UDC 575
https://doi.org/10.2298/GENSR1701063Y
Original scientific paper

APPLICATION OF COMPUTATIONAL ALGORITHMS TO ASSESS THE FUNCTIONALITY OF NON-SYNONYMOUS SUBSTITUTIONS IN MHC DRB GENE OF NIGERIAN GOATS

Abdulmojeed YAKUBU[1,2,3], Adebowale E. SALAKO[2], Marcos DE DONATO[3,4] and Ikhide G. IMUMORIN[3]

[1]Department of Animal Science, Nasarawa State University, Keffi, Shabu-Lafia Campus, Lafia, 950101, Nigeria
[2]Animal Breeding and Genetics Laboratory, Department of Animal Science, University of Ibadan, Ibadan, Nigeria
[3]Animal Genetics and Genomics Laboratory, International Programs, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY 14853. USA
[4]Laboratorio Genetica Molecular, Instituto de Investigaciones en Biomedicina y Ciencias Aplicadas, Universidad de Oriente, Cumana, Venezuela.

Yakubu A., A. E. Salako, M. De Donato, I. G. Imumorin (2017): *Application of computational algorithms to assess the functionality of non-synonymous substitutions in MHC DRB gene of Nigerian goats*.- Genetika, Vol 49, No. 1, 63 -76.

The Major Histocompatibility Complex (MHC) contains highly variable multi-gene families, which play a key role in the adaptive immune response within vertebrates. Among the Capra MHC class II genes, the expressed DRB locus is highly polymorphic, particularly in exon 2, which encodes the antigen-binding site. Models of variable non-synonymous/synonymous rate ratios among sites may provide important insights into functional constraints at different amino acid sites and may be used to detect sites under positive selection. Many non-synonymous single nucleotide polymorphisms (nsSNPs) at the DRB locus in goats are suspected to impact protein function. This study, therefore, aimed at comparing the efficiency of six computational approaches to predict the likelihood of a particular non-synonymous (amino acid change) coding SNP to cause a functional impact on the protein. This involved the use of PANTHER, SNAP, SIFT, PolyPhen-2, PROVEAN and nsSNPAnalyzer bioinformatics analytical tools in detecting harmful and beneficial effects at H57G, Y89R, V104D and Y112I substitutions in the peptide binding region of the DRB gene of Nigerian goats. The results from PANTHER analysis revealed that H57G, Y89R and Y112I substitutions ($P_{deleterious}$= 0.113, 0.204 and

Corresponding author: Abdulmojeed Yakubu, E-mail: abdulmojyak@gmail.com; Ikhide G. Imumorin, E-mail: igi2@cornell.edu

0.472, respectively) were beneficial; while that of V104D was deleterious ($P_{deleterious}$= 0.756), an indication that it was non-neutral. As regards the SNAP approach, H57G and Y89R substitutions were returned neutral with expected accuracy of 53 and 69%, respectively while V104D and Y112I substitutions were harmful. H57G and Y89R substitutions were also found harmless in the SIFT analysis. However, only H57G (PROVEAN) and V104D (nsSNPAnalyzer) amino acid substitutions were found to be beneficial. Interestingly, the predicted 3D structures of both native and mutant DRB protein appeared similar as validated by Ramachandran plots. The consensus reached by PANTHER, SNAP, SIFT and PolyPhen-2 approaches on the neutrality especially of H57G (PROVEAN inclusive) and Y89R amino acid substitutions may be used in search of disease resistant genotypes at the DRB locus of Nigerian goats.

Key words: bioinformatics tools, DRB gene, Goats, Nigeria, Non-synonymous substitutions

INTRODUCTION

With the developments in molecular biology, researchers can directly examine selection at genes that underlie functional traits. Therefore, adaptive non-neutral markers have become especially valuable. One important example comes from the genes of the major histocompatibility complex (MHC) (NEI *et al.*, 2013; YAKUBU *et al*., 2013; ASHRAFI *et al*., 2014). 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). They are known to be involved in immune function (SINGH *et al*., 2012; SHEN *et al*., 2014). The key task of the MHC is to code for specialised antigen-presenting receptor glycoproteins, also called as MHC molecules. These molecules bind processed peptide antigens and present them to T lymphocytes, thereby triggering immune responses (GOWANE *et al*., 2013). Balancing selection acting on MHC appears to be able to maintain allelic lineages for a long time, and also favour novel alleles with non-synonymous substitutions changing peptide-binding properties of MHC molecules (KUDUK *et al*., 2012). The type of genetic mutation that causes a single amino acid substitution (AAS) in a protein sequence is called a non-synonymous single nucleotide polymorphism (nsSNP) (KUMAR and HENIKOFF, 2009). Indeed, the rates of non-synonymous substitutions often exceed those of synonymous substitutions at sites involved in peptide binding (GARRIGAN and HEDRICK, 2003). It is clearly important to characterize the distribution of fitness effects of new nonsynonymous mutations. This distribution is relevant to a broad range of problems in evolutionary genetics (LOEWE *et al*., 2006). Evaluating functional effects of known non-synonymous single nucleotide polymorphisms (nsSNPs) is essential for understanding genotype/phenotype relations and for curing diseases (BROMBERG *et al*., 2008). The pattern of an elevated rate of non-synonymous substitutions at antigen binding sites is considered as clear evidence for 'positive Darwinian selection shaping genetic variation. This polymorphism enables the immune system to recognise an extensive range of pathogens and is therefore crucial for the immunological fitness of individuals and, thus, animal populations (BERNATCHEZ and LANDRY, 2003).

