





Article

# First DNA Sequencing in Beninese Indigenous Cattle Breeds Captures New Milk Protein Variants

Sèyi Fridaïus Ulrich Vanvanhossou <sup>1</sup>, Isabella Jasmin Giambra <sup>1</sup>, Tong Yin <sup>1</sup>, Kerstin Brügemann <sup>1</sup>, Luc Hippolyte Dossa <sup>2</sup> and Sven König <sup>1,\*</sup>

- <sup>1</sup> Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, 35390 Gießen, Germany; Seyi.F.Vanvanhossou-2@agrار.uni-giessen.de (S.F.U.V.); Isabella.J.Giambra@agrار.uni-giessen.de (I.J.G.); Tong.Yin@agrار.uni-giessen.de (T.Y.); Kerstin.Bruegemann@agrار.uni-giessen.de (K.B.)
- <sup>2</sup> School of Science and Technics of Animal Production, Faculty of Agricultural Sciences, University of Abomey-Calavi, Abomey-Calavi, 03 BP 2819 Jéricho Cotonou, Benin; hippolyte.dossa@fsa.uac.bj
- \* Correspondence: sven.koenig@agrار.uni-giessen.de

**Abstract:** This study investigated polymorphisms in the milk protein genes *CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *LALBA*, and *LGB*, and casein haplotypes in Beninese indigenous cattle. Considering 67 animals, DNA sequencing of the genes' exons, flanking regions and parts of the 5'-upstream regions identified 1058 genetic variants including 731 previously unknown. In addition, four novel milk protein variants were detected, including *CSN3<sup>K</sup>* (p.Ala66Val), *LALBA<sup>F</sup>* (p.Arg58Trp), *LGB<sup>B1</sup>* (p.Ala134Val) and *LGB<sup>K</sup>* (p.Thr92Asnfs\*13). *CSN3<sup>K</sup>* is caused by a novel SNP (BTA6:85656526C>T, exon 4) whereas *LALBA<sup>F</sup>* and *LGB<sup>B1</sup>* are due to rs714688595C>T (exon 1) and rs109625649C>T (exon 4), respectively. Regarding *LGB<sup>K</sup>*, a frameshift insertion of one adenine residue at BTA11:103257980 (exon 3) induces a premature translation termination resulting in a 46% reduction of the reference protein sequence. The casein polymorphisms formed five main *CSN1S1-CSN2-CSN1S2-CSN3* haplotypes including B-A1-A-B, B-A1-A-A and C-A2-A-B which are predominant in the investigated cattle breeds. Moreover, in silico analyses of polymorphisms within the 5'- and 3'- untranslated regions of all six milk proteins revealed effects on microRNA and transcription factor binding sites. This study suggests a large genetic variation of milk protein genes in Beninese cattle, which should be investigated in further studies for their effects on milk production, including quality and yield traits.

**Keywords:** casein and whey protein genes; genetic variants; casein haplotype; African taurine and indicine breeds; microRNA; transcription factor binding sites



**Citation:** Vanvanhossou, S.F.U.; Giambra, I.J.; Yin, T.; Brügemann, K.; Dossa, L.H.; König, S. First DNA Sequencing in Beninese Indigenous Cattle Breeds Captures New Milk Protein Variants. *Genes* **2021**, *12*, 1702. <https://doi.org/10.3390/genes12111702>

Academic Editor: Emilia Bagnicka

Received: 13 October 2021  
Accepted: 25 October 2021  
Published: 26 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The dairy sector in Benin, as in many African countries, is characterized by low productivity and strong dependence on imports to meet the growing demand for dairy products [1,2]. To address this situation, governments and farmers are opting for exogenous breeds, perceived as more productive. For example, the Girolando and Azawak breeds were introduced in Benin for crossbreeding in order to improve herd productivity [2,3]. Also, crossbreeding between local taurine and zebu breeds is increasingly practiced by small-scale farmers [4]. However, neither selection strategies nor proper documentation on the characteristics of the local breeds exist as a basis for optimized crossbreeding practices [5,6]. Imported exotic breeds are challenged due to their inferior adaptation to local harsh environments, implying increased disease frequencies and ongoing declines in productivity. A focus on exotic breeds also implies a growing risk to lose specific values and genetic diversity characterizing the local breeds [1,2]. Indeed, the taurine cattle (Somba, Lagune) are typically acknowledged for the appreciable organoleptic and cheesemaking qualities of their milk implying high economic values [7,8]. In contrast, the zebu cattle are known for higher milk yield, but of poorer milk quality [7]. As argued by Said Ahmed et al. [9], conservation and breeding strategies to improve milk yield should consider the

maintenance of genetic diversity and conservation of specific traits of interest, including nutritional, immunological and technological quality of milk.

Six milk protein genes influence milk production traits. They include four casein genes (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*) on *Bos taurus* autosome (BTA) 6 and two whey protein genes (*LALBA/LAA* and *LGB/PAEP*) on BTA5 and BTA11, respectively [10,11]. The genes encode the proteins  $\alpha_{s1}$ -casein (CN),  $\beta$ -CN,  $\alpha_{s2}$ -CN,  $\kappa$ -CN,  $\alpha$ -Lactalbumin ( $\alpha$ -LA), and  $\beta$ -Lactoglobulin ( $\beta$ -LG), respectively. The four caseins form a 250-kb cluster resulting in a haplotype on BTA6 [12]. Protein electrophoresis, mRNA- and DNA-sequence analyses identified enormous variability within the six main milk proteins [13,14]. The protein polymorphisms described in bovine milk include to date 10 protein variants for *CSN1S1* (A, B, C, D, E, F, G, H, I, J), 15 protein variants for *CSN2* (A1, A2, A3, B, C, D, E, F, G, H1, H2, I, J, K, L), five protein variants for *CSN1S2* (A, B, C, D, E), 14 protein variants for *CSN3* (A, A1, B, B2, C, D, E, F1, F2, G1, G2, H, I, J), five protein variants for *LALBA* (A, B, C, D, E), and 11 protein variants for *LGB* (A, B, C, D, E, F, G, H, I, J, W) [14–16]. Some protein variants, including the *CSN1S1<sup>H</sup>* and *CSN3<sup>J</sup>*, are exclusively described in the West-African taurine breed Baoulé and the crossbreed Kuri, respectively [17]. Regarding the Beninese cattle breeds, only two studies addressed milk protein gene variability [18,19]. Both studies considered samples collected in 1997 and 1998, and described divergent distribution of milk protein variants in the breeds, especially between the taurine (Lagune and Somba) and the crossbreed (Borgou) [18,19]. An unknown variant  $\alpha_{s1}$ -CN<sup>X</sup> identified by Moazami-Goudarzi et al. [18] in Borgou animals is not clarified on DNA or mRNA level to date. Several investigations on cattle breeds in other countries associated milk protein variants with milk yield, composition and coagulation traits [20–22]. For instance, the  $\beta$ -CN<sup>A2</sup>, predominant in zebu cattle presented positive effects on calf growth [23]. This variant is also of high interest for human health because of its ability to promote better digestion efficiency and antibacterial activity [14,17,24]. In contrast, the  $\beta$ -CN<sup>A1</sup> and  $\kappa$ -CN<sup>B</sup> variants, observed in high frequency in the Beninese taurine breeds, have been linked to better cheesemaking properties (curd consistency, milk or rennet coagulation, micelle size). These descriptions match the appreciable features of Beninese taurine milk as described by the farmers [25,26]. Therefore, in Beninese cattle, we hypothesize a high correlation between milk production with milk protein genetic variability.

In addition to single milk protein variants, casein clusters significantly affected milk characteristics in ruminants [27]. Researchers reported significant relationships between different haplotypes and milk protein yield, coagulation, cheesemaking properties and offspring performances [28,29]. Braunschweig et al. [30] compared the parental haplotypes B-B-A-A and B-B-A-B for *CSN1S1-CSN2-CSN1S2-CSN3* in Swiss Brown Cattle and associated B-B-A-A with lower casein content and higher whey protein content in daughter milk. Similarly, the maternal composite casein genotype *CSN1S1<sup>BB</sup>-CSN2<sup>A2A2</sup>-CSN1S2<sup>AA</sup>-CSN3<sup>AB</sup>* have been recently associated with pre-weaning offspring growth in German cattle [23].

Furthermore, polymorphisms occurring in the promoter regions of the milk protein genes are known to affect the binding sites of transcription factors (TF) or microRNA (miRNA) [31]. Transcription factors are proteins that bind to a specific DNA sequence (transcription factor binding sites, TFBS) in the cis-regulatory/intergenic regions to modulate the transcription or expression of a gene [32,33]. Similarly, miRNA, small non-coding RNA, interact with the 5'- and 3'-untranslated regions (UTR) and with gene promoter regions to suppress or activate expression, or control the translation and transcription rate of a gene [34–36]. Various single nucleotide polymorphisms (SNP) are located in TFBS or miRNA target sites flanking the milk protein genes, and are implicated in the alteration or reduction of milk protein gene functions (e.g., resulting in lower protein content) [9,37–39]. For example, the g.-475C>G SNP positioned within the activator protein-2 (AP-2) TFBS was associated with differential expression of *LGB<sup>A</sup>* and *LGB<sup>B</sup>* [40,41]. Likewise, SNP-haplotypes in the promoter regions of the *CSN1S1* are linked with different promoter alleles affecting milk production traits [42,43]. Hence, investigating the exons and flanking

regions of milk protein genes in Beninese indigenous cattle breeds is beneficial to increase knowledge on existing genetic variation and on ongoing evolutionary patterns, to unravel their potential impacts in milk production, and to establish appropriate management strategies [9,44].

Screening milk protein polymorphisms and genotype composites and casein haplotypes is particularly important in Benin given the increasing crossbreeding and environmental pressures [45,46]. Caroli et al. [14] recommended frequent monitoring of milk protein genetic variability in cattle populations to detect and prevent the expansion of undesirable genetic mutations with unfavorable effects on milk traits. Gallinat et al. [15] recently identified new milk protein variants, reflecting the evolutionary pattern in different cattle breeds. With regard to the variety of tools to assess genetic polymorphisms in milk protein genes and haplotypes (at protein or DNA level), genomic methods provide a wider overview of chromosome variations [14,23]. The latter approach is increasingly being used to characterize known and novel polymorphisms in cattle milk protein genes [9,14,47]. Consequently, in the present study, we performed the first DNA sequencing of Beninese cattle, focusing on exons and flanking intron regions and parts of the 5'-upstream regions of milk protein genes, and we explored their functional consequences *in silico*. Milk protein polymorphisms and casein haplotypes were subsequently characterized in the Beninese indigenous cattle breeds.

## 2. Materials and Methods

### 2.1. Animal Sampling, DNA Extraction and Sequencing

Sixty-seven animals consisting of 20 Lagune, 20 Somba, 27 crossbreeds (including Borgou and Pabli animals) were selected within a large dataset of Beninese indigenous cattle sampled between 2016 and 2017 [45]. The Lagune and the Somba are two representatives of West African taurine, while the Borgou and Pabli are the products of crossbreeding, i.e., Somba × White Fulani (Zebu) and Somba × Borgou, respectively [6]. The relationships between the animals were minimized by considering only one sample per herd and per village. In addition, purebred Lagune and Somba animals with respective Lagune and Somba ancestry proportions  $\geq 87\%$  were selected, considering the values estimated with Admixture by Scheper et al. [45].

The DNA of the selected animals was extracted from hair samples using the NucleoSpin Tissue Kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany), following manufacturer's instructions. A minimum of 20  $\mu\text{L}$  of DNA with a concentration of  $\sim 30 \text{ ng}/\mu\text{L}$  were submitted to LGC Genomics GmbH (Berlin, Germany) for the sequencing of all exons and flanking intron regions as well as of parts of the 5'-regions of the casein and whey protein genes. The primer sequences described in Said Ahmed et al. [9] were used to amplify all exon sequences of the six milk protein genes under investigation. Moreover, additional primer pairs were established to amplify and subsequently sequence slightly larger 5'-upstream regions of the *CSN1S1*, *CSN3*, *LALBA* and *LGB* genes, respectively (see Table S1). Sequencing was made via Genotyping-in-Thousands by sequencing (GT-seq) using the Illumina MiSeq V3 platform with  $2 \times 300 \text{ bp}$  paired-end read (LGC Genomics GmbH, Berlin, Germany). Using this method, next-generation sequencing of multiplexed polymerase chain reaction (PCR) products generated genotypes for a panel of DNA polymorphisms, in our case in the six main milk proteins in cattle.

