Global Science Research Journals

ISSN: 2408-5502 Vol. 5 (4), pp. 397-402, July, 2017

http://www.globalscienceresearchjournals.org/

Global Journal of Animal Breeding and Genetics

Full Length Research Paper

Computational algorithm to assess the functionality of non-synonymous substitution and genetic relationship of DRA II gene of cattle, goat and sheep

[*1]Dauda, A., [2]Duwa, H. and [1]Ebegbulem, V. N.

[1]Department of Animal Science, University of Calabar, Nigeria
[2]Department of Animal Science, University of Maiduguri, Nigeria

Accepted 20 June, 2017

The study was conducted to examine amino acid substitution of Non-synonymous polymorphism and phylogenetic analysis of DRA II gene in cattle, goat and sheep. Twenty (20) nucleotide sequences comprising cattle (8), goat (5) and sheep (7). The following are the sequence accession numbers; AAI51472, AAI02117, AAI10003, AAI03364, AAI42073, AAI33510, AAI02565, AAI05234 (cattle), NP_001301117, XP_005696583, AEJ87223, AEJ87221, AEJ87220 (goat) and NP_001267646, XP_011955819, AEJ36201, AEJ36200, CAA77679, AEV89921, AAA16794 (sheep) were retrieved from gene bank (NCBI). Out of the Sixteen (16) amino acid substitution of cattle, (P10G, Q14V, V16I, R19A, K35Q, R38S, I40K and A41T) appeared beneficial while (F20C, K23L, A25C, N30I and H32P) appeared deleterious. Eighteen (18) amino acid substitution of goat (G10A, I13L, S18L, Q20G, W23G, E37S, Q33I, N39R, Q42G M47L and F72Q) appeared beneficial while (E45R, M60A, K63Q, T65M, L69G, and A76A) appeared deleterious. Sixteen amino acid substitutions of sheep (G10W, F12I, V15D, S18I, E21A, K26N and L38D) appeared beneficial while (V30S, F46D, D49K, D53V, H57P, L69G, E71L, F75A and F78K). The result revealed some intermingling between cattle and sheep which is in accordance with *bovidea* speciation

Keywords: Computational, Functional, Non-synonymous, Genetic and Relationship

INTRODUCTION

The major histocompatibility complex (MHC) class II molecules are cell-surface glycol proteins that are comprised of non-covalently linked α and β chains. These molecules bind antigen peptides and present them to lymphocytes for immune recognition and hence play a pivotal role in the initiation of specific immune responses to exogenous antigens (Zhou et al., 2006). The major histocompatibility complex DRA locus is noteworthy among the major histocompatibility complex class II loci for the little or no variation reported in many species (Zhou et al., 2006).

The immune system is the adaptive defense system that has evolved in vertebrates to protect them from invading pathogens and also carcinomas. It is remarkable in the sense that it is able to generate an enormous variety of cells and biomolecules which interact with each other in numerous ways to form a complex network that helps to recognize, counteract, and eliminate the apparently limitless number of foreign invading pathogens/molecules (Behl et al., 2012). The major histocompatibility complex which is found to occur in all mammalian species plays a central role in the development of the immune system.

*Corresponding author. E-**mail**: ayubadauda87@gmail.com

It is an important candidate gene involved in susceptibility/resistance to various diseases. It is associated with intercellular recognition and with self/nonself discrimination. It plays major role in determining whether transplanted tissue will be accepted as self or rejected as foreign (Behl et al., 2012). Recent advances in technologies have generated a large amount of genome sequence and genotype data for many of species. The method to identify functional single neucleotide polymorphism (SNPs) from a pool, containing both functional and neutral SNPs is challenging by experimental protocols George et al. (2008).

Therefore, computational predictions have become significant for evaluating the disease-related impact of non-synonymous single-nucleotide variants discovered in exome sequencing Liu and Kumar (2013). A number of 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 Zemla et al. (2014). Many such methods have their roots in molecular evolution, as they use information derived from multiple sequence alignments. Most computational prediction tools for amino acid variants rely on the assumption that protein sequences observed among living organisms have survived natural selection. Therefore, evolutionarily conserved amino acid positions across multiple species are likely to be functionally important, and amino acid substitutions observed at conserved positions will potentially lead to deleterious effects on gene functions Choi et al. (2012).

The objective of the work was to carry out study on the amino acid substitution of Non-synonymous polymorphism and phylogenetic analysis of DRA II gene in cattle, goat and sheep.