The method to identify functional SNPs from a pool, containing both functional and neutral SNPs is challenging by experimental protocols (GEORGE *et al*., 2008). If a marker is found to be associated with a disease and the marker is nsSNP, prediction tools can provide independent evidence as to whether the nsSNP itself contributes to disease. This is because carrying out the appropriate assays may be time-consuming, whereas these tools can filter out nsSNPs that are unlikely to affect protein function before experimentation. The computational tools like SIFT,

SNAP, PROVEAN, PolyPhen-2, I-Mutant, PANTHER are used nowadays for detecting impact of amino acid substitution especially in coding exonic region (RAJASEKARAN *et al*., 2007; YAKUBU *et al*., 2013; YAKUBU *et al*., 2014; UGBO *et al*., 2015; PATEL *et al*., 2015).

In Nigeria, sub-saharan Africa, there is dearth of information on the use of computational algorithms to detect the presence of beneficial or harmful amino acid substitutions in MHC genes. Therefore, the present study was undertaken to predict the functional effects of non-synonymous substitutions at the peptide binding region of MHC DRB gene of Nigerian goats. The information obtained could assist in better understanding of the host parasite interaction for breeding purpose, and to devise better prophylactic measures for associated livestock diseases.

MATERIALS AND METHODS

PCR amplification and sequencing of caprine DRB gene

Genomic DNA was obtained from blood samples taken from the jugular vein using a catheter with a 21G needle, following the ethical guidelines of the International Council for Laboratory Animal Science and Cornell University, Ithaca, NY, USA. Of a total of 220 sampled animals, 80 were from the breed West African Dwarf (WAD), 90 from Red Sokoto (RS) and 50 from Sahel (SH), including goats of both sexes, sampled across various farms in Nigeria. A 284-bp fragment of exon 2 of DRB was amplified in Nigerian goats for polymorphism identification. Primers designed were identical to those described by AMILLS *et al.* (1995). Primer sequences were DRB 1.1/FW, 5′- TATCCCGTCTCTGCAGCACATTTC -3′ and DRB 1.2/REV, 5′- TCGCCGCTGCACACTGAAACTCTC -3′. PCR amplifications were carried out in a C1000™ Thermal Cycler (Bio-Rad, USA) with a total reaction volume of 20µL containing 20-30ng DNA, 10pmol of each primer in 10µL Syd Lab PCR Premix (Syd Labs, Inc., Malden, USA) containing Taq DNA polymerase, dNTPs, $MgCl_2$, reaction buffer, PCR stabilizer and enhancer at optimal concentrations. The thermal profile for amplifying the DRB exon 2 involved 30 cycles of initial denaturation at 94° for 5mins, denaturation at 94° for 60s, annealing at 60° for 90s, extension at 72° for 90s and elongation at 72° for 10mins. PCR products were detected on 1.0% agarose gel including 0.5µg/ml of GelRed nucleic acid stain and scored using GENEMateQuanti-Marker 100 bp DNA ladder (BioExpress, UT, USA. Sequencing of the amplified fragments was carried out using the Applied Biosystems Automated 3730XL DNA Analyzer.

Sequence analysis of DRB gene

Sequence alignments, translations and comparisons were carried out using ClustalW (LARKIN *et al*., 2007). The BLAST algorithm was used to search the NCBI GenBank (http://www.ncbi. nlm.nih.gov/) databases for homologous sequences. The deduced DRB amino acid sequences of Nigerian goats were compared with published caprine amino acid sequences with GenBank accession nos., AB008345.1, AB008347.1, AB008348.1, AB008349.1, AB008350.1, AB008351.1, AB008352.1, AB008353.1, AB008354.1, AB008355.1, AB008356.1, AB008357.1, AB008358.1, AB008359.1, AB008360.1, AB008361.1 and AB008362.1, respectively. MEGA version 5 (TAMURA *et al*., 2011) was employed in the analysis.

Functional analysis of coding nsSNPs at the peptide binding region of DRB gene

The functional analysis of coding nsSNPs at the peptide binding region of MHC DRB gene involved the use of six computational tools such as PANTHER, SNAP, SIFT, PolyPhen-2 , PROVEAN and nsSNPAnalyzer:

In silico functional analysis of missense mutations was obtained using PANTHER (THOMAS *et al*., 2003) whose predictions have been experimentally validated (BRUNHAM *et al*., 2005). PANTHER estimates the likelihood of a particular non-synonymous (amino acid changing) coding SNP to cause a functional impact on the protein. It calculates the substitution position-specific evolutionary conservation (subPSEC) score based on an alignment of evolutionarily related proteins (THOMAS and KEJARIWAL, 2004). The probability that a given variant will cause a deleterious effect on protein function is estimated by Pdeleterious, such that a subPSEC score of -3 corresponds to a Pdeleterious of 0.5 (THOMAS *et al*., 2006). 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 are continuous values from 0 (neutral) to about -10 (most likely to be deleterious).

