Pakistan J. Zool., pp 1-10, 2021. **DOI**: https://dx.doi.org/10.17582/journal.pjz/20191202121250

Polymorphisms of the *CSN1S1* Gene and its Protein Variants in River and Swamp Buffalo (*Bubalus bubalis*)

Yongyun Zhang[1, 2], Xinyang Fan[1], Fangting Zhou[1], Weizhen Li[3], Yina Ouyang[1,4] and Yongwang Miao[1]*

[1]*Faculty of Animal Science and Technology, Yunnan Agricultural University, Kunming, Yunnan, 650201, China*
[2]*Teaching Demonstration Center of the Basic Experiments of Agricultural Majors, Yunnan Agricultural University, Kunming, Yunnan, 650201, China*
[3]*College of Veterinary Medicine, Yunnan Agricultural University, Kunming, Yunnan, 650201, China*
[4]*Yunnan Animal Science and Veterinary Institute, Kunming, Yunnan, 650224, China*

Yongyun Zhang and Xinyang Fan are the co-first authors.

Article Information
Received 02 December 2019
Revised 11 January 2020
Accepted 14 January 2020
Available online 13 August 2020

Authors' Contribution
YM and YZ conceived and designed the research. YZ and XF performed the material preparation and experiments. FZ and WL performed the data collection and analysis. YZ and YM drafted the manuscript. All authors read and approved the final manuscript.

Key words
Buffalo, *CSN1S1*, Polymorphisms, Variant, Caseins

ABSTRACT

Alpha S1-casein (α_{S1}-CN) is a major casein in milk, which exerts a crucial role in casein transport and is related to individual milk components, nutritive value and production traits of milk. So far, α_{S1}-CN coding gene (*CSN1S1*) has been widely studied in dairy cattle, but the polymorphisms of the *CSN1S1* gene have not been fully understood in buffalo. In this study, the polymorphisms in coding sequence (CDS) of the *CSN1S1* gene for river and swamp buffalo were detected using PCR product direct sequencing. The CDS for both types of buffalo was the same in length, which contained 645 nucleotides and encoded a peptide composing of 214 residues. A total of 5 single nucleotide polymorphisms (SNPs) was identified in two types of buffalo. Among them, c.516T>C and c.578C>T were observed only in river buffalo, while c.175A>G, c.580T>C and c.609T>G were found only in swamp buffalo. The c.175A>G, c.578C>T and c.580T>C were non-synonymous, which led to substitutions of p.I44V, p.L178S and p.F179L. The prediction showed that the p.F179L may affect the function of buffalo α_{S1}-CN. Eight buffalo *CSN1S1* haplotypes were defined in this study, and accordingly, 6 protein variants and 2 synonymous variants of α_{S1}-CN were inferred and named. The variants A, B', B'', C, E and F were observed only in river buffalo, whereas variant D was found only in swamp buffalo. The variant B was shared by both types of buffalo with high frequencies. The buffalo variants determined here did not exist in *Bos* genus. In addition, 9 amino acid differential sites of α_{S1}-CN between buffalo and *Bos* genus were identified, of which p.42T and p.115S were located at phosphorylation sites, which may lead to differences in the physicochemical properties of α_{S1}-CN between buffalo and *Bos* genus.

INTRODUCTION

Animal milk is essential to human nutrition and is rich in protein, fat, vitamins and minerals. Around the world, about 13% of protein needs are met from milk and dairy products (Reinhardt *et al*., 2012). Caseins were composed of α_{S1}-, α_{S2}-, β- and κ-casein, which contain almost all kinds of amino acids, accounting for about 80% of the total milk protein (Gustavsson *et al*., 2014). As members of the phosphoprotein family, they are nutritional carriers of amino acids, calcium, phosphate and minerals in milk (Ginger and Grigor, 1999). Caseins can also provide several bioactive peptides which have anti-inflammatory activities and play an important role in immune regulation (Hatori *et al*., 2008; Bicer *et al*., 2009). In the animals of *Bos* genus and goat, four caseins are encoded by four tightly linked genes of the *CSN1S1*, *CSN2, CSN1S2*, and *CSN3* which are located on chromosome 6 within a 250-kb genomic DNA region (Rijnkels *et al*., 1997; Bevilacqua *et al*., 2006). Similarity, buffalo milk also contains the above four casein components and encoded by the same 4 genes which are located as a cluster on BBU 7 (Iannuzzi *et al*., 2003). Previous studies in dairy cows have indicated that the casein genes are closely related to milk production and processing traits, and the nutritive value of milk (Wedholm *et al*., 2006). Several QTLs affecting yield traits and protein contents near the casein region were also reported in *Bos taurus* (Kučerová *et al*., 2006). Extensive studies

* Corresponding author: yongwangmiao1@126.com
0030-9923/2021/0001-0001 $ 9.00/0

have revealed that the variation of casein gene can be used as molecular markers for milk yield, milk composition and milk processing characteristics (Bonfatti *et al*., 2012a; Huang *et al*., 2012).

Given that α_{S1}-CN constitutes up to 39-46% of total caseins in bovine milk and 21% in buffalo milk, it has aroused great research interest (Bonfatti *et al*., 2012a; Mir *et al*., 2014). Previous studies have shown the content of α_{S1}-CN in milk is highly correlated with individual milk composition. Milk with higher amount of α_{S1}-CN has higher total solid and protein, and better cheese making properties (Stocco *et al*., 2018). In addition, α_{S1}-CN plays an important role to deliver the caseins from the endoplasmic reticulum to the Golgi compartment (Chanat *et al*., 1999). In 1991, the full sequence of bovine *CSN1S1* was cloned and the results showed it consists of 19 exons (Koczan *et al*., 1991). Buffalo *CSN1S1* gene also contains 19 exons and encodes a precursor of 214 amino acids with a signal peptide of 15 amino acid residues, which is highly similar to its bovine counterpart (Sukla *et al*., 2007). The polymorphisms in the *CSN1S1* gene have been investigated extensively in different bovine breeds in recent years (Caroli *et al*., 2010). Until now, ten protein variants of α_{S1}-CN including A, B, C, D, E, F, G, H, I and J have been reported based on the polymorphisms in *Bos* genus (Caroli *et al*., 2009; Gallinat *et al*., 2013). Among them, variants B and C are the most popular. However, few α_{S1}-CN variants have been reported in water buffalo (Cosenza *et al*., 2015). So far, only three α_{S1}-CN variants, A, B and B^{RV}, have been named in buffalo (Balteanu *et al*., 2007; Chianese *et al*., 2009).

