Biotechnology Journal International

19(1): 1-8, 2017; Article no.BJI.33125
ISSN: 2456-7051
(Past name: British Biotechnology Journal, Past ISSN: 2231–2927, NLM ID: 101616695)

Computational Molecular Analysis of the Sequences of PAPPA2 Gene of Selected Ruminants and Non Ruminants

M. Omolara Akinyemi[1*], H. Osamede Osaiyuwu[1] and Ismail A. Adegoke[1]

[1]Animal Breeding and Genetics Unit, Department of Animal Science, Faculty of Agriculture and Forestry, University of Ibadan, Ibadan, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. Author MOA designed the study, performed the prediction analysis and wrote the final draft. Author HOO managed the analyses of the study and literature searches author IAA downloaded aligned the sequences and wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJI/2017/33125
Editor(s):
(1) Ghousia Begum, Toxicology Unit, Biology Division, Indian Institute of Chemical Technology, Hyderabad, India.
Reviewers:
(1) Wuyi Liu, Fuyang Normal University, China.
(2) Enrique Wulff, ICMAN-CSIC, Marine Sciences Institute of Andalusia (CSIC), Cádiz, Spain.
(3) Bibinu Bwaseh Saleh, Nigerian Institute of Animal Science, University of Agriculture Makurdi, Nigeria.
Complete Peer review History: http://www.sciencedomain.org/review-history/20267

Original Research Article

Received 31st March 2017
Accepted 2nd May 2017
Published 29th July 2017

ABSTRACT

Pregnancy associated plasma protein A2 (PAPPA2) is an Insulin-like growth factor binding protein (IGFBP) protease of the pappalysin family. This gene has been reported to be associated with prenatal growth, postnatal growth, skeletal growth, calving interval, milk yield, fertility and parity in cattle. The present study was undertaken to computationally investigate the attendant effects of the genetic variants of the PAPPA2 gene on its function and to gain insight into the evolutionary proximity and divergence in ruminants and non-ruminants at the studied locus. A total of fourteen (14) PAPPA2 nucleotide sequences comprising cattle (3), sheep (4), goat (1), pig (3), chicken (1) and horse (2) were retrieved from the GenBank. Functional analysis of non-synonymous single nucleotide polymorphism showed that eight amino acid substitutions (C31I, R60V, Y71G, S119L, S181Y, R190I, Q361M, P178_E180del) in goats, seven in sheep (V9E, L44D, T185N, A125W, Y78S, R194V, R240L), seven in cattle (L11F, C50_W51insTDAPK, E100L, A250T, G257L,

**Corresponding author: E-mail: larakinyemi@gmail.com;*

M850V), nine in chickens (A60G, S104_R170del, A190F, G128A, I10C, S309G, F48L, Q1630L) and eight in pigs (A61L, P72D, L11Q, K184T, D110C, S193_R194insTQD, Q481_E484del and T170_V172delinsVA) were returned neutral suggesting their beneficial effect. The phylogenetic trees from nucleotide sequences revealed the close relatedness of members of the *Bovidae* family (sheep, cattle and goat). The present information may be exploited in research into the association between PAPPA2 genotypes and some important economical traits in farm animals.

Keywords: PAPPA2; functional analysis; ruminants; non-ruminants; phylogenetic tree; polymorphism.

1. INTRODUCTION

Pregnancy associated plasma protein A2 (PAPPA 2) is one of the member of Pappalysin family of metzincin metalloproteinases, a protease of Insulin-like Growth Factor Binding Protein 5(IGFBP5) and Insulin-like Growth Factor Binding Protein 3(IGFBP3) [1,2] This gene is receiving increasing attention for its roles in pregnancy and post natal growth and has been reported to be similar to a related protein, PAPPA. Both genes have similar physiological function; while PAPPA2 is found at high levels in pregnancy complicated by pre-eclampsia, PAPPA is found at lower levels in various pregnancy complication which indicate that these two protein coding gene play different physiological role despite having similar biochemical actions. PAPPA2 and PAPPA share approximately 45% identity at the amino acid level and both have similar substrates [1], PAPPA2 cleaves IGFBP-5 while PAPPA cleaves IGFBP-4 [3]. PAPPA2 has been reported to be responsible for the bioavailability of insulin like growth factor (IGF I and II) in pregnancy serum [4,5] and candidate gene for QTL affecting growth.

