Received: 28 August 2025 | Revised: 29 September 2025 | Accepted: 18 October 2025

https://doi.org/10.56946/jzs.v3i2.809

Journal of Zoology and Systematics

Research Article

Evolutionary Dynamics of the Casein Gene Family in Goat and Sheep: In-Silico Analysis of Structural, Functional and Regulatory Signatures

Muhammad Abdul Manan[1†], Muhammad Hussain[2†], Ayesha Khan[1], Fariha Latif[3], Areeba Safdar[3], Wardah Hassan[4,*], Muhammad Farhan Khan[5,*]

[1]Department of Animal Science Department, Faculty of Animal and Agricultural Sciences, Universitas Diponegoro, Semarang, Indonesia.
[2]Department of Veterinary and Animal Sciences, University of Veterinary and Animal Sciences, Lahore, Punjab, Pakistan.
[3]Department of Zoology, Bahauddin Zakariya University, Multan, Pakistan.
[4]Department of Biological and Environmental Sciences, Emerson University, Multan, Pakistan.
[5]Department of Chemistry, Gomal University, Dera Ismail Khan 29050, Pakistan.

†These authors contributed equally to this work.
*Corresponding Author: Wardah Hassan (wardah.hassan@eum.edu.pk) and Muhammad Farhan Khan (farhankhanbgu@gmail.com)

Abstract

Caseins, the primary milk proteins in ruminants, critically influence dairy quality and functionality, yet their evolutionary and regulatory divergence between goats (*Capra hircus*) and sheep (*Ovis aries*) remains unresolved. Here, we performed the first comparative genomics analysis of the casein gene family (CSN1S1, CSN1S2, CSN2, CSN3) in these species, integrating phylogenetics, promoter profiling and structural biology. Phylogenetic reconstruction confirmed shared ancestry between goats and sheep, with CSN1S1 and CSN2 clustering into distinct clades (bootstrap >85%). Caseins exhibited intrinsic disorder (39-72% disordered regions) and hydrophilicity (GRAVY <0), yet goats showed enhanced thermal resilience (aliphatic index >100 for CSN2), a likely adaptation to heat stress. Promoter analysis revealed goats elevated CSN1S2 and CSN2 activity scores and reduced YY1 repressor sites compared to sheep, which is novel regulatory divergence potentially enhancing lactation efficiency. These findings provide a genomic blueprint for optimizing milk traits in small ruminants, offering actionable targets for precision breeding in tropical agricultural economics.

Keywords: Goat, sheep, milk, phylogenetic analysis, regulatory genes.

1. Introduction

Milk production is a cornerstone of global agriculture, with small ruminants like goats and sheep contributing over 20% of the world's dairy supply [1, 2]. Their milk is increasingly sought for its hypoallergenic properties, nutritional density, and industrial adaptability, as these traits are governed by caseins, a family of phosphorylated proteins critical to micelle formation, nutrient transport, and dairy processing efficiency [3, 4]. However, the genetic mechanisms underlying interspecies differences in milk composition remain unresolved, limiting advances in precision breeding for enhanced dairy yield and climate resilience [5-7].

Casein genes (CSN1S1, CSN1S2, CSN2, CSN3) exhibit remarkable evolutionary plasticity, with polymorphisms directly modulating milk protein content, micelle stability, and cheese yield [8, 9]. For instance, goat CSN1S1 variants correlate with cheese yield, while sheep CSN3 alleles affect curd texture [10-13]. Recent advances in genomics have highlighted casein gene diversity as a driver of milk adaptation [14]. CSN1S2 is a little closer to the presupposed structure and has been found to have less polymorphism than CSN1S1 [15]. It has a functional protein role in the binding of calcium, aiding in the stability of milk and enhancement of its processing characteristics [16]. In Goat, CSN1S1 exhibits high polymorphism, with null alleles reducing protein content, whereas sheep CSN1S1 shows conserved expression associated to high-protein milk [11, 17, 18]. Similarly, CSN2 variants affect micelle structure and digestibility, with potential health

implications for consumers [19 , 20].

Despite these insights, no study has systematically compared the structural, regulatory, and evolutionary divergence of the entire casein gene family between goats and sheep.

Such comparisons are critical to identify species-specific signatures that could guide marker-assisted selection. We hypothesized that lineage-specific differences in gene structure, promoter elements, and protein properties underpin divergent milk traits between these species. By an in-silico approach, integrating phylogenetic, structural, and regulatory data, we identified conserved and divergent features of casein genes in both species. Furthermore, we analyzed the genetic link variations to functional properties and proposed candidate genes for improving milk quality in dairy breeding programs.

2. Materials and methods

2.1. Sequence Retrieval and Dataset Curation

Casein gene family sequences (CSN1S1, CSN1S2, CSN2, CSN3) for goats and sheep were obtained from the NCBI genome database, which provided information on the protein, genome, and coding sequence (CDS) (Supplementary Table S1). The datasets have been carefully selected for use in subsequent investigations of gene architecture and function.

2.2 Phylogenetic and evolutionary analysis

The evolutionary history of the different species was inferred using the Maximum Likelihood approach based on models of nucleotide substitution, which was the JTT model [21]. Using 28 amino acid sequences, we recreated a phylogenetic tree with 1000 bootstrap values. In order to improve the precision of the study, the sequence gaps and errors were removed, and the phylogenetic tree was developed with the support of the MEGA7 tool [22].

2.3 Structural and functional characterization

Structural characterization of the casein gene family obtained by using the Tb tool [23]. To assess the physicochemical properties, the ProtParam tool was employed to calculate various parameters, including the instability index, amino acid count, aliphatic index, isoelectric point, molecular weight, and grand average hydropathy (GRAVY) [24].

2.4 Promoter and regulatory element analysis

In the context of promoter analysis, gene sequences were submitted to the Promoter 2.0 Prediction Server for the purpose of identifying potential transcription factor binding sites. Binding sites exhibiting a score exceeding 1.0 were identified as probable candidates, with subsequent analysis concentrating on sequences located within 1,000 bp upstream of these sites [25].

2.5 Secondary structure analysis

The secondary structure of the casein proteins was predicted using Phyre2, and the TM-align tool was employed to align the query protein structures with established protein structures to investigate 3D spatial similarities [26].

2.6 Recombination analysis

The GARD (Genetic Algorithm Recombination Detection) tool was employed to identify recombination breakpoints within the aligned sequences, evaluating segment-specific phylogenies utilizing the Akaike Information Criterion (AIC) derived from a maximum likelihood model [27].

