www.roavs.com EISSN: 2223-0343
RESEARCH OPINIONS IN ANIMAL & VETERINARY SCIENCES

Bioinformatics analysis of beta-casein gene in some selected mammalian species

Vincent S.T.[1*], Momoh O.M.[1] and Yakubu, A.[2]

[1]Department of Animal Breeding and Physiology, College of Animal Science, University of Agriculture, Makurdi, Benue State, Nigeria; [2]Department os Animal Science, Faculty of Agriculture, Nasarawa State University, Keffi, Shabu-Lafia Campus, Nasarawa State, Nigeria

Abstract

This study investigates in silico the genetic diversity of CSN2 on its evolutionary and differentiation within and among species and also examines the attendant effects of polymorphism on the functionality of CSN2. A total of 22 CNS2 gene sequences with corresponding amino acids belonging to 8 species [cattle (3), buffalo (3), camel (3), goat (3), horse (1), rabbit (3), rat (3) and mice (3) were retrieved from Genbank (www.ncbi.nlm.nih.gov). All sequences were trimmed to equal length (501bp) corresponding to the same region. Sequences alignment, translation and comparison were done with ClustalW using IUB substitution matrix, gap open penalty of 15 and gap extension penalty of 6.66. The alignment revealed high polymorphism of CNS2 sequences within and among species. The Dxy inferred using p-distance revealed a maximum value of 0.68 between horse and goat and a minimum value of 0.03 between cattle and buffalo. The hypothesis of strict neutral evolution was rejected for both genes as $P<0.05$ for species. Also allelic sequence evolution was entirely driven by positive selection. In silico functional analysis of non-synonymous mutations using PANTHER revealed that, for CSN2, 11 of the amino acid substitutions in the peptide binding region (cattle 5, goat 3 and sheep 1) did not impair protein function. However, 3 substitutions in cattle were predicted to be harmful to protein function, also 2 potentially deleterious SNPs; 1 (Glu36Lys) for cattle and 1 (Leu206Pro) for horse were found. The NJ phylogeny revealed trans-species evolution however, UPGMA tree topology was species-wise. In conclusion, all identified deleterious SNPs should be taken into account while selecting a stock for milk production.

Keywords: Polymorphism, Beta-casein, In silico, synonymous substitution, non-synonymous substitution

To cite this article: Vincent ST, Momoh OM and Yakubu A, 2014. Bioinformatics analysis of beta-casein gene in some selected mammalian species. Res. Opin. Anim. Vet. Sci., 4(10): 564-570.

Introduction

Milk proteins have been grouped into casein and whey. Caseins are acid precipitates at pH 4.6 and at 20^{o}C (Farrell et al., 2004) and consists of four autosomal (i.e. α1-casein (CSN1S1), α2-casein (CSN1S2), β-casein (CSN2) and K-casein (CSN3) genes with the relative proportion of 40, 10, 35 and 20 (Chin, 1999) that are responsible for the synthesis of α1-casein, α2-casein, β-casein and K-casein respectively (Farrell et al., 2004; Edward et al., 2008; Baker, 2011). In bovine, casein gene is located on chromosome 6 (Kucerove et al., 2006) and account for about 80% (Hoffman and Falvo, 2004). CNS2 is associated with stabilization of calcium phosphate in the body fluids. In their natural state, they associate with calcium citrate, phosphate and other inorganic ions in the formation of a colloidal particle which ranges from 20-300µm in diameter and is called micelle (Schmidt, 1982). The calcium phosphate in micelle serves as a cementing agent (Choi et al., 2011) while the micelle itself ensures high turbidity of skimmed milk.

Milk proteins exhibit genetic polymorphism at both protein and DNA levels which might either be due to

***Corresponding author:** Vincent Samuel Ter, Department of Animal Breeding and Physiology, College of Animal Science, University of Agriculture, Makurdi, Benue State, Nigeria
E-mail: Vincent.samuelter@gmail.com; Tell: (+234)7063350551

amino acid substitution or deletion of small peptides along the polypeptide chain (Chin, 1999). Depending on whether or not there is a generation of net charge, variants could either be termed electrophoretic (net charge) or silent variants. Electrophoretic variants are reflected on the amino acid profile however, not all changes are beneficial as some changes might have deleterious effects.