### 2.2. Processing of Sequence Data

The raw sequence data consisting of 5 million reads were de-multiplexed with the Illumina bcl2fastq v2.20 [48]. Quality control using the FastQC v0.11.9 software filtered out low quality reads (containing ambiguous base, with final length inferior to 65 bases or with an average Phred quality score inferior to 30 over a window of ten bases). Subsequently, the BWA-MEM algorithm [49] was used to align the quality trimmed sequence reads against the respective reference sequences of cattle milk protein genes from the Ensembl genome database ([http://www.ensembl.org/Bos\\_taurus/](http://www.ensembl.org/Bos_taurus/), accessed on 13 October 2021,

see Table 1). Finally, variants and genotypes covered by a minimum coverage of 8 reads were called using the Bayesian haplotype-based method implemented in Freebayes v1.2.0, setting ploidy equal to 2. The outputs were checked and multiple nucleotide polymorphisms were manually processed to retrieve SNP, insertions and deletions (InDel).

**Table 1.** Reference gene sequences used to investigate milk protein gene polymorphisms in the Beninese indigenous cattle breeds.

Gene Name	Gene Sequence			
	Sequence ID	Location	Strand <sup>1</sup>	Exons (n)
<i>CSN1S1</i>	ENSBTAG00000007695	6:85410518:85429868	+	19
<i>CSN2</i>	ENSBTAG00000002632	6:85448564:85458344	−	9
<i>CSN1S2</i>	ENSBTAG00000005005	6:85529305:85549156	+	18
<i>CSN3</i>	ENSBTAG000000039787	6:85645254:85659526	+	5
<i>LALBA</i>	ENSBTAG00000005859	5:31182832:31213745	+	4
<i>LGB</i>	ENSBTAG00000014678	11:103255224:103264876	+	7

<sup>1</sup> + = forward strand, − = reverse strand.

### 2.3. Functional Annotation of DNA Sequence Variants, Frequencies of SNP Alleles and Casein Haplotypes

The positions of the identified sequence variants were inferred considering the gene sequence positions in the ARS-UCD1.2 genome assembly. Functional annotation, performed with the Ensembl Variant Effect Predictor online tool [50], retrieved known variants (referenced in the Ensembl database) and their potential consequence. The previously undescribed DNA variants were submitted to the European Variation Archive (EVA) which assigned individual variant accessions (ss#) number to them [51]. To avoid redundancies and discrepancies in the functional annotation of the polymorphisms, only the most conserved transcript was investigated for each milk protein gene (i.e., ENSBTAT00000010119.3 for *CSN1S1*, ENSBTAT00000003409.6 for *CSN2*, ENSBTAT00000006590.6 for *CSN1S2*, ENSBTAT00000028685.5 for *CSN3*, ENSBTAT00000007701.2 for *LALBA* and ENSBTAT00000019538.6 for *LGB*). Variants detected in the coding regions were considered to annotate milk protein variants following the existing nomenclature [9,14,15,52] and referring to the amino acid (AA) position in the mature proteins.

SNP allele frequencies in the respective breeds were calculated with the vcftools v0.1.16 software [53]. In addition, haplotypes were constructed across the casein cluster *CSN1S1-CSN2-CSN1S2-CSN3*, and their frequencies were estimated using the haplo.group function in the haplo.stats (version 1.8.7) R package [54]. In this regard, we estimated the maximum likelihood of haplotype probabilities in the different breeds as well as for the whole population. Only the casein variants with a minimum allele frequency of 5% in the overall population were considered for haplotype constructions.

Potential effects of the splice site variants were investigated in silico using the NNSPLICE 0.9 tool [55]. In addition, the polymorphisms located in the 5'- and 3'-UTR of the milk protein genes were examined for the prediction of TFBS and miRNA targets. TFBS were identified applying the sequence analysis workflow with the Gene Transcription Regulation Database (GTRD) [56,57] in the geneXplain platform 6.4 web server (<https://platform.genexplain.com/biounlweb/>, accessed on 13 October 2021). For the investigation of miRNA target sites, the seed regions of bovine miRNA were downloaded on the TargetScan website (<http://www.targetscan.org>, accessed on 13 October 2021, see Table S2) and used applying the “targetscan\_60.pl” algorithm [58,59].

## 3. Results

### 3.1. Distribution of DNA Sequence Variants

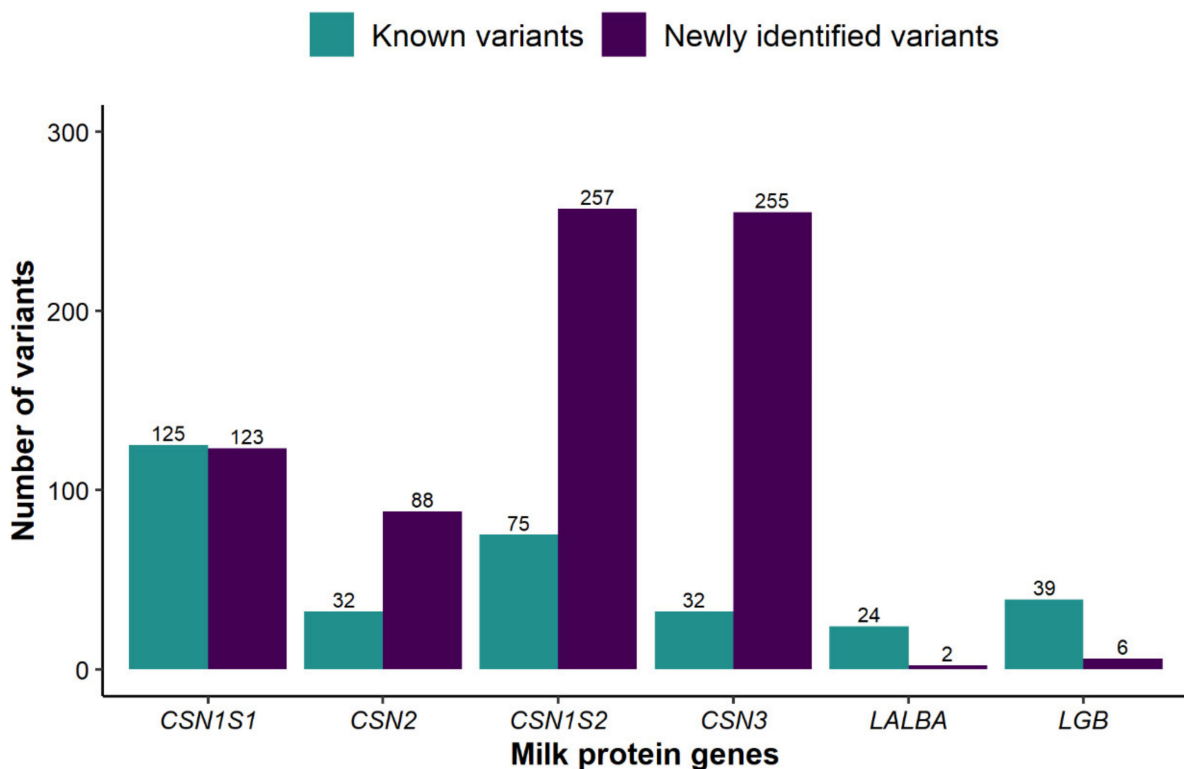
Sequencing of all exon sequences, flanking intron regions and parts of the 5'-upstream regions of the six milk protein genes, and subsequent comparisons with the Ensembl bovine reference sequence (Table 1), displayed a total of 987 SNP and 71 InDel. With regard to

the total number of variants identified in the respective gene, the *CSN1S1*, *CSN1S2* and *LALBA* presented the highest proportions of InDel (Table 2). Only 30.91% (327 variants) of the detected polymorphisms were referenced in the Ensembl variant table, whereas 731 variants have not been reported before (Table S3). The latter variants include 21 SNP occurring at the same position of a known SNP (in the Ensembl variant table), but with new alleles (Table S3). In comparison to the other genes, *CSN3* displayed the highest proportion (88.85%) of newly identified variants (Figure 1). In contrast, the majority of the identified variants in *LALBA* and *LGB* are already referenced in the Ensembl variant table.

**Table 2.** Number and percentage distribution (in brackets) of the sequence variants (SNP, InDel) detected within the exon sequences, flanking intron sequences and parts of the 5'-flanking region of the six milk protein genes (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *LALBA*, *LGB*) in Beninese indigenous cattle breeds.

	<i>CSN1S1</i>	<i>CSN2</i>	<i>CSN1S2</i>	<i>CSN3</i>	<i>LALBA</i>	<i>LGB</i>
	<i>n</i> (%)					
SNP	20 (8.06)	8 (5.83)	26 (7.53)	15 (5.23)	3 (11.54)	1 (2.22)
InDel <sup>1</sup>	228 (91.94)	114 (94.17)	313 (92.47)	275 (94.77)	23 (88.46)	44 (97.78)
<b>Total</b>	<b>248</b>	<b>120</b>	<b>332</b>	<b>287</b>	<b>26</b>	<b>45</b>

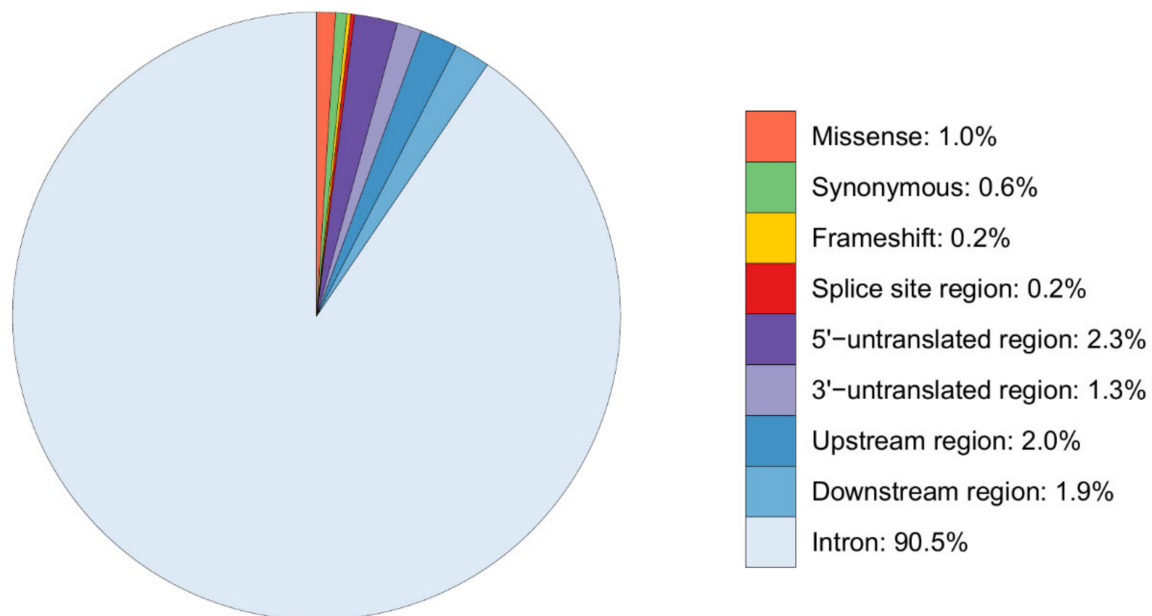
<sup>1</sup> InDel = insertions/deletions.



**Figure 1.** Distribution of known and novel DNA variants within the exon sequences, flanking intron sequences and parts of the 5'-upstream regions of milk protein genes (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *LALBA*, *LGB*) in Beninese indigenous cattle breeds.

