diversity

MDPI

Article

Comparative Analysis of the Polymorphism of the Casein Genes in Camels Bred in Kazakhstan

Makpal Amandykova [1,2,*], Kairat Dossybayev [1,2], Aizhan Mussayeva [2], Bakytzhan Bekmanov [1,2] and Naruya Saitou [3]

[1] Faculty of Biology and Biotechnology, Al-Farabi Kazakh National University, Al-Farabi Ave., 71, Almaty 050040, Kazakhstan; kairat1987_11@mail.ru (K.D.); bobekman@rambler.ru (B.B.)
[2] Laboratory of Genetics and Cytogenetics, Institute of Genetics and Physiology SC MES RK, Al-Farabi Ave., 93, Almaty 050060, Kazakhstan; aimus_@mail.ru
[3] Division of Population Genetics, National Institute of Genetics, 1111 Yata, Mishima 411-8540, Shizuoka, Japan; saitounr@nig.ac.jp
* Correspondence: makpal_30.01@mail.ru; Tel.: +7-7473609475

Abstract: Caseins play an important role in determining the technological properties and quantitative characteristics of camel milk. To date, only a few studies on the genetic polymorphism of casein genes have been reported in the camel populations of Kazakhstan. Therefore, this work aimed to identify the genetic polymorphism level of casein genes among camel populations of the Almaty region of Kazakhstan. The PCR-RFLP method was used for this purpose and the following genotypes were revealed as a result: *CSN3* gene—CC, CT, TT, where the T allele predominated in all populations, with a frequency of 0.60; *CSN2* gene—AA, AG, GG, with the predomination of A allele (0.64); and *CSN1S1* gene—GG and GT, with the predomination of G allele (0.94). Statistical analysis was carried out using the POPGENE and GenAlEx software. The $\chi 2$ values were equal to 12.1 (*CSN3*), 8.6 (*CSN2*), and 14.5 (*CSN1S1*). As a result, three out of 53 animals were designated as the "core" of the population—animals with the desired genotypes: CC genotype for the *CSN3* gene and AA genotype for the *CSN2* gene. Such animals can be selected for further use with an increase in the number of livestock with high productivity rates.

Keywords: camel; milk productivity; casein genetic variability; *CSN3*; *CSN2*; *CSN1S1*; PCR-RFLP

Citation: Amandykova, M.; Dossybayev, K.; Mussayeva, A.; Bekmanov, B.; Saitou, N. Comparative Analysis of the Polymorphism of the Casein Genes in Camels Bred in Kazakhstan. *Diversity* **2022**, *14*, 285. https://doi.org/10.3390/d14040285

Academic Editors: Luc Legal and Tatiana Deniskova

Received: 24 February 2022
Accepted: 7 April 2022
Published: 11 April 2022

1. Introduction

One-humped (*Camelus dromedarius*) and two-humped (*Camelus bactrianus*) camels are widely spread in Africa and Asia. Namely, the dromedary inhabits in the northern part of Africa, the Middle East part of Asia, and the Indian subcontinent, while the Bactrian camel occupies inner, central, and East Asia: China, Mongolia, Kazakhstan, Kyrgyzstan, and Afghanistan, northern areas of Iran, India, Pakistan and up to eastern Turkey [1]. Recently, it was revealed that the Bactrians have Central Asian origin from where they immigrated eastward to Mongolia, where the wild Bactrian camels inhabit today [2]. On the other hand, the dromedaries were domesticated in the southeast of the Arabian Peninsula [3].

Previously, the use of camels in the production of dairy products was a non-specialized and secondary activity. It is worth noting that camel milk appears to contain more common proteins such as immunoglobulins, lactoferrins, and the useful enzymes lysozyme and lactoperoxidase, which aid the body in fighting infection [4,5]. There are a large number of studies currently underway looking into whether camel milk can be an effective preventive measure against diabetes and heart diseases [6,7]. According to research in India, there are no cases of type 1 diabetes among populations that consume camel milk regularly, compared to 5.5% in other populations that do not consume this milk [8]. Camel milk does not clot in an acidic environment and has a high buffering capacity, controlling

hyperglycemia in type 1 diabetes [9,10]. Camel milk fat contains very little or no β-lactoglobulin and is rich in immunoglobulins compatible with breast milk [11]. Therefore, its products are positioned as medicinal, hypoallergenic, and dietary.

Increased interest in camel milk in a more urbanized world has prompted research into the breeding of dairy camels [12]. Along with the investigation of non-genetic factors influencing milk composition, research was conducted on the evaluation of genetic factors [13]. Considering the significance of camels for dairy production in many desert regions, genomic tools are being developed to assess milk yield, milk composition, and udder morphology as key targets for future genomic breeding [14].

Advanced proteomic methods were recently used to analyze the whey protein complex of one-humped and two-humped camels. Caseins (CN) account for roughly 80% of cow milk proteins [15], while the casein content of camel milk is 52–87% [16], consisting of αs1-, αs2-, β-, γ- and κ-CN [17–19]. While β- and γ-CN are monomorphic, four distinct genetic variants (A, B, C, and D) have been identified in αs1-CN [20]. Some studies are looking into the genetic factors that influence milk yield increase. For example, oxytocin, a neurohypophyseal peptide involved in a variety of biological functions, affects milk secretion, temperament, and reproduction [21].

Kappa-caseins are important in the stabilization of casein micelles, which determine the specific properties of milk [22]; its breakdown by chymosin is responsible for milk coagulation. They prevent casein micelle aggregation and, thus, help to keep calcium phosphate in solution, as well as provide milk calcium and phosphorus bioavailability, which is especially important for human nutrition [23]. The study of genetic polymorphisms in casein proteins can be used to increase population diversity, preserve the genetic structure, and understand its relationships with the characteristics of dairy production.