2.7 Transcription factor binding sites

Furthermore, the TFBIND software was utilized to forecast transcription factor binding sites by employing the TRANSFAC weight matrix. The study focused on 100 base pairs upstream of highly predicted transcription sites, which included 4 transcription binding sites (GATA, TATA, STAT, and OCT-1) and one repressor site (YY1). The transcription factors in question are known to bind certain DNA sequences and thereby participate in an array of processes, including gene expression and growth regulation [28].

3. Results

3.1 Phylogenetic analysis and sequence alignment

The phylogenetic analysis delineated the evolutionary relationships among the casein genes using MEGA 11.0. In order to describe the evolutionary connections of the casein gene family across various species, such as goats, sheep, cattle, buffalo, pigs, and camels, the phylogenetic tree was built as shown in figure 1 and supplementary table S1.

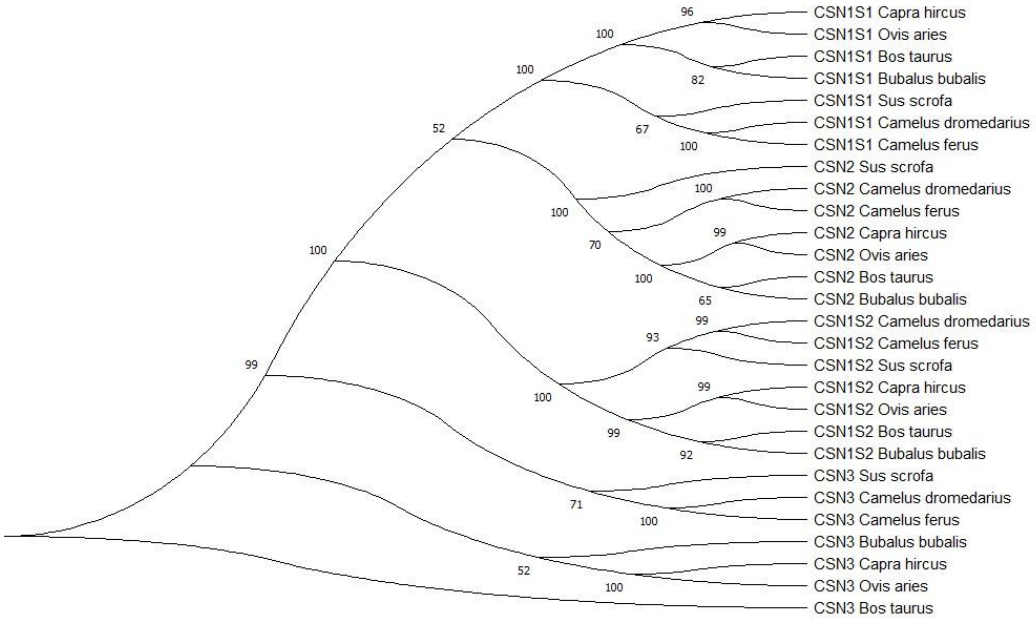

Figure 1. Phylogenetic analysis of the casein gene family in goat and sheep.

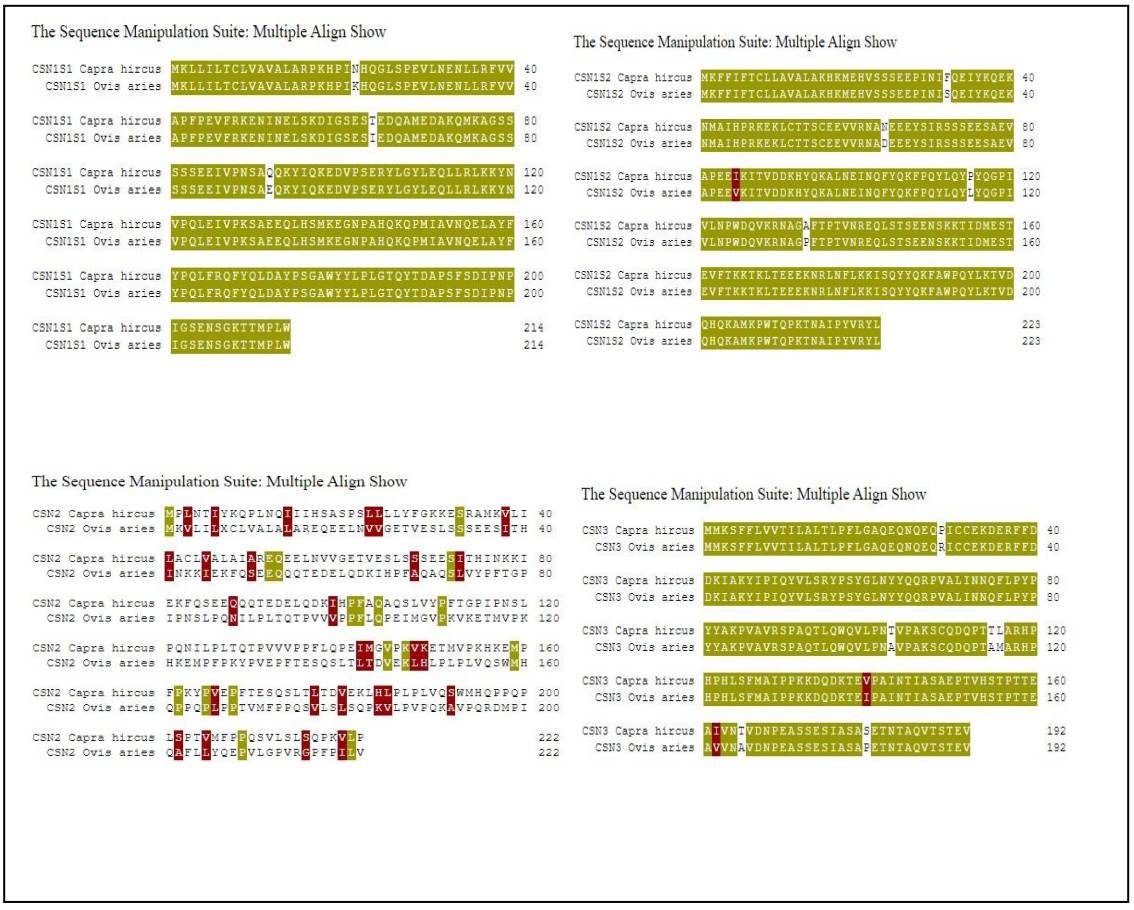

Figure 2. Multiple sequence alignment of the casein gene family in goat and sheep.

The phylogenetic tree demonstrated the casein gene evolution, clustering into 4 groups: CSNISI, CSNIS2, CSN2, and CSN3. The sequences of goats and sheep consistently cluster into all four represented 4 groups. The characteristics of casein genes or proteins in terms of evolutionary relationships and patterns that are related to the functionality of related genes. We used multiple sequence alignment (MSA). For this, we used 8 protein sequences (4 sequences of each species) of casein genes of two species, which are goat and sheep, to delve into the sequence features of the casein gene family in goat and sheep. The greenish tint represents the conserved regions of amino acids, indicating protein structure and function, while red and white tints indicate non-conserved amino acids (Figure 2 and supplementary table S2).