Insulin-like growth factors (IGFs) play a pivotal role in growth and development. The bioavailability of the IGFs is regulated by six IGF binding proteins (IGFBPs), and the release of IGFs and subsequent IGF signaling is achieved through cleavage of the IGFBPs by proteases [6]. PAPPA2 has been identified as a candidate gene responsible for the effects of a quantitative trait locus (QTL) affecting adult body size in mice [7]. A number of recent studies have found associations between the bovine PAPPA2 gene and breeding interval [8,9] pregnancy rate [10], and age at first and second calving.

Since PAPPA 2 plays and controls important economic traits such as prenatal growth, postnatal growth, calving interval, milk yield, fertility and parity in cattle, it is important to examine the effect of mutations and its consequence on this gene in different livestock species. The method to identify functional SNPs from a pool, containing both functional and neutral SNPs by experimental protocols is challenging and expensive [11]. Therefore, computational predictions have become essential for evaluating the disease-related impact of nonsynonymous single-nucleotide variants discovered in exome sequencing [12]. According to [13] the prediction of SNPs status is promising in modern genetics analysis and breeding programmes as it can be used to identify animals with high breeding value. Several computational methods have been developed to predict the functional effect of a non-synonymous single-nucleotide polymorphism (nsSNP), a single-nucleotide change in a protein-coding region of a gene that causes an amino acid substitution (AAS) in the resulting protein [14]. Such methods have their roots in molecular evolution and are based on information derived from multiple sequence alignments available in specific databases. Most computational prediction tools for amino acid variants rely on the assumption that protein sequences observed among living organisms have survived natural selection. Thus, evolutionarily conserved amino acid positions across multiple species are likely to be functionally important, and substitutions observed at conserved positions may have deleterious effects [15,16].

The objective of the study was to investigate computationally. the evolution and differentiation of PAPPA 2 gene in selected mammalian species (goat, swine, sheep, cattle and avian) and the functional effect of mutations of the gene.

2. MATERIALS AND METHODS

2.1 Sequence of Species

A total of fourteen (14) PAPPA2 nucleotide sequences comprising cattle (3), sheep (4), pig (3), Chicken (1), goat (1) and horse (2) were retrieved from the Genbank

(https://www.ncbi.nlm.nih.gov/genbank/). the Genbank accession numbers of the sequence were XM015466918.1, XM005905526.1, XM010813486.2 (bovine); XM013989645.1, XM013989644.1, XM003130330.5 (porcine); XM012187595.1, XM012165660.2, XM012165659.2, XM004013823.3 (Ovine), XM005690911.3 (Caprine), XM015290339.1, (Avian), XM014845335.1, XM001498222.2 (Equine).

2.2 Sequence Alignment and Translation

Sequence alignment, translation and comparison were done with ClustalW [17] using IUB substitution matrix, gap open penalty of 15 and gap extension penalty of 6.66.

2.3 Estimation of Evolutionary Distance

Estimation of evolutionary distance of nucleotide sequence of (cattle, sheep, pig, chicken, goat and horse) were done using the pairwise method and p-distance model with a bootstrap of 1000 replicates The analysis involve 14 (fourteen) nucleotides sequences. Codon positions included were $1^{st}+2^{nd}+3^{rd}$ + noncoding. All positions containing gaps and missing data were eliminated. There were 5263 positions in the final dataset using MEGA 7 software [18].

2.4 Functional Analysis

Protein Variation Effect Analyser (PROVEAN) was used to predict the functional effect of protein sequence variations including single amino acids substitutions and small insertions and deletions [15]. The prediction based on the change caused by a given variation in the similarity of the query sequence to a set of its related protein sequence Variants with a PROVEAN score equal to or below -2.5 are considered "DELETERIOUS," also variants with a PROVEAN score above -2.5 are considered "NEUTRAL."

2.5 Phylogenetic Trees Analysis

Neighbour-joining tree on the basis of genetic distances, depicting phylogenetic relationships among PAPPA 2 nucleotide sequences of the investigated species was constructed using the complete deletion and p-distance options [19]. The reliability of the tree was estimated by bootstrap confidence values [20], with 1000 bootstrap replications. Similarly, a consensus sequence of each of the published cattle, sheep, swine, chicken and horse sequences was used to obtain a phylogenetic tree using the UPGMA method of MEGA 7.0 software [18].