As a result of the study on the putative regulatory regions of the camel κ-casein gene, two polymorphic sites were detected in camel promoters. The first is a transition g.975A > G in position −112 which does not affect any putative regulatory regions. The latter, at nucleotide −17, is a transition g.1029T > C. The presence of thymine is responsible for the creation of an extra putative consensus sequence for the hepatocyte nuclear factor-1 (HNF-1). Another member of hepatocyte nuclear factors (HNF-3) is well known to be a potential regulator of casein genes expression [24]; thus, it is suggested that there is the influence of HNF-1 allelic variant in camel κ-casein gene regulation. Though the influence of HNF-1 allelic variants on *CSN3* needs to be further substantiated, the presence of C allele at g.1029 locus in the 5′UTR of *CSN3* gene provides the opportunity for the rapid directional selection in favour of such alleles [25].

β-casein is the most abundant protein in camel milk, and its coding gene (*CSN2*) is considered the "main" gene in other species, the alleles of which are associated with various levels of protein gene expression [26]. It has been reported that the presence of various levels of β-casein phosphorylation affects the availability and distribution of calcium, as well as the stability of micelles [27], implying that β-CN plays an important role in nutrition and the development of the technological properties of milk and dairy products. Several studies in ruminants have found an association between β-casein gene polymorphism and economically important characteristics of milk [28].

Previously, Pauciullo et al. identified a transition g.2126A > G in *CSN2* promoter region of four Sudanese *Camelus dromedarius* populations [26]. This 2126A/G SNP is located three nucleotides downstream of the TATA box and might modify the binding affinity of RNA polymerase and change the gene expression. Association among these three genotypes with milk composition traits suggests a positive effect of A/A genotype on acidity and protein percentage. Higher protein and acidity values were observed in the milk of individuals carrying the A/A genotype [29].

In ruminants, the αS1-CN fraction has been extensively studied and characterized at both the protein and DNA levels. Many alleles associated with different rates of protein synthesis have been identified in the corresponding coding gene (*CSN1S1*) [30,31]. Two genetic variants of αS1-CN have been described in one-humped camels [32]. αs1-casein

polymorphism affects the milk lipids and protein compositions, so it has a strong impact on the nutritional quality and technological properties of milk. Currently, at least four protein variants (A, B, C, and D) have been detected in the dromedary αs1-CN, and the molecular event responsible for these phenotype variations was clarified in three out of four cases. The allele C is due to a single nucleotide polymorphism (c.150G > T) occurring at the exon 5 and resulting in the amino acid substitution p.30Glu > Asp in mature protein [33]. Recently, Erhardt et al. reported a new variant, named D. Even though the *CSN1S1* gene polymorphism is actively studied in camels, its effect on the technological properties of milk has not yet been identified, and the role of these SNPs remains to be studied in comparison with the phenotypic data of animals [17].

αs2-casein is the most hydrophilic of all milk casein proteins. The proportion of s2-CN in milk varies greatly between species, and this protein is completely absent in the composition of human milk [34]. The *CSN1S2* gene, which encodes this protein, has rarely been studied in most mammals, including camels. To date, no information has been found about any single-nucleotide polymorphisms (SNP) found in this gene affecting the productive qualities of camel milk.

SNP genotyping technologies are valuable resources for animal breeding programs. Genomic selection by PCR-RFLP analysis is a simple and accessible method for selecting the best breeding animals, as well as a useful tool in small-scale basic research studies of the economic properties associated with single nucleotide polymorphisms [35].

This work aimed to identify the genetic polymorphism of casein genes and to determine the allelic variants frequencies in *Camelus dromedarius*, *Camelus Bactrianus*, and hybrids of the Kazakh camel populations, as well as identifying animals with desirable genotypes according to the studied genes. For the first time, we conducted genotyping of Kazakhstan camels for three SNPs occurring in *CSN3, CSN2*, and *CSN1S1* casein genes, the results of which can be used for breeding management and the selection of highly productive animals. This paper aims to present and summarize the results of genotyping of camels and identify the "core" of populations: animals that have the «desired» genotypes for the studied SNPs.

2. Materials and Methods

2.1. Collection of Materials

All procedures performed in studies involving animals followed the ethical standards of the Institute of Genetics and Physiology (Almaty, Kazakhstan), at which the studies were conducted. Blood samples were collected using Venosafe tubes with K2EDTA (Leuven, Belgium) in the Ile district of the Almaty region of Kazakhstan (Figure 1). Camel breeding has been a traditional branch of productive livestock breeding in this region for centuries, as a result, there are a large number of camel breeding farms (Otemis Makan Farm, Daulet-Beket LLP, etc.). Thus, 53 camels of the age of 2–3 years from four distinct herds were selected for our study. Moreover, family ties of camels bred in the same herd are possible. There are eight one-humped camels (n = 6 ♀; n = 2 ♂)—dromedaries (*Camelus dromedarius*) in the first population; eight hybrids (n = 5 ♀; n = 3 ♂) and one dromedary (♀) in the second population; fourteen hybrids (n = 10 ♀; n = 4 ♂) and six two-humped camels (n = 4 ♀; n = 2 ♂)—Bactrians (*Camelus bactrianus*) in the third population; and sixteen Bactrians (n = 8 ♀; n = 8 ♂) in the fourth population. The samples were refrigerated in special containers and transported to the laboratory of Animal Genetics and Cytology, where they were stored in a freezer until further use for DNA extraction.

Figure 1. Location of sample collection from Kazakhstan.

