instability index (II)	101.49	98.48
Grand average of hydropathicity (GRAVY)	-0.339	-0.362
Aliphatic index	88.95	88.47
Transmembrane helix region	None	None
Casein domain	1 (AA123-207)	1 (AA123-207)

Table 3. Modification sites in the mature peptides of buffalo and cattle *β*-CN.

Modification	Buffalo[1]	Cattle
O-glycosylation	80T, 96S, 120T, 122S, 154T, 161S, *164S*, 168S	*57S*, 80T, 96S, 120T, 122S, 154T, 161S, 168S
Phosphorylation	15S, 17S, 18S, 19S, 22S, 35S, 55T, 78T, 80T, 122S, 124S, 126T, 128T, 154T, 164S, 166S, 168S	15S, 17S, 18S, 19S, 22S, 35S, *41T*, 55T, *69S*, 78T, 80T, 122S, 124S, 126T, 128T, 154T, 164S, 166S, 168S

[1] The specific modification sites of species are italicized..

Table 4. Allele and genotype frequencies for the SNPs found in two types of buffalo.

Population	SNP	Genotype frequency		Allele frequency			*P*[1] value
		Genotype	Frequency, %	Number	Allele	Frequency, %	
River buffalo	c.72C>T	CC	87.04	47	C	93.52	1.00000
		CT	12.96	7	T	6.48	
		TT	0.00	0			
	c.167T>C	TT	50.00	27	T	68.52	0.57270
		TC	37.04	20	C	31.48	
		CC	12.96	7			
	c.168G>A	GG	50.00	27	G	75.00	0.42503
		GA	50.00	27	A	25.00	
		AA	0.00	0			
	c.249G>C	GG	50.00	27	G	72.22	0.73315
		GC	44.44	24	C	27.78	
		CC	5.56	3			
	c.391G>T	GG	94.44	51	G	97.22	1.00000
		GT	5.56	3	T	2.78	
		TT	0.00	0			
	c.498G>T	GG	50.0	27	G	72.22	0.73315
		GT	44.44	24	T	27.78	
		TT	5.56	3			
	c.624C>T	CC	50.00	27	C	72.22	0.73315
		CT	44.44	24	T	27.78	
		TT	5.56	3			
Swamp buffalo	c.72C>T	CC	75.00	54	C	87.50	0.69180
		CT	25.00	18	T	12.50	
		TT	0.00	0			
	c.73G>A	GG	91.67	66	G	91.67	0.00000
		GA	0.00	0	A	8.33	
		AA	8.33	6			
	c.167T>C	TT	0.00	0	T	4.17	1.00000

		TC	8.33	6	C	95.83	
		CC	91.67	66			
	c.168G>A	GG	0.00	0	G	4.17	1.00000
		GA	8.33	6	A	95.83	
		AA	91.67	66			
	c.249G>C	GG	0.00	0	G	2.09	1.00000
		GC	4.17	3	C	97.91	
		CC	95.83	69			
	c.391G>T	GG	75.00	54	G	85.42	0.34200
		GT	20.83	15	T	14.58	
		TT	4.17	3			
	c.498G>T	GG	0.00	0	G	2.09	1.00000
		GT	4.17	3	T	97.91	
		TT	95.83	69			
	c.624C>T	CC	0.00	0	C	2.09	1.00000
		CT	4.17	3	T	97.91	
		TT	95.83	69			

[1]*P* value for Chi-square test of Hardy-Weinberg equilibrium.

Population genetic analysis: In the samples of this study, a total of eight SNPs were identified in the *CSN2* CDS of two types of buffalo (all of them were confirmed at the DNA level), in which the substitution c.72C>T and c.73G>A were located in exon III, c.167T>C and c.168G>A in exon VI, and c.249G>C, c.391G>T, c.498G>T and c.624C>T in exon VII. Except that c.73G>A was only detected in swamp buffalo (all the river buffalo were homozygous GG), the other substitutions were shared by river and swamp buffalo. Among the eight SNPs, c.72C>T, c.167C>T, c.168A>G, c.249C>G and c.391G>T were non-synonymous, leading to changes of p. Gly10Ser, p. Thr41Met, p. Asn68Lys and p. Val116Phe in the mature peptide.

The genotype frequency, allele frequency and Hardy-Weinberg balance test for each SNP locus are presented in Table 4. The c.72C and c.391G were high frequency alleles in two types of buffalo. In swamp buffalo, c.73G allele had high frequency (0.917), while SNP73 was homozygous in river buffalo with homozygote of GG type. It is noteworthy that allele frequencies in river buffalo at SNP167, SNP168, SNP249, SNP498 and SNP624 were significantly different from those in swamp buffalo. Hardy-Weinberg balance test showed that only SNP73 in swamp buffalo was in an unbalanced state ($P<0.05$).

Combined the data of this study with published buffalo data in NCBI database, we found that the number of polymorphic sites increased to 15 in the CDS, among which seven were shared by river and swamp buffalo, one (SNP73) was found only in swamp buffalo, and other seven SNPs were observed only in river buffalo.