SNAP is a neural network based method for identifying from sequence functionally disruptive single amino acid substitutions (BROMBERG and ROST, 2007). The inputs to SNAP include secondary structure and solvent accessibility predictions, evolutionary and family information, biophysical differences between the wild type and mutant amino acids, statistical likelihoods of observing residue triplets around the mutation site, SIFT (NG and HENIKOFF, 2003) and SwissProt annotation if available. Users submit the wild-type sequence along with their mutants. A comma-separated list gives mutants as: XiY, where X is the wild type amino acid, Y is the mutant and iis the number of the residue (i= 1 for N-terminus). X is not required and a star (*) can replace either ior Y. Any combination of characters following these rules is acceptable; e.g. X** = replace all residues X in all positions by all other amino acids, *Y = replace all residues in all positions by Y. Users may provide a threshold for the minimal RI [Equation (1)] and/or for the expected accuracy of predictions that will be reported back. These two values correlate; when both are provided, the server chooses the one yielding better predictions. For each mutant, SNAP returns three values: the binary prediction (neutral/non-neutral), the RI (range 0–9) and the expected accuracy that estimates accuracy [Equation (1)] on a large dataset at the given RI (i.e. accuracy of test set predictions calculated for each neutral and non-neutral RI (BROMBERG *et al*., 2008).

$$RI=INT(OUT_{non-neutral} - OUTneutral)/10 \quad (1)$$

SIFT is a multi-step algorithm that uses a sequence homology-based approach to classify amino acid substitutions (NG and HENIKOFF, 2001; NG and HENIKOFF, 2003). For a given protein sequence, SIFT compiles a dataset of functionally related protein sequences by searching a protein database using the PSI-BLAST algorithm. It then builds an alignment from the homologous sequences with the query sequence. In the second step of the algorithm, SIFT scans each position in the alignment and calculates the probabilities for all possible 20 amino acids at that position. These probabilities are normalized by the probability of the most frequent amino acid and are recorded in a scaled probability matrix. SIFT predicts a substitution to affect protein function if the scaled probability, also termed the SIFT score, lies below a certain threshold value. Generally, a highly conserved position is intolerant to most substitutions, whereas, a poorly conserved position can tolerate most substitutions (KUMAR and HENIKOFF, 2009).

PolyPhen-2 is an automatic tool for prediction of possible impact of an amino acid substitution on the structure and function of a protein. This prediction is based on a number of features comprising the sequence, phylogenetic and structural information characterizing the substitution. PolyPhen-2 calculates the naive Bayes posterior probability that a given mutation is

damaging and reports estimates of false positive (the chance that the mutation is classified as damaging when it is in fact non damaging) and true positive (the chance that the mutation is classified as damaging when it is indeed damaging) rates. A mutation is also appraised qualitatively, as benign, possibly damaging or probably damaging (ADZHUBEI *et al*., 2010).

PROVEAN analysis consists of two main steps: In the first step, PROVEAN collects a set of homologous and distantly related sequences from the NCBI NR protein database (released August 2011) using BLASTP (ver.2.2.25) with an E-value threshold of 0.1. The sequences are clustered based on a sequence identity of 80% to remove redundancy using the CD-HIT program (ver.4.5.5) (LI and GODZIK, 2006). Starting from the sequence cluster most similar to the query sequence, the clusters are added to the supporting sequence set one by one until there is sufficient number of clusters in the supporting set. A cutoff of 45 clusters is currently used, that is, all sequences from up to 45 clusters are used as the supporting sequence set. In the second step, for each sequence in the supporting sequence set, a delta score is computed using the BLOSUM62 substitution matrix and gap penalties of 10 for opening and 1 for extension. Within each cluster, an average delta score is computed. The averaged delta scores are again averaged among all clusters so that each cluster is weighted equally. This unbiased averaged delta score is the final PROVEAN score. If the PROVEAN score is smaller than or equal to a given threshold, the variation is predicted as deleterious (CHOI *et al*., 2012)

nsSNPAnalyzer uses a machine learning method called Random Forest to classify the nsSNPs. It was trained using a curated SNP dataset prepared from the SwissProt database. nsSNPAnalyzer calculates three types of information from the user's input: 1) the structural environment of the SNP, including the solvent accessibility, environmental polarity and secondary structure (BOWIE *et al*., 1991; 2) the normalized probability of the substitution in the multiple sequence alignment; 3) the similarity and dissimilarity between the original amino acid and mutated amino acid.

Modelling of native and mutant structures of DRB protein

The Phyre2 server was used to predict the 3D structures of native and mutant DRB protein. This server predicts the three-dimensional structure of a protein sequence using the principles and techniques of homology modelling (KELLY *et al*., 2015).

Model quality and structure assessment

Model quality was checked for both native and altered DRB proteins by Ramachandran plot using the software, RAMPAGE (LOVELL *et al*., 2002) which analyzed residue-by-residue geometry and overall structure geometry. Total energy after energy minimization was calculated for both native and altered models using Swiss PDB viewer. Computations were done in vacuo with the GROMOS96 43B1 parameters set without reaction field.