The variants of the *CSN1S1* gene have been studied extensively in *Bos taurus*, *Capra hircus* and *Ovis aries* (Corral *et al*., 2010; Huang *et al*., 2012; Mestawet *et al*., 2013). However, the research on the polymorphisms of the *CSN1S1* gene is only found in river buffalo, while that in swamp buffalo is rarely reported. In this study, the coding sequence (CDS) polymorphisms of the *CSN1S1* gene in two types of buffalo were detected, the α_{S1}-CN variants in two types of buffalo was determined, and the α_{S1}-CN variants of buffalo and *Bos* genus were compared. It can provide a basis for revealing the molecular characteristics and function of buffalo *CSN1S1* gene and the effect of polymorphisms in this gene on lactation traits.

MATERIALS AND METHODS

Animals and tissue sampling

The procedures for sample collection were approved by the Institutional Animal Care and Use Committee of Yunnan Agricultural University (Kunming, Yunnan, China). In order to detect the polymorphisms of the *CSN1S1* coding region, 73 mammary gland samples were collected by puncturing from 33 river buffalo (Binglangjiang buffalo) and 40 swamp buffalo (Dehong buffalo, Guizhou buffalo and Enshi buffalo). All samples were collected at random and there was no direct consanguinity among the individuals. In order to compare with buffalo, the published *CSN1S1* gene sequences of *Bos* genus in NCBI database were downloaded and used for data analysis in this study.

RNA extraction, cDNA synthesis and genotyping

Total RNA was extracted adopting RNA extraction kit (TaKaRa, Dalian, China) based on the manufacturer instructions. The quality and quantity of RNA was detected by 1% agarose gel and the NANODROP LITE spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The RNA (3μg) was reverse-transcribed into cDNA and was stored in a refrigerator at -80 °C.

The CDS of buffalo *CSN1S1* gene was amplified with a pair of primers (*CSN1S1*-F: CTTGCTGCTTCTTCCCAGTCTTG and *CSN1S1*-R: CTATTCTAAAACAGCAGTTGAAGCCT) which were designed according to the mRNA sequence of buffalo *CSN1S1* (Accession no. HE573919) using Primer Premier 5.0 (Lalitha, 2000).

PCR was performed in a final volume of 25 μL containing 100 ng of template cDNA, 0.5 μM of each primer, 12.5 μL of 2×PCR Master Mix (CWBIO, Beijing, China). The PCR protocol consisted of an initial denaturing step at 94°C for 10 min, followed by 35 cycles of 94°C for 30 s, annealing of 55°C for 30 s, 72°C for 50 s and a final extension at 72ºC for 5 min. The PCR product was analyzed by 2% agarose gel electrophoresis, then were sequenced in both directions using above PCR primers.

Sequence data analysis

The sequences of buffalo *CSN1S1* obtained in this study were checked, proofread and outputted via Lasergene 7.0 software package (DNAStar, Inc., USA). Allele and genotype frequency, heterozygosity and Hardy-Weinberg equilibrium test were carried out adopting PopGen32 software (Yeh and Boyle, 1997). The mutation sites were exported with MEGA 6 (Tamura *et al*., 2013) and haplotypes were inferred by PHASE software with the number of iterations is ≥ 100 (Stephens *et al*., 2001). The function influence of non-synonymous substitutions was presumed by program PROVEAN (Choi and Chan, 2015). Genetic relationship among the haplotypes was constructed by Network 5 (Bandelt *et al*., 1999). The theoretical molecular weight and isoelectric point, signal peptide, transmembrane region, hydropathy and subcellular localization of buffalo α_{S1}-CN were analyzed by the ProtParam tool (http://web.expasy.org/protparam/),

Signal P 5.0 server (http://www.cbs.dtu.dk/services/SignalP/), TMHMM version 2.0 (http://www.cbs.dtu.dk/services/TMHMM/), ProtScale (http://web.expasy.org/protscale/), and ProtComp 9.0 (http://linux1.softberry.com/berry.phtml), respectively. Post translational modification site were predicted by NetPhos 3.1 Server (http://www.cbs.dtu.dk/services/NetPhos/) and NetOGlyc 4.0 Server (http://www.cbs.dtu.dk/services/NetOGlyc/), respectively. The conserved domains of buffalo α_{S1}-CN were analyzed using the Conserved Domain Architecture Retrieval Tool in BLAST at the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST). Alignments of the sequences were performed using ClustalX 2.0 (Larkin *et al*., 2007). The optimal maximum likelihood model was determined by MEGA 6, and then the phylogenetic tree was constructed based on Jones-Taylor-Thornton model (JTT model) with a bootstrap test of 10,000 replicates.

RESULTS

Molecular characteristics of buffalo α_{S1}-CN

The full-length CDS of buffalo *CSN1S1* gene was amplified adopting RT-PCR using the cDNA as templates. PCR product of 821 bp was obtained consistent with expectations. The open reading frame (ORF) in the sequences obtained was determined by EditSeq program in Lasergene 7.0 software package. Then, taken the obtained CDS as the query sequence, its homologous sequences were searched by using the BLAST program. The consistency between the sequences in this study and bovine *CSN1S1* gene (accession no. NM_181029) was more than 97.4%. Therefore, the sequences was identified as that of buffalo *CSN1S1* gene. The CDS of the *CSN1S1* for both types of buffalo was the same in length, which contained 645 nucleotides and encoded a peptide containing 214 amino acid residues.