The variants were mainly detected in the intron regions (90.5%), 5'-UTR (2.3%), 5'-upstream (2%) and 3'-downstream (1.9%) regions (Figure 2). Only 2% of the detected variants were positioned in coding regions (missense, synonymous and frameshift variants). The *CSN3* and *LALBA* genes displayed the majority of the missense variants (three SNP for each of the genes). A frameshift deletion and insertion were identified in *LALBA*

and *LGB*, respectively. In addition, a total of six synonymous SNP were detected in *CSN2*, *CSN3*, *LALBA* and *LGB* (Table S4).



**Figure 2.** Overall percentage of the DNA variant types detected within the exon sequences (missense, synonymous, frameshift), flanking intron sequences and parts of the 5'-upstream regions of the six milk protein genes (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *LALBA*, *LGB*) in Beninese indigenous cattle breeds. # The respective distributions of variant types for each gene (see above Figure 1) are presented in Table S4.

### 3.2. Causal Genetic Polymorphisms of Milk Protein Variants

A total of 11 missense SNP and one frameshift insertion have been identified in the milk protein genes in the investigated population. Table 3 presents the identified sequence differences, their locations within the respective gene, their effects on the protein sequence, the protein variant designations and frequencies within the different breeds and across all Beninese cattle. More detailed descriptions of the sequence differences found in the respective genes are given in Section 3.2.1 and in the following.

**Table 3.** Description (ID, location and allele frequencies) of the variants causing protein polymorphisms of the six milk protein genes (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *LALBA*, *LGB*), their effects on the protein sequence and the designations of consequent protein variants, and frequencies within the different Beninese local breeds and across all investigated cattle.

Genes	SNP			Gene Transcript				Variant Frequency			
	ID <sup>1</sup>	BTA:bp <sup>2</sup>	Allele	Exon	Protein <sup>3</sup>	Amino Acid <sup>4</sup>	Variant <sup>5</sup>	Crossbreed (n = 27)	Lagune (n = 20)	Somba (n = 20)	Total (n = 20)
<i>CSN1S1</i>	rs43703010	6:85427427	A	17	192 (207)	Glu	B	0.62	0.97	0.93	0.81
			G			Gly	C	0.38	0.03	0.08	0.19
<i>CSN2</i>	rs715383373	6:85450908	T	7	197 (247)	Val	A1	0.97	1.00	1.00	0.99
	rs43703011	6:85451298	C	7	67 (117)	Ala	L	0.03	0.00	0.00	0.01
			A			His	A1	0.39	0.80	0.63	0.57
C	Pro	A2	0.61	0.20	0.37	0.43					
<i>CSN1S2</i>	rs441966828	6:85533780	C	3	8 (23)	Ser	A	0.98	1.00	1.00	0.99
			T			Phe	B	0.02	0.00	0.00	0.01
<i>CSN3</i>	ss7626433428	6:85656526	C	4	66 (87)	Ala	B	1.00	0.75	1.00	0.97
			T			Val	K	0.00	0.25	0.00	0.03
	rs43703015	6:85656736	T	4	136 (157)	Ile	B	0.50	0.75	0.80	0.72
			C			Thr	A	0.50	0.25	0.20	0.28
			C			Ala	B	0.50	0.50	0.80	0.69
			A			Asp	A	0.50	0.50	0.20	0.31

Table 3. Cont.

Genes	SNP			Gene Transcript				Variant Frequency			
	ID <sup>1</sup>	BTA:bp <sup>2</sup>	Allele	Exon	Protein <sup>3</sup>	Amino Acid <sup>4</sup>	Variant <sup>5</sup>	Crossbreed (n = 27)	Lagune (n = 20)	Somba (n = 20)	Total (n = 20)
LALBA	rs714688595	5:31184282	C T	1	10 (58)	Arg Trp	B F	0.96 0.04	1.00 0.00	0.88 0.13	0.94 0.06
	rs722550244	5:31184283	G A	1	10 (58)	Arg Gln	B A	0.88 0.12	1.00 0.00	1.00 0.00	0.95 0.05
	rs465119286	5:31184696	A G	2	41(89)	Ile Val	B E	0.88 0.12	1.00 0.00	0.95 0.05	0.93 0.07
LGB	rs109625649	11:103259232	C T	4	118 (134)	Ala Val	B B1	0.89 0.11	0.50 0.50	0.81 0.19	0.80 0.20
	ss7626433430	11:103257980	- A	3	Thr92 Thr92Asnfs*13	- K	B K	0.96 0.04	1.00 0.00	1.00 0.00	0.98 0.02

<sup>1</sup> Variant ID: the rs# accession number of known variants were retrieved from the Ensembl database; the ss# accession number of the new variants identified from our sequencing were assigned by the European Variation Archive; <sup>2</sup> bp = SNP position according to the ARS-UCD1.2 genome map; <sup>3</sup> Amino acid (AA) position in mature protein. The corresponding AA positions in the protein sequence including signal peptide are given in parentheses; <sup>4</sup> Ala: Alanine, Arg: arginine, Asp: Aspartic acid, Gln: glutamine, Glu: glutamic acid, Gly: Glycine, His: histidine, Ile: isoleucine, Phe: Phenylalanine, Pro: proline, Ser: serine, Thr: Threonine, Trp: Tryptophan, Val: Valine; <sup>5</sup> Milk protein variants were inferred following previous nomenclature [9,14,15,52]. The underlined variants have been newly annotated.

### 3.2.1. CSN1S1

The sequencing of the 19 exons of the *CSN1S1* gene identified the already known casein variants *CSN1S1<sup>B</sup>* and *CSN1S1<sup>C</sup>*. The latter protein variant is due to the SNP rs43703010C>T located in exon 17, inducing the substitution of the glutamic AA by glycine (p.Glu192Gly) in the mature protein. In our animal material, 81% of all animals, especially the taurine, had the reference allele rs43703010A, leading to *CSN1S1<sup>B</sup>*. In contrast, the alternative allele rs43703010G, causing *CSN1S1<sup>C</sup>*, was mainly present in the crossbreed animals with an intra-breed frequency of 38% (Table 3).

### 3.2.2. CSN2

We detected the casein variants *CSN2<sup>A1</sup>*, *CSN2<sup>A2</sup>* and *CSN2<sup>L</sup>*. *CSN2<sup>A2</sup>* is caused by rs43703011A>C in exon 7 (SNP), inducing the replacement of histidine by proline at position 67 of the mature protein (p.His67Pro, Table 3). The allele rs43703011C was dominant in the crossbreeds (intra-breed frequency of 61%). In contrast, the reference allele rs43703011A, leading to the casein variant *CSN2<sup>A1</sup>*, was prevalent in the taurine, especially for Lagune (intra-breed frequency equal to 80%). The casein variant *CSN2<sup>L</sup>* is characterized by a replacement of valine by alanine (p.Val197Ala), due to the SNP rs715383373T>C in exon 7. The allele rs715383373C was identified in one crossbreed animal. Additionally, we identified the SNP rs468218273C>T affecting the AA 42 (p.Ala42) of the immature protein, located within the signal peptide.

### 3.2.3. CSN1S2

Only the SNP rs441966828C>T (exon 3) differing the protein variants *CSN1S2<sup>A</sup>* and *CSN1S2<sup>B</sup>* was found in the investigated Beninese cattle population. Indeed, p.Ser8Phe is the resulting AA exchange differentiating *CSN1S2<sup>B</sup>* from *CSN1S2<sup>A</sup>*. The reference SNP allele rs441966828C encoding for *CSN1S2<sup>A</sup>* was observed in the majority (99%) of the animals, while the allele rs441966828T producing the protein variant *CSN1S2<sup>B</sup>* was detected in one crossbreed animal (Table 3).

### 3.2.4. CSN3

The analysis of the five exons of *CSN3* detected two known (*CSN3<sup>A</sup>*, *CSN3<sup>B</sup>*) and one new casein variant (*CSN3<sup>K</sup>*). The casein variants *CSN3<sup>A</sup>* and *CSN3<sup>B</sup>* are caused by the SNP rs43703015T>C and rs43703016C>A in exon 4, inducing the AA replacements p.Ile136Thr and p.Ala148Asp, respectively. Most of the animals carried the reference SNP alleles rs43703015T and rs43703016C, coding for *CSN3<sup>B</sup>* (Table 3). In contrast, the respective alleles rs43703015C and rs43703016A, coding for *CSN3<sup>A</sup>*, were mainly observed in heterozygous genotypes in the crossbreed animals (50%). Additionally, we detected the synonymous

SNP rs110014544G>A (p.Ala168) also differentiating the casein variants *CSN3<sup>A</sup>* and *CSN3<sup>B</sup>*. The novel kappa-casein variant (*CSN3<sup>K</sup>*) detected in this study is due to a previously undescribed SNP (BTA6:85656526C>T) in exon 4 leading to an AA exchange from alanine to valine (p.Ala66Val). The new variant *CSN3<sup>K</sup>* was observed in one crossbreed animal.

### 3.2.5. LALBA

Across the four exons of *LALBA*, three missense SNP causing four milk protein variants including three known (*LALBA<sup>A</sup>*, *LALBA<sup>B</sup>*, *LALBA<sup>E</sup>*) and one previously undescribed (*LALBA<sup>F</sup>*), were identified. The SNP rs722550244G>C in exon 1 is responsible for the change of arginine to glutamine (p.Arg10Gln), differentiating *LALBA<sup>B</sup>* from *LALBA<sup>A</sup>*. The reference SNP allele rs722550244G causing *LALBA<sup>B</sup>* was present in most of the animals (95%), whereas the alternative SNP allele rs722550244A leading to *LALBA<sup>A</sup>* was only observed in crossbreed animals (Table 3). The protein variant *LALBA<sup>E</sup>* is caused by an A to G nucleotide transition in exon 2 (SNP rs465119286A>G). The SNP allele rs465119286G found in 7% of the animals (including Somba and crossbreed) causes the replacement of isoleucine to valine (p.Ile41Val). In addition to the SNP rs722550244G>C that distinguishes *LALBA<sup>B</sup>* from *LALBA<sup>A</sup>*, the SNP rs7144688595C>T also affects the AA 10 of the mature protein  $\alpha$ -LA. Here, the SNP leads to the AA exchange p.Arg10Trp. Therefore, we have introduced a new naming *LALBA<sup>F</sup>* for this protein variant. The SNP allele rs7144688595T was observed in 6% of the animals. Moreover, a synonymous SNP rs477959124C>T (p.Asp64) was identified in exon 2.

### 3.2.6. LGB

Concerning the *LGB* gene, we identified three milk protein variants (*LGB<sup>B</sup>*, *LGB<sup>B1</sup>*, *LGB<sup>K</sup>*) but only *LGB<sup>B</sup>*, being the reference allele (Table 1) was previously described. Indeed, only one causal genetic polymorphism, i.e., the missense SNP rs109625649C>T differing the protein variants *LGB<sup>B</sup>* and *LGB<sup>A</sup>*, was detected in exon 4 (Table 3). The SNP allele rs109625649T observed in 20% of the animals induces an exchange of alanine to valine (p.Ala118Val). With the absence of the second missense SNP characterizing *LGB<sup>A</sup>*, we considered the observed sequence as an intermediate and suggested a new variant name *LGB<sup>B1</sup>*. The second novel protein variant *LGB<sup>K</sup>* is caused by a frameshift insertion of an adenine nucleotide at position BTA11:103257980 in exon 3. This new polymorphism was detected in one animal and induced the emergence of a premature stop codon, reducing the coding sequence of the *LGB* gene from a total of 178 to 104 AA with a complete exchange of the protein sequence from AA 92 to 104 (p.Thr92Asnfs\*13, see Table S5). Additionally, three synonymous SNP were mapped in the *LGB*, including rs109116595T>C (p.Ile2) in exon 1, rs110641366T>C (p.Asn88) and rs715512468C>T (p.Thr97) in exon 4.