Milk protein genetic polymorphisms have evoked considerable research interest in recent years because of possible association between milk protein genotypes and economically important traits. A number of studies have indicated that milk production, composition and quality are affected by genetic variants of milk protein (Ng-Kwai-Hang and Grosclaude, 1992; Chin, 1999). For instance, it has been reported that β-casein A2 and A3 and K-casein are associated with higher milk yield when compared with other variants (Ng-Kwai-Hang et al., 1986). Therefore identification of milk protein variants provides an important tool for improving the yield and quality of milk and dairy production (Chin, 1999).

Furthermore, the statue of any given SNPs is a pointer to whether it is beneficial or not. The status of various non synonymous SNPs has been reported to be linked with economically important traits and disease development (Yakubu et al., 2013a), thus, it is important to identify and distinguished potentially beneficial production/reproduction linked SNPs from deleterious ones. Such beneficial SNPs hold great promise for indigenous breeds. Many *in silico* methods have provided valuable information on the effect of SNPs on protein structure and function. *In silico* analyses have been used to predict the functions of non synonymous SNP for Lf (Yakubu et al., 2013b), MHC (Yakubu et al. 2013c) genes.

In this present study the aim is to explore the genetic diversity of β-casein *in silico* especially on their evolution and differentiation within and among species and also examining the attendant effects of polymorphism on its functionality using MEGA 5.0 and PANTHER.

Materials and Methods

Sequences

A total of twenty two beta casein sequences from eight (8) species [cattle (3), buffalo (3), camel (3), goat (3), horse (1), rabbit (3), rat (3) and mice (3)] were retrieved from the GenBank (www.ncbi.nlm.nih.gov). The GenBank accession numbers of the sequences are: U47012, U47013, AJ973327 (Cattle); GQ259485, GQ259483, GQ176290 (Buffalo); AY311384, M90559, AH001195 (Goat); AM259943, AJ409279, AF108120 (Camel); AY579425 (Horse); X16484, GU734712, NM_009972 (Mice); M11178, NM_017120, J00711 (Rat) and NM_001082759, X15735, X13043 (Rabbit) respectively.

MSA and translation

Sequences alignment, translation and comparison were done with ClustalW as described by Larkin et al. (2007) using IUB substitution matrix, gap open penalty of 15 and gap extension penalty of 6.66.

Test of selection

The relative proportion of non-synonymous substitution per non synonymous site (dN) and the number of synonymous substitutions per synonymous site (dS) was estimated from the deduced amino sequences of beta casein using modified Nei-Gojobori (proportion) Pair wise deletion method. The ratio of non-synonymous to synonymous divergence (dN/dS) was also tested for departure from the neutral expectation of unity using the codon based Z-distribution as by Nei-Gojobori (MEGA 5.1), applying proportion correction. The hypothesis of strict neutrality is rejected for all sequences with $P<0.05$.

Functional analysis

In silico functional analysis of non-synonymous mutation was estimated using PANTHER (Thomas et al., 2003). PANTHER tools estimates the likelihood of a given non-synonymous (mutation) Coding SNP to initiate a functional change in protein. The subPSEC (substitution position-specific evolutionary conservation) score was also estimated based on alignment of evolutionarily related proteins. The probability that a given variant will cause a deleterious effect on protein function is estimated by $P_{deleterious}$, such that a subPSEC score of -3 corresponds to a $P_{deleterious}$ of 0.5 (Brunham et al., 2005). The subPSEC score is the negative logarithm of the probability ratio of the wild-type and mutant amino acids at a particular position. PANTHER subPSEC scores derived from the probabilities of observing the variant amino acids in a PANTHER Hidden Markov Model (HMM), are continuous values from 0 to -10. When subPSEC = 0, the substitution is interpreted as functionally neutral, whereas more negative values of subPSEC predict more deleterious substitutions (Brunham et al., 2005).

Phylogenetic analysis

Neighbor-Joining NJ trees were constructed each using P-distance model and pair wise deletion gap/missing data treatment. The construction was on the basis of genetic distances, depicting phylogenetic relationships among the beta lactoglobulin and beta casein nucleotide sequences of the investigated species. The reliability of the trees was calculated by bootstrap

confidence values (Felsenstein, 1985), with 1000 bootstrap iterations using MEGA 5.1 software (Tamura et al., 2011). Similarly, UPGMA trees for each gene was also constructed with consensus sequences; using same model as that of the NJ tree. All sequences were trimmed to equal length corresponding to same region before generating the tree.