Basic physicochemical characteristics of buffalo and cattle α_{S1}-CN (accession no. CCC54662 and CAA42516) were predicted in Table I. Bioinformatics predictions showed that buffalo α_{S1}-CN is a hydrophilic protein with a 15 amino acids (AAs) signal peptide at N-terminus and without a transmembrane region. This protein contains a casein domain (AA114 to AA187), which belongs to casein superfamily. According to the function described in UniProt (https://www.uniprot.org), α_{S1}-CN exerts a key role in the transport of calcium phosphate in buffalo and cow milk. The results of cytoplasmic/nuclear discrimination suggested that the buffalo α_{S1}-CN is secreted to extracellular with high reliability (100%). The predicted modification sites of α_{S1}-CN mature peptide in buffalo and cattle are shown in Table II. No O-glycosylation site was found in both types of buffalo.

Polymorphisms in the CDS of the CSN1S1 gene

A total of five SNPs in the *CSN1S1* gene were identified in the two types of buffalo, of which one SNP (c.175A>G) was located in exon 7 and four SNPs (c.516T>C, c.578C>T, c.580T>C and c.609T>G) in exon 17 (Table III). The c.516T>C and c.578C>T were only found in river buffalo, while the c.175A>G, c.580T>C and c.609T>G were found in swamp buffalo. And alleles c.175A, c.516T, c.578C, c.580T and c.609T had high frequencies in the population. The test of Chi-square for Hardy-Weinberg Equilibrium (HWE) showed that all the five SNPs were in an unbalanced state ($P<0.05$).

Among the five SNPs, c.175A>G, c.578C>T and c.580T>C were non-synonymous which led to corresponding amino acid changes of p.I44V, p.L178S, and p.F179L, respectively. The effects of these non-synonymous substitutions on the function of α_{S1}-CN were predicted, and the results showed that the substitution p.F179L had a significant impact on the function of α_{S1}-CN in buffalo (Table IV).

Haplotype analysis and their genetic relationship

According to the five SNPs found in this study, seven possible haplotypes were inferred, of which the frequencies of four haplotypes were more than 0.05, named B1, B2, B3 and B4 (Table V). The haplotype sequences of this study were submitted into the NCBI database under accession no. MK067366-MK067369. Among the four haplotypes, B2 was shared by two types of buffalo with the highest frequency, B1 and B3 were found only in river buffalo, and B4 in swamp buffalo. B3 and B4 were the new haplotypes found in this study. The amino acid sequence encoded by B3 was the same as that of B2. Previous studies have found that haplotypes B1 and B2 exist in river buffalo (Chianese *et al*., 2009).

At present, there are eight complete sequences of buffalo *CSN1S1* in NCBI database (accession nos.: XM_006071125, AJ005430, AY948385, DQ111783, HE573919, HE573920, XM_006071126 and FJ392261), including six haplotypes. Among the six haplotypes, two haplotypes are consistent with the haplotypes in this study (B1: XM_006071125 and HE573920; B2: AY948385 and HE573919), and the other four haplotypes belong to the sequences of river buffalo, named B5-B8 (AJ005430, DQ111783, XM_006071126 and FJ392261). B1, B2 and B8 are equivalent to the previously reported alleles A, B and B^{RV} (Balteanu *et al*., 2008; Cosenza *et al*., 2015). Thus, combining the data of this study with the published data, a total of eight haplotypes of the *CSN1S1* gene in buffalo were identified (Fig. 1). The CDS length of haplotype B1-B6 was 645bp, while the CDS length of B7 and B8 was 621bp due to skipping exon 6 (deletion of 24 nucleotides).

Table I. Physicochemical characteristics of α_{S1}-CN mature peptide for buffalo and cattle.

Basic physical and chemical properties	Water buffalo	Cattle
Formula	$C_{1029}H_{1577}N_{261}O_{313}S_5$	$C_{1035}H_{1587}N_{265}O_{317}S_5$
Number of amino acids	199	199
Molecular weight	22.77KD	22.97KD
Isoelectric point (pI)	4.70	4.85
Strongly acidic amino acid (D, E)	29	32
Strongly basic amino acid (K, R)	17	20
Polar amino acid (N, C, Q, S, T, Y)	55	53
Hydrophobic amino acid (A, I, L, F, W, V)	60	58
Instability index (II)	60.08	57.99
Grand average of hydropathicity (GRAVY)	-0.592	-0.704
Aliphatic index	79.35	75.43

Table II. Modification sites in the mature peptide of buffalo and cattle α_{S1}-CN.

Modification	Buffalo (B variant)	Cattle (B variant)
Phosphorylation	41S, 42T, 46S, 48S, 49T, 64S, 66S, 67S, 68S, 75S, 88S, 94Y, 122S, 144Y, 171T, 178S, 180S, 191S	41S, 46S, 48S, 49T, 64S, 66S, 67S, 68S, 75S, 88S, 94Y, 115S, 122S, 144Y, 171T, 173Y, 178S, 180S, 191S, 194T
O-glycosylation		75S

Table III. Genotype and allele frequencies, HE and *P* value of Chi-square test for the SNPs.

Population	SNPs	Genptype frequency		Allele frequency		$_H$E	P value
		Genotype	Frequency	Allele	Frequency		
River buffalo	c.516T>C (Exon17)	TT	0.818	T	0.879	0.216	0.007318
		TC	0.121	C	0.121		
		CC	0.061				
	c.578C>T (Exon17)	CC	0.606	C	0.697	0.429	0.000724
		CT	0.182	T	0.303		
		TT	0.212				
Swamp buffalo	c.175A>G (Exon7)	AA	0.889	A	0.889	0.209	0.000036
		AG	0.000	G	0.111		
		GG	0.111				
	c.580T>C (Exon17)	TT	0.900	T	0.900	0.182	0.000000
		TC	0.000	C	0.100		
		CC	0.100				
	c.609T>G (Exon17)	TT	0.900	T	0.900	0.182	0.000000
		TG	0.000	G	0.100		
		GG	0.100				

SNP, means single nucleotide polymorphism; H_E, expected heterozygosity; *P* value, the probability of Chi-square test under Hardy-Weinberg equilibrium.