### 3.3. Casein Haplotypes

A total of 24 casein haplotypes (*CSN1S1-CSN2-CSN1S2-CSN3*) were built from the identified casein variants (with individual allele frequency  $\geq 5\%$ ). However, only five haplotypes display a frequency of 5% in at least one breed (Table 4). The haplotype B-A1-A-B was the most frequent, especially in Somba (47%) and Lagune (50%). In contrast, the haplotypes B-A1-A-A and C-A2-A-B were predominant in the crossbreeds. Moreover, the B-A2-A-B haplotype was observed in all breeds with comparable frequency.



**Table 4.** Frequencies of the bovine casein haplotypes (*CSN1S1-CSN2-CSN1S2-CSN3*) in the different Beninese breeds and across all investigated cattle.

Haplotypes <sup>1</sup>	Crossbreed (n = 27)	Lagune (n = 20)	Somba (n = 20)	Total (n = 20)
<b>B-A1-A-B</b>	0.01	0.50	0.47	0.39
<b>B-A2-A-B</b>	0.20	0.18	0.25	0.22
<b>B-A1-A-A</b>	0.39	0.29	0.20	0.20
<b>C-A2-A-A</b>	<0.01	<0.01	0.02	0.19
<b>C-A2-A-B</b>	0.38	<0.01	0.02	<0.01

<sup>1</sup> Only haplotypes with a minimum frequency of 5% in one breed are considered.

### 3.4. Functional Effects of Variants in the Splice Sites, 5'- and 3'-Untranslated Regions of Milk Protein Genes

In silico evaluations of the polymorphisms in the non-coding exonic regions of the six milk protein genes displayed various potential effects. The SNP rs452830840C>T and rs208412793G>T are positioned in the splice site regions of *CSN3* and *LGB*, respectively. However, they presented no impact on the binding score of the splice enzymes using NNSPLICE 0.9. Furthermore, no further identified SNP or InDel affected the 5'-donor or 3'-acceptor splice sites of the exons of the six genes.

The *CSN1S1* and *LALBA* genes presented several SNP in the 5'- and 3'-UTR and many of them influenced TFBS or miRNA target sites (Table 5). In contrast, we mapped only one SNP (rs440770944C>A) in the 3'-UTR of *CSN2*. This SNP altered the binding site of the bta-miR-2464-3p miRNA, but favored the p63 TFBS. Overall, 11 and 15 SNP located in the 5'-UTR of the six milk protein genes affected TFBS or miRNA target sites, respectively. Six SNP were responsible for the loss of TFBS, whereas seven SNP promoted the gain of TFBS. They included one 5'-UTR SNP (rs109900747T>C) causing the loss and the gain of TFBS simultaneously. Five SNP abrogated miRNA target sites in the 5'-UTR, while five other SNP introduced new miRNA target sites. In addition, simultaneous loss and gain of miRNA target sites were observed in five SNP located in the 5'-UTR. Within the 3'-UTR of the milk protein genes, five and 10 SNP altered TFBS and miRNA target sites, respectively. In contrast, two and five SNP introduced new TFBS and miRNA target sites, respectively.

**Table 5.** Potential effects of SNP positioned within the 5'- and 3'-untranslated regions (UTR) of bovine milk protein genes on the binding of micro-RNA (miRNA) and transcription factors (TF) in Beninese cattle breeds.

Gene	Location (BTA:bp <sup>1</sup> )	Allele (ref/alt) <sup>2</sup>	ID <sup>3</sup>	Position	miRNA		TF	
					Loss <sup>4</sup>	Gain <sup>4</sup>	Loss <sup>4</sup>	Gain <sup>4</sup>
<i>CSN1S1</i>	6:85411136	C/T	rs379920406	5'-UTR	-	bta-miR-2291	-	-
	6:85411147	G/A	rs517257790	5'-UTR	-	-	-	-
	6:85411192	C/T	rs520777414	5'-UTR	-	-	-	-
	6:85411221	A/G	rs133040184	5'-UTR	bta-miR-133c	bta-miR-9-5p	-	-
	6:85411236	A/G	rs137119956	5'-UTR	bta-miR-2446	-	E2F-6	-
	6:85411427	A/C	rs135615686	5'-UTR	-	-	RAR	-
	6:85411578	A/G	rs109817504	5'-UTR	-	-	ATF-3; B-ATF; c-Jun; SMARCC1; SMARCC2	-
	6:85411618	C/G	rs109757609	5'-UTR	bta-miR-2284e	bta-miR-7864	-	Sp1; IRF-4
	6:85411677	G/A	rs134534951	5' UTR	-	-	STAT5A	-
	6:85411780	T/C	rs110163054	5'-UTR	bta-miR-15a; bta-miR-562	-	-	-
	6:85413195	C/T	rs110899610	5'-UTR	bta-miR-2325b; bta-miR-2420; bta-miR-453	bta-miR-376a	-	-
	6:85428962	G/A	rs466704456	3'-UTR	bta-miR-874	bta-miR-219-5p; bta-miR-2355-3p	-	-
	6:85429024	C/T	ss7626432754	3'-UTR	-	-	-	-
	6:85429077	A/G	rs435231328	3'-UTR	bta-miR-338	-	-	-
	6:85429241	A/G	rs211141048	3'-UTR	bta-miR-380-3p	-	SMAD4	-
6:85429252	C/A	rs716970086	3'-UTR	bta-miR-371	-	SMAD4	-	
<i>CSN2</i>	6:85449252	C/A	rs440770944	3'-UTR	bta-miR-2464-3p	-	-	p63

Table 5. Cont.

Gene	Location (BTA:bp <sup>1</sup> )	Allele (ref/alt) <sup>2</sup>	ID <sup>3</sup>	Position	miRNA		TF	
					Loss <sup>4</sup>	Gain <sup>4</sup>	Loss <sup>4</sup>	Gain <sup>4</sup>
CSN1S2	6:85531897	T/C	rs109900747	5'-UTR	-	bta-miR-452	Sp1; HNF-4alpha; HNF-4gamma; FOXO1	c-Ets-1
	6:85548443	A/G	rs211156498	3'-UTR	bta-miR-2318; bta-miR-2479; bta-miR-2480	bta-miR-2463; bta-miR-452	CP2-L1	-
	6:85548520	A/G	rs109274107	3'-UTR	-	bta-miR-1388-3p	-	-
CSN3	6:85656841	T/A	rs109787476	3'-UTR	bta-miR-496	bta-miR-2284w	ZNF274	-
	6:85658779	C/A	ss7626433382	3'-UTR	bta-miR-193a-3p	-	-	-
	6:85658873	G/A	rs134516686	3'-UTR	-	-	-	-
LALBA	5:31183709	G/C	rs444727593	5'-UTR	-	bta-miR-2475	-	-
	5:31183736	G/A	rs458466372	5'-UTR	bta-miR-2284r; bta-miR-2284s	-	-	-
	5:31183766	T/C	rs471361585	5'-UTR	-	bta-miR-2370-5p	-	-
	5:31183789	A/G	rs440012037	5'-UTR	bta-miR-7864	-	-	-
	5:31183806	C/T	rs717249686	5'-UTR	-	-	-	KLF4
	5:31183848	C/A	rs460157851	5'-UTR	-	bta-miR-2472; bta-miR-6517	-	AP-2alpha; IRF-9; STAT2;STAT1
	5:31183921	A/G	rs448925171	5'-UTR	bta-miR-2325c; bta-miR-6120-3p	bta-miR-218; bta-miR-2452	-	-
	5:31183924	T/C	rs719631407	5'-UTR	bta-miR-6120-3p	bta-miR-218	-	-
	5:31184082	T/C	rs462561324	5'-UTR	-	-	-	-
	5:31185981	C/T	rs718675014	3'-UTR	-	bta-miR-2425-5p; bta-miR-2467-3p	N-Myc; N-Myc	c-Myc
5:31186027	A/G	rs110359174	3'-UTR	bta-miR-143	-	-	-	
LGB	11:103255847	G/A	rs516356159	5'-UTR	-	-	-	RCO2
	11:103255918	C/T	rs41255682	5'-UTR	bta-miR-1777a; bta-miR-2454-5p	-	SIX5	-
	11:103255930	G/C	rs41255683	5'-UTR	-	-	-	STAT1; STAT3
	11:103260824	G/A	rs1116405113	3'-UTR	bta-miR-7860	-	-	-

<sup>1</sup> bp = SNP position; <sup>2</sup> ref/alt = reference/alternate allele, <sup>3</sup> Variant ID: the rs# accession number of known variants were retrieved from the Ensembl database; the ss# accession number of the new variants identified from our sequencing were assigned by the European Variation Archive; <sup>4</sup> loss of binding, gain of binding.

#### 4. Discussion

In this study, we sequenced the exons, flanking intron regions, and parts of the 5'-flanking regions of the six milk protein genes from 67 animals, and identified a total of 1058 DNA polymorphisms. This value is higher than those reported in previous studies targeting sequence data of cattle milk protein genes [9,60]. For instance, Meier et al. [60] considered the 1000 Bull Genomes with 1821 animals from 14 different breeds and reported only 892 SNPs within the four casein genes (exons and flanking intron regions) and their 1000 bp upstream regions. In contrast, the four casein genes display 987 variants in our study. This difference implies the high level of polymorphisms in the milk protein gene segments of the Beninese breeds compared to other cattle breeds. In addition, the large number of novel variants is typical in the context of the pronounced genetic diversity of the Beninese cattle, and concurs with the need for deeper investigation of African livestock genomes [45,61].

Similar to Meier et al. [60], we observed the majority of the polymorphisms in the intron regions. However, several variants are also located in the coding regions as well as in the 5'- and 3'-UTR. We only focus on the polymorphisms in 5'- and 3'-UTR for the screening of TFBS or target sites for miRNA because of the large number of detected SNP. The identified effects of the 5'- and 3'-UTR SNP on TFBS and miRNA sites (see Table 5) are consistent with previous findings and suggest further investigations on the SNP located in the 5'-upstream and 3'-downstream region of the milk protein genes [9,38]. The analyses of the SNP impacts on the splice site recognition of the milk protein genes using NNSPLICE revealed no effect. However, supplementary tools (e.g., ESEfinder) can be applied to investigate exonic splicing enhancers in the milk protein gene [62]. Moreover, further research is needed to clarify the real effects of these polymorphisms in the TFBS, miRNA binding sites and splice site regions, for example using PCR-based expression studies

(like real-time quantitative PCR or droplet digital PCR), gene expression array or RNA sequencing (RNA-Seq) methods [63–66].

Within the coding regions, a total of 12 DNA polymorphisms caused milk protein variants. They include two previously unknown DNA variants, i.e., one SNP in *CSN3* and one frameshift insertion in *LGB*. The total number of missense variants is appreciable, but less than those (a total of 20 missense SNP in casein genes) reported in the 1000 Bull Genomes consisting of a large variety of cattle breeds [60]. Nevertheless, the higher number of missense variants identified in *CSN2*, *CSN3* and *LALBA* in the present study are correlated with the high polymorphisms of these milk protein genes compared to the others [14,60]. Several other segregating variants of the bovine milk protein genes were also not identified [14,15]. Overall, 17 milk protein variants were characterized and included two for *CSN1S1* (B, C), three for *CSN2* (A1, A2, L), two for *CSN1S2* (A, B), three for *CSN3* (A, B, K), four for *LALBA* (A, B, E, F), and three for *LGB* (B, B1, K). *CSN3<sup>K</sup>*, *LALBA<sup>F</sup>*, *LGB<sup>B1</sup>* and *LGB<sup>K</sup>* are undescribed so far.