RESULTS AND DISCUSSION

Nonsynonymous amino acid substitutions

Four amino acid substitutions (H57G, Y89R, V104D and Y112I) of the wild type alleles located in the peptide coding region of the DRB exon 2 of Nigerian goat alleles were obtained from the alignment of deduced amino acid sequences of indigenous goats and the published caprine sequences. Peptide binding sites were identified following the description of NASKAR *et al*. (2012). Genetic variation in parasite and host and its relative distribution across space and time is

of interest as the basis for adaptive change and it has long been recognized that spatial structure can strongly influence the process of co-adaptation between host and parasite and the evolution of virulence (BIEK and REAL, 2010). MHC has a determinant role in deciding the fate of antigen and initiating the immune response, therefore studying host pathogen relationship and intensifying research on variation at host genomic level combined with allelic diversity for immune response genes of MHC is required to support disease control strategy (GOWANE *et al*., 2013). Phylogenetic evidence supports the notion that the generation of new DRB genes is a dynamic and steadily ongoing process and there are indications that ancient peptide binding motifs are frequently reshuffled among duplicated members of the DRB multigene family (DOXIADIS *et al*., 2008). Therefore, the four amino acid replacements at the peptide binding sites of Nigerian DRB gene could add to the body of knowledge on DRB diversity in caprine species.

Functional analysis of Nonsynonymous amino acid substitutions

PANTHER analysis revealed that H57G, Y89R and Y112I substitutions ($P_{deleterious}$= 0.113, 0.204 and 0.472, respectively) were beneficial; while that of V104D was deleterious ($P_{deleterious}$= 0.756) (Table 1). As regards the SNAP approach, H57G and Y89R substitutions were returned neutral with expected accuracy of 53 and 69%, respectively while V104D and Y112I substitutions were harmful (Table 2). H57G and Y89R substitutions were also found harmless in the SIFT analysis (Table 3). SIFT scores<0.05 are predicted by the algorithm to be intolerant or deleterious amino acid substitutions, whereas scores >0.05 are considered tolerant (NG and HENIKOFF, 2001). The higher a tolerance index, the less functional impact a particular amino acid substitution is likely to have. Similarly, PolyPhen-2 returned H57G and Y89R as benign (neutral) while V104D and Y112I were found to be probably damaging (Table 4). However, only H57G (PROVEAN) (Table 5) and V104D (nsSNPAnalyzer) (Table 6) amino acid substitutions were found to be beneficial. Single Nucleotide Polymorphisms (SNPs) are being intensively studied to understand the biological basis of complex traits and diseases (GEORGE *et al*., 2008). The enormous amount of sequence variation data generated from large-scale projects necessitates computational approaches to assess the potential impact of amino acid changes on gene functions (CHOI *et al*., 2012). The prediction of "functional SNPs" is an active and evolving area of SNP bioinformatics (JOHNSON, 2009).

Table 1. Functional analysis of coding nsSNPs of the DRB gene of Nigerian goats using PANTHER

Substitution	$P_{deleterious}$	subPSEC	MSA position	P_{wt}	$P_{substituted}$	NIC
H57G	0.11318	-0.94137	55	0.05476	0.03138	1.558
Y89R	0.20432	-1.64046	89	0.17299	0.04462	1.558
V104D	0.75647	-4.13345	104	0.48891	0.00757	1.558
Y112I	0.47193	-2.88761	112	0.39357	0.02477	1.558

H=histidine; G= glycine; Y= tyrosine; R= arginine; V= valine; D= aspartic acid; I= isoleucine
The probability that a given variant will cause a deleterious effect on protein function is estimated by Pdeleterious, such that a subPSEC score of -3 corresponds to a Pdeleterious of 0.5.

Table 2. Functional analysis of coding nsSNPs of the DRB gene of Nigerian goats using SNAP

nsSNP	Prediction	Reliability Index	Expected Accuracy
H57G	Neutral	0	53%
Y89R	Neutral	2	69%
V104D	Non- Neutral	2	70%
Y112I	Non- Neutral	3	78%

Including only predictions with: RI >= 0
Expected Accuracy >= 50%

Table 3. Functional analysis of coding nsSNPs of the DRB gene of Nigerian goats using SIFT

Substitution	Prediction	Score	Median Info	No of sequences at position
H57G	Tolerated	0.34	3.28	45
Y89R	Tolerated	0.30	3.28	45
V104D	Damaging*	0.00	3.28	45
Y112I	Damaging*	0.00	3.28	45

*WARNING!! This substitution may have been predicted to affect function just because the sequences used were not diverse enough. There is LOW CONFIDENCE in this prediction.

Table 4. Functional analysis of coding nsSNPs of the DRB gene of Nigerian goats using Polyphen-2

Substitution	Prediction	Score	Sensitivity	Specificity
H57G	Benign	0.001	0.99	0.15
Y89R	Benign	0.197	0.92	0.87
V104D	Probably Damaging	1.000	0.00	1.00
Y112I	Probably Damaging	0.996	0.55	0.98

Table 5. Functional analysis of coding nsSNPs of the DRB gene of Nigerian goats using PROVEAN

Variant	PROVEAN Score	Prediction
H57G	-0.440	Neutral
Y89R	-6.222	Deleterious
V104D	-6.173	Deleterious
Y112I	-7.512	Deleterious

1. Default threshold is -2.5, that is
-Variants with a score equal to or below -2.5 are considered ‘’deleterious’’
-Variants with a score above -2.5 are considered ‘’neutral’’