2.2. Genomic DNA Extraction and PCR-RFLP Analysis

Genomic DNA was extracted according to the protocol supplied by the manufacturer in the DNA-sorb-B kit for DNA isolation from blood and tissues (AmpliSens, Moscow, Russia). Prepared DNA samples were stored in a freezer at −20 °C. The DNA quality was determined using agarose gel electrophoresis, and the concentration of DNA was determined using Nanodrop One (ThermoScientific, Waltham, MA, USA). The genomic DNA were amplified through the polymerase chain reaction (PCR) method in the thermal cycler Mastercycler nexus gradient (Eppendorf, Hamburg, Germany). PCR was performed in 10 µL final volume solution using the Master Mix (PCR Master Mix, ThermoScientific, Waltham, MA, USA). Table 1 shows for genes *CSN3*, *CSN2* and *CSN1S1*: PCR fragment sizes, oligonucleotide primers, annealing temperature, and restriction enzymes.

Table 1. PCR-RFLP analysis conditions, restriction enzymes and oligonucleotide primers for casein genes genotyping.

№	SNP	Fragment Length	Primers	Annealing Temperature	Restriction Enzymes	References
1	*CSN3* g.1029T > C	488 bp	5′-CACAAAGATGACTCTGCTATCG-3′ 5′-GCCCTCCACATATGTCTG-3′.	60 °C	*AluI*	[6]
2	*CSN2* g.2126A > G	659 bp	5′-GTTTCTCCATTACAGCATC-3′ 5′-TCAAATCTATACAGGCACTT-3′.	53 °C	*HphI*	[26]
3	*CSN1S1* c.150G > T	930 bp	5′-TGAACCAGACAGCATAGAG-3′ 5′-CTAAACTGAATGGGTGAAAC-3′	55 °C	*SmoI*	[17]

The PCR products were visualized using 2% agarose gel electrophoresis, which was stained with ethidium bromide. Restriction products were analyzed using 5% polyacrylamide gel electrophoresis stained with SYBR Gold Nucleic Acid Gel Stain (ThermoScientific, Waltham, MA, USA) in 1 × TBE buffer. The results were visualized by using a gel-documenting system (Quantum-ST5-1100 Vilber Lourmat, France).

2.3. Statistical Data Processing

The allele frequency and Hardy–Weinberg equilibrium ($\chi 2$) were calculated in all studied populations. Statistical analysis of such indicators as the frequency of occurrence of genotypes, the observed and effective number of alleles, Nei's gene diversity, and Shannon's Information index was carried out using the GeneAlex [36] and POPGENE Software [37]. These programs calculate the above statistics based on Nei's standard genetic distance, which has the property that if the rate of genetic change is constant per year or generation

then Nei's standard genetic distance (D) increases in proportion to the divergence time. This measure assumes that genetic differences are caused by mutation and genetic drift [38]. A phylogenetic tree based on the standard Nei's genetic distance was constructed by MEGA-X, which is based on the principle of the Neighbor-joining Method, to find pairs of neighbors that minimize the total branch length at each stage of clustering of neighbors, starting with a star-like tree [39,40].

3. Results

Four populations of camels bred in the Almaty region were genotyped on polymorphic SNPs in three genes (g.942G > T, *CSN1S1* gene; g.2126A > G, *CSN2* promoter; and g.1029T > C, *CSN3* promoter). As a result, the following genotypes were identified among the studied populations: CC, CT, and TT for *CSN3*; AA, AG, and GG for *CSN2*; GG and GT for *CSN1S1* gene (Table 2).

Table 2. *CSN3*, *CSN2*, and *CSN1S1* genes polymorphism study.

Population	Overall	*CSN3*					*CSN2*					*CSN1S1*				
		Identified Genotypes			Allele Frequency		Identified Genotypes			Allele Frequency		Identified Genotypes			Allele Frequency	
		CC	CT	TT	C	T	AA	AG	GG	A	G	GG	GT	TT	G	T
1	8	1	3	4	0.30	0.70	5	2	1	0.75	0.25	6	2	0	0.88	0.12
2	9	1	2	6	0.18	0.82	4	2	3	0.60	0.40	8	1	0	0.94	0.06
3	20	2	10	8	0.37	0.63	18	2	0	0.95	0.05	20	0	0	1.00	0
4	16	8	3	5	0.44	0.56	1	6	9	0.25	0.75	14	2	0	0.94	0.06
Overall	53	12	18	23	0.40	0.60	28	12	13	0.64	0.36	48	5	0	0.94	0.06
				$\chi2 = 12.1$					$\chi2 = 8.6$					$\chi2 = 14.5$		

As a result of camel genotyping for the *CSN3* gene, the T allele predominated in all four populations, with a frequency of 0.61 ranging from 0.56 to 0.82 (Table 1). Regarding *CSN2* gene polymorphism, the frequency of A allele was higher in the 1st, 2nd, and 3rd populations, with values of 0.75, 0.6, and 0.95, respectively. In the fourth population, in contrast, allele G accounted for the majority of genotypes (0.75). For *CSN1S1* gene, it can be seen that G allele predominated in all four populations, while T allele had a minor frequency.

These research results were used to calculate the Hardy–Weinberg equilibrium ($\chi2$). The $\chi2$ values in the studied camel groups were equal to 12.1 (g.1029T > C, *CSN3* promoter, $p = 0.146798$; non-significant at 5% probability level of significance), 8.6 (g.2126A > G, *CSN2* promoter, $p = 0.377154$; non-significant at 5% probability level of significance), and 14.5 (g.942G > T, allele C of *CSN1S1*, $p = 0.069629$; non-significant at 5% probability level of significance).