Table IV. The functional effect of non-synonymous substitutions on buffalo α_{S1}-CN.

SNP	Substitution	PROVEAN score	Prediction (cutoff=-2.5)
c.175A>G	I44V	0.184	Neutral
c.578C>T	L178S	0.376	Neutral
c.580T>C	F179L	-3.010	Deleterious

Table V. Frequencies of *CSN1S1* haplotypes in two types of buffalo.

Haplotype ID	Base composition of haplotype	Actual frequency	Expected frenquency	Number	AFR	AFS
B1	ATTTT	0.1370	0.1367	20	0.3030	0.0000
B2	ATCTT	0.7534	0.7517	110	0.5758	0.9000
B3	ACCTT	0.0548	0.0546	8	0.1212	0.0000
B4	GTCCG	0.0548	0.0528	8	0.0000	0.1000

Note: Haplotypes with frequency lower than 0.05 were not considered, and the frequency is estimated by program PHASE. AFR, actual frequency in river buffalo; AFS, actual frequency in swamp buffalo.

Table VI. Amino acid positions and differences in the variants of buffalo α_{S1}-CN.

Variants (haplotype)	Positon and amino acid in the mature peptide								
	31	35-42	44	97	157	178	179	188	192
Buffalo A (B1)	Val GTG		Ile ATT	Gln CAG	Asp GAT	Leu TTA	Phe TTC	Ser TCT	Gly GGA
Buffalo B (B2)						Ser TCA			
Buffalo B'(B3)					Asp GAC	Ser TCA			
Buffalo B''(B5)				Gln CAA		Ser TCA			
Buffalo C (B6)					Asp GAC	Ser TCA			Glu GAA
Buffalo D (B4)			Val GTT			Ser TCA	Leu CTC	Ser TCG	
Buffalo E (B7)		deleted				Leu TTA			
Buffalo F (B8)	Met ATG	deleted	Val GTT			Ser TCA			

```
                111 111 111 111 111 111 111 111 111 111 333 555 555 555 666 666
                333 445 555 555 555 666 666 666 677 777 333 111 777 888 000 122
                678 890 123 456 789 012 345 678 901 567 456 456 789 012 789 901
Buffalo B1  GTG GAG AAG GTC AAT GAA CTG AGC ACG ATT CAG GAT TTA TTC TCT GGA
Buffalo B2  ... ... ... ... ... ... ... ... ... ... ... ... .C. ... ... ...
Buffalo B3  ... ... ... ... ... ... ... ... ... ... ... ..C .C. ... ... ...
Buffalo B4  ... ... ... ... ... ... ... ... ... G.. ... ... .C. C.. ..G ...
Buffalo B5  ... ... ... ... ... ... ... ... ... ... ..A ... .C. ... ... ...
Buffalo B6  ... ... ... ... ... ... ... ... ... ... ... ..C .C. ... ... .A.
Buffalo B7  ... --- --- --- --- --- --- --- --- ... ... ... ... ... ... ...
Buffalo B8  A.. --- --- --- --- --- --- --- --- G.. ... ... .C. ... ... ...
```

Fig. 1. Nucleotide of the haplotype sequences in buffalo. Dots (.) denote identity with the Buffalo_hap1. Nucleotide substitutions are denoted by different letters. Missing information is demonstrated by a blank (-). The same hereinafter.

The possible genetic relationships among the eight haplotypes of the *CSN1S1* gene in buffalo were investigated by employing median-joining network. As shown in the Figure 2, haplotype B2 was the dominant haplotype which was widely distributed in two types of buffalo. Other haplotypes may originate from B2, that is, B1, B3, B5 and B7 (exon 6 skipping) may evolve from B2 through a single mutation, while B6 and B8 (exon 6 skipping) evolved from B2 through two mutations, and the outermost B4 evolved through three mutations from B2.

Variants of buffalo α_{S1}-CN and their phylogenetic analysis

According to eight haplotypes of buffalo *CSN1S1* gene, eight α_{S1}-CN variants were identified. According to the existing nomenclature of *Bos* genus, we named the α_{S1}-CN variants in buffalo as A, B, B', B'', C, D, E and F, respectively (Figs. 2 and 3), in which the B' and B'' were synonymous variants of the B. Amino acid positions and differences for the genetic variants of buffalo α_{S1}-CN are presented in Table VI. Variant B were shared by both types of buffalo, and variants A, B', B'', C, E and F were found in river buffalo, whereas variant D was found in swamp buffalo. Variants A, B and F correspond to the previously named α_{S1}-CN variant A, B and B^{RV}. It is worth noting that α_{S1}-CN variants E and F have eight amino acid deletions compared with other variants, which is caused by exon 6 skipping. Among all buffalo α_{S1}-CN variants, variant B is a widely distributed variant in two types of buffalo, and other variants of buffalo α_{S1}-CN may have evolved from it.

Based on the literatures (Caroli *et al*., 2009; Gallinat *et al*., 2013), we have reconstructed amino acid sequences of all α_{S1}-CN variants found in *Bos* genus for comparative analysis. It was found that the α_{S1}-CN variants identified in buffalo were obviously different from those in *Bos* genus, and there were 9 amino acid differences (six of which involving charged amino acid residues) between buffalo and bovine α_{S1}-CN variants (Fig. 3) which included p.Q4H, p.G14E, p.T42K, p.I74N, p.N105K, p.L115S, p.Q119R, p.Q148E and p.P174T. Compared to *Bos genus*, buffalo

α_{S1}-CN had less negative charge due to some amino acid residues in the α_{S1}-CN, such as p.14G, p.115L, p.148Q and so on.