Haplotypes division and their genetic relationship: A total of 14 buffalo haplotypes (B1-B14) were defined based on the observed 15 SNPs (Fig. 2, 3). Among them, 7 haplotypes (accession numbers MN560174-MN560180) were determined by the data of this study and verified by clonal sequencing, and the other 7 haplotypes were from online data (accession numbers AJ005432, DQ317447, DW007989, GQ176288, DW007978, EF115306, KC577240). The expected and actual frequencies of these haplotypes are shown in Table 5. Among the 14 haplotypes, B1-B3 were shared by river and swamp buffalo. The B1 was the main haplotype of swamp buffalo with a frequency of 0.73, while its frequency in river buffalo was only 0.22. The frequency of B2 was 0.69 in river buffalo and 0.02 in swamp buffalo. The haplotypes B4 and B5 were specific to swamp buffalo, while B6-B14 were exclusive to river buffalo. Haplotype alignments showed that there are no nucleotide sites that could distinguish the two types of buffalo.

The genetic relationship among 14 haplotypes of buffalo *CSN2* was constructed by the median-joining network method (Fig. 2). Fourteen buffalo haplotypes were divided into two groups on the median-joining network. One group, including haplotypes B1, B3, B4, B5, B6, B11 and B14, mainly distributed in swamp buffalo with B1 as central, while the other group, including haplotypes B2, B7, B8, B9, B10, B12 and B13, were mainly found in river buffalo with B2 as central. There were one or three mutations between B1 and B3, B4, B5, B6, B11, B14. According to the network, it is concluded that B3, B4, B5, B6, B14 probably have evolved from a single mutation of B1, and B11 may have evolved from three mutations of B1. Except for only one mutation between B2 and B7, B8, B9, B10 and B13, there were two mutations between B2 and B12. Therefore, B7, B8, B9, B10 and B13 may have evolved from a single mutation of B2, and B12 may be evolved from B2 through two mutations.

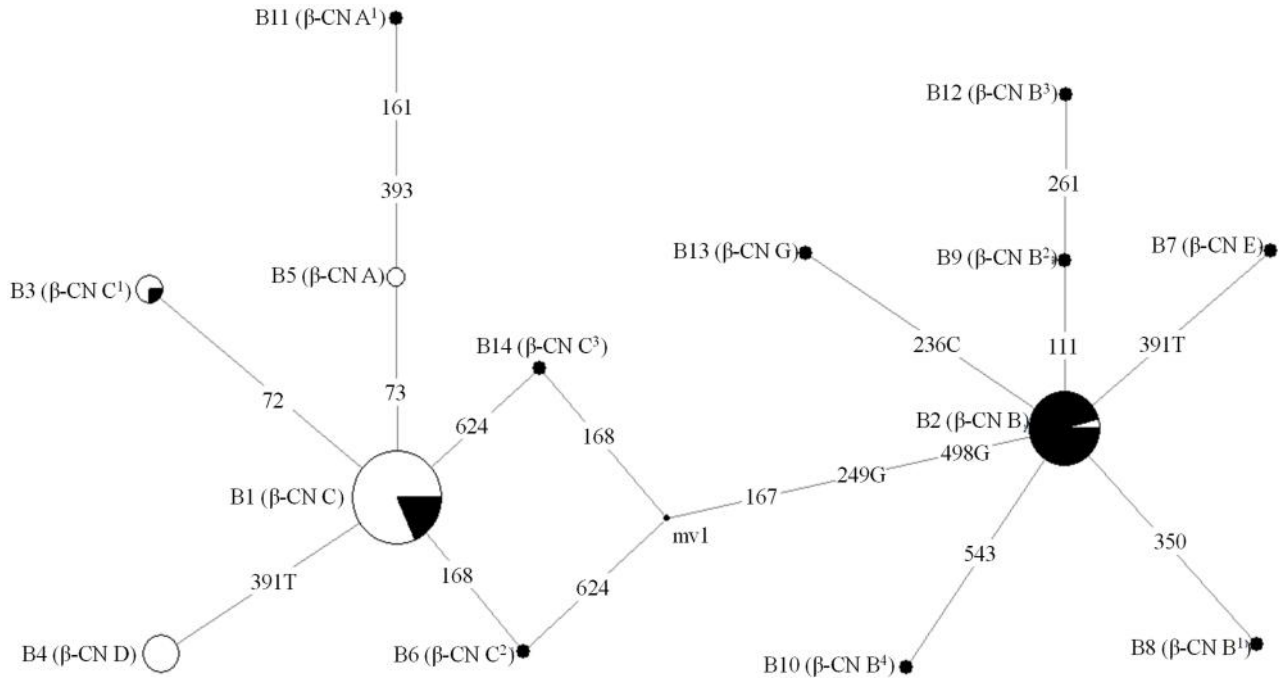

Fig. 2. Network profile of 14 buffalo haplotypes of *CSN2* gene. B1-B14, the haplotypes defined in water buffalo. Mutations along the branch are labelled by the nucleotide positions, transversions are specified by further adding suffixes A, G, C, and T. Each haplotype is represented by a circle, with the area of the circle proportional to its frequency. Samples from river and swamp buffalo are indicated by black and white color, respectively. The mv1 is a median vector.

In terms of haplotype composition and frequency distribution of the *CSN2* gene, there were obvious differences between the two types of buffalo, revealing that the *CSN2* gene in two types of buffalo had different evolutional patterns. In view of the fact that B1 is widely distributed in two types of buffalo and B2 is mainly distributed in river buffalo, it can be inferred that the haplotype B1 might represent the ancestral haplotype of buffalo *CSN2*, while the other haplotypes may originate from this haplotype. Whether this is the case, it is necessary to further expand the sample for experimental verification.