Next, an identical analysis of the frequency of occurrence of various genotypes and alleles for the studied genes was performed, dividing the studied individuals by species: one-humped camels—dromedary (*Camelus dromedarius*), two-humped camels—Bactrian (*Camelus bactrianus*), and their hybrids.

As shown in Table 3, the «desired allele» C for the kappa-casein gene was most frequently found in Bactrians (0.52), while in dromedaries and hybrids, this indicator had nearly the same value (0.3 and 0.33). In the case of the β-casein gene, the «desired allele» A was found most frequently in hybrids (0.8), then in dromedaries (0.66), and least frequently in Bactrians (0.48).

Table 3. CSN3, CSN2, and CSN1S1 genotype distribution in one-humped, two-humped camels and their hybrids.

Populations	Overall	*CSN3*					*CSN2*					*CSN1S1*				
		Identified Genotypes			Allele Frequency		Identified Genotypes			Allele Frequency		Identified Genotypes			Allele Frequency	
		CC	CT	TT	C	T	AA	AG	GG	A	G	GG	GT	TT	G	T
One-humped camels (*Camelus dromedarius*)	9	1	4	4	0.33	0.67	5	2	2	0.66	0.33	7	2	0	0.89	0.11
Two-humped camels (*Camelus bactrianus*)	22	8	7	7	0.52	0.48	7	7	8	0.48	0.52	20	2	0	0.95	0.05
Hybrids	22	3	7	12	0.30	0.70	16	3	3	0.80	0.20	21	1	0	0.98	0.02
Overall	53	12	18	23	0.40	0.60	28	12	13	0.64	0.36	48	5	0	0.94	0.06
		$\chi2 = 0.40$					$\chi2 = 0.98$					$\chi2 = 36.32$				

The obtained data of all studied animals were analyzed to determine several statistical indicators: the identified and effective number of alleles, Nei's gene diversity, Shannon's information index, and original measures of Nei's genetic identity and genetic distance (Table 4).

Table 4. Summary of the genetic variation statistics for all loci.

Locus		Na	Ne	h	I
CSN3	CC	2.0000	1.5392	0.3503	0.5349
	CT	2.0000	1.8134	0.4486	0.6408
	TT	2.0000	1.9657	0.4913	0.6844
CSN2	AA	2.0000	1.9824	0.4956	0.6887
	AG	2.0000	1.5879	0.3702	0.5571
	GG	2.0000	1.4902	0.3289	0.5107
CSN1S1	GG	2.0000	1.2061	0.1709	0.3125
	GT	2.0000	1.2061	0.1709	0.3125
	TT	1.0000	1.0000	0.0000	0.0000
Mean		1.8889	1.5323	0.3141	0.4713
Statistical error		0.3333	0.3483	0.1684	0.2255

Na = Observed number of alleles [41]. Ne = Effective number of alleles [42]. h = Nei's gene diversity [43]. I = Shannon's Information index [44].

The average value of the observed number of alleles (Na), which represents the actual number of alleles found in the studied populations, was the same (2.0000) for all genotypes identified in this study, except the TT genotype for the *CSN1S1* gene (1.0000 ± 0.3333), which is attributable to the complete absence of this genotype in the studied populations.

For the abovementioned reasons, the average effective number of alleles (Ne) was highest for the AA genotype of the *CSN2* gene (1.9824 ± 0.3483) and lowest for the TT genotype of the *CSN1S1* gene (1.0000 ± 0.3483). This indicator describes the number of alleles with the same frequency required to achieve the same expected heterozygosity as in the population under study. In terms of gene diversity, this indicator had the highest value for the AA genotype of the *CSN2* gene (0.4956 ± 0.1684) and the lowest value for the TT genotype of the *CSN1S1* gene (0.4956 ± 0.1684), indicating the complete absence of genetic diversity for this gene. The Shannon's index, which measures gene diversity, was also high for the *CSN2* AA genotype and equal to zero for the *CSN1S1* TT genotype.

Next, we constructed a phylogenetic tree to assess the related proximity of populations in terms of the diversity of *CSN3*, *CSN2*, and *CSN1S1* genes (Figure 2). It can be seen that,

according to the studied SNPs, the 1st and 3rd populations are the most closely related to each other and are similar to the 2nd population, while the 4th population is located on the branch of the phylogenetic tree, which is the most distant from all populations.

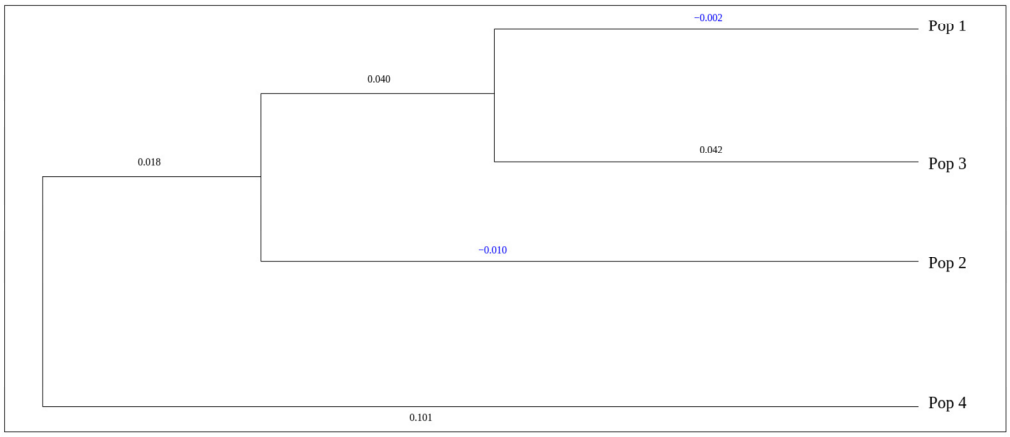

Figure 2. Phylogenetic tree based on the standard Nei's genetic distance between four camel populations.