MATERIALS AND METHOD

Twenty (20) nucleotide sequences comprising cattle (8), goat (5) and sheep (7) were retrieved from the NCBI gene bank and are used for the study. The following are the sequence accession numbers; AAI51472, AAI02117, AAI10003, AAI03364, AAI42073, AAI33510, AAI02565, AAI05234 (cattle), NP_001301117, XP_005696583, AEJ87223, AEJ87221, AEJ87220 (goat) and NP_001267646, XP_011955819, AEJ36201, AEJ36200, CAA77679, AEV89921, AAA16794 (sheep)

Sequence alignment and translations was done using Cluster W (Larkin et al., 2007). Gap open penalty of 15 and gap extension penalty of 6.66 were used**.**

Computational functional analysis, non-synonymous mutations was carried out using PROVEAN (Protein Variant Effect Analyzer) with threshold value of -2.5. PROVEAN collects a set of homologous and distantly related sequences from the NCBI NR protein database using BLASTP (ver.2.2.25) with an E-value threshold of 0.1. The sequences were clustered based on a sequence identity of 80% to remove redundancy using the CD-HIT program (ver.4.5.5) (Li and Godzik, 2006). If the PROVEAN score is smaller than or equal to a given threshold, the variation is predicted as deleterious (Choi et al., 2012).

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the evolutionary history of the protein sequence analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed (Felsenstein, 1985). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. The analysis involved 20 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 189 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).

RESULTS

The functional analysis of coding non-synonymous polymorphism (nsSNP) of DRA II gene of cattle, goat and sheep are presented in Table 1, 2 and 3 respectively. Amino acid substitutions of the wild type alleles located in the coding region of cattle, goat and sheep were obtained from the alignment of deduced amino acid sequences of Cattle, goat and sheep. Out of the Sixteen (16) amino acid substitution of cattle, (P10G, Q14V, V16I, R19A, K35Q, R38S, I40K and A41T) appeared beneficial while (F20C, K23L, A25C, N30I and H32P) appeared deleterious. Eighteen (18) amino acid substitution of goat (G10A, I13L, S18L, Q20G, W23G, E37S, Q33I, N39R, Q42G M47L and F72Q) appeared beneficial while (E45R, M60A, K63Q, T65M, L69G, and A76A) appeared deleterious. Sixteen amino acid substitutions of sheep (G10W, F12I, V15D, S18I, E21A, K26N and L38D) appeared beneficial while (V30S, F46D, D49K, D53V, H57P, L69G, E71L, F75A and F78K) appeared deleterious. The beneficial is an indication that they did not impair protein function while the deleterious does the opposite. The genetic relationship of the amino acid nucleotides of cattle, goat and sheep is presented in Figure 1. The result revealed by the phylogenetic tree showed that cattle and sheep intermingle.

Table 1: Functional analysis of coding nsSNP of the DRA II gene of cattle using PROVEAN

Variant	PROVEAN Score	Prediction
P10G	0.855	Neutral
Q14V	-0.804	Neutral
V16I	-0.172	Neutral
R19A	0.007	Neutral
F20C	-3.890	Deleterious
K23L	-3.931	Deleterious
A25C	-3.026	Deleterious
N30I	-3.114	Deleterious
H32P	-4.594	Deleterious
K35Q	-2.404	Neutral
R38S	-1.074	Neutral
I40K	-2.459	Neutral
A41T	-1.973	Neutral
F43G	-6.545	Deleterious
K45G	-5.139	Deleterious
H52F	-6.832	Deleterious
F56T	-6.096	Deleterious

Default threshold is −2.5, that is; Variants with a PROVEAN score equal to or below −2.5 are considered "deleterious" while Variants with PROVEAN score above −2.5 are considered "neutral". G = glycine, A = Alanine, L = leucine, M = methionine, F = phenylalanine, W = tryptophan,Q = glutamine, E = glutamic acid, S = serine, P = proline, V = valine, Y = tyrosine, R = arginine, N = asparagine, T = threonine, C = cysteine

Table 2: Functional analysis of coding nsSNP of the DRA II gene of goat using PROVEAN

Variant	PROVEAN Score	Prediction
G10A	-0.491	Neutral
I13L	0.646	Neutral
S18L	-2.103	Neutral
Q20G	-2.295	Neutral
W23G	0.874	Neutral
E37S	0.185	Neutral
Q33I	-2.299	Neutral
N39R	-1.149	Neutral
Q42G	-0.895	Neutral
E45R	-2.899	Deleterious
M47L	-1.467	Neutral
M60A	-2.944	Deleterious
K63Q	-4.439	Deleterious
T65M	-7.475	Deleterious
L69G	-8.402	Deleterious
F72Q	0.000	Neutral
A76A	-4.480	Deleterious