Table 5. Haplotype information of buffalo *CSN2*.

	Haplotype	Actual frequency	Expected frequency
B1	CGCACGTT	0.512	0.440
B2	CGTGGGGC	0.321	0.310
B3	TGCACGTT	0.036	0.073
B4	CGCACTTT	0.083	0.077
B5	CACACGTT	0.024	0.041
B6	CGCGCGTT	0.012	0.029
B7	TGCACTTT	0.012	0.016

Nucleotide differences between buffalo and the species of *Bos* genus: In order to reveal the differences in the CDS of *CSN2* gene between buffalo and the species of *Bos* genus, the homologous sequences of *Bos* genus published were included in this study. A total of 9, 2, 4 and 1 haplotypes were defined in these species, respectively (Fig. 3). The alignment of all haplotype sequences in buffalo, cattle, yak, zebu and bison displayed 8 differences between buffalo and the species of *Bos* genus, which were found in the sequences at positions c.72, c.119, c.270, c.319, c.488, c.519, c.555 and c.639.

Based on the haplotype sequences of buffalo, cattle, yak, zebu and bison, the phylogenetic tree was constructed by ML method with goat and sheep as outgroup (Fig. 4). The phylogenetic analysis showed that the species of *Bos* genus and buffalo gather on different clades with high support rates, indicating that there is a great difference in *CSN2* between buffalo and the species in *Bos* genus.

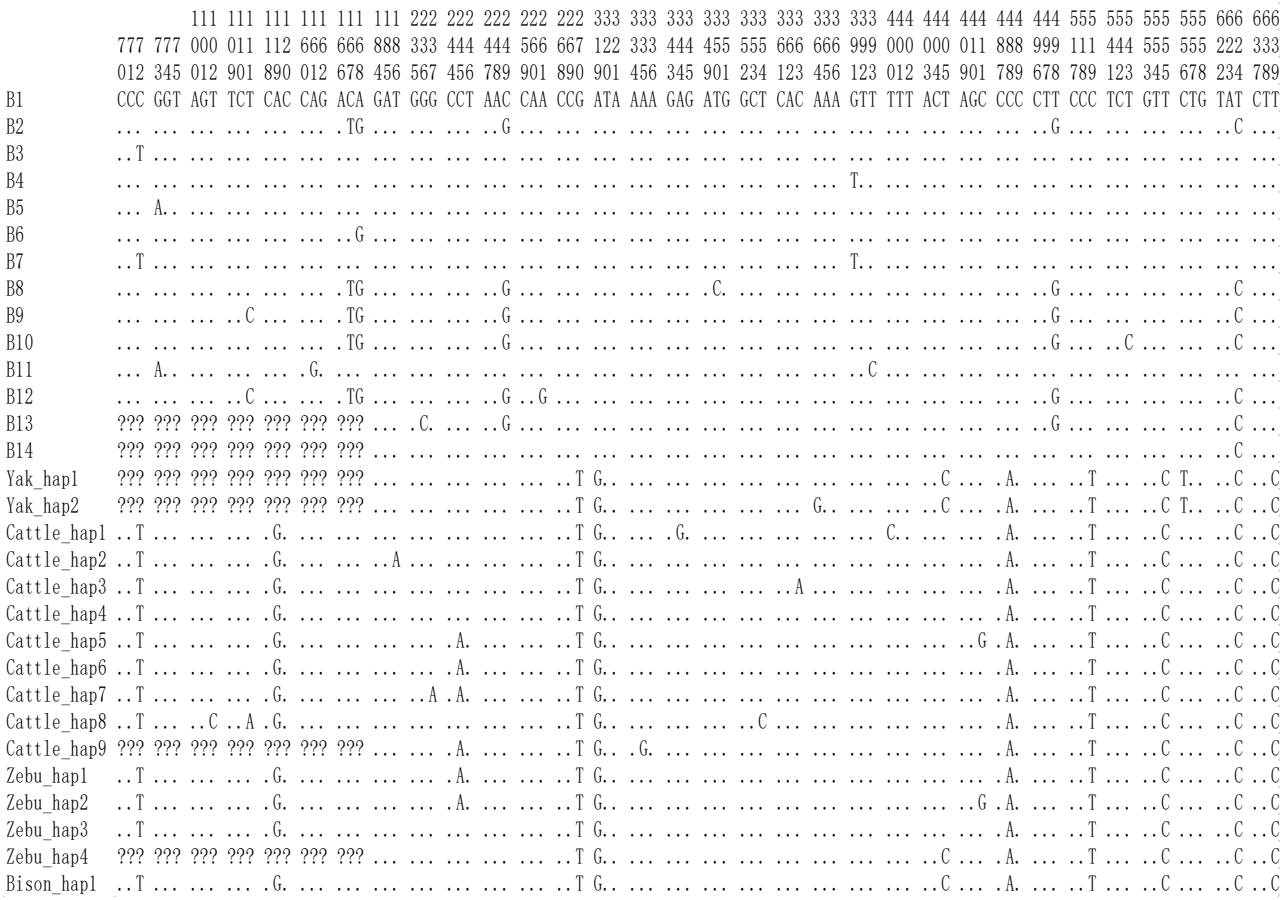

Fig. 3. Nucleotide differences in the haplotype sequences of *CSN2* between buffalo and the species of *Bos* genus. Number represents the position of coding region. Dots (.) represents the identity with the B1. Nucleotide substitutions are denoted by different letters. Missing information is demonstrated by a question marker (?).