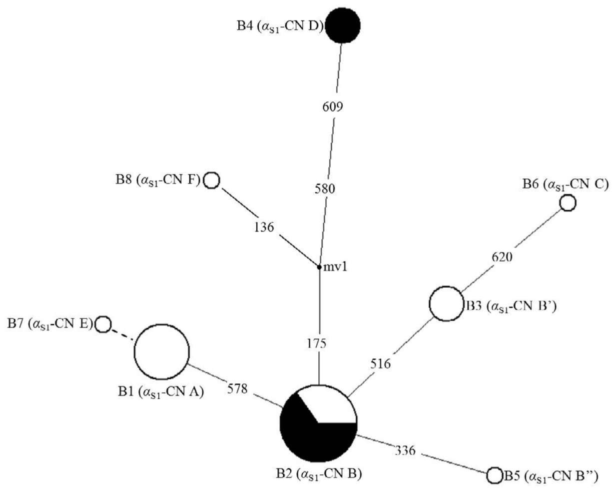

Fig. 2. Network profile of 8 buffalo haplotypes of the *CSN1S1* gene. The links are labeled by nucleotide positions to designate transition or transversions. Dotted line represents the haplotype generated from skipping exon 6 after transcription. The *CSN1S1* haplotype frequency in buffalo was proportional to the circle area. Samples from river and swamp buffalo are indicated by white and black color, respectively.

```
                                                 111111111
                 11111122222223333334444555555555678011467779
                445678901234561567890124123456789644559874892
Buffalo A       QGVLNENLLRFFVAVEKVNELSTIDQAMEDIKQSIENLQQVPLFG
Buffalo B       ..........................................S..
Buffalo B'      ..........................................S..
Buffalo B"      ..........................................S..
Buffalo C       ..........................................S.E
Buffalo D       .......................V..................SL.
Buffalo E       ...............--------......................
Buffalo F       ..............M--------V..................S..
Bovine  A       H-------------........K...........N.KSRE.TS.E
Bovine  B       HE....................K...........N.KSRE.TS.E
Bovine  C       HE....................K...........N.KSRE.TS..
Bovine  D       HE....................K...T.......N.KSRE.TS.E
Bovine  E       HE....................K.........K.N.KSRE.TS..
Bovine  F       HE....................K..........LN.KSRE.TS.E
Bovine  G       HE....................K...........N.KSRE.TS.E
Bovine  H       HE....................K.--------..N.KSRE.TS.E
Bovine  I       HE....................K...........NDKSRE.TS.E
Bovine  J       HE....................K...........N.KSREDTS.E
```

Fig. 3. Sequence differences of α_{S1}-CN variants between buffalo and the species of *Bos* genus.

A phylogenetic tree of maximum-likelihood was constructed with goat and sheep as outgroup based on the amino acid sequences of α_{S1}-CN variants of buffalo and *Bos* genus (Fig. 4). The tree showed that the α_{S1}-CN variants of buffalo and bovine species were clustered on different branches, respectively. This shows that there is a close genetic relationship between the species of *Bos* genus, while the genetic relationship between buffalo and the species of *Bos* genus is relatively far away.

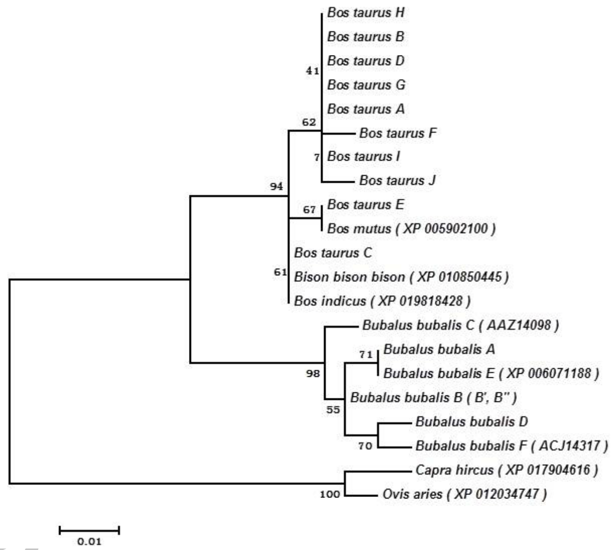

Fig. 4. Maximum-likelihood phylogenetic trees constructed by using goat and sheep as outgroups. Support rates are estimated based on bootstrap test with 10,000 replicates.

DISCUSSION

α_{S1}-casein is a phosphorylated protein, in addition to being an important nutrient, it also exerts a key role in transportation of calcium phosphate (Ginger *et al.*, 1999). Though α_{S1}-CN accounting for 21% of the total protein in buffalo milk (Bonfatti *et al.*, 2012a), the study on the variation of buffalo *CSN1S1* gene is not enough. In this study, the polymorphisms in the CDS of the *CSN1S1* gene for river and swamp buffalo were detected by direct sequencing of RT-PCR products. As a result, a total of five SNPs were found in two types of buffalo. It is worth noting that there are no shared SNPs between river and swamp buffalo, and the SNPs found is mainly distributed in exon 17. Previous studies have shown that there are four SNPs in the CDS of river buffalo *CSN1S1*, that is, c.136G>A in exon 5, c.175A>G in exon 7, c.578C>T and c.620G>A in exon 17 (Balteanu *et al.*, 2013; Cosenza *et al.*, 2015). The SNPs of buffalo *CSN1S1* gene found in this study is not completely consistent with the previous study, and there are three new SNPs here. In previous studies, it has been reported that there are two kinds of the *CSN1S1* transcripts in river buffalo, one covering 19 exons with a CDS length

of 645 bp, and the other has a exon 6 skipping (B^{RV}) with a CDS length of only 621 bp. In this study, only the former *CSN1S1* transcript was found in the mammary gland of both types of buffalo, indicating that this transcript is the main form to express the *CSN1S1* in the mammary gland.