Table 6. Functional analysis of coding nsSNPs of the DRB gene of Nigerian goats using nsSNPAnalyzer

SNP	Phenotype	Environment	Area Buried	FracPolar	Secondstr	ScopLink	Sift_Score
H57G	Disease	B2S	0.694	0.417	S	d1fv1b2	NA
Y89R	Disease	P2H	0.266	0.833	H	d1a6ab2	NA
V104D	Neutral	P2H	0.432	0.573	H	d1a6ab2	NA
Y112I	Disease	B1H	0.744	0.302	H	d1a6ab2	NA

FracPolar= Environmental polarity score
Secondstr=secondary structure; H: alpha-helix, S: beta-sheet
NA= not available

The ability to differentiate disruptive mutations from neutral ones is necessary for a better understanding of protein function. A given nsSNP may disrupt function in two ways: (1) by directly changing the "active" residue (e.g. by replacing the amino acid for a residue involved in ligand binding, catalysis, allosteric regulation, or post-translational modification), or (2) by affecting the scaffolding of the protein (e.g. by deforming and/or destabilizing the binding site or the entire protein structure) (BROMBERG and ROST, 2009). Although no particular algorithm could be said to have given 100% results in the present study, the consensus reached by PANTHER, SNAP, SIFT and PolyPhen-2 on H57G (including PROVEAN) and Y89R substitutions could be an indication of the possibility of promising genotypes from these substitutions that could serve as marker genes. Obtaining an estimate of genetic variability such as heritability for the vaccine mediated immune response in the population will be very much helpful for designing the future breeding plans (GOWANE *et al*., 2013) and exploiting the existing genetic variance component for developing better future breeding stock. Therefore, the genotypes that may emanate from the beneficial amino acid substitutions at the peptide binding region of the DRB gene of Nigerian goats in this study could be exploited in future to formulate appropriate breeding and selection strategies for the control of some livestock diseases (YAKUBU, 2014).

3-D models of native and mutant DRB protein

A total of 197 residues (74% each of native and mutant DRB amino acid sequence) were modeled with 100.0% confidence by the single highest scoring template (Figures 1and 2). The models showed good geometry: a Ramachandran plot of chains A and B showed that residues lie mainly in the most favored regions of the plot with 86.7% in core, 9.2% in allowable regions and 4.1% outliers (Figures 3 and 4). The Ramachandran plot is a two-dimensional graph of the phi (f) and psi (y) backbone angles for each amino acid residue of a protein; it is a simple method of assessing the quality of a protein by providing an indispensable summative description of backbone conformation (SOLIS, 2015; TRAN *et al*., 2015). Energy minimization for the mutant protein (-7410.142kj/mol) was slightly higher than that of the native protein (-7543.407kj/mol). This may not be unconnected with the presence of few deleterious nsSNPs in the mutant protein. However, this finding with the results of the Ramachandran plots suggests that the 3-D models of the native and mutant DRB amino acid sequences could be accepted as reliable with good confidence.

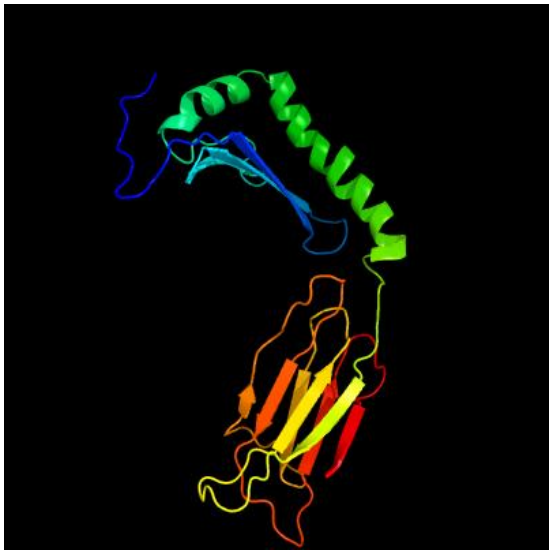

Figure 1. Schematic 3D structure of native caprine DRB protein
Image coloured by rainbow N → C terminus
Model dimensions (Å): X:52.973 Y:61.794 Z:55.413

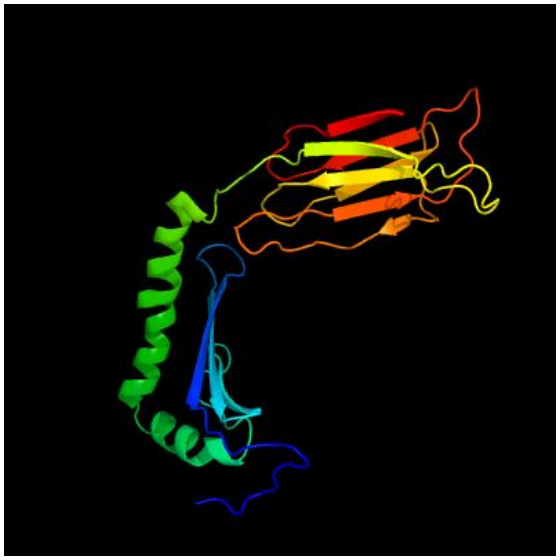

Figure 2. Schematic 3D structure of mutant caprine DRB protein
Image coloured by rainbow N → C terminus
Model dimensions (Å): X:53.441 Y:62.115 Z:55.413