Default threshold is −2.5, that is; Variants with a PROVEAN score equal to or below −2.5 are considered "deleterious" while Variants with PROVEAN score above −2.5 are considered "neutral". G = glycine, A = Alanine, L = leucine, M = methionine, F = phenylalanine, W = tryptophan,Q = glutamine, E = glutamic acid, S = serine, P = proline, V = valine, Y = tyrosine, R = arginine, N = asparagine, T = threonine, C = cysteine

Table 3: Functional analysis of coding nsSNP of the DRA II gene of sheep using PROVEAN

Variant	PROVEAN Score	Prediction
G10W	-2.300	Neutral
F12I	-0.355	Neutral
V15D	-1.130	Neutral
S18I	-2.433	Neutral
E21A	-0.112	Neutral
K26N	-2.129	Neutral
V30S	-3.296	Deleterious
L38D	-0.025	Neutral
F46D	-6.748	Deleterious
D49K	-2.822	Deleterious
D53V	-7.502	Deleterious
H57P	-5.972	Deleterious
L69G	-7.469	Deleterious
E71L	-5.697	Deleterious
F75A	-3.530	Deleterious
F78K	-8.314	Deleterious

Default threshold is −2.5, that is; Variants with a PROVEAN score equal to or below −2.5 are considered "deleterious" while Variants with PROVEAN score above −2.5 are considered "neutral". G = glycine, A = Alanine, L = leucine, M = methionine, F = phenylalanine, W = tryptophan,Q = glutamine, E = glutamic acid, S = serine, P = proline, V = valine, Y = tyrosine, R = arginine, N = asparagine, T = threonine, C = cysteine

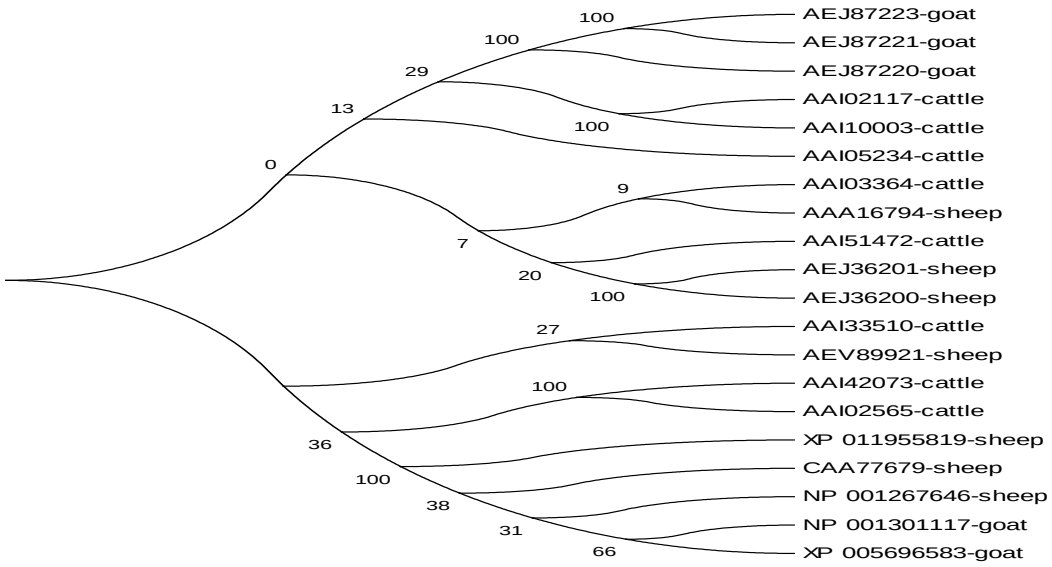

Figure 1: Evolutionary relationship of cattle, goat and sheep using the Neighbor-Joining method