3. RESULTS AND DISCUSSION

Fertility is an important trait which affects the efficiency of farm animal production and ensures food security. Genetic improvement of fertility traits in indigenous domestic animals will enhance productivity and food security especially in developing countries. The GH-IGF1 signaling pathway is an important mediator of the mechanism that regulates fertility. PAPPA2 gene is a major component of this pathway as it codes for a protease responsible for the increase in bioavailability of IGF1 for reproduction. Single nucleotide polymorphism (SNP) within the gene has been reported to be a predictor of postpartum fertility in cows [8]. The results obtained in this study indicate that PAPPA2 gene is highly polymorphic in all the species sampled. The average nucleotide substitution per site (D_{xy}) indicating evolutionary divergence between selected species nucleotide sequence from mammalian species (ruminant and non-ruminant) are presented in Table 1. The larger the D_{xy} value, the greater the genetic distance while the smaller the D_{xy} value the closer the genetic distance between the species [21].

Table 1. Average nucleotide substitutions per site (D_{xy})

Species	Cattle	Sheep	Goat	Horse	Pig	Chicken
Cattle		0.003	0.003	0.004	0.005	0.014
Sheep	0.040		0.001	0.005	0.005	0.014
Goat	0.039	0.010		0.005	0.005	0.014
Horse	0.122	0.127	0.126		0.005	0.013
Pig	0.104	0.106	0.106	0.099		0.013
Chicken	0.480	0.479	0.481	0.473	0.483	

Among the ruminant *Bos Ovis* and Capra, the highest D_{xy} of 0.040 was found between cattle and sheep while the value between sheep and goat was 0.010. In non-ruminants, the highest D_{xy} value (0.483) was found between *Gallus gallus* and *Sus scrofa*. This is further illustrated in the dendogram (Figs. 1 and 2) drawn from the consensus sequences of the PAPPA2 gene in selected ruminant and non-ruminant animals. The UPGMA consensus trees based on nucleotide sequences suggests that cattle, sheep and swine, horse were more closely related while chicken was farther apart. This is in accordance with classical classification sheep and cattle are members of the *Bovidae* family, and swine shares the order *Artiodactyla* with the ruminants. The level of genetic differentiation among the studied ruminant species may be due to the recent separation in evolutionary process among ruminants [22]. The findings of this study also agree with the submissions of [23] and [16] who observed similar clustering of members of the *Bovidae* family and order *Artiodactyla*.

The results of functional analysis of coding nonsynonymous single nucleotide polymorphism (nsSNP) of PAPPA2 gene for goats, sheep, cattle chicken and pig are presented in Tables 2-6.

Table 2. Functional analysis of coding nsSNP of the PAPPA2 gene of goats using PROVEAN

Variant	PROVEAN score	Prediction (cutoff= -2.5)
C3I	-0.388	Neutral
C50D	-2.778	Deleterious
R60V	-1.519	Neutral
Y71G	-1.520	Neutral
S119L	0.040	Neutral
S181Y	-0.384	Neutral
R421P	-6.151	Deleterious
R190I	-0.626	Neutral
Q361M	-0.813	Neutral
G610K	-6.350	Deleterious
P178_E180del	-0.072	Neutral
G611Q	-6.991	Deleterious
P961_V962insVA	-6.622	Deleterious
P726T	-2.907	Deleterious

C = Cysteine, I = Isoleucine, D = Aspartic acid, R = Arginine, V = Valine, Y = Tyrosine, G = Glycine, S = Serine, L = Leucine, P = Proline, Q = Glutamine, M = Methionine, K = Lysine, E = Glutamic acid, A = Alanine, T = Threonine

Table 3. Functional analysis of coding nsSNP of the PAPPA2 gene of chicken (Gallus) using PROVEAN

Variant	PROVEAN score	Prediction (Cutoff= -2.5)
A60G	0.286	Neutral
S104_R107del	-0.474	Neutral
A190F	-0.213	Neutral
G128A	0.260	Neutral
I10C	-0.170	Neutral
S309G	-0.885	Neutral
F48L	0.148	Neutral
P851_I852insQRV	-9.768	Deleterious
G541Q	-3.012	Deleterious
G430A	-5.615	Deleterious
D969L	-3.575	Deleterious
V1268_F1270del	-14.454	Deleterious
Q1630L	-2.184	Neutral
A1750Y	-3.720	Deleterious
R241_G244del	-1.200	Neutral