Results

The variation in sequence length within and across species ranges between 624 and 4931. Two sequences each for rat and rabbit have 1114bp with slight variation in the third as shown above. Mice sequences reveal great variability in that one having 847bp and the others 1120bp and 13054bp.

Estimated distance matrix for β-casein between consensus sequences of 8 mammalian species. In ruminants, a maximum Dxy of 0.21 between camel and buffalo and minimum Dxy of 0.03 between cattle and buffalo were obtained. Amongst non-ruminants, maximum Dxy value of 0.68 and a minimum value of 0.21 were obtained for horse and camel and rat and mice respectively. Whereas on a generalized note, a maximum value of 0.68 was gotten for horse and goat and a minimum value of 0.03 was obtained for cattle and buffalo.

There are nine, three and one SNPs for cattle, goats and buffalo respectively. In cattle, only five (Wide type P, H, E, C and E) of the nine are beneficial as they do not impair protein function. Also all reported mutation at the peptide binding site for goat (wide type S, P and P and buffalo (wide type L) are also beneficial. After an exhaustive search, however, no mutation was found for camel as shown in Table 3. In horse, there are four amino SNPs out of which only one has deleterious effect, all others are beneficial. SNPs were not found for rabbit, rat and mice.

Mean numbers of nucleotide substitutions per synonymous site (dS) and per non-synonymous site (dN) with their ratio among the studied species is represented on Table 4. The comparison pattern of dS and dN substitution (dN>dS) reveals a balance selection for the polymorphism reported. In the entire sequence, the hypothesis of strict neutrality was rejected as all P-values were <0.05.

From the NJ tree, ruminants and non-ruminants were somewhat separated. However, rabbit (X15735) and Mice (X16484) tend to cluster with goat (AH001195) in a separate clade.

The UPGMA consensus tree shows the clustering of all ruminants separate from non-ruminants. Cattle and buffalo are in a single clade in the ruminant cluster whereas rat and mice are also in a single clade in the non-ruminant cluster.

Table 1: Beta casein sequence length variation within and among several species

Species	Number	Sequence length variation (bp)
Cattle	3	1720, 1721, 1723
Buffalo	3	1734, 1739, 1741
Goat	3	1184, 2009, 4931
Camel	3	624, 1799, 2555
Horse	1	2668
Rabbit	3	1114, 1114, 2157
Rat	3	1114, 1114, 1846
Mice	3	847, 1120, 13054

bp represents 'base pair'

Discussion

Length variation have been reported to affect the functionality of protein (Ebhardt et al., 2010), as evident with frameshift mutations. There are cases where variability might results from DNA duplication, DNA rearrangement, short tandem repeat (STR), insertions or deletion (indel) of sequences. In β-casein, the length variation observed in cattle (U47012 and U47013) might be as a result of indel as revealed by the graphics features information (Table 1). This frameshift mutation often results in a completely different translation from the original protein and is also likely to cause a stop codon to be read which truncates the further synthesis of protein (Williams and Eernegreen, 2013). It is also possible that the variability might results from regional variation; U47012 and U47013 are both of exons 1 while AJ973327 is of the promoter region. The variability in the length of camel sequences are clearly due to regional variability, as sequence AF108120 with 624bp is from intro 7; AM259943 with 1799bp is from promoter region while AJ409279 with 2555bp is from the flanking region. The variability between sequence and across species might initiate unique structures between individual members in conferring different biological activities (Bhattacharya et al., 2008).

The average genetic distance Dxy is an index of divergence between and among species, where Dxy=distance between sequence x and sequence y. The higher the value of Dxy the far apart the two species are, by implication, higher values have lesser ortholog and more paralog and vice versa. In the ruminant group, cattle are closer to buffalo than any other member of the group. The minimum value obtained between cattle and goat is consistent with the findings of earlier workers (Pareek and Arora, 2012). Similarly, the least value also obtained between rat and mice further support their both membership of the *rodentia* family, its shows that they share more orthologs than any other member of the non-ruminant class.

Several SNPs have been reported to serve as biomarkers for exploring the genetic bases of disease conditions and production traits (Tariq et al., 2013).