DISCUSSION

Major histocompatibility (MHC) genes are the most polymorphic genes described in vertebrates, with polymorphisms occurring predominantly at residues involved in peptide binding (antigen binding sites) (Zhao et al., 2011). Variation at these sites may affect the antigen binding groove and antigenic-peptide binding ability, and hence peptide specificity (Zhou et al., 2005). It is one of the most important genetic systems for infectious disease resistance in vertebrates (Hill, 1998; Hedrick et al., 2000). Therefore, defining the structure, function, and diversity of this system is very important to understand immune response in vertebrate species (Zhou et al., 2006). The variation in DRA is a highly conserved polypeptide in mammals in general and especially in ruminants. Single nucleotide polymorphism (SNP) in cattle was synonymous, and no amino acid polymorphism was predicted in the *α*1 domain of the bovine DRA molecule (Zhou et al., 2006). However, the functional significance of these SNP cannot be ignored because it has been reported that synonymous mutations in coding regions may act alone or in combination with other mutations in the same transcript to influence mRNA stability and translation, thereby causing functional effects (Duan et al., 2003). Additionally, the SNP may be linked to variation in other DRA gene regions with functional or structural significance. Alternatively, the lack of polymorphism at the amino acid level may be as a result of critical structural or functional constraints in the exon 2 region (Chu et al., 1994).

The present result indicated that bovine DRA II gene is highly polymorphic which is also observed in caprine and ovine species. The amino acid substitution observed in this study revealed variants that are both neutral/beneficial and deleterious/harmful for cattle, goat and sheep. The neutral or beneficial amino acid substitutions are those substitutions that help in maintaining the structural integrity of cells and tissues. Also, they affect positively the functional roles of proteins involved in signal transduction of visual, hormonal, and other stimulants. Therefore, development of disease resistant animal should take this into account as selection for resistance allele may increase the frequency of harmful allele. Therefore effort should be made to increase the disease resistant allele with an aim of producing population of cattle, goat and sheep with disease resistant allele which has a global demand. However, the harmful amino acid substitutions could cause amino acid change further altering protein function which may lead to susceptibility to disease. They may modify enzyme activity, destabilize protein structures or disrupt protein interactions (Bibinu et al., 2016). To identifying single gene markers associated with resistance to gastro-intestinal parasites is difficult as resistance to parasites is considered to be polygenic with hundreds to thousands mutations responsible for the resistant phenotype (Hickford et al., 2011; Kemper et al., 2011), research continues in the area of genetic markers as they have the advantage over phenotypic markers of measurement prior to birth, meaning that producers can make productivity decisions early (Preston et al., 2011). A high level of diversity in MHC genes allows populations to survive despite exposure to rapidly evolving pathogens (Ellis and Hammond, 2014). It plays major role in determining whether transplanted tissue will be accepted as self or rejected as foreign. Also the study of the MHC can aid in the development and design of vaccines based on synthetic peptides comprising one or more T-cell epitopes of the pathogen (Wuliji et al., 2014).

The evolutionary relationship presented in Figure 1. Revealed that the appearance of many alleles at a particular MHC locus is evidence of long-term evolutionary persistence of the locus (Ugbo et al., 2015). This is suggested by the frequency with which alleles in one species are more closely related to the alleles in a closely related species than to the other alleles in the same species (Wuliji et al., 2014). This could be exploited in the development and the design of vaccines as well as drug production. The close similarity of a gene among ruminants may be termed as to recent separation in evolutionary process and/or similar selection pressure which the ruminants have suffered during evolution (Sun et al., 2015). This is in accordance with the well-known evolutionary history of Bovidae subfamily speciation (Floudas, 2007).

CONCLUSION

The study observed both beneficial and harmful amino acid substitution for all the three species. The beneficial amino acid substitutions did not impair the function of protein while the harmful amino acid substitution appeared to have a negative effect on the function of protein of cattle, goat and sheep. The phylogeny analysis based on nucleotide and amino acid sequences of DRA II gene showed some level of intermingling between cattle and sheep. Information from this study is relevant for performing pharmacogenetic, mutagenesis and genotype-phenotype study and also, used for development of molecular makers for selection.

REFERENCE

Behl JD, Verma NK, Tyagi N, Mishra P, Behl R, Joshi BK (2012) The Major Histocompatibility Complex in Bovines: A Review. *International Scholarly Research.*

Bibinu, B. S., Yakubu, A., Ugbo, S. B. and Dim, N. I. (2016). Computational Molecular Analysis of the Sequences of BMP15 Gene of Ruminants and Non-Ruminants. *Open Journal of Genetics*. 6:39 – 50

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, **7**, e46688.

Chu ZT, Carswell-Crumpton C, Cole BC, Jones PP (1994) The minimal polymorphism of class II E*α* chains is not due to the functional neutrality of mutations. *Immunogenetics* 40:9–20.

Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, Gejman PV (2003) Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet*. 12:205–216.

Ellis SA, Hammond JA (2014) The Functional Significance of Cattle Major Histocompatibility Complex Class I Genetic Diversity. *Annual Review of Animal Biosciences*, **2**, 285-306.