#### 4.1. Sequence Polymorphisms and Milk Proteins in Beninese Cattle

##### 4.1.1. *CSN1S1*

The detection of *CSN1S1<sup>B</sup>* and *CSN1S1<sup>C</sup>* substantiates the previous identification of  $\alpha_{s1}$ -CN<sup>B</sup> and  $\alpha_{s1}$ -CN<sup>C</sup> at protein level in Beninese cattle [18]. The lower distribution of the protein variant *CSN1S1<sup>C</sup>* in the taurine breeds (in contrast to the crossbreeds) is consistent with results in other African cattle breeds [9,17,19]. The association of the *CSN1S1<sup>C</sup>* variant with impaired immunoreactivity in humans (in contrast to *CSN1S1<sup>B</sup>*) suggests its conservation through the valorization of milk from zebu and crossbreed animals for human health [67]. We are not able to elucidate the rare allele  $\alpha_{s1}$ -CN<sup>X</sup> identified by Moazami-Goudarzi et al. [18] by protein electrophoresis in the Borgou cattle. Therefore, analyses on protein level using milk samples is necessary to infer the protein allele  $\alpha_{s1}$ -CN<sup>X</sup> and to clarify its DNA sequence difference. The isoelectric focusing method described by Giambra et al. [68] with adjacent mRNA or DNA sequencing can be a useful approach in this regard.

Seven SNP within the 5'-UTR of *CSN1S1* represent the causal polymorphisms of the *CSN1S1* promoter alleles 1, 2, 3 and 5, as previously described by Prinzenberg et al. [43] and Ibeagha-Awemu et al. [42]. However, only the *CSN1S1Prom<sup>1</sup>* and *CSN1S1Prom<sup>2</sup>* were fully identified in our samples. The *CSN1S1Prom<sup>3</sup>* and *CSN1S1Prom<sup>5</sup>* diverge from our sequences with the absence of some characteristic polymorphisms.

Moreover, we observed various effects of the SNP located in the 5'- and 3'-UTR of the *CSN1S1* gene on TFBS and miRNA targets. Similar impacts of SNP in the *CSN1S1* flanking regions on TFBS and milk protein content have been previously reported in cattle [42,43] and in small ruminants [69].

##### 4.1.2. *CSN2*

The distributions of *CSN2<sup>A1</sup>* and *CSN2<sup>A2</sup>* in the Somba and crossbreed populations are in line with previous findings [18,19]. Some differences in the allele frequencies of the two alleles may be due to a greater local limitation in the samples of Moazami et al. [18]. In contrast to our study, the authors only studied Lagune and Borgou animals from the regions of Porto Novo and Borgou, respectively. Due to a wider distribution of the animals sampled in our study in Benin and an associated low relationship between the animals, it is possible that we were able to present the more current allele frequency for these casein variants in Beninese cattle. However, we did not detect the variant *CSN2<sup>I</sup>* reported by Ceriotti et al. [19] in the Borgou. Interestingly, we found *CSN2<sup>L</sup>* in a crossbreed animal. The protein variant has recently been described at low frequency (0.02) in Gyr cattle by Gallinat et al. [15]. Said Ahmed et al. [9] similarly observed this protein variant in 10% of their genotyped Butana cattle. However, recent analysis of the 1000 Bull Genomes considering 14 different taurine breeds did not detect *CSN2<sup>L</sup>*. In consequence, our findings reinforce the specificity of *CSN2<sup>L</sup>* to indicine cattle. Further rare variants *CSN2<sup>B</sup>* and *CSN2<sup>D</sup>* previously

reported in other African taurine (Baoulé, N'Dama) and indicine breeds (Sudanese Fulani, zebu Shuwas Arab) [17] were not identified in our study.

In contrast to the other investigated milk protein genes, we detected no SNP in the 5'-UTR and only one SNP in the 3'-UTR of *CSN2*. The sequence analyses suggested that the latter SNP (rs440770944C>A) induces the binding of the p63 TF. However, no influence of the p63 TF on milk protein gene transcription has been previously reported. In addition, the impact of SNP positioned in non-coding regions of the *CSN2* gene has been rarely reported in cattle. Only one SNP located in the 5'-upstream region (g.1311T>C) of the caprine *CSN2* was associated with a lack of  $\beta$ -CN in milk [70].

#### 4.1.3. *CSN1S2*

In contrast to the other milk protein genes, no previous study reported the distribution of the protein variants of *CSN1S2* in the Beninese cattle breeds. The gene has a low polymorphic level in cattle, and consequently, it was only marginally addressed [71]. The detection of the *CSN1S2*<sup>A</sup> in 99% of the investigated animals confirm its prevalence in cattle [14]. In addition, the protein variant *CSN1S2*<sup>B</sup> is detected in only one crossbred animal, in agreement with its specificity to indicine cattle [71]. Ibeagha-Awemu et al. [71] observed this variant in European cattle breeds influenced by zebu introgression. In addition, *CSN1S2*<sup>D</sup> and *CSN1S2*<sup>E</sup> reported in low frequencies in some European and Iranian cattle breeds, respectively, were not found in this study [14,15,71].

Only one SNP (rs109900747T>C) was observed in the 5'-UTR of *CSN1S2*, whereas Said Ahmed et al. [9] reported two other SNP in this region. The alternative allele rs109900747C alters several TFBS, but promotes the c-Ets-1 TFBS and the target site of the bta-miR-452 miRNA. Similar results were reported by Kishore et al. [39] in the 5'-flanking region of *CSN1S2* in Indian zebu cattle. Furthermore, two SNP in the 3'-UTR induce the deletion and/or the substitution of several miRNA binding sites. Some of these miRNA were associated with milk production traits including milk fat metabolism [72,73]. Therefore, the impact of the alteration/substitution of miRNA target sites from SNP in 3'-UTR of *CSN1S2* would be interesting for investigations in an experimental study.

#### 4.1.4. *CSN3*

The identification of *CSN3*<sup>A</sup> and *CSN3*<sup>B</sup> in Beninese cattle breeds is in accordance with previous results [18,19]. In addition, the high frequencies of the *CSN3*<sup>B</sup> causative SNP supports the known prevalence of this variant in African taurine [14,17]. Further variants *CSN3*<sup>A1</sup>, *CSN3*<sup>H</sup> and *CSN3*<sup>I</sup> as previously reported in African indicine and taurine breeds, respectively, were not detected in our samples. However, the SNP (BTA6:85656526C>T) causing p.Ala66Val is interesting, as it has never been reported before. The *CSN3* is highly polymorphic with several variants (e.g., A<sup>1</sup>, E, F<sup>2</sup>, H, I) due to only one SNP, respectively [14]. We name this new variant *CSN3*<sup>K</sup> following the previous nomenclature in cattle [13,14].

We identified a SNP (BTA6:85658779C>A) in the 3'-UTR of *CSN3*, which has not been described before. This SNP abrogates the target site of the bta-miR-193a-3p. Additionally, the 3'-UTR of *CSN3* displays two other SNP including the rs109787476T>A, which affect miRNA target sites. Said Ahmed et al. [9] similarly reported several SNP affecting miRNA binding in this region in the Butana cattle, but none of these SNP was identified in our study. Moreover, a SNP in the non-coding region of the *CSN3* gene was associated with milk production traits [14,74]. In consequence, the detected polymorphisms in the 3'-UTR of *CSN3* are of interest as they may affect the expression of the gene, and consequently, they influence the technological quality of milk in Beninese cattle. Indeed, the importance of the *CSN3* gene in cheesemaking properties of milk is well established [14].

#### 4.1.5. *LALBA*

In *LALBA*, we observed the predominance of the variant *LALBA*<sup>B</sup>, while *LALBA*<sup>A</sup> was only found in some crossbred animals. These findings are in agreement with reports

addressing Beninese cattle as well as other African cattle breeds [9,14,17,18]. No taurine animal presents the *LALBA*<sup>A</sup> variant in our samples. This variant is more prevalent in indicine cattle, but has been previously reported in low frequency in African taurine [17,18]. The identification of the rare variant *LALBA*<sup>E</sup> in the Beninese cattle supports its recent description in the African zebu Butana as well as in some European and Asian breeds [9,52]. Tetens et al. [52] associated the presence of this variant in the Hinterwalder taurine breed with prior indicine introgression. The presence of this variant in some Somba cattle originated from zebu, considering the existing indicine introgression in the Somba cattle breed [45]. Nevertheless, the evaluation of the *LALBA*<sup>E</sup> distribution in other breeds is recommended. Additionally, we identified the missense SNP rs714688595C>T in Somba and crossbreed cattle. This polymorphism is known and annotated in the Ensembl variant table for the bovine *LALBA* gene. However, we found no study describing the p.Arg10Trp substitution in  $\alpha$ -LA caused by this SNP. Indeed,  $\alpha$ -LA<sup>A</sup> diverges from  $\alpha$ -LA<sup>B</sup> by the AA replacement p.Arg10Gln. Further known AA exchanges in  $\alpha$ -LA include the variants C (p.Gln>Glu, unknown position), D (p.Gln65His) and E (p.Ile41Val) [14,75,76], whereas we could not identify the causing SNP on DNA-level in our animal material. Therefore, we believe that SNP rs714688595C>T leading to p.Arg10Trp characterizes a new protein variant *LALBA*<sup>F</sup>.

The detection of the SNP rs718675014C>T and rs110359174A>G in the 3'-UTR corroborates previous findings in the Butana cattle [9]. In addition, the functional effects of these polymorphisms in the bta-miR-143 and bta-miR-2467-3p, respectively, concur with the findings of Said Ahmed et al. [9]. However, we retrieved less miRNA binding sites or targets/seeds affected by the SNP. Several SNP were observed in the 5'-UTR of the *LALBA* gene, but they were not addressed in previous studies (Table S3). Many of these SNP are noteworthy as they affect TFBS or miRNA-binding sites, with potential influences on the differential expression of *LALBA* in the Beninese cattle.

#### 4.1.6. LGB

The high frequency of *LGB*<sup>B</sup> (corresponding to the reference sequence, see Table 1) supports previous findings in Beninese as well as in other African cattle breeds [17,18]. In contrast to Moazami-Goudarzi et al. [18] and Ceriotti et al. [19], we cannot confirm the presence of *LGB*<sup>A</sup>. This variant is due to two missense SNP rs110066229G>A (p.Gly64Asp) and rs109625649C>T (p.Ala118Val), and the two synonymous SNP rs110180463C>T (p.Asn63) and rs110641366T>C (p.Asn88) [40,77]. As we could only identify the SNP rs109625649C>T and rs110641366T>C in our animal material, we propose that this is an intermediate *LGB* form, not described until now. Moazami-Goudarzi et al. [18] used isoelectric focusing to segregate  $\beta$ -LG<sup>A</sup> from  $\beta$ -LG<sup>B</sup>, whereas it is not mentioned if the differences in the electrophoretic separation is caused by changes in the isoelectric point due to both amino acid exchanges p.Gly64Asp and p.Ala118Val, or only due to one of the amino acid exchanges. In the study by Ceriotti et al. [19], the typing of the protein variant *LGB*<sup>A</sup> was based on the visualization of the SNP rs109625649C>T using the PCR-restriction fragment length polymorphism (RFLP) method. The second SNP rs110066229G>A was not analyzed using this method [78]. Therefore, it is possible that the animal material examined by Ceriotti et al. [19] did not contain the actual variant *LGB*<sup>A</sup> but, as in our case, the allele *LGB*<sup>B1</sup>. Ganai et al. [40] observed a complete linkage disequilibrium (LD) between the causal SNP of *LGB*<sup>A</sup> and *LGB*<sup>B</sup> in Dutch Holstein Friesian cows. Therefore, the omission of the SNP rs110066229G>A in the Beninese cattle breeds is unprecedented and may suggest the breakage of LD between the two major causal SNP of *LGB*<sup>A</sup> and *LGB*<sup>B</sup> in the Beninese cattle. Considering the existing information, we annotate the intermediate form of the *LGB* gene sequence caused by the missense SNP rs109625649C>T (p.Ala118Val) and the synonymous SNP rs110641366T>C (p.Asn88) as *LGB*<sup>B1</sup>.