3.2 Motif and gene structure

Furthermore, to undertake the evolutionary links in the casein gene family, we performed motif, conserved domain, and gene structure analysis to analyze the structural configuration of the casein gene family in goat and sheep (Figure 3A and 3B). From both represented species 10 conserved motifs were identified with MEME-1. Using Pfams, it was confirmed that the identified motifs matched the casein gene domain. The conserved domains within the casein gene family were verified through comparison with NCBI CDD, ensuring the identification of structurally and functionally essential regions. Gene structure analysis further revealed variations in exon-intron arrangements and non-coding UTRs, which reflect evolutionary changes within the casein gene family.

3.3 Physicochemical Properties and Promoter Region prediction

Physiochemical characteristics of casein genomic family were evaluated on their molecular weight (MW in Da), amino acid (A.A), aliphatic index, isoelectric point (pI), instability index (II), and grand average of hydropathicity index (Gravy) as shown in (Table 1). Each variant of the casein gene family across both species described minor differences, such as MW ranged from 21441Da to 28860Da in goat, and in sheep it ranged from 21438Da to 26331Da. The pI values spanned from 5.30 to 8.24 in goat and 5.26 to 7.66 in sheep, respectively. The AI values were all above 66.91 in both species, which demonstrates the thermal stability of these proteins, with goat CSN2 representing the highest value, 101.91, signifying enhanced stability in extreme conditions. All casein proteins in both species had negative GRAVY values, which implied that these proteins have a hydrophilic nature. Despite their thermal stability, all the casein proteins in both species seemed to be unstable based on their instability index exceeding 40. A promoter region prediction analysis was also performed (Table S2).

3.4 Secondary structure and RMSD value analysis

Secondary structure analysis was performed across the genomes of the goat and sheep. Predictions were made for each casein gene identified in both the represented species. An online tool, Phyre2 was employed to compute the secondary structure characteristics (Table 2). The secondary structure of casein protein in goat consists of beta sheets ranging from 0%-13% and 0%-12%, alpha helices were observed from 10%-37% and 6%-38%, and disorder areas ranging from 39%-72% and 39%-63% in goat and sheep, respectively. Furthermore, in figure 4, the superimposition of casein protein was performed to compare and align the structure and sequences with a view to identifying similarities or differences. In this analysis, the TM score provides the topological similarity between query and casein proteins, while the RMSD value provides the average distance between alpha-carbon backbones of the two models. Additionally, the lower RMSD values predicted greater uniformity between models. Notably, Figure 5 also exhibited that the RMSD value and TM score of our query and superimposed models were 0.30 and 1. This result proved that the compared models were exactly the same.

3.5 Recombination analysis

Recombination analysis was employed by utilizing GARD and 11394 models at the potential breakpoint 970 was computed for the identification of recombination breakpoints, in which up to 4 inferred break points, markedly, the genetic algorithms calculated just 0.00% in goat (Supplementary Figure S1a). Recombination analysis was performed using the GARD and 11408 models to identify potential breakpoints, with a potential

Figure 3. (A) Casein Gene Characterization in Goat and (B) Casein Gene Characterization in Sheep

breakpoint at position 969. Up to 5 inferred break points were detecetd, withthe genetic algorithms calculating a value of 0.00% in sheep (Supplementary Figure S1b).

To calculate the frequency of obtaining a breakpoint at a particular position, we used the standardized Akaike weights of the models and observed breakpoint sites through the model-averaged and best-fit models; the analysis was approved. All casein gene nucleotide sequences via multiple sequence alignment, we utilized multiple break points methods facilitated by a genetic algorithm through which 4 in goat and 5 in sheep main recombination breakpoints were identified, along with further minor breakpoints at different positions (Figure 6a and 6b).

3.6 Analysis of transcription factor binding sites

The TF binding sites to a specific genomic position are the key point for transcriptional regulation in cells. On the basis of five transcription sites, which are TATA, YY1, GATA, OCT1, and STAT, we analyzed TF binding sites of the casein gene family in goat and sheep.

Table 1. Casein Genes Physicochemical Properties

No	Gene	Molecular Weight (Da)	Amino Acids (A.A.)	pI	Instability Index (II)	Aliphatic Index (AI)	GRAVY
Goat							
1	**CSN1S1**	24276.59	214	5.3	58.76	82.06	-0.504
2	**CSN1S2**	26362.99	223	8.24	53.6	66.91	-0.829
3	**CSN2**	28860.87	257	5.89	92.41	101.91	-0.095
4	**CSN3**	21441.32	192	5.53	46.72	79.27	-0.328
Sheep							
5	**CSN1S1**	24303.7	214	5.32	59.52	83.88	-0.482
6	**CSN1S2**	26331.93	223	7.66	54.16	67.76	-0.837
7	**CSN2**	24915.38	222	5.26	95.52	98.2	-0.137
8	**CSN3**	21438.38	192	5.78	48.44	78.8	-0.318

Table 2. Secondary structure analysis.

No	Gene	Disorder (%)	Alpha Helix (%)	Beta Strand (%)
Goat				
1	CSN1S1	57	37	0
2	CSN1S2	39	32	0
3	CSN2	72	8	6
4	CSN3	39	10	13
Sheep				
1	CSN1S1	56	38	0
2	CSN1S2	39	32	0
3	CSN2	63	6	2
4	CSN3	43	10	12