Table 2: Evolutionary divergence between species

	Cattle	Buffalo	Goat	Camel	Horse	Rat	Rabbit	Mice
Cattle		0.01	0.01	0.02	0.02	0.02	0.02	0.02
Buffalo	0.03		0.01	0.02	0.02	0.02	0.02	0.02
Goat	0.05	0.05		0.02	0.02	0.02	0.02	0.02
Camel	0.20	0.21	0.19		0.02	0.02	0.02	0.02
Horse	0.66	0.65	0.68	0.65		0.02	0.02	0.02
Rat	0.62	0.62	0.63	0.62	0.63		0.02	0.02
Rabbit	0.63	0.63	0.63	0.63	0.63	0.44		0.02
Mice	0.62	0.62	0.62	0.61	0.60	0.21	0.35	

NB: the upper diagonal represents standard error estimate(s) while the lower diagonal is the average genetic distances between species

GOAT AH001195	F	F	S	C	L	G	G	S	S	G	S	Y	L	H	F	P	F	L	M	Y	L	I	L	F	S	K	[175]
GOAT AY311384	S	L	.	E	H	S	I	L	.	S	.	S	I	S	P	D	L	F	V	V	.	P	S	Y	C	P	[175]
GOAT M90559	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[175]
CAMEL AF108120	.	.	Q	N	F	Y	Y	R	F	Y	.	C	R	N	.	Q	L	F	V	R	.	S	.	.	G	L	[175]
CAMEL AM259943	L	.	L	.	.	.	.	.	C	D	.	.	.	.	L	H	.	.	V	.	.	M	.	.	.	.	[175]
CAMEL AJ409279	L	Y	F	.	F	P	D	N	F	F	Y	Q	F	G	L	.	W	.	V	K	V	.	F	S	L	G	[175]
CATTLE U47012	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.	.	[175]
CATTLE U47013	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.	.	[175]
CATTLE AJ973327	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.	.	[175]
BUFFALO GQ259485	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[175]
BUFFALO GQ259483	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[175]
BUFFALO GQ176290	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[175]
HORSE AY579425	P	.	F	T	P	A	F	Y	I	Q	C	S	F	G	Q	A	L	.	G	P	W	.	P	L	.	P	[175]
MICE X16484	S	L	L	T	S	R	F	G	I	H	C	P	S	.	S	S	I	.	R	F	S	R	.	G	R	R	[175]
RAT M11178	S	.	.	I	Y	A	I	N	I	P	L	P	.	C	P	I	.	.	I	S	.	A	P	T	L	.	[175]
RAT NM 017120	S	.	.	I	Y	A	I	N	I	P	L	P	.	C	P	I	.	.	I	S	.	A	P	T	L	.	[175]
RAT J00711	S	.	.	I	Y	A	I	N	I	P	L	P	.	C	P	I	.	.	I	S	.	A	P	T	L	.	[175]
Rabbit NM 001082759	S	I	.	F	F	A	L	N	I	N	L	S	T	C	P	L	.	.	G	R	P	A	P	N	L	Q	[175]
RABBIT X13043	S	I	.	F	F	A	L	N	I	N	L	S	T	C	P	L	.	.	G	R	P	A	P	N	L	Q	[175]
RABBIT X15735	S	L	L	L	F	S	F	N	F	S	C	S	.	K	R	N	L	V	.	T	T	T	N	I	L	A	[175]
MICE GU734712	S	V	.	F	Y	A	V	N	I	P	L	P	.	C	P	V	.	.	I	S	P	A	P	T	L	.	[175]
MICE NM 009972	S	V	.	F	Y	A	V	N	I	P	L	P	.	C	P	V	.	.	I	S	P	A	P	T	L	.	[175]

Fig. 1: Amino acid prediction of mammalian β-casein, deduced by CLUSTALW excluding sites with missing/ ambiguous data and gaps. Dots indicated identical amino acids and numbers on the RHS represent site number