Another new protein variant detected in this study is the  $\beta$ -LG<sup>K</sup>. This variant is due to an insertion at BTA11:103257980 in exon 3 altering the reading frame, and inducing a premature stop codon (p.Thr92Asnfs\*13). The mutated protein is shortened by 74 AA

(46%) and probably lacks domains that are essential for normal protein function. Although peptide deletions or insertions characterize some milk protein variants in ruminants, the loss of a large peptide sequence as observed in this study is uncommon [14,79,80]. Firstly, to our knowledge, no peptide skipping has been previously reported in the ruminant  $\beta$ -LG protein. Secondly, the largest peptide deletion identified in cattle associated the  $\alpha_{s1}$ -CN<sup>A</sup> with the loss of 13 AA caused by a single T>A SNP [81]. Further deletions of peptide sequence in cattle were associated with the  $\alpha_{s1}$ -CN<sup>H</sup> (peptide 51 to 58) in Kuri (an African crossbreed) and  $\alpha_{s2}$ -CN<sup>D</sup> (peptide 51 to 59) in European breeds due to exon skipping events [17,79]. In goats, the deletions of larger peptide sequences are more frequent with significant effects on milk protein genes. For instance, the caprine  $\alpha_{s1}$ -CN<sup>F</sup> (37 AA deleted) as well as several null alleles ( $\alpha_{s1}$ -CN<sup>O1</sup>,  $\alpha_{s1}$ -CN<sup>O2</sup>,  $\alpha_{s2}$ -CN<sup>O</sup>,  $\beta$ -CN<sup>O</sup>) were associated with low or null expression of casein genes, respectively [79,82]. Therefore, we presume a high impact of the observed peptide skipping in the  $\beta$ -LG protein due to the described insertion in *LGB* exon 3. With regard to the indirect effect of the *LGB* gene on the total casein content [83,84], the variant *LGB*<sup>K</sup> could affect animal milk composition and yield. The low frequency of the variant in the investigated population suggest a relative recent occurrence of the variant or indirect negative selection against animals carrying this variant. Further investigations should evaluate the functional impact of the *LGB*<sup>K</sup> variant. Moreover, this protein variant is interesting given the high allergenic potential of  $\beta$ -LG, a milk protein missing in human milk [14]. For instance, camel milk containing no  $\beta$ -LG is recommended as an alternative milk protein source for cow milk protein allergy patients [85–87]. Similar observations were also made for the bovine *CSN1S1*<sup>G</sup> and several null casein alleles in goats [83,88].

In addition to the missense and frameshift variants, we detected two previously not described synonymous SNP. Further investigations are imperative to clarify the impact of the new synonymous variants in the *LGB* gene. Indeed, Cartegni et al. [89] revealed potential effects of synonymous SNP on splicing. Various synonymous SNP are also linked to specific protein variants (e.g., *CSN2*<sup>A2</sup>, *CSN2*<sup>A3</sup>, *CSN3*<sup>A1</sup>) [14].

The unique polymorphism observed in the 3'-UTR of *LGB* induces the loss of a bta-miR-7860 target site (Table 5). Regarding the polymorphisms in 5'-UTR of the *LGB* gene, two SNP promote the gain of TFBS, whereas the third alters a TFBS and the target sites of several miRNA. The SNP rs41255683G>C is in close proximity (at 34 bp) to SNP g.-215C>A (BTA5:103255964) causing a low expression level of *LGB* in cattle [90]. We were not able to detect the g.-215C>A SNP, but it seems to be rare, as it was similarly not found by Ganai et al. [40], and is also not referenced in the Ensembl variant table. Nevertheless, we mapped the SNP rs41255679C>G in the 5'-upstream region of *LGB*. This polymorphism was reported within the TF activator protein-2 (AP-2), and associated with differential expression of the *LGB*<sup>A</sup> and *LGB*<sup>B</sup> protein variants [40,41]. These findings evidence the necessity to demonstrate the potential effects of SNP positioned in 5'-upstream regions on milk protein genes.

#### 4.2. Casein Haplotypes

The presence of the haplotype B-A1-A-B (corresponding to *CSN1S1*-*CSN2*-*CSN1S2*-*CSN3*) in about 50% of the Lagune and Somba cattle (Table 4), is in agreement with previous findings in the Beninese and African taurine breeds [17,18,42]. Bonfatti et al. [21] related haplotypes including *CSN2*<sup>A1</sup> and *CSN3*<sup>B</sup> with increased  $\alpha_{s1}$ - and  $\kappa$ -CN milk contents, and reduced  $\beta$ - and  $\alpha_{s2}$ -CN concentrations. *CSN2*<sup>A1</sup> is negatively correlated with milk digestibility and human health [11,91]. In addition, haplotypes with the *CSN2*<sup>A1</sup> were associated with low milk and protein contents [20,22]. Therefore, the low milk performance of the Beninese taurine cattle is possibly due to the high frequency of the haplotype B-A1-A-B [7]. Inversely, a high frequency of the haplotype B-A2-A-A was described for Holstein and several European taurine breeds [20,60], which favorably supports the importance of the *CSN2*<sup>A2</sup> in bovine milk production. The latter haplotype was identified in low frequency in African taurine and the Sudanese Butana cattle [9,17,42],

but not in the present study. Nevertheless, the distribution of the haplotype B-A2-A-B in moderate frequencies over all breeds, is interesting as it may be associated with appreciable milk performances and should be included in future Beninese cattle improvement strategies. This haplotype is similarly predominant in the Brown Swiss cattle population [20,60]. Moreover, we estimated a high frequency of the haplotypes B-A1-A-A and C-A2-A-B in the crossbreeds. Remarkably, the distribution of the latter haplotype in crossbreeds is linked to the high occurrence of *CSN1S1<sup>C</sup>* in indicine cattle. However, this haplotype differs with *CSN3<sup>B</sup>* from the most frequent haplotypes in indicine cattle (C-A2-A-A) [9,17]. These results suggest the presence of both indicine and taurine imprints in caseins regions in the crossbreeds. With regard to their association with milk performances, Boettcher et al. [20] related haplotypes comparable to B-A1-A-A and C-A2-A-B (i.e., excluding the *CSN1S2* gene) with higher milk protein and fat contents, and with lower milk yield in Italian Holstein and Brown Swiss Cattle. Considering that the *CSN1S2<sup>A</sup>* protein variant is the most prevalent in European taurine breeds, we assume that the haplotypes described by the latter authors correspond to B-A1-A-A and C-A2-A-B, and in consequence, we expect the same effects [13]. Finally, the casein haplotype B-A1-A-A was rare in African cattle, but dominant in British Friesian and German Black Pied Cattle [17,42,44,60].

## 5. Conclusions

The current study confirmed milk protein variants in Beninese cattle breeds as previously identified in commercial or outbred populations. As a novelty based on the analyses of sequence data, unknown polymorphisms including SNP and InDel, were detected. In this regard, the identified new milk protein variants are located in the *CSN3*, *LALBA* and *LGB* genes. Hence, the Beninese cattle carry favorable alleles as well as haplotypes, which can be used in genetic improvement programs for milk yield and milk quality of various breeds.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/genes12111702/s1>, Table S1: Primer sequences utilized for the sequencing of the six milk protein genes (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *LALBA*, *LGB*) in the Beninese indigenous cattle breeds; Table S2: List of the bovine miRNA and their corresponding seed sequences retrieved from the TargetScan website and used in “targetscan\_60.pl” tool for the detection of miRNA target sites; Table S3: List of the polymorphisms detected within the exon sequences, flanking intron sequences and parts of the 5′-upstream regions of milk protein genes (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *LALBA*, *LGB*) in Beninese indigenous cattle breeds; Table S4: Distributions of variant types (according to their positions and consequence) detected within each of the six milk protein genes (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *LALBA*, *LGB*) and total in Beninese indigenous cattle breeds; Table S5: Functional consequence of the frameshift insertion of nucleotide A at BTA11:103257980 on the translational reading frame and on mature protein.

**Author Contributions:** Conceptualization, methodology and analysis, S.F.U.V. and I.J.G.; writing—original draft preparation, S.F.U.V.; writing—review and editing, I.J.G., T.Y., K.B., L.H.D. and S.K.; supervision, L.H.D. and S.K.; funding acquisition S.F.U.V. and L.H.D.; project administration, K.B. and L.H.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Volkswagen Stiftung, Hannover, Germany (grant Az 94 829). We acknowledge funding of S.F.U.V. through the German Academic Exchange Service (DAAD) scholarship program: Research Grants—Doctoral Programmes in Germany.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study because there is no specific legislation for animal welfare or animal hair sampling in Benin [92]. Nevertheless, the data were collected in accordance with the FAO guidelines for the characterization of animal genetic resources and following the criteria for the assessment of animal welfare as identified by the Welfare Quality Project (WQP) [93,94].

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All the DNA polymorphisms identified from our sequencing are submitted to the European Variation Archive (EVA) and are openly accessible at <https://wwwdev.ebi.ac.uk/eva/?eva-study=PRJEB47999>, accessed on 13 October 2021, [51]. Further data supporting the results of this article are presented in the Supplementary files.