A = Alanine, G = Glycine, S = Serine, R = Arginine, F = Phenylalanine, I = Isoleucine, C = Cysteine, L = Leucine, P = Proline, Q = Glutamine, V = Valine, D = Aspartic acid I = Isoleucine, Y = Tyrosine, R = Arginine

Table 4. Functional analysis of coding nsSNP of the PAPPA2 gene of cattle using PROVEAN

Variant	PROVEAN score	Prediction (Cutoff= -2.5)
L11F	-0.725	Neutral
C50_W51insTDAPK	-1.547	Neutral
E100L	-0.716	Neutral
A250T	-0.824	Neutral
P301_A302del	-13.334	Deleterious
H421T	-3.102	Deleterious
G257L	-0.763	Neutral
V379G	-5.838	Deleterious
L537_N538insWGDY	-10.626	Deleterious
H601L	-8.413	Deleterious
M850V	-0.927	Neutral
Q961_D963del	-8.268	Deleterious
I1690P	-3.483	Deleterious
V1021L	-0.185	Neutral
C1150T	-8.560	Deleterious

L = Leucine, F = Phenylalanine, C = Cysteine, W = Tryptophan, T = Threonine, D = Aspartic acid, A = Alanine P = Proline, K = Lysine, H = Histidine, G = Glycine, V = Valine, N = Asparagine, Y = Tyrosine, M = Methionine, Q = Glutamine, I = Isoleucine