Inference and Nomenclature of *β*-CN Variants: Among the 15 SNPs identified in buffalo *CSN2* (Fig. 5), c.72C>T, c.161A>G, c.167C>T, c.168A>G, c.236G>C, c.249C>G, c.350T>C and c.391G>T were non-synonymous, which leads to substitutions of p.Gly10Ser, p.Gln39Arg, p.Thr41Met, p.Gly64Ala, p.Asn68Lys, p.Met102Thr and p.Val116Phe in mature peptide of buffalo *β*-CN. Based on these substitutions, we defined eight protein variants and six synonymous variants in the *β*-CN (Fig. 5).

Based on the literatures (Caroli *et al*., 2009; Gallinat *et al*., 2013), we have reconstructed the amino acid sequences of the various *β*-CN variants found in *Bos* genus. All the variants of *Bos* genus identified in previous reports have not been found in two types of buffalo. The sequence alignment of *β*-CN variants demonstrated that there are 11 amino acid differential sites between *Bubalus* and *Bos*. In addition to 4 known variants, 10 new variants, including four non-synonymous and six synonymous variants, were identified via comparing with the reference sequence of buffalo *β*-CN B (accession no. GQ176287). The variants identified in buffalo *β*-CN were respectively designated as variant B^2, B^3, B^4, C, C^1, C^2, C^3, D, E and F, among which the variants B^2, B^3, B^4, C^1, C^2 and C^3 are synonymous (Table 6). In the samples of this study, the variants B, C, C^1, C^2 and E were found in river buffalo with the frequencies of 69.4%, 22.2%, 2.8%, 2.8% and 2.8%, respectively, whereas the variants A, B, C, C^1 and D were found in swamp buffalo with the frequencies of 4.2%, 2.1%, 72.9%, 6.3% and 14.6%, respectively..

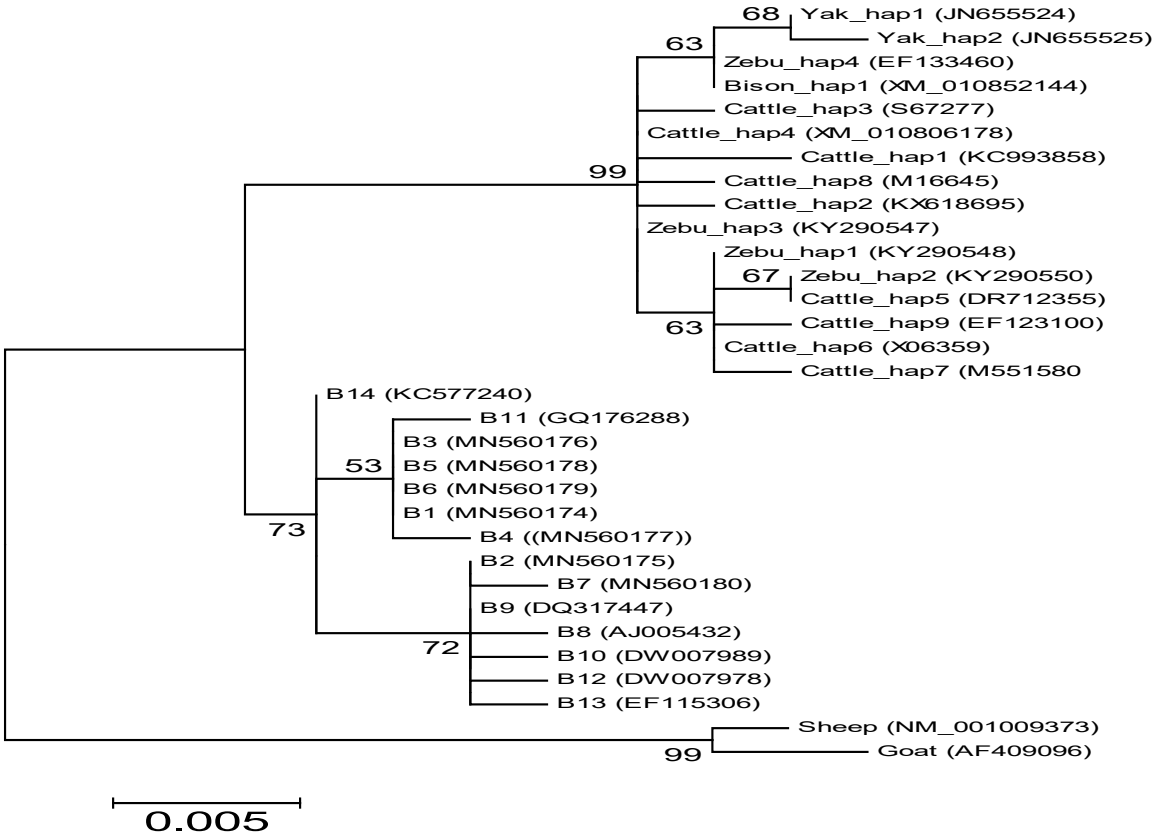

Fig. 4. Phylogenetic tree of haplotype sequences constructed by using maximum likelihood method (Kimura 2-parameter model). Bootstrap values are based on 10,000 replicates and are adjacent to nodes.