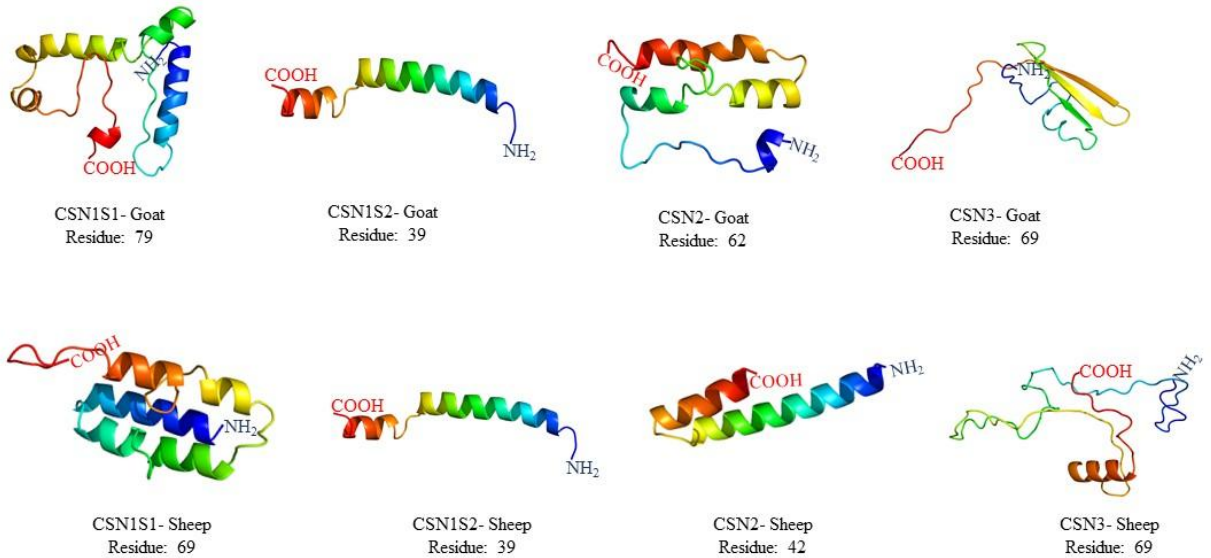

Figure 4. Secondary structure analysis in goat and sheep.

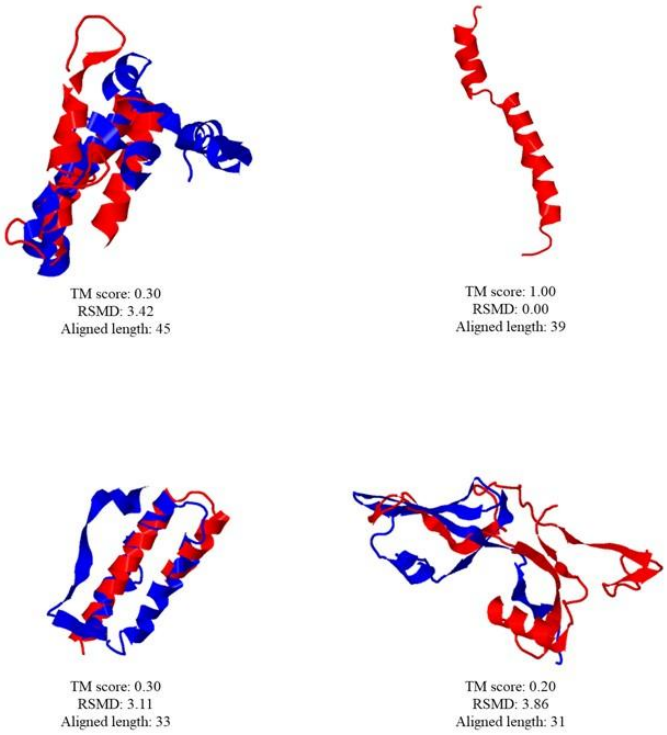

Figure 5. RMSD Value Prediction (The figure displays structural alignments of casein proteins across species, with ribbons showing the reference structure in red and aligned structures in blue. TM-score, RSMD, and aligned length are used to assess alignment quality, with higher TM-scores and lower RSMD values indicating better alignment accurac**y)**

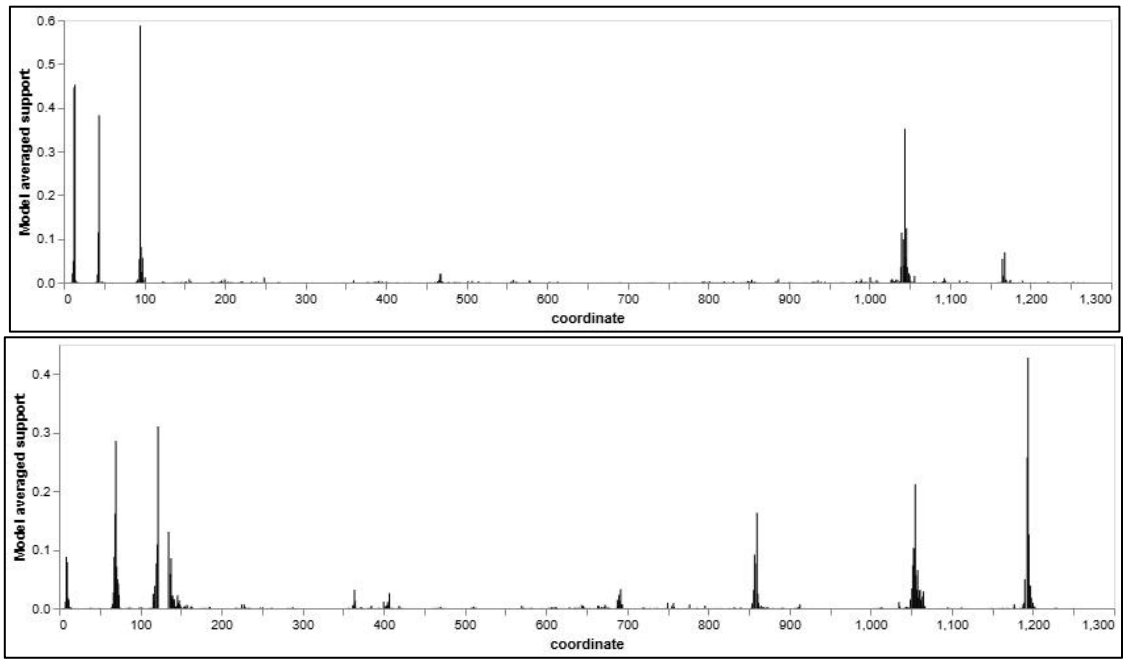

Figure 6. (A) Recombination analysis in goat (B) Recombination analysis in sheep

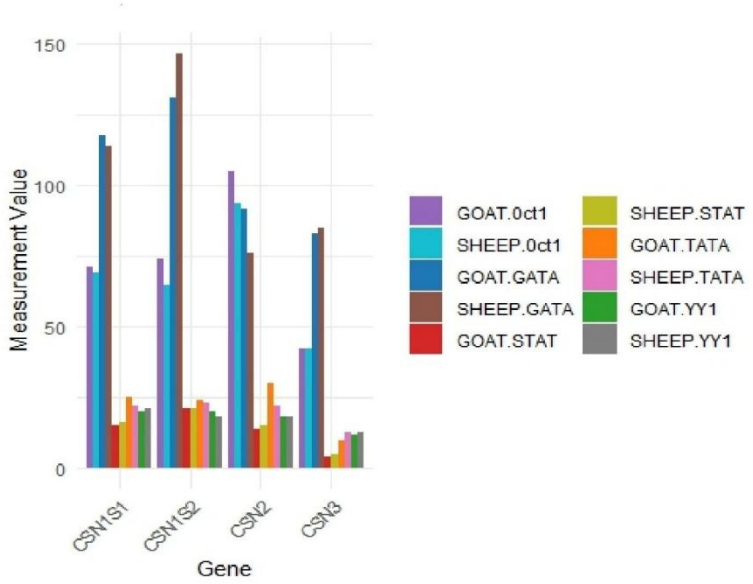

Figure 7. Transcription binding sites of casein gene family in goat and sheep

The pattern of TF binding sites obtained in both represented species were: GATA>YY1>OCT1>STAT>TATA. YY1 repressor binding sites were more in sheep compared to goat which might subsequently affect milk production (Figure 7).