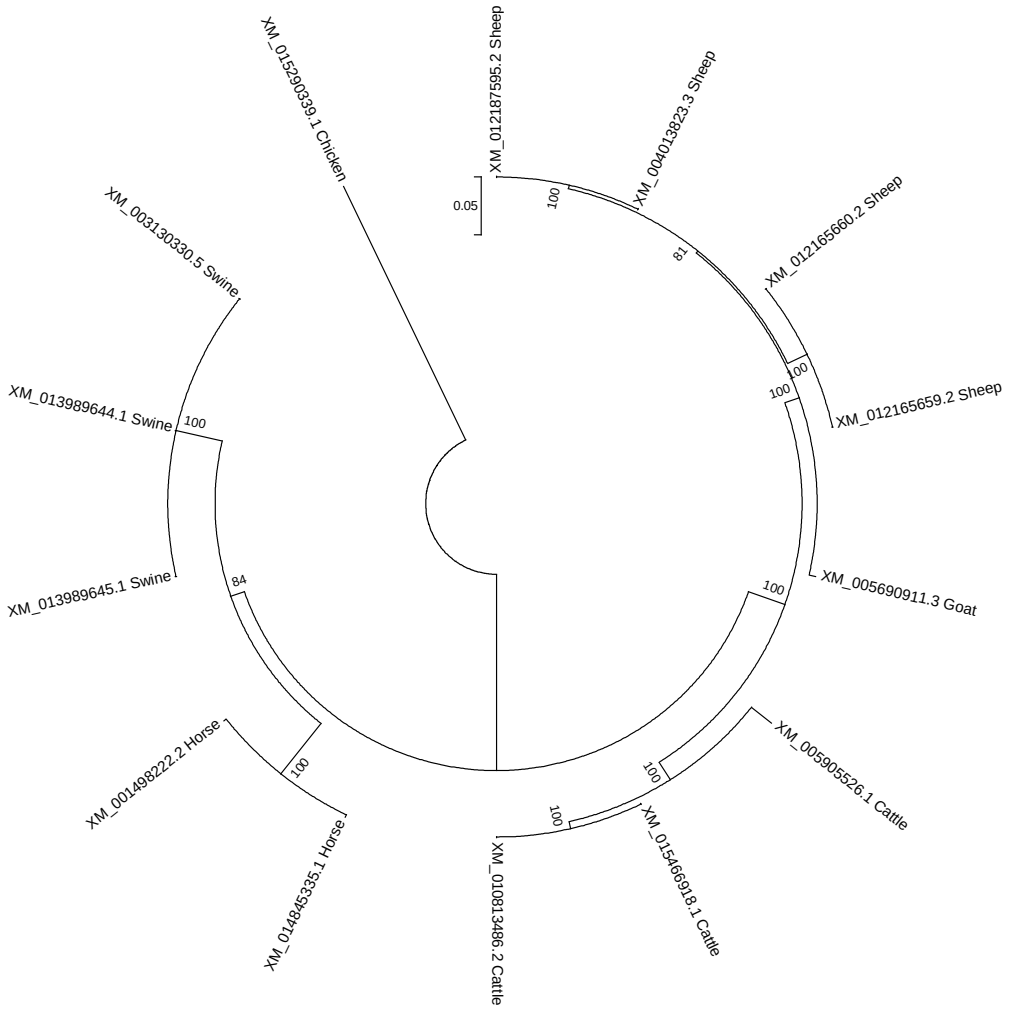

Fig. 1. Neighbour-joining tree obtained from PAPPA2 gene in some ruminants and non-ruminants using the neighbor joining method

Fourteen (14) amino acid substitutions were obtained in the alignment of deduced amino acid sequences of goats. Out of these, eight (8) amino acid substitutions (C3I, R60V, Y71G, S119L, S181Y, R190I, Q361M, P178_E180del) were returned neutral an indication that the amino acids substitutions did not cause any damage to the protein function However the remaining six (6) amino acids substitutions (C50D, R421P, G610K, G611Q, P961_V962insVA, P720T) were returned as deleterious indicating that such substitutions or mutations may be harmful. In Chicken (*Gallus gallus*) fifteen (15) amino acids substitutions (A60G, S104_R107del, A190F, G128A, I10C, S309G, F48L, Q1630L and R241_G244del) were returned neutral while the remaining amino acids substitutions (P851_I852insQRV, G541Q, G430A, D969L, V1268_F1270del and A1750Y) were deleterious. Out of the fifteen (15) amino acid substitutions in cattle, seven (L11F, C50_W51insTDAPK, E100L, A250T, G257L, M850V and V1021L) were neutral while the remaining eight (8) (P301_A302del, H421T, V379G, L537_N538insWGDY, H601L, Q961_D963del, I1690P and C1150T) were deleterious. For sheep, of the fifteen (15) amino acid substitutions seven (7) were deleterious: (V9E, L44D, T185N, A125W, Y78S, R194V and R240L). In pigs of the ten amino acid substitution predictions eight were returned as neutral suggesting that these did not impair protein function while the remaining two were harmful. The deleterious or harmful amino acid