		111111111
		112233346667899001234579
		080567914782823266288297
	β-CN_A1	GSEREEQTGHNQLVMMHVSPHPPV
	β-CN_A2	.........P..............
	β-CN_A3	.........P......Q.......
	β-CN_B	..................R.....
	β-CN_C	.....K..................
	β-CN_D	.K.......P..............
	β-CN_E	....K....P..............
	β-CN_F	....................L..
	β-CN_G	....................L....
	β-CN_H1	...C.....P..I...........
	β-CN_H2	.........P.E..L.........
	β-CN_I	.........P....L.........
	β-CN_J	..K.....................
	β-CN_K	....................A.
	β-CN_L	....................A
Buffalo	β-CN_A	S..H.....P...I......P...
Buffalo	β-CN_A1	S..H..R..P...I......P...
Buffalo	β-CN_B	...H...M.PK..I......P...
Buffalo	β-CN_B1	...H...M.PK..I.T....P...
Buffalo	β-CN_C	...H.....P...I......P...
Buffalo	β-CN_D	...H.....P...I...F..P...
Buffalo	β-CN_E	...H...M.PK..I...F..P...
Buffalo	β-CN_F	????????APK..I......P...

Fig. 5. Sequence differences of *β*-CN variants between buffalo and the species of *Bos* genus. Number represents the position of the mature peptide. Dots (.) represents the identity with the *β*-CN A[1]. Amino acid substitutions are denoted by different letters. Missing information is indicated by question marks (?).

Table 6. Amino acid differences and its corresponding positions in genetic variants of buffalo *β*-CN.

Position	Buffalo *β*-CN variants (Buffalo haplotypes)													
	B (B2)	B[1] (B8)	B[2] (B9)	B[3] (B12)	B[4] (B10)	A (B5)	A[1] (B11)	C (B1)	C[1] (B3)	C[2] (B6)	C[3] (B14)	D (B4)	E (B7)	F (B13)
9	CCC								CCT					
10	Gly GGT					Ser AGT	Ser AGT							???
22	TCT		TCC	TCC										
39	Gln CAG						Arg CGG							???
41	Met ATG					Thr ACA	Thr ACA	Thr ACA	Thr ACA	Thr ACG	Thr ACA	Thr ACA		???
64	Gly GGG													Ala GCG
68	Lys AAG					Asn AAC	Asn AAC	Asn AAC	Asn AAC	Asn AAC	Asn AAC	Asn AAC		
72	CAA			CAG										
102	Met ATG	Thr ACG												
116	Val GTT						Val GTC					Phe TTT	Phe TTT	
151	CTG					CTT	CTT	CTT	CTT	CTT	CTT	CTT		
166	TCT				TCC									
193	TAC					TAT	TAT	TAT	TAT	TAT		TAT		

Table 7. Effect of non-synonymous substitutions on the function of buffalo *β*-CN.

SNP	Substitution	Preservation time	Message
c.72C>T	Gly10Ser	2	Probably benign
c.161A>G	Gln39Arg	176	Probably benign
c.167C>T c.168A>G	Thr41Met	1	Probably benign
c.236G>C	Gly64Ala	1	Probably benign
c.249C>G	Asn68Lys	1	Probably benign
c.350T>C	Met102Thr	1	Probably benign
c.391G>T	Val116Phe	1	Probably benign

Functional effect of non-synonymous substitution: The functional effect of non-synonymous substitutions found in buffalo *CSN2* gene was presumed by the program PANTHER (Table 7). The result showed that the amino acid substitutions found in coding region did not impair protein function.

DISCUSSION

β-CN is a milk protein participated in many important physiological processes in mammals. The genetic polymorphisms of *CSN2* gene are of importance, since some polymorphisms could be related to composition, production and technological properties of milk (Boland *et al*., 2001; Martin *et al*., 2002). In this research, the polymorphisms of *CSN2* gene in river and swamp buffalo were investigated and analyzed in combination with published data of this gene. No alternative splicing was found in the *CSN2* gene of two types of buffalo. And 15 SNPs were identified in water buffalo, including two at exon III, one at exon V, three at exon VI, and nine at exon VII. Previous findings have shown that the polymorphisms of *CSN2* in other mammals also exists mainly in exon VII (Jann *et al*., 2002; Gallinat *et al*., 2013). The exon VII of *CSN2* encodes 79% of mature *β*-CN in two types of buffalo and other mammals, which constitutes the Casein domain. Consequently, the studies in recent years on genetic variation of the *CSN2* gene in some species mainly focused on exon VII, and named *β*-CN variants based on the polymorphisms of this exon (Farrell *et al*., 2004; Caroli *et al*., 2009; Gallinat *et al*., 2013). In light of this, to explore the association between polymorphisms of *CSN2* CDS and the traits in buffalo and other mammals, it is suggested that exon VII should be taken as the main analysis region.

In this study, allele frequencies in river buffalo at SNP167, SNP168, SNP249, SNP498 and SNP624 were significantly different from those in swamp buffalo.