4. Discussion

Technological advancements in genome sequencing, specifically in next-generation sequencing have been brought researchers to comprehend the sequenced genome across different species at the molecular level. The study on comparative genomics enables novel gene discovery and their functionality. Understanding of genetics and evolutionary processes is required to delve into the regulatory mechanisms of various significant genes, such as the casein gene family in mammals [29].

The casein gene family consists of CSNISI, CSN2, CSNIS2, and CSN3, which play an important role in mammals regarding their continuous evolution [30]. In goat and cattle, around 250 kb of casein genes have been identified on chromosome 6 [31]. Furthermore, these genes are diverse, transcriptomic, and possess various levels of protein at the genetic level [32]. In many species, it has been reported that these genes have genetic variations and polymorphisms, such as cattle, goat and sheep [33], goat and cattle have been reported with the highest genetic variability. The diverse variations have been related to varied gene expression and rate of protein biosynthesis [34]. Moreover, recent literature advocates that ther casein gene variation might be related to composition of milk and ratio [35].

On the basis of our findings of phylogenetic analysis, the casein genes in goat and sheep are closely related and provide many sequence similarities. Moreover, we identified a concise evolutionary link between the represented species and the attributes of casein genes, with respect to evolutionary connection and distribution is associated with functions of related genes. As per previous study reported, it was observed that a concise evolutionary connection with rat and mouse as well as karan fires cattle, are related to *Bos taurus*, *Bos indicus*, and buffalo [36]. Moreover, the overall phylogenetic associates revealed that *Bos mutus*, *Bos taurus,* and *Bos indicus* are closely linked with the casein gene family in buffalo [32]. Studies have shown that casein gene polymorphisms significantly influence milk composition, including protein and fat content, which are crucial for dairy production and processing. For instance, in Sarda goats, variations in the CSN1S1 and CSN2 genes have been associated with milk protein and fat content, while the CSN1S2 gene has been linked

to milk yield and other traits [37]. These associations suggest that casein gene divergence between goats and sheep may reflect adaptations to different dairy production systems and environmental conditions. Additionally, the high variability of casein genes in goats, particularly in breeds like the Sarda, indicates a strong selection pressure for milk quality traits. This variability is less pronounced in sheep, suggesting that goats have undergone more intense selection for specific milk characteristics [37].

Multiple sequence alignments are considered to very important tool for identification and analyzing phylogeny inference, function prediction, protein structure, and other related tasks. Recent advances in current systems have the capability to achieve maximum accuracy, capacity to gauge thousands of proteins, and scale and facility in contrasting proteins that do not directly share the same structure. Modern multiple alignment standard databases such as IRMBASE, SABMARK, PREFAB, and OXBENCH; although CLUSTALW is still considered to be a prevalent alignment tool as current methods provide improved alignment quality and low cost [38]. For this, we used 8 protein sequences (4 sequences of each species) of casein genes of two species, which are goat and sheep to delve into sequence features of the casein gene family in goat and sheep. The greenish tint represents the conserved regions of amino acids, indicating protein structure and function, while red and white tints indicate non-conserved amino acids (Figure 2). Likewise, studies have been reported with the same trend [39, 40]. Multiple sequence alignment (MSA) functions protein sequences into a rectangular arrangement, aiming to align residues in each column on the basis of homology (originating from the same location ancestral sequence), structurally superposable (aligning rigid local structure) or shared functional roles. While these criteria align closely for closely related proteins, structure, sequence, and evolutionary divergence, which can result in variations in alignment depending on the selected criteria [38].

Conserved motifs manifested by distinct patterns within the sequences of amino acids provide significant understanding regarding functional areas and evolutionary limitations determining the casein gene family in goat and sheep, with MEME-1 having the highest number of amino acids (16) among the identified motifs in the referenced analysis. The gene structure analysis showed variations in exon-intron arrangements and non-coding UTRs, revealing evolutionary changes within the casein genes. Previous study reported that in BMP1 contained additional conserved domain was found with inclusion of those belonging to the astacin superfamily, the TGF β-propeptide superfamily, EGF-3, and the EGF_CA superfamily [41, 42]. The astacin family metalloproteinase BMP1 stimulates TGF- to promote bone formation [43] . Likewise, a study reported on gene structure presented that blue whale TGF-β genes have different exon and intron patterns and different upstream and downstream untranslated regions. Transcription elongation and termination alter gene structure, resulting the addition and removal of protein domains which influence protein functionality [44]. Additionally, motifs of transcription factors and domains suggestively affected protein interaction and DNA binding [45]. Intron-rich genes existed in the ancestors of each eukaryotic supergroup, as depicted by the expansion of genomic data and refinement of more reliable models. The subsequent evolution of most eukaryotes was involved the removal of introns [46].

Understanding the functionalities and characteristics of proteins encoded by genes under study, their physical and chemical attributes are essential to understand across various species [47]. In the present study, we computed the essential parameters of the casein gene family across goat and sheep in which pI showed that all members possessed a basic nature. AI of globular proteins related to thermostability. The capacity of a protein engaged by aliphatic side chains like alanine, valine, leucine, and isoleucine is called AI and considered to be a positive characteristic related to the thermostability of the globular proteins [48]. The AI values in both species showed that all four casein genes were thermostable (**Table 1**). To analyze the interaction between water and protein, the GRAVY value is applied. With the sum of all amino acid hydropathy values and then by dividing it by the length of the protein, the

resultant is called GRAVY. GRAVY is used in the description and measurement of a protein's hydrophilic or hydrophobic in totality. A negative GRAVY value of a protein is considered to be hydrophilic while a high GRAVY value of a protein is related to hydrophobic [49, 50]. GRAVY with a negative value implied that the protein is highly insoluble. All four casein proteins showed hydrophilic characteristics with negative GRAVY scores. In our study, the hydrophilic feature of casein proteins is demonstrated by negative GRAVY scores, which may enhance their ability for protein binding and oligomerization [51, 52].