To check the reliability of the constructed phylogenetic tree, we used data from the calculation of genetic identity and genetic distance between populations, presented in the table below (Table 5).

Table 5. Nei's original measures of genetic identity and genetic distance [37]. Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

Population	1	2	3	4
1	–	0.9567	0.9603	0.8684
2	0.0443	–	0.9447	0.8969
3	0.0405	0.0569	–	0.8046
4	0.1411	0.1088	0.2174	–

The genetic identity of the 1st and 3rd populations is 0.9603, and the genetic distance is 0.0405, demonstrating their close relationship, as shown in the phylogenetic tree. Furthermore, the most related populations were the 1st and 2nd, as well as the 2nd and 3rd, with genetic identities of 0.9567 and 0.9447, respectively. The genetic identity of the 1st, 2nd, and 3rd populations is approximately the same as that of the 4th population (0.8684; 0.8969, and 0.8046). Thus, the 3rd and 4th populations are the most distant from each other, as their genetic distance has the highest rate of 0.2174.

4. Discussion

Heterozygosity is considered the most widely used criterion for assessing genetic variation in a population. The effective number of alleles is assumed to be a measure of genetic diversity in both individual species and populations, and it is determined by the proportion of polymorphic loci, the number of alleles per locus, and the evenness of allele frequencies. The effective number of alleles is used to calculate the inverse measure of homozygosity. When the frequency of these alleles in the population is the same, heterozygosity will be equal to the actual one [45].

The presence of thymine in the studied *CSN3* gene SNP is responsible for creating an additional putative consensus sequence for hepatocyte nuclear factor-1 (HNF-1). According to reports, this transcription factor regulates several genes involved in innate immunity, lipid and glucose transport, metabolism, along with others [46]. However, as the g.1029T > C transition occurs in the *C. dromedarius* promoter just in front of the exon 1, allele C is

also responsible for the creation of an additional putative site for the transcription factor HNF-1, which is why this allele is considered desirable in selection of the *CSN3* gene polymorphism. However, the frequency of the desired C allele was less significant, ranging from 0.18 to 0.44 among studied populations, which is lower than that of in Sudanese camel populations, where this indicator ranged from 0.30 to 0.46 [6]. Thus, based on the data obtained, it is clear that there is a need for targeted selection in favor of the C allele in all of the studied populations. On average, the frequency of allele T was observed to be 0.60 and that of C was observed to be 0.40. The present results are in agreement with the results of the genotyping of Indian camel populations, where the T allele was equal to 0.763 and the C allele was observed to be 0.237 [25]. This trend is also observed in the Egyptian camel population, where the C allele showed the lowest frequency of occurrence (0.32) compared to the T allele (0.68) [47].

SNP (g.2126A > G) occurs three nucleotides downstream of the TATA box and modifies the putative consensus site for the non-canonical TATA box (TFmatrix: M00252), lowering the binding affinity from 87.7 to 86.2%. The lower frequency of the G allele indicates that this genetic variant, which is responsible for the low affinity of the TATA box-binding site, may be less facilitated during gene transcription [26]. As a result, we can suggest that allele A is "desirable" in this SNP. The frequency of the desired A allele was 0.64 on average in all studied populations, while the frequency of the G allele was 0.36. These indicators coincide with the results of genotyping of 180 Sudanese camels, where the A allele frequency was estimated to be 0.65 and the G allele frequency was 0.35 [26], as well as with the genotyping data of Egyptian camels, where the frequency of the A allele indicated 0.63, and the G allele was 0.37, respectively [29].

According to the results of several studies, the PCR-RFLP method for typing SNP g.942G > T can be used to determine the variability in the camel *CSN1S1* gene, regardless of age, sex, or lactation stage [33]. This can be useful for a broader analysis of the variability of camel milk protein and enables the analysis of the relationship between milk protein variability and milk production characteristics in camels, as has already been carried out for several farm animals [48,49]. According to the findings of our study, the TT genotype did not occur in all four populations; thus, the frequency of the T allele is much lower than the frequency of the G allele (0.06 and 0.94, respectively). The T allele did not exist in the third population, even as a heterozygous genotype. It should be noted that the "desired" genotype for the studied SNP was not established yet, and future research should focus on determining the relationship of the identified genotypes with the technical properties or other characteristics of milk.

The use of $\chi 2$ allows us to determine the extent to which the actual genotype distribution corresponds to its theoretical value. According to the Hardy–Weinberg equilibrium law, any equilibrium ratio of alleles can occur in infinite genetic populations in the absence of mutation, migration, and selection, and the relative frequencies of each allele remain constant from generation to generation at the same time. It is well known that shifts in dynamic equilibrium in favor of a specific allele or genotype are caused by the combined action of four factors: mutations, migrations, selection, and scholastic fluctuations in allele concentration due to small population size (genetically automatic processes or genetic drift) [26]. The $\chi 2$ values in the studied camel groups indicate a deviation in the actual allele frequency of the studied genes from the theoretical one. It demonstrates the lack of gene balance for these loci. The latter can be explained by the action of artificial selection in the population.

Summarizing the data on the frequency of occurrence of alleles and genotypes for the *CSN2* and *CSN3* genes, the so-called «core» of populations: animals with the desired genotypes for both SNPs under consideration, can be identified. Thus, three out of fifty-three animals can be assigned to the «core» of populations, as they have the corresponding «desirable» CC genotype for the *CSN3* gene and AA genotype for the *CSN2* gene. All three animals are hybrids; two of them are from the third population and one is from the second.

Such animals can be selected for further use with an increase in the number of livestock with high productivity rates.