In recent decades, casein genes have been widely studied in the family of *Bovidae* (Huang *et al*., 2012; Mir *et al*., 2014; Stocco *et al*., 2018), especially in dairy cattle. Previous studies in dairy cow have revealed that the polymorphisms of the *CSN1S1* gene are not only related to micelle formation and casein transport (Martin *et al*., 1999; Le Parc *et al*., 2010), but also related to protein percentage, milk yield, and fat content (Seefried, 2007). In recent years, some studies have also shown that the polymorphisms of the *CSN1S1* gene was closely related with production and milk components traits of river buffalo. Among the five SNPs found in this study, c.175A>G, c.578C>T and c.580T>C were non-synonymous, which led to the changes of p.I44V, p.S178L and p.F179L in α_{S1}-CN mature peptide, respectively. It is predicted that the change of p.F179L may have a significant effect on the function of buffalo α_{S1}-CN, which changes the hydrophilicity of α_{S1}-CN. In this study, whether the SNPs found in buffalo *CSN1S1*, especially c.580T>C, affect the function of α_{S1}-CN and whether they are related to milk yield and milk quality need to be further studied.

So far, ten α_{S1}-CN variants have been found in *Bos* genus. However, due to the limited previous studies on the polymorphisms of buffalo *CSN1S1* gene, only three α_{S1}-CN variants, including A, B and B^{RV}, have been identified in buffalo. In this study, we studied the variation of the *CSN1S1* gene in two types of buffalo in an attempt to achieve a comprehensive understanding of its variants. Sequence alignment showed that there was a significant difference in α_{S1}-CN variant sequences between buffalo and *Bos* genus, and the α_{S1}-CN variants identified in *Bos* genus did not exist in buffalo. In view of the great differences in α_{S1}-CN sequences between buffalo and bovine, it is necessary to name the variants of buffalo α_{S1}-CN separately. According to the existing naming conventions, six α_{S1}-CN variants and two synonymous variants were named here in buffalo based on α_{S1}-CN haplotype. Among them, variants A, B and F, are consistent with the previously named α_{S1}-CN A, B and B^{RV}. Variant B is widely distributed in two types of buffalo. From the median-joining network of the haplotypes and the sequence difference between variants, it can be indicated that variant B is probably the ancestral form of buffalo α_{S1}-CN. Variants B' and B'' may originate from B through a synonymous mutation, variants A and C may be directly derived from variant B by one amino acid substitution at residue p.178 and p.192, respectively. And variants D may be derived from variant B by two substitutions at residues p.44 and p.179, while variant E may directly originate from A through exon 6 skipping, and variant F may generate from B through two substitutions at residues p.31 and p.44 plus exon 6 skipping.

The comparison of α_{S1}-CN variants showed that there were 9 amino acid differences between buffalo and *Bos* genus, indicating that there were great genetic differences between buffalo and *Bos* genus. The analysis of the molecular characteristics of α_{S1}-CN showed that there were differences in the basic physicochemical properties, especially in post-translational modification, between buffalo and cattle α_{S1}-CN. Post-translational modification of proteins exerts important roles in the realization of their biological functions, including phosphorylation, O-glycosylation and N-glycosylation (Jensen *et al*., 2015). The predicted results showed that the post-translational modification of α_{S1}-CN is mainly phosphorylation, which mostly occurred on the serine and threonine of the polypeptide chain. The phosphorylation level of α_{S1}-CN will affect its physicochemical properties, which is not only related to the formation of casein micelles and milk coagulation properties (Poulsen *et al*., 2016), but also affect the mineral binding ability of mature peptides (Sukla *et al*., 2007). Since the α_{S1}- and β-casein fractions constitute up to 70% of the micelle protein network (Ferranti *et al*., 1998), the localization of the phosphorylation sites of α_{S1}-CN may exert a crucial role in micelle aggregation. The loss of the phosphoserine 115 in buffalo α_{S1}-CN, substituted by a leucine residue, strengthens the nonpolar characteristics of the protein domain and could partly explain the difference observed in the composition of the casein micelles from buffalo and cattle milk (Ferranti *et al*., 1998). In this study, the phosphorylation site p.42T was found in buffalo, but not in cattle. Another phosphorylation site, p.115S, was found in cattle, but not in buffalo. These two sites are also the amino acid differential sites of α_{S1}-CN between buffalo and cattle. In addition, there are two more phosphorylation sites of α_{S1}-CN in cattle than in buffalo, i.e. 173Y, 194T. According to previous studies, buffalo milk has shorter clotting time and higher cheese production than bovine milk (Bonfatti *et al*., 2012b). Whether the difference of post-translational modification of α_{S1}-CN between buffalo and cattle is related to the characteristics of their milk remains to be further studied.

CONCLUSIONS

In this study, a total of five SNPs were found in the *CSN1S1* gene of two types of buffalo, including two SNPs in river buffalo and three SNPs in swamp buffalo. Based on the data of this study and the published data, eight haplotypes of buffalo *CSN1S1* were defined, from which six

α_{S1}-CN variants and two synonymous variants were inferred and named in buffalo. Variant B is widely distributed in two types of buffalo, which is probably the ancestral form of buffalo α_{S1}-CN. The sequences of α_{S1}-CN variants were significantly different between buffalo and *Bos* genus. In addition, this study also indicated that there were disparities in the basic physicochemical properties, especially the post-translational modification, between buffalo and bovine α_{S1}-CN. This study can provide a basis for revealing the different physicochemical and processing properties of milk between buffalo and *Bos* genus. Furthermore, this study can also provide a basis for exploring the association between the *CSN1S1* polymorphism and milk yield traits.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 31760659 and No. 31460582), the Natural Science Foundation Key Project of Yunnan Province, China (No. 2014FA032).