In addition, seven SNPs were detected only in river buffalo and one SNP was detected only in swamp buffalo, which indicated that the variation of *CSN2* gene in two types of buffalo had different population genetic characteristics. Except for SNP72 and SNP624, the other SNPs observed in buffalo did not exist in the species of *Bos* genus, and the SNPs observed in *Bos* genus also did not exist in water buffalo, indicating that the mutation pattern of *CSN2* gene in water buffalo was obviously different from that in *Bos* genus. The comparison of *β*-CN variants showed that there are 11 amino acid differences between *Bubalus* and *Bos*, which indicates that there exist large genetic divergences between them. Phylogenetic analysis also supports that the buffalo *CSN2* is different from that of the species in *Bos* genus.

In recent years, the development of molecular biology techniques has enhanced the study of milk protein variants, with a large number of milk protein variants being identified. So far, the *β*-CN variants found in *Bos* genus have been named (Gallinat *et al*., 2013). However, due to insufficient study on polymorphisms of *CSN2* gene in water buffalo, the nomenclature of *β*-CN variants has not been fully carried out in buffalo. In this research, we detected the variation of the *CSN2* gene in water buffalo, trying to reach a comprehensive understanding of its variants. In view of the great difference in the sequences of *CSN2* gene between buffalo and *Bos* genus, it is necessary to name the *β*-CN variants of buffalo independently. So far, four *β*-CN variants have been named in water buffalo. According to the existing nomenclature convention, we determined 4 new *β*-CN protein variants and 6 new synonymous variants in this study based on buffalo *CSN2* haplotypes. It is worth noting that the *β*-CN variants with high frequencies in river and swamp buffalo were different. The distribution frequency of variant B is the highest in river buffalo, while the distribution frequency of variant C is the highest in swamp buffalo.

From the median-joining network of the buffalo haplotypes, *β*-CN variants A, A^1, D and F each differ from variant C with one or two amino acid change and they directly derived from buffalo variant C. Buffalo variants C^1, C^2 and C^3 probably evolved from variant C by a synonymous substitution. Variant A evolved from variant C by amino acid exchange p.Gly10Ser and variant D by exchange p.Val116Phe. Variant A^1 evolved from variant C by amino acid exchange p.Gly10Ser and p.Gln39Arg with an extra synonymous substitution. There is only one amino acid difference between the *β*-CN variants B^1, E and F and variant B sequences, which may be directly derived from the variant B. Variant B^1 evolved from variant B through amino acid exchange p.Met102Thr, variant E by exchange p.Val116Phe, and F by exchange p.Gly64Ala. Buffalo variant B^2, B^3 and B^4 evolved from variant B by one or two synonymous substitutions. As for variants B and C, variant B can be transformed into variant C by amino acid exchanges p.Thr41Met and p.Asn68Lys with 2 extra synonymous substitutions.

Protein phosphorylation is the most common post-translational modification (Knight *et al*., 2003). Phosphorylation of *β*-CN has a significant effect on stability of casein micelle and milk coagulation properties (Buitenhuis *et al*., 2016; Poulsen *et al*., 2016). Casein micelle size has a direct effect on milk processing characteristics. Previous studies have indicated that the phosphorylation occurs mainly on the serine residues of *β*-CN (Jensen *et al*., 2012). The results indicated that phosphorylation sites of buffalo *β*-CN were mainly located on serine, which agree with the results of previous studies in cattle. According to the predicted results, there were differences in the number and type of glycosylation sites and phosphorylation sites of *β*-CN between water buffalo and cattle. In addition, 11 amino acid differential sites of *β*-CN between water buffalo and *Bos* genus were identified in this study, of which one was located at phosphorylation sites (41T). Therefore, we infer that there are certain differences in post-translational modification of *β*-CN between *Bubalus* and *Bos*, which may lead to differences in the physicochemical properties of their *β*-CN.

The *β*-CN variants A^1 and A^2 (with His 67 for A^1 variant and Pro 67 for A^2 variant) are the most common in the species of *Bos* genus. Some reports showed that variant A^1 was a risk factor of ischemic heart disease and type 1 diabetes mellitus (Elliott *et al*., 1999; McLachlan, 2001). It yields the BCM-7 that may have an unclear effect on human diseases. In addition, variant A^1 has hypothetical correlation with milk allergy (Chatchatee *et al*., 2001; Gobbetti *et al*., 2002). The BCM-7 can be produced in gastrointestinal digestion, milk fermentation and cheese ripening (Kamiński *et al*., 2007). The BCM-7 is released by cleavage of peptide bond between 59Val and 60Tyr by leucine aminopeptidase and pepsin, and peptide bond between 66Ile and 67His by elastase (Jinsmaa and Yoshikawa, 1999). However, metabolism of variant A^2 is different and reduces the serum cholesterol which plays a key role in prevention of human vascular diseases (Hanusová *et al*., 2010). The difference of amino acid sequence between "A^1-like" and "A^2-like" leads to conformational change of secondary structure and therefore has an impact on the physicochemical properties of casein micelles (Ganguly *et al*., 2013). Because the "A^2-like" variant has the peptide bond of 66Ile and 67Pro, thus it cannot release the BCM-7. The results of this study manifested both types of buffalo contain 67Pro in the protein sequences. Therefore, *β*-CN in buffalo will not release the bioactive peptide BCM-7. As a result, the buffalo milk is healthier than that of having a high frequency of "A^1-like" variant such as Holstein Friesian (Vincent *et al*., 2014).