The prediction of SNPs status is promising in modern genetics analysis and breeding programmes but it is still a great challenge to identify functional SNPs in production or disease related gene. However, computational approach has helped in overcoming this challenge and this has increased the success rate of genetic association studies. Major interest in both human and animal genetics is to distinguish mutations that increase fitness from those that reduced fitness (Tariq et al., 2013). In this study, a total of 17 SNPs were sourced, 3 deleterious SNPs (Pro67His, Glu37Lys, and Pro152Leu) for cattle were predicted, also 2 potentially deleterious SNPs; 1 (Glu36Lys) for cattle and 1 (Leu206Pro) for horse were found. A single nucleotide polymorphism (SNP) from C**C**T to C**A**T leads to an amino acid change in the mature protein that is from Pro to His. Among the fifteen β-CN variants identified, A1 and A2 are the common types while others are very rare. A1 variant has His at position 67 of the amino acid sequence while A2 possess Pro at the same position. This single amino acid change causes the release of bioactive peptides upon gastro intestinal digestion. Morphine like opioid beta casomorphine-7 (BCM-7) released from A1 milk is reported to cause various illness like diabetes mellitus, heart diseases, atherosclerosis, schizophrenia and sudden infant death syndrome (SIDS) (Mohammed, 2011). The original variant A2 does not produce BCM-7 and thus is safe for human consumption. Major breeds such as Holstein Friesian and Ayrshire have a high frequency of A1 allele whereas breeds such as Guernsey and Jersey have more of A2 allele (Mohammed, 2011). It is also interesting to note that most of the Indian breeds of animals have only β-CN A2 allele. Selection for improvement must take this into account as selection for improve milk production may increase the frequency of harmful A1 allele in our bovine population. So efforts should be made to enhance the A2 allele with a view of transforming our cattle population capable of producing A2 variant of β-CN which has a global demand. A study by Farag et al. (2013) reported that, A1 allele was implicated in the mortality of embryo in

Table 3: Functional analysis of beta-casein

Mutated amino acid cattle	SubPSEC score	P deleterious	P Substituted	Wild type
Pro67His	-3.739	0.676	0.001	N/A
His106Gln	-	-	-	P
Ser122Arg	-	-	-	H
Glu37Lys	-3.608	0.647	0.025	N/A
Ser18Lys	-	-	-	E
Glu36Lys	-2.665	0.417	0.048	N/A
Arg25Cys	-	-	-	C
Pro152Leu	-4.742	0.851	0.008	N/A
Pro137Leu	-	-	-	E
Goat				
Ala177Val	-	-	-	S
Val207Asn	-	-	-	P
Ser166Tyr	-	-	-	P
Buffalo				
Lys61Asp	-	-	-	L
Horse				
Gln203Arg	-	-	-	A
Leu206Pro	-1.833	0.237	0.049	N/A
Phe210Leu	-	-	-	P
Pro219Ala	-	-	-	D

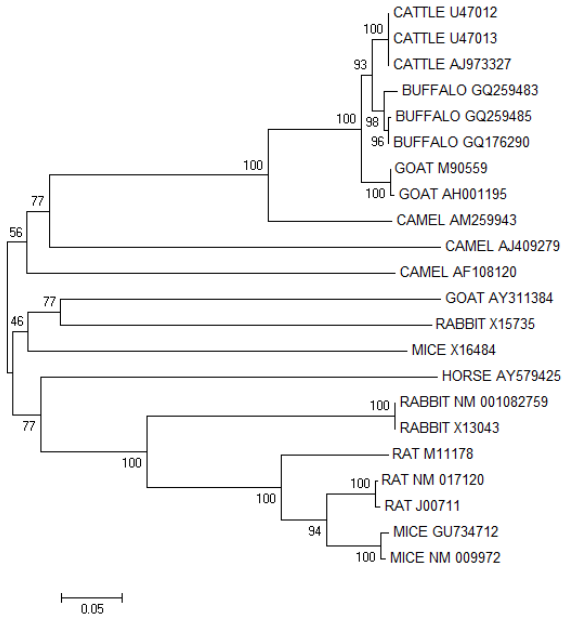

Fig. 2: Phylogenetic tree of mammalian beta-casein computed using Neighbor-Joining method with Bootstrap of 1000 replicate and P-distance Model of evolution.

vivo thus, lowering the litter size of A1A1 female zaraiby and Damascus female goats. The SNPs Glu37Lys is deleterious probably because of the opposing biochemical status. While Glu is an acid, Lys is a base therefore it would be all most impossible for such substitution to be productive.