Furthermore, it was interesting for us to compare the results of our study with camels belonging to certain species: one-humped, two-humped camels, and hybrids. According to the literature, dromedary milk production is quite high—over 2000 kg per lactation period (up to 16–18 months). Bactrians are distinguished by a lower milk yield but a higher fat content [50]. Nar-Maya and Kospak hybrids of the first generation are intermediate between Bactrians and dromedaries [51]. As a result, we can assume a relationship between the frequency of occurrence of the «desirable» genotypes of certain species based on our research findings, and their milk yield indicators based on literature data. As our study results show, the data on genotyping does not fully correspond to the data on quantitative indicators of camel productivity. However, even though milk from two-humped camels has a higher fat content than milk from one-humped camels and hybrids, it can be assumed that the high content of the C allele of kappa-casein in two-humped camels affects this milk quality trait. This assumption can be confirmed by performing a combined genotyping and biochemical analysis of milk from the same camel population.

Concerning the phylogenetic tree, the observed discrepancy can be explained by the similarity of allele frequencies for the studied genes. For example, the first and third populations closest to each other in the phylogenetic tree have the closest allele frequencies for all studied SNPs (Table 2), which shows their similarity at genetic polymorphism level of these genes. This may also indicate the presence of closely related relationships between these populations. A similar explanation can be given regarding the divergence of the remaining populations on the phylogenetic tree. The measurement of genetic changes is determined by two parameters: genetic identity (I), which estimates the proportion of genes that are identical in two populations, and genetic distance (D), which estimates the proportion of gene changes that have occurred during the evolution of two populations. The I value can range from 0 to 1, which corresponds to extreme situations in which none or all genes are identical, respectively; the D value can range from zero to infinity. D can exceed 1, as each gene can change more than once in one or both populations, since evolution continues for many generations.

5. Conclusions

Camel genotyping by milk productivity genes is assumed to be a more effective method of targeted breeding for improving milk quantitative and qualitative parameters. The results of this work show that the genotyping of camels for casein protein genes using the PCR-RFLP method can be successfully used to determine the potential of the studied animals for use in breeding work with an increase in the milk productivity of populations. The results of our research can be used to describe the genetic potential of the studied populations for the *CSN3*, *CSN2*, and *CSN1S1* genes of milk productivity, allowing us to make recommendations to camel farms for future breeding management within the populations. Furthermore, the obtained data can be used as a basis for comparative analysis with the data from subsequent studies conducted in this area. Further research should be carried out by way of a comparative analysis of the obtained genetic data with the phenotypic parameters of animals (milk yield, biochemical properties of milk, etc.), as well as a description of the studied genes using high-throughput sequencing methods, to obtain more extensive data on the relationship of the studied SNPs with the technological properties of milk.

Author Contributions: Conceptualization, M.A. and K.D.; methodology, M.A.; software, M.A.; validation, A.M., B.B., and N.S.; formal analysis, M.A. and K.D.; investigation, M.A.; resources, A.M. and B.B.; data curation, M.A.; writing—original draft preparation, M.A.; writing—review and editing, K.D.; visualization, A.M. and B.B.; supervision, A.M., B.B., and N.S.; project administration, M.A.; funding acquisition, B.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Education and Science of the Republic of Kazakhstan, grant number OR11465435 «Development and application of new genomic technologies for protecting organisms from mutagenic effects, increasing the productivity of natural resources and improving the quality of life of the population».

Institutional Review Board Statement: The animal study protocol was approved by the Ethics Committee of the Non-profit joint-stock company «Kazakh National Agrarian University» (14 September 2020).