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

Floudas CA (2007). Computational methods in protein structure prediction. *Biotechnology and Bioengineering*, 97 (2): 207-213

George PDC, Rajasekaran R, Sudandiradoss C, Ramanathan K, Purohit R, Sethumadhavan R (2008). A Novel Computational and Structural Analysis of nsSNPs in *CFTR* Gene. *Genomic Medicine* **2**: 23-32.

Hedrick PW, Parker KM, Gutierrez-Espeleta GA, Rattink A, Lievers K (2000). Major histocompatibility complex variation in the Arabian oryx. *Evolution*. 54(6):2145–2151.

Hickford JGH, Forrest RHJ, Zhou H, Fang Q, Frampton CM (2011) Association between Variation inFaecal Egg Count for a Mixed Field-Challenge of Nematode Parasites and Ovine MHC-DQA2 Polymorphism. *Veterinary Immunology and Immunopathology*, 144: 312-320.

Hill AVS (1998) The immunogenetics of human infectious diseases, Annual Review of Immunology. 16:593–617.

Kemper KE, Emery DL, Bishop SC (2011) The Distribution of SNP Marker Effects for Faecal Worm Egg Count in Sheep, and the Feasibility of Using These Markers to Predict Genetic Merit for Resistance to Worm Infections.*Genetics Research*, **93**, 203-219.

Kumar S, Stecher G, Tamura K (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33:1870-1874.

Larkin MA, Blackshields G, Brown, NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X Version 2.0. *Bioinformatics*, 23:2947-2948

Li W, Godzik A (2006) Cd-Hit: A Fast Program for Clustering and Comparing Large Protein or Nucleotide Sequences. *Bioinformatics*, **22**, 1658-1659.

Liu L, Kumar S (2013) Evolutionary Balancing Is Critical for Correctly Forecasting Disease Associated Amino Acid Variants. *Molecular Biology and Evolution*, **30**: 1252-1257.

Preston SJM, Sandeman M, Gonzalez J, Piedrafita D (2014) Current Status for Gastrointestinal Nematode Diagnosis in Small Ruminants: Where Are We and Where Are We Going? *Journal of Immunology Research*.4:77-81

Saitou N, Nei M (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Sun, Y., Zhang, X., Xi, D., Li, G., Wang, L., Zheng, H., Du, M., Gu, Z., Yang, Y. and Yang, Y. (2015) Isolation and cDNA Characteristics of MHC-DRA Genes from Gayal (*Bos frontalis*) and Gaytle (*Bos frontalis* × *Bostaurus*). *Biotechnology & Biotechnological Equipment*, **29**, 33-39**.**

Ugbo SB, Yakubu A, Omeje JN, Bibinu BS, Musa IS, Egahi JO, Dim NI (2015) Assessment of Genetic Relationship and Application of Computational Algorithm to Assess Functionality of Non-SynonymousSubstitutions in DQA2 Gene of Cattle, Sheep and Goats. *Open Journal of Genetics*, **5**, 145-158.

Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X Version 2.0. *Bioinformatics*, **23**, 2947-2948.

Wuliji T, Hickford JGH, Lamberson WR, Shanks BC, Azarpajouh S (2014) Ovine Footrot Gene MarkerScreening in a Katahdin Sheep Flock. *Proceedings of Joint Annual Meeting of ASAS*, Kansas City, 20-24 July 2014.

Zemla D, Kostova T, Gorchakov R, Volkova E, Beasley DWC, Cardosa J, Weaver SC, Vasilakis N, Naraghi-Arani P (2014) Genesv—An Approach to Help Characterize Possible Variations in Genomic and Protein Sequences. *Bioinformatics and Biology Insights* **8**: 1-16.

Zhao Y, Xu H, Shi L, Zhang J (2011) Polymorphisms in Exon 2 of MHC Class II DRB3 Gene of 10 Domestic Goats in Southwest China. *Asian-Australasian Journal of Animal Science*, **24**, 752-756.

Zhou H, Hickford JGH, Fang Q (2005) Polymorphism of the DQA2 Gene in Goats. *Journal of Animal Science*.**83**, 963-968.

Zhou H, Hickford JGH, Fang Q, Byun SO (2006). Identification of Allelic Variation at the Bovine *DRA* Locus by Polymerase Chain Reaction-Single Strand Conformational Polymorphism. *J. Dairy Sci*. 90:1943–1946.

Zuckerkandl E, Pauling L (1965). Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.