Conclusion: In this study, a total of 15 SNPs was identified in buffalo *CSN2* gene, of which seven caused changes in amino acids. The amino acid composition and physicochemical characteristics of buffalo *β*-CN are slightly different from those of cattle. The mutation pattern of *CSN2* gene in buffalo was obviously different from that in *Bos* genus. We defined 14 haplotypes in buffalo *CSN2* gene. Accordingly, 14 variants were identified in buffalo *β*-CN, including four new protein variants and six new synonymous variants. The types and frequencies of the variants were distinct in the two types of buffalo. The main variant in river buffalo was variant B, whereas in swamp buffalo was variant C. Furthermore, the variants observed in buffalo did not exist in *Bos* genus. Whether the SNPs found in this study have any effect on the milk yield and composition of buffalo remains to be further studied.

Acknowledgments: This study was financially supported by the National Natural Science Foundation of China (no. 31760659 and no. 31460582) and the Natural Science Foundation Key Project of Yunnan Province, China (no. 2014FA032 and no. 2007C0003Z).

REFERENCES

Bhattacharya, T. K., P. Kuma, A. Sharma (2008). Molecular cloning and characterization of the beta-casein gene in an indian riverine buffalo (bubalus bubalis). Buffalo Bull. 27(3): 222-230.

Bionaz, M., J. J. Loor (2007). Identification of reference genes for quantitative real-time PCR in the bovine mammary gland during the lactation cycle. Physiol. Genomics. 29(3): 312-319.

Boland, M., A. MacGibbon, J. Hill (2001). Designer milks for the new millennium. Livest. Prod. Sci. 72(1-2): 99-109.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, L. B. Larsen (2016). Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. BMC Genet. 17(1): 114.

Caroli, A. M., S. Chessa, G. J. Erhardt (2009). Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. J. Dairy Sci. 92(11): 5335-5352.

Chatchatee, P., K. M. Jarvinen, L. Bardina, L. Vila, K. Beyer, H. A. Sampson (2001). Identification of IgE and IgG binding epitopes on beta- and kappa-casein on cow's milk allergic patients. Clin. Exp. Allergy. 31(8): 1256–1265.

Choi, J., D. S. Horne, J. A. Lucey (2011). Determination of molecular weight of a purified fraction of colloidal calcium phosphate derived from the casein micelles of bovine milk. J. Dairy Sci. 94(7): 3250-3261.

Elliott, R. B., D. P. Harris, J. P. Hill, N. J. Bibby, H. E. Wasmuth (1999). Type I (insulindependent) diabetes mellitus and cow milk: casein variant consumption. Diabetologia. 42(3): 292–296.

Farrell, H. M. J., R. Jimenez-Flores, G. T. Bleck, E. M. Brown, J. E. Butle, L. K. Creamer, C. L. Hicks, C. M. Hollar, K. F. Ng-Kwai-Hang, H. E. Swaisgood (2004). Nomenclature of the proteins of cows' milk--sixth revision. J. Dairy Sci. 87(6): 1641-1674.

Gallinat, J. L., S. Qanbari, C. Drögemüller, E. C. G. Pimentel, G. Thaller, J. Tetens (2013). DNA-based identification of novel bovine casein gene variants. J. Dairy Sci. 96(1): 699-709.

Ganguly, I., S. Kumar, G. K. Gaur, U. Singh, A. Kumar, S. Kumar, S. Mann, A. Sharma (2013). Status of β-casein (*CSN2*) polymorphism in Frieswal (HF X Sahiwal Crossbred) cattle. International J. Biotechnology and Bioengineering Research. 4(3): 249-256.

Gobbetti, M., L. Stepaniak, A. M. De, A. Corsetti, C. R. Di (2002). Latent bioactive peptides in milk proteins: proteolytic activation and significance in dairy processing. Crit. Rev. Food Sci. Nutr. 42(3): 223–239.

Hanusová, E., J. Huba, M. Oravcová, P. Polák, I. Vrtková (2010). Genetic variants of beta-casein in Holstein dairy cattle in sloVaKia. Slovak J. Anim. Sci. 43(2): 63-66.

Jann, O., G. Ceriotti, A. Caroli, G. Erhardt (2002). A new variant in exon VII of bovine *β*-casein gene (*CSN2*) and its distribution among European cattle breeds. J. Anim. Breed. Genet. 119(1): 65-68.

Jensen, H. B., N. A. Poulsen, K. K. Andersen, M. Hammershøj, H. D. Poulsen, L. B. Larsen (2012). Distinct composition of bovine milk from Jersey and Holstein-Friesian cows with good, poor, or noncoagulation properties as reflected in protein genetic variants and isoforms. J. Dairy Sci. 95(12): 6905-6917.

Jinsmaa, Y., M. Yoshikawa (1999). Enzymatic release of neocasomorphin and beta-casomorphin from bovine beta-casein. Peptides. 20(8): 957–962.

Kamiński, S., A. Cieslińska, E. Kostyra (2007). Polymorphism of bovine beta-casein and its potential effect on human health. J. Appl. Genet. 48(3): 189-198.

Knight, Z. A., B. Schilling, R. H. Row, D. M. Kenski, B. W. Gibson, K. M. Shokat (2003). Phosphospecific proteolysis for mapping sites of protein phosphorylation. Nat. Biotechnol. 21(9): 1047-1054.