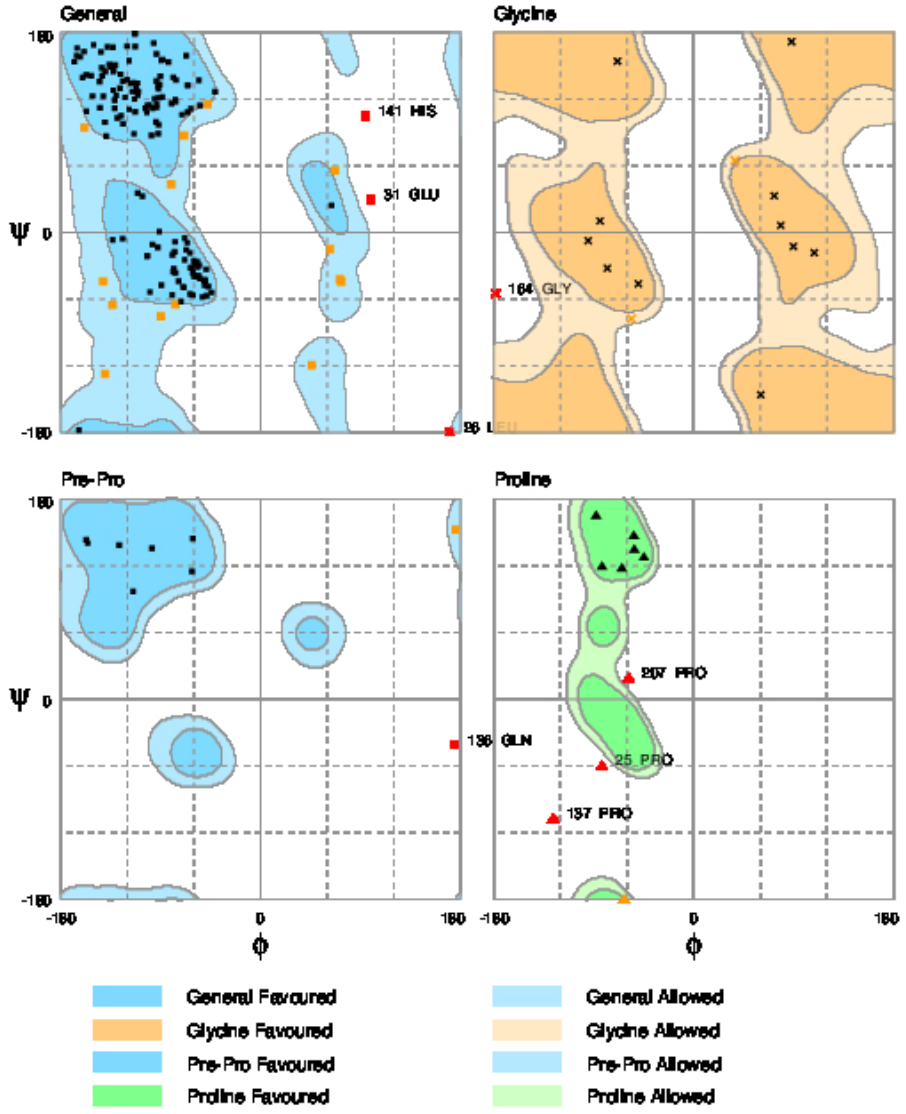

Figure 3. Ramachandran plot of native DRB protein

Number of residues in favoured region (~98.0% expected) : 169 (86.7%)
Number of residues in allowed region (~2.0% expected) : 18 (9.2%)
Number of residues in outlier region : 8 (4.1%)