The promoter region of a gene, which is responsible for regulating the onset of transcription is considered to be a crucial step in gene regulation. The identification of promoter regions using in-silico methods, which is applicable in discovery of genes and comprehending the regulation of gene expression [53]. The results obtained through promoter prediction between sheep and goat are also significantly different, which can also be ascribed to the differential expression levels of casein genes. Goats had higher prediction scores for the CSN1S2 and CSN2 than the sheep, implying that these particular genes were transcribing more voyages, shell produce proteins, and probably superior goat quality than sheep quality. Similar studies have been reported on the promoter prediction analysis [54, 55, 56].

Over last few decades, computational studies has been trending in exploring the mechanism underlying secondary structure formation and have been fairly effective, in spite the fact that the prediction of native protein structure from sequence is still challenging. Substantial efforts have been consumed in that particular area, resulting a considerable progress regarding secondary prediction methods [57]. Specifically, the combination of optimally configured machine learning algorithms, including details of homologous sequence has resulted in an increased accuracy prediction up to 80% [58]. In our study, the Phyre tool was employed for secondary structure prediction. By comparing goat and sheep secondary structure of casein genes, different residues were found that may change protein structure and function. Previous literature has reported secondary structure for visualizing different genes and functions in different species [59 , 60].

Implication of phylogenetic and evolutionary processes can be compromised if recombination analysis is not present; therefore its role is very important in almost every comparative study. There is a possibility that through a single phylogenetic tree, recombinant sequences might be unable to understand up to the mark. On the other hand, several phylogenies can provide more details and deeper insight when they correct model the evolution of non-recombinant. Using recombination analysis, our study revealed that 5 major recombination breakpoints at different location have been identified. Similar findings have been reported in previous literature where recombination analysis was performed on CLCN genes in buffalo [61]. To search multiple sequence alignments for indication of recombination breakpoints and for identification of presumed recombinant sequences [62].

The binding of TF to particular locations in genomics is vital to the orchestration of transcription regulation across the study of cells. The relationships between TF genes and DNA are fundamental to transcriptional regulation, which is coordinated process, responsible to environmental aspects to attain temporal and tissue particularly [63, 64]. Thus, the facility to assess and identify TF binding sites across various genes is crucial to comprehend the details of gene regulation and deduce the regulatory network. According to our findings, the distribution pattern of TF binding sites within casein proteins of goat was; GATA>YY1>OCT1>STAT>TATA, and in sheep it was as follows: GATA>YY1>OCT1>STAT>TATA which is similar in both species but TF binding sites were greater in sheep than in goat. Similar studies reported trends of TF binding sites that are in line with the present study, in which YY1 repressor site and TF binding sites such as GATA, STAT, OCT1and TATA were present in represented genes of cattle and buffalo [65].

5. Conclusion

This study provides the first comparative genomic framework for the casein gene family in goats and sheep, resolving structural, regulatory and evolutionary mechanisms underlying

divergent milk traits. Tight clustering of the goat and sheep casein genes in phylogenetic clades confirms shared ancestry but masks lineage-specific innovations, such as goat-specific STAT5 promoter motifs linked to the elevated milk protein synthesis. Goats exhibited reduced YY1 repressor sites compared to sheep, a novel finding suggesting relaxed transcriptional repression that may enhance lactation efficiency. This aligns with goat's superior milk yield in arid tropical regions. These findings have substantial practical implications for optimizing milk production in small ruminants, which sustain over 500 million households in developing and under-developing countries. By linking promoter polymorphisms (CSN1S2, CSN2) and recombination breakpoints to milk traits, this work provides actionable targets for marker-assisted selection. Future studies should validate YY1 repression dynamics in mammary tissue and explore CRISPR editing to amplify favorable alleles in high-yielding breeds, potentially revolutionizing milk production strategies.

Author contribution

M.A.M conducted research and M.H wrote the original manuscript. A.K performed analysis. F.L and A.S wrote the methodology. M.F.K conceived the idea and revised the manuscript. W.H supervised the study and Project Administration. All authors have read and agreed with the published version of the manuscript.

Ethical approval

Not applicable

Conflicts of Interest

The authors report no conflicts of interest.

Acknowledgment

Not applicable

Data Availability statement

The data presented in this study are available on request from the corresponding author.

Funding

Not Applicable (N/A)

Supplementary File

The Supplementary Material for this article can be found online at:

https://www.jspae.com/index.php/jzs/article/view/809/327

REFERENCES

1. Kirgiafini, D., Kyrgiafini, M.-A., Gournaris, T., & Mamuris, Z. (2024). Understanding circular RNAs in health, welfare, and productive traits of cattle, goats, and sheep. Animals, 14(5), 733.
2. National Research Council. (2007). Nutrient requirements of small ruminants: Sheep, goats, cervids, and new world camelids. National Academy Press.
3. Boland, M., & Singh, H. (2019). Milk proteins: From expression to food. Academic Press.
4. Piazenski, I. N., Candelário, J. P. M., Soccol, V. T., de Souza Vandenberghe, L. P., de Melo Pereira, G. V., & Soccol, C. R. (2024). From lab to table: The path of recombinant milk proteins in transforming dairy production. Trends in Food Science & Technology, 104562.
5. Rahmatalla, S. A., Arends, D., & Brockmann, G. A. (2022). Genetic and protein variants of milk caseins in goats. Frontiers in Genetics, 13, 995349.
6. Martin, P., Cebo, C., & Miranda, G. (2013). Interspecies comparison of milk proteins: Quantitative variability and molecular diversity. In Advanced dairy chemistry: Volume 1A: Proteins: Basic aspects (4th ed., pp. 387–429).
7. Van Eenennaam, A. L., De Figueiredo Silva, F., Trott, J. F., & Zilberman, D. (2021). Genetic engineering of livestock: The opportunity cost of regulatory delay. Annual Review of Animal Biosciences, 9(1), 453–478.
8. Selvaggi, M., Laudadio, V., Dario, C., & Tufarelli, V. (2014). Investigating the genetic polymorphism of sheep milk proteins: A useful tool for dairy production. Journal of the Science of Food and Agriculture, 94(15), 3090–3099.