The comparison of the number of non-synonymous substitution per non-synonymous sites (dN; amino acid altering) to the number of synonymous mutations per synonymous sites (dS; silent mutation) also known as omega (w = dN/dS); is a useful estimate of gene selective pressure (Yakubu et al., 2012). Zhang et al. (2005) pines that omega >1 implies positive selection i.e., selection has caused some amino acid substitution that are non deleterious and that the operative effect of purifying selection is not strong enough to overcome the effect of positive selection. In this study, within sequence and across species there was significant access of non-synonymous mutations over synonymous mutation. The estimated omega values which ranges from 1.046-1.3 for β-casein symbolizes that non-synonymous site evolved faster than the synonymous sites and positive selection effect over-shadows purifying selection except for horse as only a sequence cannot be used. High '*w*' signifies that balancing selection favored new variants and increased allelic polymorphism (Bergstrom and Gyllensten, 1995) which may in turn enhanced calcium phosphate transport in milk thereby providing suckling infant with a source of Ca and P for bone formation (Stewart et al., 1987) for β-casein.

Insight into the mechanism by which natural selection drives gene functional diversification across different species and lineages is a key issue in biology (Toll-Riera et al., 2011; Yakubu et al., 2013a). The varying substitutions of amino acids between and across species might be as a result of separate divergence from their common ancestor. According to Marini et al. (2010), as orthologs diverge from their most recent common ancestor, their different evolutionary trajectories leads to divergence in the selective constraints on homologous sites. Figure 1 depicts the deduced amino acid by CLUSTALW excluded sites with missing/ambiguous data and gaps. A greater level of polymorphism was shown within and across species in both genes.

From the phylogenetic NJ-tree, the trend of evolutionary relationship among β-casein extant was not completely species-wise. In figure 1, the clustering of goat (AY311384) and rabbit (X15735) and mice (Xi6484) indicated some degree of trans-species evolution, reflecting horizontal gene transfer among these species. However, on a general note, the clustered of most ruminants is distinctive of that of all non ruminants. Cattle and buffalo formed an embranchment confirming most recent shared ancestry as reported by Watanabe et al. (2000). Also the NJ-tree shows a clustering of rodents', aside mouse (X16484) that falls out, distinctive of horse and rabbit. Rats and mouse form a single clade, verifying a more recent ancestry. This is in line with classical classification which also group rat and mouse in the order *rodentia* family *miridae*. Also buffalo, cattle and goat formed a clade, because they are all members of the *bovidae* family. It

Table 4: Mean numbers of nucleotide substitutions per synonymous site (DS) and per non-synonymous site (DN) with their ratio in β-CN among the studied species

Species	No. of codons	Sd±(SE)	dN (±SE)	W	dN-dS	Z-Statatistics	P-value
Cattle	50	0.28±0.05	0.36±0.03	1.30	5.60	4.54	0.00
Buffalo	48	0.95±0.08	1.08±0.04	1.13	8.02	2.83	0.01
Goat	45	1.28±0.86	1.55±0.03	1.21	10.37	10.40	0.00
Camel	44	1.32±0.07	1.61±0.04	1.22	13.22	13.21	0.00
Rabbit	51	0.62±0.05	0.80±0.02	1.28	6.785	6.79	0.01
Rat	47	0.67±0.04	0.70±0.02	1.04	3.48	2.33	0.01
Mice	48	0.64±0.05	0.83±0.02	1.29	6.68	6.82	0.00

W = omega, dN = relative proportion of non-synonymous substitution per non-synonymous site, dS = the number of synonymous substitutions per synonymous site.

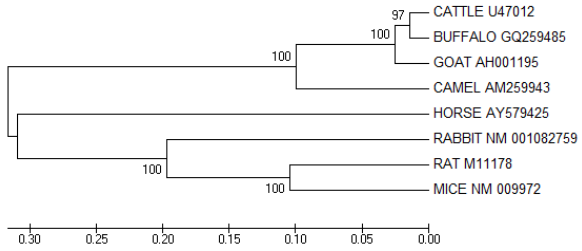

Fig. 3: Phylogenetic consensus tree of mammalian β-CN calculated via UPGMA method and P-distance evolutionary model.

is no surprised for them (*bovidae)* to have the closest relationship in the evolutionary pathway distinct of their closest relative camel which belongs to the camelidae family despite they are all members of the order artiodactyla. This finding is consistent with those of Pareek et al. (2012), Bhattacharya et al. (2008) and Watanabe et al. (2000). The species–wise clustering exhibited by some non-ruminants maybe explained by species-specific residues (Takahashi and Nei, 2000) and such sequence pattern maybe the result of gene conversion and balance selection. The UPGMA tree rules out trans-species evolution. There is a clear bifurcation of ruminants from non ruminants indicating that the phylogenetic relationship is in accordance with well-known evolutionary history of speciation (Fig. 3).