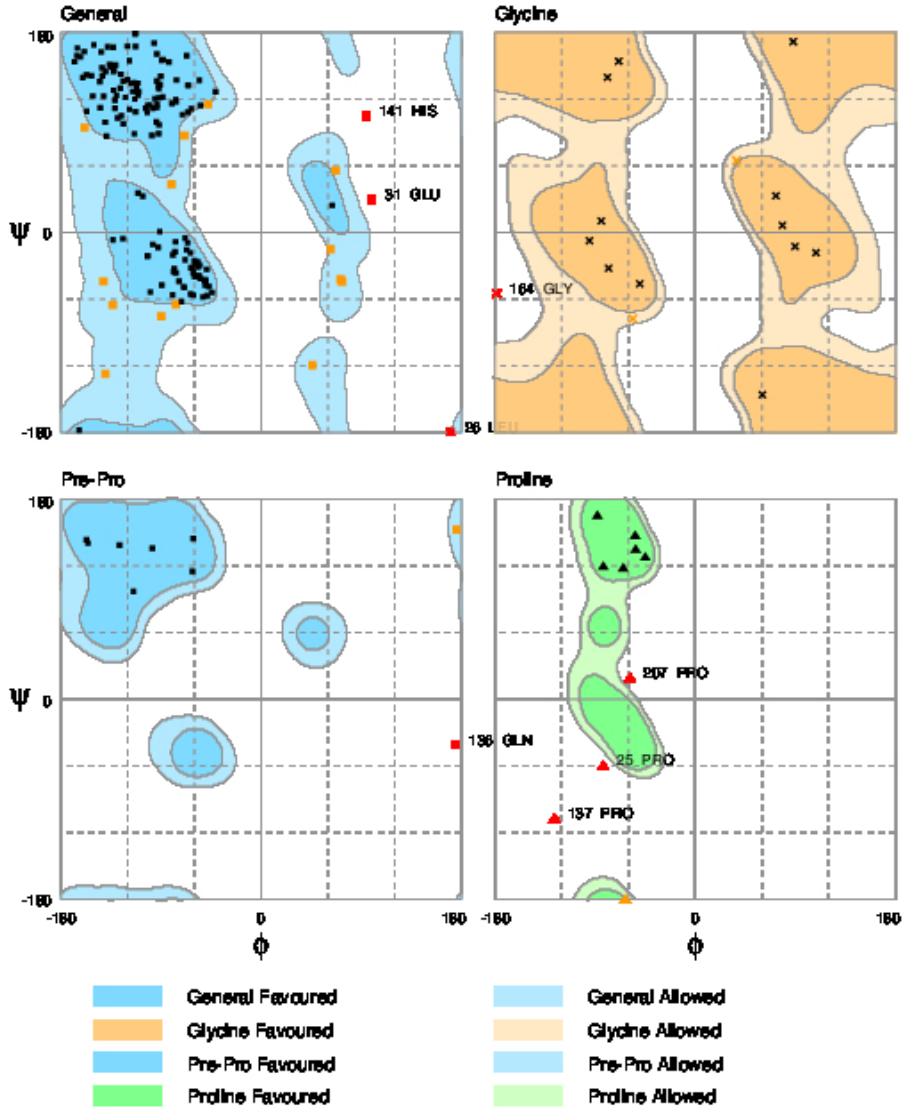

Figure 4. Ramachandran plot of DRB Mutant protein

Number of residues in favoured region (~98.0% expected) : 169 (86.7%)
Number of residues in allowed region (~2.0% expected) : 18 (9.2%)
Number of residues in outlier region : 8 (4.1%)

CONCLUSIONS

Research on Nigerian goat DRB gene has yielded some markers, which would be helpful for further genotype–phenotype research as well as pharmacogenetics studies. However, more efforts on allele mining in this part of the caprine genome are required. This might help in changing the future of animal husbandry practices especially in the tropics as the upcoming generations will be less inflicted with dreaded diseases on account of presence of disease resistant genes.

ACKNOWLEDGEMENT

Funding for this study was provided by the College of Agriculture & Life Sciences and Mario Einaudi Center for International Studies of Cornell University, Ithaca, NY. Additional support from the Nigerian Government Tertiary Education Trust Fund (TETFUND) fellowship to AY is gratefully acknowledged.

Received April 06th, 2016
Accepted December 15th, 2016

REFERENCES

ADZHUBEI, I.A., S., SCHMIDT, L. PESHKIN, V.E. RAMENSKY, A. GERASIMOVA, P. BORK, A.S. KONDRASHOV, S.R. SUNYAEV (2010): A method and server for predicting damaging missense mutations. Nature Protocols, *7*: 248-249.

AMILLS, M., O. FRANCINO, A. SANCHEZ (1995): Nested PCR allows the characterization of TaqI and PstI RFLPs in the second exon of the caprine MHC class II DRB gene. Veterinary Immunology and Immunopathology, *48*: 313-321.

ASHRAFI, F., A. HASHEMI, K. MARDANI, R. DARVISHZADEH (2014): Study on genetic variability in MHC-DRB1 second exon in Makuie sheep breed population. Genetika, *46*: 269-275.

BERNATCHEZ, L. and C. LANDRY (2003): MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J. Evol. Biol., *16*: 363-377.

BIEK R. and L.A. REAL (2010): The landscape genetics of infectious disease emergence and spread. Molecular Ecology, *19*: 3513-3531.

BOWIE, J.U., R. LUTHY, D. EISENBERG (1991): A method to identify protein sequences that fold into a known three-dimensional structure. Science, *253*: 164-170.

BROMBERG, Y. and B. ROST (2009): Correlating protein function and stability through the analysis of single amino acid substitutions. BMC Bioinformatics, *10* (Suppl 8): S8 doi:10.1186/1471-2105-10-S8-S8

BROMBERG, Y. and B. ROST (2007): SNAP: predicts effect of non-synonymous polymorphisms on function. Nucleic Acids Research, *35*: 3823-3835.

BROMBERG, Y., G. YACHDAV, B. ROST (2008): SNAP predicts effect on protein function. Bioinformatics Applications Note, *24*: 2397–2398.

BRUNHAM, L.R., R.R., SINGARAJA, T.D. PAPE, A. KEJARIWAL, P.D. THOMAS, M.R. HAYDEN (2005): Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene. PLoS Genetics, *1*:e83.

CHOI, Y., G.E. SIMS, S. MURPHY, J.R. MILLER, A.P. CHAN (2012): Predicting the functional effect of amino acid substitutions and indels. PLoS One, *7*: e46688.

DOXIADIS, G.G., N. DE GROOT, N.G. DE GROOT, I.I. DOXIADIS, R.E. BONTROP (2008): Reshuffling of ancient peptide binding motifs between HLA-DRB multigene family members: old wine served in new skins. Molecular Immunology, *45*: 2743-2751.