9. Salgado Pardo, J. I., Delgado Bermejo, J. V., González Ariza, A., León Jurado, J. M., Marín Navas, C., Iglesias Pastrana, C., & Navas González, F. J. (2022). Candidate genes and their expressions involved in the regulation of milk and meat production and quality in goats (Capra hircus). Animals, 12(8), 988. .
10. Calvo, J. H., Dervishi, E., Sarto, P., González-Calvo, L., Berzal-Herranz, B., Molino, F., ... & Joy, M. (2013). Structural and functional characterisation of the αS1-casein (CSN1S1) gene and association studies with milk traits in Assaf sheep breed. Livestock Science, 157(1), 1-8.
11. O'Callaghan, T. F. (2018). Effect of cow feeding system on the quality and constituents of raw milk, dairy products and the rumen.
12. Abbas, A. M., Issa, A. R., Tayib, N. N., & Jubrael, J. (2024). Kappa-casein gene (CSN3) polymorphisms detection in three indigenous Iraqi goat breeds using PCR-RFLP and SNP markers. Egyptian Journal of Veterinary Sciences, 55(3), 643–649.
13. ALKaisy, Q. H., Al-Saadi, J. S., Al-Rikabi, A. K. J., Altemimi, A. B., Hesarinejad, M. A., & Abedelmaksoud, T. G. (2023). Exploring the health benefits and functional properties of goat milk proteins. Food Science & Nutrition, 11(10), 5641–5656.
14. Borş, A., Borş, S.-I., & Floriștean, V.-C. (2024). Health-related outcomes and molecular methods for the characterization of A1 and A2 cow's milk: Review and update. Veterinary Sciences, 11(4), 172.
15. He, Q., Zhang, F., Zhang, X., Yao, W., Wu, J., Niu, H., & Luo, J. (2024). Chromosome-level dairy goat genome reveals the regulatory landscape of lactation. International Journal of Biological Macromolecules, 280, 135968.
16. Dettori, M. L., Pazzola, M., Noce, A., Landi, V., & Vacca, G. M. (2023). Variations in casein genes are associated with milk protein and fat contents in Sarda goats (Capra hircus), with an important role of CSN1S2 for milk yield. Animals, 14(1), 56.
17. Widodo, H. S., Murti, T. W., Agus, A., & Pertiwiningrum, A. (2023). Identification of CSN1S1 gene variations between dairy goat breeds and its influence on milk protein fractions in Indonesia. Advances in Animal and Veterinary Sciences, 11(12), 1936–1944.
18. Cosenza, G., Albarella, S., D'Anza, E., Iannuzzi, A., Selvaggi, M., Pugliano, M., ... & Peretti, V. (2023). A New AS-PCR Method to Detect CSN2 01 Allele, Genotyping at Ca-Sensitive Caseins Loci and Milk Traits Association Studies in Autochthonous Lazio Goats. Animals, 13(2), 239.
19. Jawasreh, K. I., & Al-Amareen, A. H. (2023). The implications of various gene variant combinations on breeding values for Awassi ewe milk production. Veterinary World, 16(12), 2411.
20. Parveen, S., Zhu, P., Shafique, L., Lan, H., Xu, D., Ashraf, S., & Liu, Q. (2023). Molecular characterization and phylogenetic analysis of casein gene family in Camelus ferus. Genes, 14(2), 256.
21. Mehra, V. K., & Kumar, S. (2023). Genetic profiling and in silico sequence analysis of CSN2 (β-casein) and CSN3 (κ-casein) genes in the Indian buffalo (Bubalus bubalis). bioRxiv, 2023–2025.
22. Mehra, V. K., Malakar, D., & Kumar, S. (2021). Cloning and characterization of α-S1-casein (CSN1S1) gene and protein in Indian buffalo (Bubalus bubalis). Indian Journal of Veterinary Sciences and Biotechnology, 17(4), 7–12.
23. Hassanin, A. A., et al. (2022). Phylogenetic comparative analysis: Chemical and biological features of caseins (alpha-S-1, alpha-S-2, beta, and kappa) in domestic dairy animals. Frontiers in Veterinary Science, 9, 952319.
24. Umarov, R., Kuwahara, H., Li, Y., Gao, X., & Solovyev, V. (2019). Promoter analysis and prediction in the human genome using sequence-based deep learning models. Bioinformatics, 35(16), 2730–2737.
25. Singh, R., et al. (2019). Comparative modeling and mutual docking of structurally uncharacterized heat shock protein 70 and heat shock factor-1 proteins in water buffalo. Veterinary World, 12(12), 2036.

26. Khan, M. F., Parveen, S., Sultana, M., Zhu, P., Xu, Y., Safdar, A., & Shafique, L. (2025). Evolution and comparative genomics of the transforming growth factor-β-related proteins in Nile tilapia. Molecular Biotechnology, 67(9), 3517-3531..
27. Zater, Z. Y., Merzoug, M., Baltaci, M. O., Todorov, S. D., Adiguzel, A., & Roudj, S. (2024). Impact of a novel caseinolytic protease single mutation on Lactiplantibacillus pentosus growth performance. Process Biochemistry, 145, 145–152.
28. Rehman, S. U., Feng, T., Wu, S., Luo, X., Lei, A., Luobu, B., ... & Liu, Q. (2021). Comparative genomics, evolutionary and gene regulatory regions analysis of casein gene family in Bubalus bubalis. Frontiers in genetics, 12, 662609
29. Pauciullo, A., Versace, C., Gaspa, G., Letaief, N., Bedhiaf-Romdhani, S., Fulgione, A., & Cosenza, G. (2023). Sequencing and characterization of αs2-casein gene (CSN1S2) in the Old-World camels have proven genetic variations useful for the understanding of species diversification. Animals, 13(17), 2805.
30. Luo, X., Zhou, Y., Zhang, B., Zhang, Y., Wang, X., Feng, T.,& Liu, Q. (2020). Understanding divergent domestication traits from the whole-genome sequencing of swamp-and river-buffalo populations. National Science Review, 7(3), 686-701.
31. Redwan, E. M., Korim, S., Samra, A., Saad, Y., Amhedar, H. A., & Uversky, V. N. (2018). Variability of some milk-associated genes and proteins in several breeds of Saudi Arabian camels. Protein Journal, 37, 333–352.
32. Ramunno, L., Cosenza, G., Rando, A., Illario, R., Gallo, D., Di Berardino, D., & Masina, P. (2004). The goat αs1-casein gene: gene structure and promoter analysis. Gene, 334, 105-111.
33. Madende, M., Kemp, G., Stoychev, S., & Osthoff, G. (2018). Characterisation of African elephant beta casein and its relevance to the chemistry of caseins and casein micelles. International Dairy Journal, 85, 112–120.
34. Zervas, G., & Tsiplakou, E. (2013). Goat milk. In Milk and dairy products in human nutrition: Production, composition and health (pp. 498–518).
35. Kumar, R., Gupta, I. D., Verma, A., Kumari, R., & Verma, N. (2017). Molecular characterization and SNP identification in HSPB6 gene in Karan Fries (Bos taurus × Bos indicus) cattle. Tropical Animal Health and Production, 49, 1059–1063.
36. Dettori, M. L., Pazzola, M., Noce, A., Landi, V., & Vacca, G. M. (2023). Variations in casein genes are associated with milk protein and fat contents in Sarda goats (Capra hircus), with an important role of CSN1S2 for milk yield. Animals, 14(1), 56.
37. Edgar, R. C., & Batzoglou, S. (2006). Multiple sequence alignment. Current Opinion in Structural Biology, 16(3), 368–373.
38. Lazăr, C., Gras, M. A., Pelmu, R. S., Rotar, C. M., & Popa, F. (2020). Review regarding the genomic evolution in sheep milk production and their application to improve the selection criteria. Emirates Journal of Food and Agriculture, 32, 691–701.
39. Scintu, M. F., & Piredda, G. (2007). Typicity and biodiversity of goat and sheep milk products. Small Ruminant Research, 68(1–2), 221–231.
40. Delolme, F., Anastasi, C., Alcaraz, L. B., Mendoza, V., Vadon-Le Goff, S., Talantikite, M., ... & Moali, C. (2015). Proteolytic control of TGF-β co-receptor activity by BMP-1/tolloid-like proteases revealed by quantitative iTRAQ proteomics. Cellular and Molecular Life Sciences, 72(5), 1009-1027.
41. Zhang, C., Wang, W., Wang, D., Hu, S., Zhang, Q., Wang, Z., & Cui, L. (2022). Genome-wide identification and characterization of the WRKY gene family in Scutellaria baicalensis Georgi under diverse abiotic stress. International journal of molecular sciences, 23(8), 4225.
42. Mao, L., Yano, M., Kawao, N., Tamura, Y., Okada, K., & Kaji, H. (2013). Role of matrix metalloproteinase-10 in the BMP-2 inducing osteoblastic differentiation. Endocrine Journal, 60(12), 1309–1319.