**Acknowledgments:** The authors are grateful to LGC Genomics GmbH, Berlin, Germany for their collaborative support for the sequencing of the animals. We are also thankful to all farmers who give their consent to collect hair samples of their animals.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Alassane, Y.; Ahounou, S.; Toleba, S.; Adjakpa, A.; Dotche, I.; Houaga, I.; Moula, N.; AntoineMoussiaux, N.; Hornick, J.; Youssao, A. Zootechnical performance of Girolando cattle at Kpinnou Breeding Farm, South-West of Benin Republic. *J. Adv. Veter-Anim. Res.* **2018**, *5*, 123. [CrossRef]
2. Alk, I.T.; Yari, H.M.; Gbangboche, A.B.; Lokossou, R. Reproductive Performance and Milk Production of Girolando Cows in the Ranch of Kpinnou, South-West of Benin Republic. *J. Anim. Veter-Adv.* **2011**, *10*, 2588–2592. [CrossRef]
3. Kévin, K.S.; Charles, D.G.-K.; Valentine, Y.-G.; Souleymane, S.; Maurice, K.; Issaka, Y.A.K. Genetic Diversity of Benin Cattle Populations Using Microsatellite Markers. *Int. J. Anim. Sci. Technol.* **2019**, *3*, 7. [CrossRef]
4. Vanvanhossou, S.F.U.; Koura, I.B.; Dossa, L.H. The implications of herd entrustment practice for the sustainable use of cattle genetic resources in the (agro)-pastoral systems of West Africa: A case study from Benin. *Pastoralism* **2021**, *11*, 1–14. [CrossRef]
5. Houessou, S.O.; Dossa, L.H.; Assogba, C.A.; Diogo, R.V.C.; Vanvanhossou, S.F.U.S.; Schlecht, E. The role of cross-border transhumance in influencing resident herders' cattle husbandry practices and use of genetic resources. *Animal* **2020**, *14*, 2378–2386. [CrossRef] [PubMed]
6. Vanvanhossou, S.F.U.; Dossa, L.H.; König, S. Sustainable Management of Animal Genetic Resources to Improve Low-Input Livestock Production: Insights into Local Beninese Cattle Populations. *Sustainability* **2021**, *13*, 9874. [CrossRef]
7. Dossa, L.H.; Vanvanhossou, S.F.U.S. The indigenous Somba cattle of the hilly Atacora region in North-West Benin: Threats and opportunities for its sustainable use. *Trop. Anim. Health Prod.* **2015**, *48*, 349–359. [CrossRef] [PubMed]
8. Hiernaux, P.; Adamou, K.; Zezza, A.; Ayantunde, A.A.; Federighi, G. Milk offtake of cows in smallholder farms of semiarid Sahel: Low yields with high value! *Rev. Elev. Med. Vet. Pays Trop.* **2016**, *69*, 143–153. [CrossRef]
9. Ahmed, A.S.; Rahmatalla, S.; Bortfeldt, R.; Arends, D.; Reissmann, M.; Brockmann, G.A. Milk protein polymorphisms and casein haplotypes in Butana cattle. *J. Appl. Genet.* **2016**, *58*, 261–271. [CrossRef] [PubMed]
10. Lemay, D.; Lynn, D.J.; Martin, W.F.; Neville, M.C.; Casey, T.M.; Rincon, G.; Kriventseva, E.V.; Barris, W.C.; Hinrichs, A.S.; Molenaar, A.J.; et al. The bovine lactation genome: Insights into the evolution of mammalian milk. *Genome Biol.* **2009**, *10*, R43. [CrossRef] [PubMed]
11. Sebastiani, C.; Arcangeli, C.; Ciullo, M.; Torricelli, M.; Cinti, G.; Fisichella, S.; Biagetti, M. Frequencies Evaluation of  $\beta$ -Casein Gene Polymorphisms in Dairy Cows Reared in Central Italy. *Animals* **2020**, *10*, 252. [CrossRef]
12. Ferretti, L.; Leone, P.; Sgaramella, V. Long range restriction analysis of the bovine casein genes. *Nucleic Acids Res.* **1990**, *18*, 6829–6833. [CrossRef]
13. Farrell, H.M.; Jimenez-Flores, R.; Bleck, G.T.; Brown, E.M.; Butler, J.E.; Creamer, L.K.; Hicks, C.L.; Hollar, C.M.; Ng-Kwai-Hang, K.F.; Swaisgood, H.E. Nomenclature of the Proteins of Cows' Milk—Sixth Revision. *J. Dairy Sci.* **2004**, *87*, 1641–1674. [CrossRef]
14. Caroli, A.M.; Chessa, S.; Erhardt, G.J. Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. *J. Dairy Sci.* **2009**, *92*, 5335–5352. [CrossRef] [PubMed]
15. Gallinat, J.L.; Qanbari, S.; Drögemüller, C.; Pimentel, E.C.G.; Thaller, G.; Tetens, J. DNA-based identification of novel bovine casein gene variants. *J. Dairy Sci.* **2013**, *96*, 699–709. [CrossRef] [PubMed]
16. Visker, M.H.P.W.; Heck, J.M.L.; Van Valenberg, H.J.F.; Van Arendonk, J.A.M.; Bovenhuis, H. Short communication: A new bovine milk-protein variant:  $\alpha$ -Lactalbumin variant D. *J. Dairy Sci.* **2012**, *95*, 2165–2169. [CrossRef]
17. Mahé, M.-F.; Miranda, G.; Queval, R.; Bado, A.; Zafindrajaona, P.S.; Grosclaude, F. Genetic polymorphism of milk proteins in African *Bos taurus* and *Bos indicus* populations. Characterization of variants  $\alpha$ s1-Cn H and  $\kappa$ -Cn. *J. Genet. Sel. Evol.* **1999**, *31*, 239–253. [CrossRef]
18. Goudarzi, K.M.; Belemsaga, D.M.; Ceriotti, G.; Laloë, D.; Fagbohoun, F.; Kouagou, N.T.; Sidibé, I.; Codjia, V.; Crimella, M.C.; Grosclaude, F.; et al. Caractérisation de la race bovine Somba à l'aide de marqueurs moléculaires. *Rev. Elev. Med. Vet. Pays Trop.* **2001**, *54*, 129. [CrossRef]
19. Ceriotti, G.; Marletta, D.; Caroli, A.; Erhardt, G. Milk protein loci polymorphism in taurine (*Bos taurus*) and zebu (*Bos indicus*) populations bred in hot climate. *J. Anim. Breed. Genet.* **2004**, *121*, 404–415. [CrossRef]
20. Boettcher, P.J.; Caroli, A.; Stella, A.; Chessa, S.; Budelli, E.; Canavesi, F.; Ghiroldi, S.; Pagnacco, G. Effects of Casein Haplotypes on Milk Production Traits in Italian Holstein and Brown Swiss Cattle. *J. Dairy Sci.* **2004**, *87*, 4311–4317. [CrossRef]



21. Bonfatti, V.; Di Martino, G.; Cecchinato, A.; Vicario, D.; Carnier, P. Effects of  $\beta$ - $\kappa$ -casein (CSN2-CSN3) haplotypes and  $\beta$ -lactoglobulin (BLG) genotypes on milk production traits and detailed protein composition of individual milk of Simmental cows. *J. Dairy Sci.* **2010**, *93*, 3797–3808. [[CrossRef](#)]
22. Nilsen, H.; Olsen, H.G.; Hayes, B.; Sehested, E.; Svendsen, M.; Nome, T.; Meuwissen, T.; Lien, S. Casein haplotypes and their association with milk production traits in Norwegian Red cattle. *Genet. Sel. Evol.* **2009**, *41*, 24. [[CrossRef](#)]
23. Hohmann, L.G.; Weimann, C.; Scheper, C.; Erhardt, G.; König, S. Associations between maternal milk protein genotypes with preweaning calf growth traits in beef cattle. *J. Anim. Sci.* **2020**, *98*, skaa280. [[CrossRef](#)] [[PubMed](#)]
24. Deth, R.; Clarke, A.; Ni, J.; Trivedi, M. Clinical evaluation of glutathione concentrations after consumption of milk containing different subtypes of  $\beta$ -casein: Results from a randomized, cross-over clinical trial. *Nutr. J.* **2015**, *15*, 1–6. [[CrossRef](#)]
25. Ionel, N.R.; Elena, I.D.; Toma, C.L. Effects of Sires Genotyping for  $\kappa$ -Casein on B Allele Frequency in Romanian Brown Cattle. *Res. J. Biotechnol.* **2017**, *12*, 12.
26. Losi, G.; Capella, P.; Castagnetti, G.B.; Grazia, L.; Zambonelli, C.; Mariani, P.; Russo, V. Influenza delle varianti genetiche della caseina k sulla formazione e sulle caratteristiche della cagliata. *Sci. Tecnol. Aliment.* **1973**, *3*, 373.
27. Martin, P.; Szymanowska, M.; Zwierzchowski, L.; Leroux, C. The impact of genetic polymorphisms on the protein composition of ruminant milks. *Reprod. Nutr. Dev.* **2002**, *42*, 433–459. [[CrossRef](#)] [[PubMed](#)]
28. Comin, A.; Cassandro, M.; Chessa, S.; Ojala, M.; Zotto, R.D.; De Marchi, M.; Carnier, P.; Gallo, L.; Pagnacco, G.; Bittante, G. Effects of Composite  $\beta$ - and  $\kappa$ -Casein Genotypes on Milk Coagulation, Quality, and Yield Traits in Italian Holstein Cows. *J. Dairy Sci.* **2008**, *91*, 4022–4027. [[CrossRef](#)] [[PubMed](#)]
29. Heck, J.M.L.; Schennink, A.; van Valenberg, H.; Bovenhuis, H.; Visker, M.H.P.W.; van Arendonk, J.; Van Hooijdonk, A.C.M. Effects of milk protein variants on the protein composition of bovine milk. *J. Dairy Sci.* **2009**, *92*, 1192–1202. [[CrossRef](#)]
30. Braunschweig, M.; Hagger, C.; Stranzinger, G.; Puhan, Z. Associations Between Casein Haplotypes and Milk Production Traits of Swiss Brown Cattle. *J. Dairy Sci.* **2000**, *83*, 1387–1395. [[CrossRef](#)]
31. Szymanowska, M.; Malewski, T.; Zwierzchowski, L. Transcription factor binding to variable nucleotide sequences in 5'-flanking regions of bovine casein genes. *Int. Dairy J.* **2004**, *14*, 103–115. [[CrossRef](#)]
32. Chan, T.-M.; Leung, K.-S.; Lee, K.-H. TFBS identification based on genetic algorithm with combined representations and adaptive post-processing. *Bioinformatics* **2007**, *24*, 341–349. [[CrossRef](#)]
33. Boeva, V. Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. *Front. Genet.* **2016**, *7*, 24. [[CrossRef](#)] [[PubMed](#)]
34. O'Brien, J.; Hayder, H.; Zayed, Y.; Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front. Endocrinol.* **2018**, *9*, 402. [[CrossRef](#)] [[PubMed](#)]
35. Ha, M.; Kim, V.N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 509–524. [[CrossRef](#)] [[PubMed](#)]
36. Broughton, J.; Lovci, M.T.; Huang, J.L.; Yeo, E.; Pasquinelli, A.E. Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Mol. Cell* **2016**, *64*, 320–333. [[CrossRef](#)] [[PubMed](#)]
37. Ibeagha-Awemu, E.M.; Kgwatalala, P.; Zhao, X. A critical analysis of production-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mamm. Genome* **2008**, *19*, 591–617. [[CrossRef](#)] [[PubMed](#)]
38. Szymanowska, M.; Siadkowska, E.; Lukaszewicz, M.; Zwierzchowski, L. Association of nucleotide-sequence polymorphism in the 5'-flanking regions of bovine casein genes with casein content in cow's milk. *Le Lait* **2004**, *84*, 579–590. [[CrossRef](#)]
39. Kishore, A.; Sodhi, M.; Mukesh, M.; Mishra, B.P.; Sobti, R.C. Sequence analysis and identification of new variations in the 5'-flanking region of  $\alpha$ S2-casein gene in Indian zebu cattle. *Mol. Biol. Rep.* **2013**, *40*, 4473–4481. [[CrossRef](#)]
40. Ganai, N.; Bovenhuis, H.; Van Arendonk, J.A.M.; Visker, M.H.P.W. Novel polymorphisms in the bovine  $\beta$ -lactoglobulin gene and their effects on  $\beta$ -lactoglobulin protein concentration in milk. *Anim. Genet.* **2009**, *40*, 127–133. [[CrossRef](#)]
41. Lum, L.S.; Dovč, P.; Medrano, J.F. Polymorphisms of Bovine  $\beta$ -Lactoglobulin Promoter and Differences in the Binding Affinity of Activator Protein-2 Transcription Factor. *J. Dairy Sci.* **1997**, *80*, 1389–1397. [[CrossRef](#)]
42. Ibeagha-Awemu, E.; Prinzenberg, E.-M.; Erhardt, G. High variability of milk protein genes in *Bos indicus* cattle breeds of Cameroon and Nigeria and characterization of a new  $\alpha$ S1-casein promoter allele. *J. Dairy Res.* **2005**, *72*, 1–9. [[CrossRef](#)] [[PubMed](#)]
43. Prinzenberg, E.-M.; Weimann, C.; Brandt, H.; Bennewitz, J.; Kalm, E.; Schwerin, M.; Erhardt, G. Polymorphism of the Bovine CSN1S1 Promoter: Linkage Mapping, Intragenic Haplotypes, and Effects on Milk Production Traits. *J. Dairy Sci.* **2003**, *86*, 2696–2705. [[CrossRef](#)]
44. Jann, O.C.; Ibeagha-Awemu, E.; Özbeyaz, C.; Zaragoza, P.; Williams, J.; Marsan, P.A.; Lenstra, J.A.; Moazami-Goudarzi, K.; Erhardt, G. Geographic distribution of haplotype diversity at the bovine casein locus. *Genet. Sel. Evol.* **2004**, *36*, 243–257. [[CrossRef](#)]
45. Scheper, C.; Bohlouli, M.; Brügemann, K.; Weimann, C.; Vanvanhossou, S.F.U.; König, S.; Dossa, L.H. The role of agro-ecological factors and transboundary transhumance in shaping the genetic diversity in four indigenous cattle populations of Benin. *J. Anim. Breed. Genet.* **2020**, *137*, 622–640. [[CrossRef](#)]
46. Vanvanhossou, S.F.U.; Scheper, C.; Dossa, L.H.; Yin, T.; Brügemann, K.; König, S. A multi-breed GWAS for morphometric traits in four Beninese indigenous cattle breeds reveals loci associated with conformation, carcass and adaptive traits. *BMC Genom.* **2020**, *21*, 783. [[CrossRef](#)] [[PubMed](#)]
47. Lühken, G.; Caroli, A.; Ibeagha-Awemu, E.M.; Erhardt, G. Characterization and genetic analysis of bovine  $\alpha$ S1-casein invariant. *Anim. Genet.* **2009**, *40*, 479–485. [[CrossRef](#)]