Conclusion

In silico analysis was performed to predict the functionality of CSN2 SNPs using *PANTHER*. 11 of the amino acid substitutions in the peptide binding region (cattle 5 goat 3 and sheep 1) did not impair protein function, however, 3 substitutions in cattle were predicted to be harmful to protein function, also 2 potentially deleterious SNPs; 1 (Glu36Lys) for cattle and 1 (Leu206Pro) for horse were found. Also the ratio of dN/dS indicates that the allelic sequence evolution is driven by positive selection for CSN2. The phylogeny revealed trans-species evolution however, UPGMA tree topology shows species relatedness that is consistent with well-established evolutionary history. In view of the above, all identified deleterious SNPs should be looked out for and not tolerated when implementing animal breeding programmes. Further analysis should prioritize these deleterious SNPs to obtain more detailed information on their effects.

References

Baker, K. 2011. Backborn dynamics of bovine BLg by ^{15}NMR spectroscopy. M.Sc. thesis, Institute of Fundamental Science and Molecular Bioscience. Massey University.

Bergstrom, T. and Gyllensten, U. 1995. Evolution of the MHC class II polymorphism: the rise and fall of class II gene function in primates. *Immunology Review,* 143: 14-31.

Bhattacharya, T., Kumar, P. and Sharma, A. 2008. Molecular cloning and characterization of CSN2 in Indian riverian buffalo (*Bubalis bubalis). Buffalo Bulletin,* 27: 222-230.

Brunham, L.R., Singaraja, R.R., Pape, T.D., Kejariwal, A., Thomas, P.D. and Hayden, M.R. 2005. Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene. *PLoS Genetics,* 1: e83.

Chin, D. 1999. Charaterisation of genetic variants of milk proteins that are not identified by electrophoresis. Ph.D thesis, Dept. of Animal Science, Montreal University, Quebec, Canada

Choi, J., Horne, D.S. and Lucey, J.A. 2011. Determination of molecular weight of a purified fraction of colloidal calcium phosphate derived from the casein micelles of bovine milk. *Journal of Dairy Science*, 94: 3260 - 3261

Ebhardt, A.H., Fednak, A. and Fahlman, R. 2010. Naturally occurring variation in sequence length creates miRNA isoforms that differ in argonaute effector complex specificity. *Silence,* 1: 12.

Edwards, P.J.B., Creamer, L.K. and Jamesion, G.B. 2008. Structure and stability of whey proteins. In: milk proteins from expression to food. Boslon: Academy Press. Pp: 163-203.

Farag, I.M., Abd El-Aziz, K.B., Ahmed, E.S., Abdelsalam, A.Z.E., Soliman, Kh. A., Darwish,

A.M.M. and Asmaa, A.S. 2013. Polymorphism of BMPR-1B, BMP-15 and β-CN genes in two Egyptian domestic goats. *Journal of Applied Science Research,* 9: 3199-3211.

Farrell, H.M., Jimenez-flores, J.R., Bleck, G.T. Brown, E.M., Butler, J.E., Creamer, L.K. Hick, C.M., Ng-Kwai-Hang, K.F. and Swaisgoog, H.E. 2004. Nomenclature of proteins of cows' milk-Sixth Revision. *Journal of Dairy Science,* 87: 1641-1674.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution,* 39: 783-791.

Hoffman, J.R. and Falvo, M.J. 2004. Protein-which is best? *Journal of Sport Science and Medicine,* 3: 118-130.

Kucerova, J., Matejicek, A., Jandurova, O.M., Sørensen, P., Nemcova, E. Stipkova, M., Kott, T., Bouska, J. and Frelich, J. 2006. Milk protein genes CSN1S1, CSN2, CSN3, β-Lg and their relation to genetic values of milk production parameters in Czech Fleckvieh. *Czech Journal of Animal Science*, 51: 241-247

Larkin, M.A., Blackshields, G., Brown, N.P, Chenna R, McGettigan, P.A, McWilliam H, Valentin F, Wallace, I.M, Wilm A, Lopez R, Thompson, J.D, Gibson, T.J. and Higgins, D.G. 2007. Clustal, W. and Clustal, X. version 2.0. *Bioinformatics*, 23: 2947-8.