43. Song, N., et al. (2021). Genome-wide characterization and expression analysis of the SBP-box gene family in sweet orange (Citrus sinensis). International Journal of Molecular Sciences, 22(16), 8918.
44. Liu, M., Huang, L., Ma, Z., Sun, W., Wu, Q., Tang, Z., ... & Chen, H. (2019). Genome-wide identification, expression analysis and functional study of the GRAS gene family in Tartary buckwheat (Fagopyrum tataricum). BMC plant biology, 19(1), 342.
45. Rogozin, I. B., Carmel, L., Csuros, M., & Koonin, E. V. (2012). Origin and evolution of spliceosomal introns. Biology Direct, 7, 1–28.
46. Jasrapuria, S., Arakane, Y., Osman, G., Kramer, K. J., Beeman, R. W., & Muthukrishnan, S. (2010). Genes encoding proteins with peritrophin A-type chitin-binding domains in Tribolium castaneum are grouped into three distinct families. Insect Biochemistry and Molecular Biology, 40(3), 214–227.
47. Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. The Journal of Biochemistry, 88(6), 1895–1898.
48. Najafi, M., Mianji, G. R., & Pirsaraie, Z. A. (2014). Cloning and comparative analysis of gene structure in promoter site of alpha-s1 casein gene in Naeinian goat and sheep. Meta Gene, 2, 854–861.
49. Lanneau, D., Brunet, M., Frisan, E., Solary, E., Fontenay, M., & Garrido, C. (2008). Heat shock proteins: Essential proteins for apoptosis regulation. Journal of Cellular and Molecular Medicine, 12(3), 743–761.
50. Roy, S., Maheshwari, N., Chauhan, R., Sen, N. K., & Sharma, A. (2011). Structure prediction and functional characterization of secondary metabolite proteins of Ocimum. Bioinformation, 6(8), 315.
51. Kadel, S., Persico, M., Thibodeau, J., Laine, C., & Bazinet, L. (2019). Use of redundancy analysis and multivariate regression models to select the significant membrane properties affecting peptide migration during electrodialysis with filtration membranes. Separation and Purification Technology, 221, 114–125.
52. Narang, V., Sung, W.-K., & Mittal, A. (2005). Computational modeling of oligonucleotide positional densities for human promoter prediction. Artificial Intelligence in Medicine, 35(1–2), 107–119.
53. Yang, J., Wang, D. F., Huang, J. H., Zhu, Q. H., Luo, L. Y., Lu, R., ... & Li, M. H. (2024). Structural variant landscapes reveal convergent signatures of evolution in sheep and goats. Genome Biology, 25(1), 148.
54. Tumino, S., et al. (2023). Polymorphisms at candidate genes for fat content and fatty acids composition: Effects on sheep milk production and fatty acid profile using two dietary supplementations. Animals, 13(15), 2533.
55. Auzino, B. (2019). Expression and variability study of genes influencing quali-quantitative donkey milk traits.
56. Lee, S. M., Kim, H.-M., Moon, S. J., & Kang, M.-J. (2012). Cloning and molecular characterization of porcine β-casein gene (CNS2). Asian-Australasian Journal of Animal Sciences, 25(3), 421.
57. Pareek, K., Shah, S., & Arora, A. (2012). Casein protein: Phylogeny study of important domestic animals using computational method. International Journal of Advanced Research in Bioscience, 2, 326–337.
58. Su, J., Li, Z., Gao, P., Ahmed, I., Liu, Q., Li, R., ... & ur Rehman, S. (2023). Comparative evolutionary and molecular genetics based study of Buffalo lysozyme gene family to elucidate their antibacterial function. International Journal of Biological Macromolecules, 234, 123646.
59. Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., & Assmann, S. M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature, 505(7485), 696–700.
60. Fu, Y., Khan, M. F., Wang, Y., Parveen, S., Sultana, M., Liu, Q., & Shafique, L. (2024). In silico analysis: Molecular characterization and evolutionary study of CLCN gene family in buffalo. Genes, 15(9), 1163.
61. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). GARD: A genetic

algorithm for recombination detection. Bioinformatics, 22(24), 3096–3098.

62. Lemon, B., & Tjian, R. (2000). Orchestrated response: A symphony of transcription factors for gene control. Genes & Development, 14(20), 2551–2569.
63. Butler, J. E. F., & Kadonaga, J. T. (2002). The RNA polymerase II core promoter: A key component in the regulation of gene expression. Genes & Development, 16(20), 2583–2592.
64. Wilhelmi, S. S. (2023). Investigation of the regulatory processes of gene expression in animal and plant sciences.

How to cite this article: Manan MA, Hussain M, Khan A, Latif F, Safdar A, Hassan W, Khan MF, (2025). Evolutionary Dynamics of the Casein Gene Family in Goat and Sheep: In-Silico Analysis of Structural, Functional and Regulatory Signatures. Journal of Zoology and Systematics, 3(2), 3(2), pp. 71–85.