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We express our sincere thanks to the heads of the camel farms of the Ili district of the Almaty region for their assistance in collecting research materials and providing data on the species of camels. We are most grateful to Vsevolodov E.B. for the consultation on preparing the manuscript of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dioli, M. Dromedary (*Camelus dromedarius*) and Bactrian camel (*Camelus bactrianus*) crossbreeding husbandry practices in Turkey and Kazakhstan: An in-depth review. *Pastor. Res. Policy Pract.* **2020**, *10*, 6. [CrossRef]
2. Ming, L.; Yuan, L.; Yi, L.; Ding, G.; Hasi, S.; Chen, G.; Jambl, T.; Hedayat-Evright, N.; Batmunkh, M.; Badmaevna, G.K.; et al. Whole-genome sequencing of 128 camels across Asia reveals origin and migration of domestic Bactrian camels. *Commun. Biol.* **2020**, *3*, 1. [CrossRef] [PubMed]
3. Almathen, F.; Charruau, P.; Mohandesan, E.; Mwacharo, J.M.; Orozco-Terwengel, P.; Pitt, D.; Abdussamad, A.M.; Uerpmann, M.; Uerpmann, H.-P.; De Cupere, B.; et al. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 6707–6712. [CrossRef] [PubMed]
4. Farah, Z. Composition and characteristics of camel milk. *J. Dairy Res.* **1993**, *60*, 603–626. [CrossRef] [PubMed]
5. Mal, G.; Sena, D.S.; Jain, V.K.; Sahani, M.S. Therapeutic Value of Camel Milk as a Nutritional Supplement for Multiple Drug Resistant (MDR) Tuberculosis Patients. *Isr. J. Vet. Med.* **2006**, *61*, 88–91.
6. Pauciullo, A.; Shuiep, E.; Cosenza, G.; Ramunno, L.; Erhardt, G. Molecular characterization and genetic variability at κ-casein gene (*CSN3*) in camels. *Gene* **2012**, *15*, 22–30. [CrossRef] [PubMed]
7. Shori, A.B. Camel milk as a potential therapy for controlling diabetes and its complications: A review of in vivo studies. *J. Food Drug Anal.* **2015**, *23*, 609–618. [CrossRef]
8. Rachagani, S.; Gupta, I.D. Bovine kappa-casein gene polymorphism and its association with milk production traits. *Genet. Mol. Biol.* **2008**, *31*, 893–897. [CrossRef]
9. Agrawal, R.P.; Beniwal, R.; Kochar, D.K.; Tuteja, F.C.; Ghorui, S.K.; Sahani, M.S.; Sharma, S. Camel milk as an adjunct to insulin therapy improves long-term glycemic control and reduction in doses of insulin in patients with type-1 diabetes: A 1 year randomized controlled trial. *Diabetes Res. Clin. Pract.* **2005**, *68*, 176–177. [CrossRef]
10. Beg, O.U.; von Bahr-Lindström, H.; Zaidi, Z.H.; Jörnvall, H. Characterization of a camel milk protein rich in proline identifies a new β-casein fragment. *Regul. Pept.* **1986**, *15*, 55–61. [CrossRef]
11. Shamsia, S.M. Nutritional and therapeutic properties of camel and human milks. *Int. J. Genet. Mol. Biol.* **2009**, *1*, 52–58.
12. Faye, B. The enthusiasm for camel production. *Editor. Emir. J. Food Agric.* **2018**, *30*, 249–250.
13. Nagy, P.; Fabri, Z.N.; Varga, L.; Reiczigel, J.; Juhasz, J. Effect of genetic and nongenetic factors on chemical composition of individual milk samples from dromedary camels (*Camelus dromedarius*) under intensive management. *J. Dairy Sci.* **2017**, *100*, 8680–8693. [CrossRef] [PubMed]
14. Burger, P.A.; Clani, E.; Faye, B. Old World camels in a modern world—A balancing act between conservation and genetic improvement. *Anim. Genet.* **2019**, *50*, 598–612. [CrossRef]
15. Hipp, N.J.; Groves, M.L.; Custer, J.N.; McMeekin, T.L. Separation of α-, β- and γ-casein. *J. Dairy Sci.* **1952**, *35*, 272–281. [CrossRef]
16. Alhaj, O.A.; Al-kanhal, H.A. Compositional, technological and nutritional aspects of dromedary camel milk. *Int. Dairy J.* **2010**, *20*, 811–821. [CrossRef]
17. Erhardt, G.; Shuiep, E.S.; Lisson, M.; Weimann, C.; Wang, Z.; Ibtisam, E.M.; Zubeir, E.I.; Pauciullo, A. Alpha S1-casein polymorphisms in camel (*Camelus dromedarius*) and descriptions of biological active peptides and allergenic epitopes. *Trop. Anim. Health Prod.* **2016**, *48*, 879–887. [CrossRef]
18. Ryskaliyeva, A.; Henry, C.; Miranda, G.; Faye, B.; Konuspayeva, G.; Martin, P. Combining different proteomic approaches to resolve complexity of the milk protein fraction of dromedary, Bactrian camels and hybrids, from different regions of Kazakhstan. *PLoS ONE* **2018**, *13*, e0197026. [CrossRef]
19. Singh, K.; Jayakumar, S.; Dixit, S.P.; Malik, S.Z. Molecular characterization and genetic variability of Alpha Casein gene, *CSN1S1* in Bikaneri camel (*Camelus dromedarius*) milk. *Indian J. Anim. Res.* **2018**, *53*, 67–70. [CrossRef]