Kumar, S., G. Stecher, K. Tamura (2016). Mega7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33(7):1870-1874.

Martin, P., M. Szymanowska, L. Zwierzchowski, C. Leroux (2002). The impact of genetic polymorphisms on the protein composition of ruminant milks. Reprod. Nutr. Dev. 42(5): 433-459.

McLachlan, C. N (2001). Beta-casein A1, ischaemic heart disease mortality, and other illnesses. Med. Hypotheses. 56(2): 262–272.

Mi, H., X. Huang, A. Muruganujan, H. Tang, C. Mills, A. Kang, P. D. Thomas (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45(D1): D183-D189.

Michelizzi, V. N., M. V. Dodson, Z. Pan, M. E. Amaral, J. J. Michal, D. J. McLean, J. E. Womack, Z. Jiang (2010). Water buffalo genome science comes of age. Int. J. Biol. Sci. 6(4): 333-349.

Miluchová, M., M. Gábor, A. Trakovická (2014). Analysis of beta-casein gene (*CSN2*) polymorphism in different breeds of cattle. Scientific Papers: Animal Science and Biotechnologies. 47(2): 56-59.

Mishra, B. P., M. Mukesh, B. Prakash, M. Sodhi, R. Kapila, A. Kishore, R. R. Kataria, B. K. Joshi, V. Bhasin, T. J. Rasool, K. M. Bujarbaruah (2009). Status of milk protein, *β*-casein variants among Indian milch animals. Indian J. Anim. Sci. 79(7): 722-725.

Poulsen, N. A., H. B. Jensen, L. B. Larsen (2016). Factors influencing degree of glycosylation and phosphorylation of caseins in individual cow milk samples. J. Dairy Sci. 99(5): 3325-3333.

Sambrock, J., D. Russell (2001). Molecular cloning: a laboratory manual. 3rd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 6.4-6.12.

Song, S., Y. Ou-Yang, J. Huo, Y. Zhang, C. Yu, M. Liu, X. Teng, Y. Miao (2016). Molecular cloning, sequence characterization, and tissue expression analysis of three water buffalo (Bubalus bubalis) genes - *ST6GAL1*, *ST8SIA4*, and *SLC35C1*. Arch. Anim. Breed. 59: 363-372.

Stephens, M., N. Smith, P. Donnelly (2001). A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68(4): 978-989.

Vinesh, P. V., B. Brahma, R. Kaur, T. K. Datta, S. L. Goswami, S. De (2013). Characterization of β-casein gene in Indian riverine buffalo. Gene. 527(2): 683-688.

Vincent, S. T., O. M. Momoh, A. Yakubu (2014). Bioinformatics analysis of beta-casein gene in some selected mammalian species. Res. Opin. Anim. Vet. Sci. 4(10): 564-570.

Yeh, F. C., R. Yang, T. J. Boyle, Z. Ye, J. M. Xiyan (1999). PopGene 3.2, Microsoft Windows-Based freeware for population genetic analysis. Version 1.32. Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, Alberta, Canada.

Yuan, B., X. Y. Li, T. Zhu, L. Yuan, J. P. Hu, J. Chen, W. Gao, W. Z. Ren (2015). Antibody study in canine distemper virus nucleocapsid protein gene-immunized mice. Genet. Mol. Res. 14(2): 3098-3105.

Table S1. The published *CSN2* gene sequences of *Bubalus* and *Bos* cited in this study.

Species	Accession numbers
Buffalo	DW007989, AJ005432, DQ317447, DW007978, GQ176288, EF115306, KC577240, GQ17628, KX896651, XM_006071124, AJ005165, DQ631829, FM946182, FN424088, GQ176287, NM_001290879, GQ176292
Cattle	XM_010806178, KC993858, X06359, DR712355, M55158, KX618695, M16645, EF123100, BC111172, DR711795, DR711797, DR711801, DR711802, DR711805, DR711806, DR711814, DR711822, DR711827, DR711835, DR711839, DR711848, DR711858, DR711866, DR711867, DR711876, DR711890, DR711916, DR711927, DR711930, DR711934, DR711942, DR711943, DR711944, DR711956, DR711964, DR711979, DR711997, DR712009, DR712012, DR712018, DR712022, DR712024, DR712027, DR712038, DR712043, DR712046, DR712055, DR712073, DR712074, DR712078, DR712086, DR712089, DR712098, DR712100, DR712112, DR712115, DR712136, DR712143-DR712146, DR712186, DR712190, DR712225, DR712229, DR712240, DR712247, DR712250, DR712252, DR712255, DR712258, DR712259, DR712311, DR712324, DR712351, DR712356, DR712371, DR712374, DR712376, DR712382, DR712383, DR712395, DR712403, DR712426, DR712430, DT830908, DT831154, DT832526, DT835350, DT837336, DT838903, DT839291, DT839749, DT842643, DT842905, DT845526, DT847152, DT851074, DT851130, DT851218, DT855193, DT856772, DT859555, DT861678, EH372806, R711905, XM_015463786, XM_015471671
Yak	JN655524, JN655525, XM_014480228, XM_005902037, XM_014480227, XM_014480229
Zebu	KY290547-KY290551, EF133460, XM_019962870
Bison	XM_010852144