Statement of conflict of interest

The authors have declared no conflict of interest.

REFERENCES

Balteanu, V.A., Carsai, T.C. and Vlaic, A., 2013. Identification of an intronic regulatory mutation at the buffalo as1-casein gene that triggers the skipping of exon 6. *Mol. Biol. Rep.*, **40**: 4311-4316. https://doi.org/10.1007/s11033-013-2518-2

Balteanu, V.A., Pop, F.D., Vlaic, A. and Rusu, A.R., 2008. Multiple mutations events are characterizing α_{S1}-casein B^{RV} and beta casein C^{RV} alleles discovered in Romanian buffalo breed. *Bull. USAMV Anim. Sci. Biotechnol.*, **65**: 478.

Balteanu, V.A., Vlaic, A., Rusu, A., Pop, F.D., Odagiu, A. and Creanga, S., 2007. Identification of two possible new casein alleles in Romanian buffalo milk: α_{S1}-casein B^{RV}, β-casein C^{RV}. *Bull. USAMV-CN,* **62**: 528.

Bandelt, H.J., Forster, P. and Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.*, **16**: 37-48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Bevilacqua, C., Helbling, J.C., Miranda, G. and Martin, P., 2006. Translational efficiency of casein transcripts in the mammary tissue of lactating ruminants. *Reprod. Nutr. Dev.*, **46**: 567-578. https://doi.org/10.1051/rnd:2006028

Bicer, S., Reiser, P.J., Ching, S. and Quan, N., 2009. Induction of muscle weakness by local inflammation: an experimental animal model. *Inflamm. Res.*, **58**: 175-183. https://doi.org/10.1007/s00011-008-8093-7

Bonfatti, V., Giantin, M., Gervaso, M., Rostellato, R., Coletta, A., Dacasto, M. and Carnier, P., 2012a. Short communication: CSN1S1-CSN3 (α_{S1}-κ-casein) composite genotypes affect detailed milk protein composition of Mediterranean water buffalo. *J. Dairy Sci.*, **95**: 6801-6805. https://doi.org/10.3168/jds.2012-5601

Bonfatti, V., Giantin, M., Gervaso, M., Coletta, A., Dacasto, M. and Carnier, P., 2012b. Effect of CSN1S1-CSN3 (α_{S1}-κ-casein) composite genotype on milk production traits and milk coagulation properties in mediterranean water buffalo. *J. Dairy Sci.*, **95**: 3435-3443. https://doi.org/10.3168/jds.2011-4901

Caroli, A.M., Chessa, S. and Erhardt, G.J., 2009. Invited review: milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. *J. Dairy Sci.*, **92**: 5335-5352. https://doi.org/10.3168/jds.2009-2461

Caroli, A, Rizzi, R., Lühken, G. and Erhardt, G., 2010. Short communication: milk protein genetic variation and casein haplotype structure in the original Pinzgauer cattle. *J. Dairy Sci.*, **93**: 1260-1265. https://doi.org/10.3168/jds.2009-2521

Chanat, E., Martin, P. and Ollivier-Bousquet, M., 1999. α_{S1}-casein is required for the efficient transport of β- and-κ-casein from the endoplasmic reticulum to the Golgi apparatus of mammary epithelial cells. *J. Cell Sci.*, **112**: 3399-3412.

Chianese, L., Quarto, M., Pizzolongo, F., Calabrese, M.G., Caira, S., Mauriello, R., De Pascale, S. and Addeo, F., 2009. Occurrence of genetic polymorphism at the α_{S1}-casein locus in Mediterranean water buffalo milk. *Int. Dairy J.*, **19**: 181-189. https://doi.org/10.1016/j.idairyj.2008.09.008

Choi, Y. and Chan, A.P., 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, **31**: 2745-2747. https://doi.org/10.1093/bioinformatics/btv195

Corral, J.M., Padilla, J.A. and Izquierdo, M., 2010. Associations between milk protein genetic polymorphisms and milk production traits in Merino sheep breed. *Livest. Sci.*, **129**: 73-79. https://doi.org/10.1016/j.livsci.2010.01.007

Cosenza, G., Pauciullo, A., Macciotta, N.P.P., Apicella, E., Steri, R., Battaglia, A.La., Jemma, L., Coletta, A., Berardino, D.D. and Amunno, L., 2015. Mediterranean river buffalo *CSN1S1* gene: Search for polymorphisms and association studies. *Anim. Prod. Sci.*, **55**: 654-660. https://doi.org/10.1071/

AN13438

Ferranti, P., Scaloni, A., Caira, S., Chianese, L., Malorni, A. and Addeo, F., 1998. The primary structure of water buffalo alpha (s1)- and beta-casein identification of phosphorylation sites and characterization of a novel beta-casein variant. *J. Protein Chem.*, **17**: 835-844. https://doi.org/10.1023/A:1020786503978

Gallinat, J.L., Qanbari, S., Drögemüller, C., Pimentel, E.C., Thaller, G. and Tetens, J., 2013. DNA-based identification of novel bovine casein gene variants. *J. Dairy Sci.*, **96**: 699-709. https://doi.org/10.3168/jds.2012-5908

Ginger, M.R. and Grigor, M.R., 1999. Comparative aspects of milk caseins. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.*, **124**: 133-145. https://doi.org/10.1016/S0305-0491(99)00110-8

Gustavsson, F., Buitenhuis, A.J., Johansson, M., Bertelsen, H.P., Glantz, M., Poulsen, N.A., Lindmark Mansson, H., Stalhammar, H., Larsen, LB, Bendixen, C., Paulsson, M. and Andrén, A., 2014. Effects of breed and casein genetic variants on protein profile in milk from Swedish red, Danish Holstein, and Danish Jersey cows. *J. Dairy Sci.*, **97**: 3866-3877. https://doi.org/10.3168/jds.2013-7312