20. Mati, A.; Senoussi-Ghezali, C.; Si Ahmed Zennia, S.; Almi-Sebbane, D.; El-Hatmi, H.; Girardet, J.M. Dromedary camel milk proteins, a source of peptides having biological activities—A review. *Int. Dairy J.* **2017**, *73*, 25–37. [CrossRef]
21. Pauciullo, A.; Ogah, D.M.; Iannaccone, M.; Erhardt, G.; Stasio, L.D.; Cosenza, G. Genetic characterization of the oxytocinneurophysin I gene (*OXT*) and its regulatory regions analysis in domestic Old and New World camelids. *PLoS ONE* **2018**, *13*, e0195407. [CrossRef] [PubMed]
22. Alim, N.; Fondrini, F.; Ionizzi, I.; Feligini, M.; Enne, G. Characterization of casein fractions from Algerian dromedary (*Camelus dromedarius*) milk. *Pak. J. Nutr.* **2005**, *4*, 112–116. [CrossRef]
23. Yelubayeva, M.E.; Buralkhiyev, B.A.; Tyshchenko, V.I.; Terletskiy, V.P.; Usenbekov, Y.S. Results of *Camelus dromedarius* and *Camelus bactrianus* Genotyping by Alpha-S1-Casein, Kappa-Casein Loci, and DNA Fingerprinting. *Cytol. Genet.* **2018**, *52*, 179–185. [CrossRef]
24. Shild, T.A.; Geldermann, H. Variants within the 5′-flanking regions of bovine milk-protein-encoding genes. III. Genes encoding the Ca-sensitive Caseins alpha-s1, alpha-s2 and beta. *Theor. Appl. Genet.* **1996**, *93*, 887–893. [CrossRef] [PubMed]
25. Jadhav, S.A.; Umrikar, U.D.; Sawane, M.P.; Pawar, V.D.; Deshmukh, R.S.; Dahiya, S.S.; Mehta, S.C. Genetic polymorphism at k-casein gene in Indian camel breeds (*Camelus dromedarius*). *J. Camel Pract. Res.* **2020**, *27*, 201–206. [CrossRef]
26. Pauciullo, A.; Giambra, I.; Iannuzzi, L.; Erhardt, G. The β-casein in camels: Molecular characterization of the *CSN2* gene, promoter analysis and genetic variability. *Gene* **2014**, *547*, 159–168. [CrossRef]
27. Amigo, L.; Recio, I.; Ramos, M. Genetic polymorphism of ovine milk proteins: Its influence on technological properties of milk—A review. *Int. Dairy J.* **2000**, *10*, 135–149. [CrossRef]
28. Soyudal, B.; Ardicli, S.; Samli, H.; Dincel, D.; Balci, F. Association of polymorphisms in the *CSN2*, *CSN3*, *LGB* and *LALBA* genes with milk production traits in Holstein cows raised in Turkey. *J. Hellenic Vet. Med. Soc.* **2018**, *69*, 1271–1282. [CrossRef]
29. Nowier, A.M.; Ramadan, S.I. Association of β-casein gene polymorphism with milk composition traits of Egyptian Maghrebi camels (*Camelus dromedarius*). *Arch. Anim. Breed.* **2020**, *63*, 493–500. [CrossRef]
30. Giambra, I.J.; Chianese, L.; Ferranti, P.; Erhardt, G. Short communication: Molecular genetic characterization of ovine αS1-casein allele H caused by alternative splicing. *J. Dairy Sci.* **2010**, *93*, 792–795. [CrossRef]
31. Ramunno, L.; Cosenza, G.; Rando, A.; Pauciullo, A.; Illario, R.; Gallo, D.; Di Berardino, D.; Masina, P. Comparative analysis of gene sequence of goat *CSN1S1* F and N alleles and characterization of *CSN1S1* transcript variants in mammary gland. *Gene* **2005**, *345*, 289–299. [CrossRef] [PubMed]
32. Kappeler, S.; Farah, Z.; Puhan, Z. Sequence analysis of *Camelus dromedarius* milk caseins. *J. Dairy Res.* **1998**, *65*, 209–222. [CrossRef] [PubMed]
33. Shuiep, E.; Giambra, I.J.; El Zubeir, I.E.M.; Erhardt, G. Biochemical and molecular characterization of polymorphisms of αs1-casein in Sudanese camel (*Camelus dromedarius*) milk. *Int. Dairy J.* **2013**, *28*, 88–93. [CrossRef]
34. Kim, J.; Yu, J.; Bag, J.; Bakovic, M.; Cant, J.P. Translation attenuation via 3′ terminal codon usage in bovine csn1s2 is responsible for the difference in αs2- and β-casein profile in milk. *RNA Biol.* **2015**, *12*, 354–367. [CrossRef]
35. Kharrati, H.K.; Esmailizadeh, A.K. SNPs genotyping technologies and their applications in farm animals breeding programs: Review. *Braz. Arch. Biol. Technol.* **2014**, *57*, 87–95. [CrossRef]
36. Peakall, R.; Smouse, P.E. GenAlex 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **2012**, *28*, 2537–2539. [CrossRef]
37. Yeh, F.C.; Yang, R.C.; Boyle, T. *POPGENE Version 1.32: Microsoft Window-Based Freeware for Population Genetics Analysis*; University of Alberta: Edmonton, AB, Canada, 1999.
38. Nei, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **1978**, *89*, 583–590. [CrossRef]
39. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [CrossRef]
40. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
41. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, NY, USA, 1987; pp. 176–187. [CrossRef]
42. Kimura, M.; Crow, F.J. The number of alleles that can be maintained in a finite population. *Genetics* **1964**, *49*, 725–738. [CrossRef]
43. Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3321–3323. [CrossRef] [PubMed]
44. Lewontin, R.C. The Apportionment of Human Diversity. *Evol. Biol.* **1972**, *6*, 381–398.
45. Smekenov, I.T.; Akishev, Z.D.; Altybayeva, N.A.; Mukhitdinov, N.M.; Bisenbayev, A.K. Evaluation of genetic polymorphism of the Berberis iliensis population in the Ili-Balkhash region based on ISSR markers. *Rep. Natl. Acad. Sci. Repub. Kazakhstan* **2012**, *4*, 49–57.
46. Reiner, A.P.; Gross, M.D.; Carlson, C.S.; Bielinski, S.J.; Lange, L.A.; Fornage, M.; Jenny, N.S.; Walston, J.; Tracy, R.P.; Williams, O.D.; et al. Common coding variants of the HNF1A gene are associated with multiple cardiovascular risk phenotypes in community-based samples of younger and older European American adults: The coronary artery risk development in young adults study and the cardiovascular health study. *Circ. Cardiovasc. Genet.* **2009**, *2*, 244–254. [CrossRef] [PubMed]
47. Othman, E.O.; Nowier, A.M.; Medhat, E.E. Genetic Variations in Two Casein Genes Among Maghrabi Camels Reared in Egypt. *Biosci. Biotechnol. Res. Asia* **2016**, *13*, 473–480. [CrossRef]

48. Giambra, I.J.; Brandt, H.; Erhardt, G. Milk protein variants are highly associated with milk performance traits in East Friesian Dairy and Lacaune sheep. *Small Rumin. Res.* **2014**, *121*, 382–394. [CrossRef]
49. Caroli, A.M.; Chessa, S.; Erhardt, G.J. Invited review: Milk protein genetic variation in cattle: Impact on animal breeding and human nutrition. *J. Dairy Sci.* **2009**, *92*, 5335–5352. [CrossRef]
50. Imamura, K. Camel Production in Kazakhstan. *J. Nagoya Gakuin Univ.* **2015**, *52*, 1–13.
51. Tanana, L.A.; Karaba, V.I.; Peshko, V.V. *Breeding of Farm Animals and the Basics of Selection: A Tutorial*; Republican Institute of Professional Education: Minsk, Belarus, 2017; pp. 236–237.