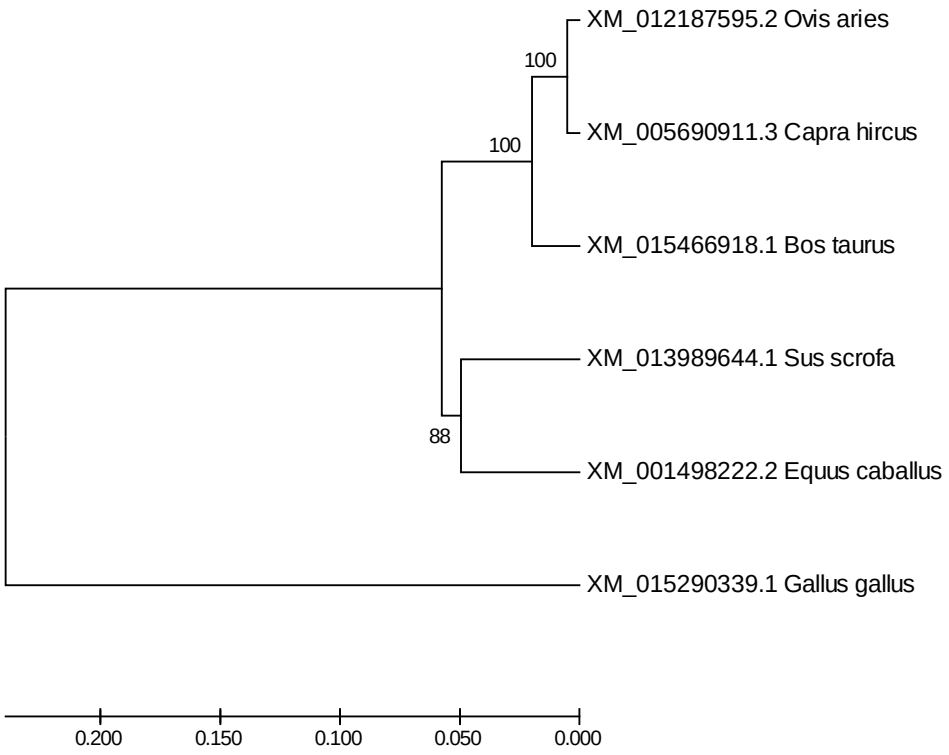

Fig. 2. Phylogenetic tree derived from selected sequences of PAPPA2 gene in some ruminants and non-ruminants species using the UPGMA method

Table 5. Functional analysis of coding nsSNP of the PAPPA2 gene of sheep using PROVEAN

Variant	PROVEAN score	Prediction (Cutoff= -2.5)
V9E	-1.121	Neutral
L44D	-0.237	Neutral
T185N	-0.197	Neutral
A125W	-0.393	Neutral
Y78S	-0.829	Neutral
L273_R276del	-10.583	Deleterious
N301_S302insTLRYC	-10.075	Deleterious
W430G	-10.333	Deleterious
Y372R	-7.669	Deleterious
R194V	-0.753	Neutral
R240L	-0.257	Neutral
H734_E735del	-20.546	Deleterious
L1000_D1001insYKGH	-10.144	Deleterious
T968N	-2.626	Deleterious
W910S	-10.885	Deleterious
C755A	-7.675	Deleterious

V = Valine, E = Glutamic acid, L = Leucine, D = Aspartic acid, T = Threonine,, N = Asparagine, A = Alanine, W = Tryptophan , Y = Tyrosine, S = Serine, R = Arginine, C = Cysteine, G = Glycine, H = Histidine, K = Lysine

substitutions in the selected species likely caused an alteration or impairment of protein functions which may modify enzyme exertion, impair protein structure or interaction and may lead to low birth weight occasioned by intra uterine growth restriction (IUGR). This study reveals high genetic variation at the PAPPA2 locus in cattle, sheep, goat, pig and chicken. The beneficial SNPs identified can aid in the selection and

genetic improvement of Nigerian livestock species.

Table 6. Functional analysis of coding nsSNP of the PAPPA2 gene of pig using PROVEAN

Variant	PROVEAN score	Prediction (Cutoff= -2.5)
A61L	-0.654	Neutral
P72D	-1.052	Neutral
L11Q	-1.293	Neutral
K184T	-0.022	Neutral
D110C	0.146	Neutral
S193_R194insTQD	-1.857	Neutral
A302_I303insPLV	-7.905	Deleterious
G113_T314del	-166.713	Deleterious
Q481_E484del	-0.033	Neutral
T170_V172delinsVA	-0.253	Neutral

M = Methionine, A = Alanine, L = Leucine, S = Serine, E = Glutamic acid, F = Phenylalanine, C = Cysteine, T = Threonine, V = Valine, P = Proline, , G = Glycine, D = Aspartic acid., Q= Glutamine, K = Lysine, R = Arginine, P= Proline