Hatori, M., Ohki, K., Hirano, S., Yang, X.P., Kuboki, H. and Abe, C., 2008. Effects of a casein hydrolysate prepared from *Aspergillus oryzae* protease on adjuvant arthritis in rats. *Biosci. Biotechnol. Biochem.*, **72**: 1983-1991. https://doi.org/10.1271/bbb.70792

Huang, W., Peñagaricano, F., Ahmad, K.R., Lucey, J.A., Weigel, K.A. and Khatib, H., 2012. Association between milk protein gene variants and protein composition traits in dairy cattle. *J. Dairy. Sci.*, **95**: 440-449. https://doi.org/10.3168/jds.2011-4757

Iannuzzi, L., Di Meo, G.P, Perucatti, A., Schibler, L., Incarnato, D., Gallagher, D., Eggen, A., Ferretti, L., Cribiu, E.P. and Womack, J., 2003. The river buffalo (*Bubalus bubalis*, 2n = 50) cytogenetic map: assignment of 64 loci by fluorescence in situ hybridization and R-banding. *Cytogenet. Genome Res.*, **102**: 65-75. https://doi.org/10.1159/000075727

Jensen, H.B., Pedersen, K.S., Johansen, L.B., Poulsen, N.A., Bakman, M., Chatterton, D.E. and Larsen, L.B., 2015. Genetic variation and posttranslational modification of bovine κ-casein: Effects on caseino-macropeptide release during renneting. *J. Dairy Sci.*, **98**:747-758. https://doi.org/10.3168/jds.2014-8678

Koczan, D., Hobom, G. and Seyfert, H.M., 1991. Genomic organization of the bovine alpha-S1 casein gene. *Nucl. Acids Res.*, **19**: 5591-5596. https://doi.org/10.1093/nar/19.20.5591

Kučerová, J., Lund, M.S., Sørensen, P., Sahana, G., Guldbrandtsen, B., Nielsen, V. H., Thomsen, B. and Bendixen, C., 2006. Multi-trait QTL mapping for milk production traits in Danish Holstein cattle. *J. Dairy Sci.*, **89**: 2245-2256. https://doi.org/10.3168/jds.S0022-0302(06)72296-2

Lalitha, S., 2000. Primer Premier 5. *Biotech Softw. Internet Rep.*, **1**: 270-272. https://doi.org/10.1089/152791600459894

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna R, McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D, Gibson, T.J. and Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**: 2947-2948. https://doi.org/10.1093/bioinformatics/btm404

Le Parc, A., Leonil, J. and Chanat, E., 2010. AlphaS1-casein, which is essential for efficient ER-to-Golgi casein transport, is also present in a tightly membrane-associated form. *BMC Cell Biol.*, **11**: 65. https://doi.org/10.1186/1471-2121-11-65

Martin, P., Ollivier-Bousquet, M. and Grosclaude, F., 1999. Genetic polymorphism of caseins: a tool to investigate casein micelle organization. *Int. Dairy J.*, **9**: 163-171. https://doi.org/10.1016/S0958-6946(99)00055-2

Mestawet, T.A., Girma, A., Adnøy,T., Devold, T.G. and Vegarud, G.E., 2013. Newly identified mutations at the *CSN1S1* gene in Ethiopian goats affect casein content and coagulation properties of their milk. *J. Dairy Sci.*, **96**: 4857-4869. https://doi.org/10.3168/jds.2012-6467

Mir, S.N., Ullah, O. and Sheikh, R., 2014. Genetic polymorphism of milk protein variants and their association studies with milk yield in Sahiwal cattle. *Afr. J. Biotechnol.*, **13**: 555-565. https://doi.org/10.5897/AJB2013.13216

Poulsen, N.A., Jensen, H.B. and Larsen, L.B., 2016. Factors influencing degree of glycosylation and phosphorylation of caseins in individual cow milk samples. *J. Dairy Sci.*, **99**: 3325-3333. https://doi.org/10.3168/jds.2015-10226

Reinhardt, T.A., Lippolis, J.D., Nonnecke, B.J. and Sacco, R.E., 2012. Bovine milk exosome proteome. *J. Proteom.*, **75**: 1486-1492. https://doi.org/10.1016/j.jprot.2011.11.017

Rijnkels, M., Kooiman, P.M., de Boer, H.A. and Pieper, F.R., 1997. Organization of the bovine casein gene locus. *Mamm. Genome*, **8**: 148-152. https://doi.org/10.1007/s003359900377

Seefried F.R., 2007. *Characterisation and polymorphism analysis of candidate genes for milk production*

traits and association studies with milk production traits in three cattle breeds. PhD thesis, University of Munich, Munich, Germany.

Stephens, M., Smith, N. and Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, **68**: 978-989. https://doi.org/10.1086/319501

Stocco, G., Pazzola, M., Dettori, M.L., Paschino, P., Bittante, G. and Vacca, G.M., 2018. Effect of composition on coagulation, curd firming, and syneresis of goat milk. *J. Dairy Sci.*, **101**: 9693-9702. https://doi.org/10.3168/jds.2018-15027

Sukla, S., Bhattacharya, T.K., Venkatachalapathy, R.T., Kumar, P. and Sharma, A., 2007. Molecular cloning and characterization of buffalo α_{S1}-casein gene. *DNA Seq.*, **18**: 334-340. https://doi.org/10.1080/10425170701388974

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.*, **30**: 2725-2729. https://doi.org/10.1093/molbev/mst197

Wedholm, A., Larsen, L.B., Lindmark-Mansson, H., Karlsson, A.H. and Andren, A., 2006. Effect of protein composition on the cheese-making properties of milk from individual dairy cows. *J. Dairy Sci.*, **89**: 3296-3305. https://doi.org/10.3168/jds.S0022-0302(06)72366-9

Yeh, F.C. and Boyle, T.B.J., 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belg. J. Bot.*, **129**: 157-163.