Marini, N.J, Thomas, P.D. and Rine, J. 2010. The use of orthologous sequences to predict the impact of amino acid substitutions on protein function. PLoS Genet 6, e1000968.

Mohammed, E.M. 2011. B-CN A2 gene in Vechur cattle. PhD Thesis: Department of Animal Husbandry, Kerala.

Ng-Kwai-Hang, K.F. and Grosclaude, F. 1992. Genetic polymorphism of milk proteins. Ed. Elsevier Applied Science, New York NY. Advance Dairy Chemistry. 405-455.

Ng-Kwai-Hang, K.F., Hayes, J.F., Moxley, J.E. and Monardes, H.G. 1986. Relationship between milk protein polymorphism and major milk constituents in Hostein Freisian Cows. *Journal of Dairy Science,* 69: 22-26

Pareek, K., Shah, S. and Arora, A. 2012. Casein phylogeny: study of important domestic animal using computed method. *International Journal of Advance Research in Pharmacology and Biosciences,* 1: 326-337.

Schmidt, D.G. 1982. Association of caseins and casein micelle structure. In: Dev. In Dairy Chem., Vol. I. Proteins ed: P.F. Fox. Applied Science Publication Ltd. London. pp: 61-86.

Stewart, A.F., Bonsing, J., Beattie, C.W., Shah, F., Willis, I.M. and Mackinlay, A.G. 1987. Complete nucleotide sequences of bovine α_{s2} and β-casein cDNAs: comparisons with related sequences in other species. *Molecular Biology Evolution,* 4: 23-241.

Takahashi, K. and Nei, M. 2000. Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Molecular Biology and Evolution*, 17: 1251-1258.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology Evolution,* 28: 2731-2739.

Tariq, A.M., Al-Shammari, A.S., Al-Muammar, M.N. and Alhamdan, A.A. 2013. Evaluation and identification of damaged SNPs in COL1A1 gene involved in osteoporosis. *Archive Medical Science,* 9: 899-905.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A. and Narechania, A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Research,* 13: 2129-41.

Toll-Riera, M., Laurie, S. and Alba, M.M. 2011. Lineage-specific variation in intensity of natural selection in mammals. *Molecular Biology Evolution*, 28: 383-398.

Watanabe, M., Sugano, S., Togashi, T., Imai, J., Uchida, K., Yamaguchi, R. and Tateyama, T. 2000. Molecular cloning and phylogenetic analysis of canine β-casein DNA sequence. 11: 295-300

Williams, L.E. and Eernegreen, J.J. 2013. Sequence context of indel mutations and their effect on protein evolution in a bacterial endosymbiolnt. *Genomics Biology and Evolution,* 5(3): 599-605.

Yakubu, A., Salako, A.E., De Donata, M., Takeet, M.I., Peters, S.O., Adefenwa, M.A., Okpeku, M., Wheto, M., Agaviezor, B.O., Sanni, T.M., Ajayi, O.O., Onasanya, G.O., Ekundayo, O.J., Ilori, B.M., Amusan, S.A. and Imumorin, I.G. 2013c Genetic diversity in Exon 2 at the MHC DQB1 locus in Nigerian indigenous goats. *Biochemical Genetics*. 51: 954-966

Yakubu, A., Musa-Azara, I.S., Yakubu, B.N.S., Daikwo, S.I., Vincent, S.T., Momoh, O.M. and Dim, N.I. 2013a Biocomputational Genome-Wide Analysis of micro RNA genetic variability in some vertebrates. *Genetika,* 45: 799-810.

Yakubu, A., Faith, E.A., Peters, S.O., Takeet, M.I., De Donato, M and Imumorin, I.G. 2013b In silico molecular analysis of the evolution and differentiation of lactoferrin gene in some ruminants ans non-ruminants animals. *Proceeding of the 73th Annual Conference held in Lafia 20th-24th October* P 661.

Zhang, J., Nielsen, R. and Yang, Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology Evolution,* 22: 2472-2479.