4. CONCLUSION

The result obtained in the present study offers hope for future genetic improvement of sheep, cattle, and chicken at the PAPPA2 locus in search of association with economically important traits such as birth weight, fertility, milk yield and disease resistance.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Overgaard MT, Boldt HB, Laursen LS, Sottrup- Jeensen L, Conover CA, Oxvig C. Pregnancy-associated plasma protein-A2 (PAPP-A2), a novel insulin-like growth factor- binding protein-5 proteinase. Journal of Biological Chemistry. 2001; 276:21849-21853.
2. Wagner PK, Christians JK. Altered placental expression of PAPPA2 does not affect birth weight in mice Reproductive Biology and Endocrinology. 2010;8:90.
3. Boldt HB, Conover CA. Pregnancy-associated plasma protein-A (PAPPA): A local regulator of IGF bioavailability through cleavage of IGFBPs. Growth Horm IGF Res. 2007;17:10-18.
4. Page NM, Butlin DJ, Lomthaisong K, Lowry PJ. The characterization of pregnancy associated plasma protein-E and the identification of an alternative splice variant. Placenta. 2001;22:681-687.
5. Yan X, Baxter RC, Firth S. Involvement of pregnancy-associated plasma protein-A2 in insulin-like growth factor (IGF) binding protein-5 proteolysis during pregnancy: a potential mechanism for increasing IGF bioavailability. J. Clin. Endocrinol. Metab. 2010;95:1412-1420.
6. Bunn RC, Fowlkes JL. Insulin-like growth factor binding protein proteolysis. Trends Endocrinology Metabolism. 2003;14:176–181.
7. Christians JK, Hoeflich A, Keightley PD. PAPPA2, an enzyme that cleaves an insulin- like growth-factor-binding protein, is a candidate gene for a quantitative trait locus affecting body size in mice. Genetics. 2006;173:1547–1553.
8. Luna-Nevarez P, Rincon G, Medrano JF, Riley DG, Chase Jr CC, Coleman SW. Single nucleotide polymorphisms in the growth hormone-insulin-like growth factor axis in straightbred and crossbred angus, brahman, and romosinuano heifers: Population genetic analyses and association of genotypes with reproductive phenotypes. Journal of Animal Science. 2011;89:926–934.
9. Hawken RJ, Zhang YD, Fortes M, Collis E, Barris WC, Corbet NJ. Genome-wide association studies of female reproduction in tropically adapted beef cattle. Journal of Animal Science. 2012;90:1398–1410.
10. Wickramasinghe S, Rincon G, Medrano JF. Variants in the pregnancy associated plasma protein-A2 gene on Bos Taurus Autosome 16 are associated with daughter calving ease and productive life in holstein cattle. Journal of Dairy Science. 2011;94: 1552–1558.
11. George PDC, Rajasekaran R, Sudandiradoss C, Ramanathan K, Purohit R, Sethumadhavan R. A novel computational and structural analysis of nsSNPs in CFTR gene. Genomic Med. 2008;2:23–32.
12. Liu L, Kumar S. Evolutionary balancing is critical for correctly forecasting disease associated amino acid variants. Mol Biol Evol. 2013;30:1252-1257
13. Tariq AM, Al-Shammari AS, Al-Muammar MN, Alhamdan AA. Evaluation and identification of damaged SNPs in

COL1A1 gene involved in osteoporosis. Archives of Medical Science. 2013;9:899-905.
14. Zemla D, Kostova T, Gorchakov R, Volkova E, Beasley DWC, Cardosa J, et al. Genesv-An approach to help characterize possible variations in genomic and protein sequences. Bioinformatics and Biology Insights. 2014;8:1-16. Available:http://dx.doi.org/10.4137/BBI.S13076
15. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoSONE. 2012:7,e46688. Available:http://dx.doi.org/10.1371/journal.pone.0046688
16. Ugbo SB, Yakubu A, Omeje JN, Bibinu BS, Musa IS, Egahi JO et al. Assessment of genetic relationship and application of computational algorithm to assess functionality of non-synonymous substitutions in DQA2 gene of cattle, sheep and goats. Open Journal of Genetics. 2015;5:145-158. Available:http://dx.doi.org/10.4236/ojgen.2015.54011
17. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X Version 2.0. Bioinformatics. 2007;23:2947-2948. Available:http://dx.doi.org/10.1093/bioinformatics/btm404
18. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 2016;33(7):1870-4. DOI: 10.1093/molbev/msw054 Epub 2016 Mar 22.
19. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987;4:406-425.
20. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 1985;39:783-791.
21. Kang JF, Li XL, Zhou RY, Li LH, Feng FJ, Guo XL. Bioinformatics analysis of lactoferrin gene for several species. Biochem Genet. 2008;46:312–322.
22. Sun Y, Zhang X, Xi D, Li G, Wang L, Zheng H, et al. Isolation and cDNA characteristics of MHC-DRA genes from gayal (*Bos frontalis*) and gaytle (*Bos frontalis* × *Bostaurus*). Biotechnology and Biotechnological Equipment. 2015;29: 33. Available:http://dx.doi.org/10.1080/13102818.2014.986128
23. Misra SS, Ganai TAS, Mir SA, Kirmani MA. Molecular characterization of partial exon-2 of the bone morphogenetic protein 15 (BMP15) gene in indian buffalo (*Bubalus bubalis*): Its Contrast with Other Species. Buffalo Bulletin. 2011;30: 24-54.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/20267