48. Illumina. bcl2fastq2 Conversion Software: Version 2.20.0. 2017. Available online: [https://support.illumina.com/sequencing/sequencing\\_software/bcl2fastq-conversion-software.html](https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html) (accessed on 24 June 2021).
49. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
50. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.S.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 122. [[CrossRef](#)]
51. European Variation Archive. Milk Protein Gene Polymorphisms in the Beninese Indigenous Cattle Breeds. Available online: <https://wwwdev.ebi.ac.uk/eva/?eva-study=PRJEB47999> (accessed on 13 October 2021).
52. Tetens, J.L. Genetic Variability of Bovine and Ovine Milk Protein Genes. Ph.D. Thesis, Christian-Albrechts-Universität zu Kiel, Kiel, Germany, 2014.
53. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)]
54. Sinnwell, J.P.; Schaid, D.J. *Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous*; R package version 1.8.7; Mayo Clinic Division of Health Sciences Research: Rochester, MN, USA, 2021.
55. Reese, M.G.; Eeckman, F.H.; Kulp, D.; Haussler, D. Improved Splice Site Detection in Genie. *J. Comput. Biol.* **1997**, *4*, 311–323. [[CrossRef](#)] [[PubMed](#)]
56. Yevshin, I.; Sharipov, R.; Kolmykov, S.; Kondrakhin, Y.; Kolpakov, F. GTRD: A database on gene transcription regulation—2019 update. *Nucleic Acids Res.* **2018**, *47*, D100–D105. [[CrossRef](#)] [[PubMed](#)]
57. Kolmykov, S.; Yevshin, I.; Kulyashov, M.; Sharipov, R.; Kondrakhin, Y.; Makeev, V.J.; Kulakovskiy, I.V.; Kel, A.; Kolpakov, F. GTRD: An integrated view of transcription regulation. *Nucleic Acids Res.* **2020**, *49*, D104–D111. [[CrossRef](#)]
58. Agarwal, V.; Bell, G.W.; Nam, J.-W.; Bartel, D.P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **2015**, *4*, e05005. [[CrossRef](#)]
59. Lewis, B.P.; Burge, C.B.; Bartel, D.P. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* **2005**, *120*, 15–20. [[CrossRef](#)] [[PubMed](#)]
60. Meier, S.; Korcuć, P.; Arends, D.; Brockmann, G.A. DNA Sequence Variants and Protein Haplotypes of Casein Genes in German Black Pied Cattle (DSN). *Front. Genet.* **2019**, *10*, 1129. [[CrossRef](#)] [[PubMed](#)]
61. Hanotte, O.; Dessie, T.; Kemp, S. Time to Tap Africa’s Livestock Genomes. *Science* **2010**, *328*, 1640–1641. [[CrossRef](#)] [[PubMed](#)]
62. Cartegni, L.; Wang, J.; Zhu, Z.; Zhang, M.Q.; Krainer, A.R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **2003**, *31*, 3568–3571. [[CrossRef](#)] [[PubMed](#)]
63. Yuan, M.; Xia, W.; Zhang, X.; Liu, Y.; Jiang, M. Identification and verification of differentially expressed genes in yak mammary tissue during the lactation cycle. *J. Dairy Res.* **2020**, *87*, 158–165. [[CrossRef](#)]
64. Sumner-Thomson, J.M.; Vierck, J.L.; McNamara, J.P. Differential expression of genes in adipose tissue of first-lactation dairy cattle. *J. Dairy Sci.* **2011**, *94*, 361–369. [[CrossRef](#)] [[PubMed](#)]
65. Hindson, C.M.; Chevillet, J.; Briggs, H.A.; Gallichotte, E.N.; Ruf, I.K.; Hindson, B.J.; Vessella, R.L.; Tewari, M. Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods* **2013**, *10*, 1003–1005. [[CrossRef](#)] [[PubMed](#)]
66. Costa-Silva, J.; Domingues, D.; Lopes, F.M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* **2017**, *12*, e0190152. [[CrossRef](#)] [[PubMed](#)]
67. Lisson, M.; Lochnit, G.; Erhardt, G. In vitro gastrointestinal digestion of bovine  $\alpha$ S1- and  $\alpha$ S2-casein variants gives rise to different IgE-binding epitopes. *Int. Dairy J.* **2014**, *34*, 47–55. [[CrossRef](#)]
68. Giambra, I.J.; Jäger, S.; Erhardt, G. Isoelectric focusing reveals additional casein variants in German sheep breeds. *Small Rumin. Res.* **2010**, *90*, 11–17. [[CrossRef](#)]
69. Najafi, M.; Mianji, G.R.; Pirsaraie, Z.A. Cloning and comparative analysis of gene structure in promoter site of alpha-s1 casein gene in Naeinian goat and sheep. *Meta Gene* **2014**, *2*, 854–861. [[CrossRef](#)] [[PubMed](#)]
70. Cosenza, G.; Pauciullo, A.; Colimoro, L.; Mancusi, A.; Rando, A.; Di Bernardino, D.; Ramunno, L. An SNP in the goat CSN2 promoter region is associated with the absence of  $\beta$ -casein in milk. *Anim. Genet.* **2007**, *38*, 655–658. [[CrossRef](#)] [[PubMed](#)]
71. Ibeagha-Awemu, E.; Prinzenberg, E.M.; Jann, O.C.; Lühken, G.; Ibeagha, A.E.; Zhao, X.; Erhardt, G. Molecular Characterization of Bovine CSN1S2B and Extensive Distribution of Zebu-Specific Milk Protein Alleles in European Cattle. *J. Dairy Sci.* **2007**, *90*, 3522–3529. [[CrossRef](#)]
72. Shen, B.; Zhang, L.; Lian, C.; Lu, C.; Zhang, Y.; Pan, Q.; Yang, R.; Zhao, Z. Deep Sequencing and Screening of Differentially Expressed MicroRNAs Related to Milk Fat Metabolism in Bovine Primary Mammary Epithelial Cells. *Int. J. Mol. Sci.* **2016**, *17*, 200. [[CrossRef](#)] [[PubMed](#)]
73. de Sousa, M.A.P.; de Athayde, F.R.F.; Maldonado, M.B.C.; de Lima, A.O.; Fortes, M.R.S.; Lopes, F.L. Single nucleotide polymorphisms affect miRNA target prediction in bovine. *PLoS ONE* **2021**, *16*, e0249406. [[CrossRef](#)] [[PubMed](#)]
74. Damiani, G.; Florio, S.; Budelli, E.; Bolla, P.; Caroli, A. Single nucleotide polymorphisms (SNPs) within Bov-A2 SINE in the second intron of bovine and buffalo k-casein (CSN3) gene. *Anim. Genet.* **2000**, *31*, 277–279. [[CrossRef](#)]
75. Tetens, J.L.; Qanbari, S.; Drögemüller, C.; Pimentel, E.C.G.; Bennewitz, J.; Thaller, G. Bos indicus introgression into (peri-)alpine cattle breeds—evidence from the analysis of bovine whey protein variants. *Anim. Genet.* **2014**, *45*, 585–588. [[CrossRef](#)]
76. Bell, K.; Hopper, K.E.; McKenzie, H.A. Bovine  $\alpha$ -Lactalbumin C and  $\alpha$ S1-,  $\beta$ - and  $\gamma$ -Caseins of Bali (Banteng) Cattle, Bos (Bihos) Javanicus. *Aust. J. Biol. Sci.* **1981**, *34*, 149. [[CrossRef](#)]

77. Pausch, H.; MacLeod, I.M.; Fries, R.; Emmerling, R.; Bowman, P.J.; Daetwyler, H.D.; Goddard, M.E. Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. *Genet. Sel. Evol.* **2017**, *49*, 1–14. [[CrossRef](#)] [[PubMed](#)]
78. Medrano, J.F.; Aguilar-Cordova, E. Polymerase chain reaction amplification of bovine  $\beta$ -lactoglobulin genomic sequences and identification of genetic variants by RFLP analysis. *Anim. Biotechnol.* **1990**, *1*, 73–77. [[CrossRef](#)]
79. Martin, P.; Ollivier-Bousquet, M.; Grosclaude, F. Genetic polymorphism of caseins: A tool to investigate casein micelle organization. *Int. Dairy J.* **1999**, *9*, 163–171. [[CrossRef](#)]
80. Formaggioni, P.; Summer, A.; Malacarne, M.; Mariani, P. Milk protein polymorphism: Detection and diffusion of the genetic variants in Bos genus. *Ann. Fac. Med. Vet. Univ. Parma* **1999**, *19*, 127–165.
81. Mohr, U.; Koczan, D.; Linder, D.; Hobom, G.; Erhardt, G. A single point mutation results in A allele-specific exon skipping in the bovine  $\alpha$ s1-casein mRNA. *Gene* **1994**, *143*, 187–192. [[CrossRef](#)]
82. Rahmatalla, S.A.; Arends, D.; Ahmed, A.S.; Hassan, L.M.A.; Krebs, S.; Reissmann, M.; Brockmann, G.A. Capture Sequencing to Explore and Map Rare Casein Variants in Goats. *Front. Genet.* **2021**, *12*, 620253. [[CrossRef](#)]
83. Rando, A.; Ramunno, L.; Masina, P. Mutations in casein genes. *Zootec. Nutr. Anim.* **2000**, *26*, 105–114.
84. Mariani, P.; Pecorari, M. Fattori genetici, attitudine alla caseificazione e resa del latte in formaggio. *Sci. Tecn. Latt.-Cas* **1987**, *38*, 286–326.
85. Merin, U.; Bernstein, S.; Bloch-Damti, A.; Yagil, R.; van Creveld, C.; Lindner, P.; Gollop, N. A comparative study of milk serum proteins in camel (*Camelus dromedarius*) and bovine colostrum. *Livest. Prod. Sci.* **2001**, *67*, 297–301. [[CrossRef](#)]
86. Shabo, Y.; Barzel, R.; Margoulis, M.; Yagil, R. Camel milk for food allergies in children. *Isr. Med Assoc. J.* **2005**, *7*, 796–798. [[PubMed](#)]
87. El-Agamy, E.-S.I. Camel Milk. In *Handbook of Milk of Non-Bovine Mammals*; Park, Y.W., Haenlein, G.F.W., Wendorff, W.L., Eds.; John Wiley & Sons, Ltd.: Oxford, UK, 2017; pp. 409–480, ISBN 9781119110316.
88. Rando, A.; Di Gregorio, P.; Ramunno, L.; Mariani, P.; Fiorella, A.; Senese, C.; Marletta, D.; Masina, P. Characterization of the CSN1AG Allele of the Bovine  $\alpha$ s1-Casein Locus by the Insertion of a Relict of a Long Interspersed Element. *J. Dairy Sci.* **1998**, *81*, 1735–1742. [[CrossRef](#)]
89. Cartegni, L.; Chew, S.L.; Krainer, A. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* **2002**, *3*, 285–298. [[CrossRef](#)] [[PubMed](#)]
90. Braunschweig, M.H.; Leeb, T. Aberrant Low Expression Level of Bovine  $\beta$ -Lactoglobulin Is Associated with a C to A Transversion in the BLG Promoter Region. *J. Dairy Sci.* **2006**, *89*, 4414–4419. [[CrossRef](#)]
91. Brooke-Taylor, S.; Dwyer, K.; Woodford, K.; Kost, N. Systematic Review of the Gastrointestinal Effects of A1 Compared with A2  $\beta$ -Casein. *Adv. Nutr.* **2017**, *8*, 739–748. [[CrossRef](#)]
92. Gautier, P.; Escobar Stuardo, L. Rapport de la Mission de suivi de L'évaluation PVS des Services Vétérinaires du Bénin, OIE—World Organisation for Animal Health, Paris, France. 2013. Available online: [https://rr-africa.oie.int/wp-content/uploads/2019/11/pvs\\_fu\\_rapport-benin.pdf](https://rr-africa.oie.int/wp-content/uploads/2019/11/pvs_fu_rapport-benin.pdf) (accessed on 20 June 2021).
93. FAO. *Phenotypic Characterization of Animal Genetic Resources*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2012.
94. Vapnek, J.; Chapman, M. *Legislative and Regulatory Options for Animal Welfare*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2011; ISBN 9251067112.