GARRIGAN, D. and P.W. HEDRICK (2003): Perspective: Detecting adaptive molecular polymorphism, lessons from the MHC. Evolution, *57*: 1707–1722.

GEORGE, P.D., R. RAJASEKARAN, C. SUDANDIRADOSS, K. RAMANATHAN, R. PUROHIT, R. SETHUMADHAVAN (2008): A novel computational and structural analysis of nsSNPs in CFTR gene. Genomic Med., *2*: 23–32.

GOWANE, G.R., A.K. SHARMA, M. SANKAR, K. NARAYANAN, S. SUBRAMANIAM, B. PATTNAIK (2013): Evaluation of genetic and environmental parameters determining antibody response induced by vaccination against Foot and Mouth Disease. Agricultural Research, *2*: 140-147.

GOWANE, G.R., R.K. VANDRE, M. NANGRE, A.K. SHARMA (2013): Major histocompatibility complex (MHC) of bovines: an insight into infectious disease resistance. Livestock Research International, *1*: 46-57.

JOHNSON, A.D. (2009): Single-Nucleotide Polymorphism Bioinformatics: A Comprehensive Review of Resources. Circ. Cardiovasc. Genet., *2*: 530-536.

KELLY, L.A., S. MEZULIS, C.M. YATES, M.N. WASS, M.J.E. STERNBERG (2015): The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols, *10*: 845-858.

KUDUK, K., W. BABIK, K. BOJARSKA, E.B. ŚLIWIŃSKA, J. KINDBERG, P. TABERLET, J.E. SWENSON, J. RADWAN (2012): Evolution of major histocompatibility complex class I and class II genes in the brown bear. BMC Evolutionary Biology, *12*: 197.http://www.biomedcentral.com/1471-2148/12/197.

KUMAR, P., S. HENIKOFF, P.C. NG (2009): Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protocols, *4*: 1073-1081.

LARKIN, M.A., G. BLACKSHIELDS, N.P. BROWN, R. CHENNA, P.A. MCGETTIGAN, H. MCWILLIAM F. VALENTIN, I.M. WALLACE, A. WILM, R. LOPEZ, J.D. THOMPSON, T.J. GIBSON, D.G. HIGGINS (2007): Clustal W and Clustal X version 2.0. Bioinformatics, *23*: 2947-2948.

LI, W. and A. GODZIK (2006): Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, *22*: 1658–1659.

LOEWE, L., B. CHARLESWORTH, C. BARTOLOME, V. NOEL (2006): Estimating selection on nonsynonymous mutations. Genetics, *172*: 1079–1092.

LOVELL, S.C., I.W. DAVIS, III W.B. ARENDALL, P.I.W. DE BAKKER, J.M. WORD, M.G. PRISANT, J.S. RICHARDSON, D.C. RICHARDSON (2002): Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function and Genetics, *50*: 437-450.

NASKAR, S., S.M. DEB, S.K. NIRANJAN, S. KUMAR, D. SHARMA, D. SAKARAM, A. SHARMA (2012): Molecular characterization of MHC-DRB cDNA in water buffalo (*Bubalus bubalis*). Genetics and Molecular Biology, *35*: 95-98.

NEI, C., Y. LI, J. ZHAO, X. WU (2013): Extremely high major histocompatibility complex class IIb gene intron 2 variation and population structure in Chinese alligator. Journal of Genetics, *91*: e86–e91.

NG P.C. and S. HENIKOFF (2003): SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research, *31*: 3812-3814.

NG, P.C. and S. HENIKOFF (2001): Predicting deleterious amino acid substitutions. Genome Res., *11*: 863–874.

PATEL, S.M., P.G. KORINGA, B.B. REDDY, N.M. NATHANI, C.G. JOSHI (2015): In silico analysis of consequences of non-synonymous SNPs of Slc11a2 gene in Indian bovines. Genomics Data, *5*: 72–79.

RAJASEKARAN, R., C. SUDANDIRADOSS, C.G. DOSS, R. SETHUMADHAVAN (2007): Identification and in silico analysis of functional SNPs of the BRCA1 gene. Genomics, *90*: 447–452.

SHEN, H., G. HAN, B. JIA, S. JIANG, Y. DU (2014): MHC-DRB1/DQB1 gene polymorphism and its association with resistance/susceptibility to Cystic echinococcosis in Chinese Merino sheep. Journal of Parasitology Research doi:10.1155/2014/272601.

SINGH, P.K., S.V. SINGH, M.K. SINGH, V.K. SAXENA, A.V. SINGH, J.S. SOHAL (2012): Genetic analysis of MHC Class 11 DRB gene in an endangered Jamunapari goat breeds. Indian Journal of Biotechnology, *11*: 220-223.

SOLIS, A.D. (2015): Deriving High-Resolution Protein Backbone Structure Propensities from All Crystal Data Using the Information Maximization Device. PLoS One, *9*: e94334.

TAMURA, K., D. PETERSON, N. PETERSON, G. STECHER, M. NEI, S. KUMAR (2011): MEGA 5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution. *28*: 2731-2739.

THOMAS, P.D., M.J. CAMPBELL, A. KEJARIWAL, H. MI, B. KARLAK, R. DAVERMAN, K. DIEMER, A. MURUGANUJAN, A. NARECHANIA (2003): PANTHER: A library of protein families and subfamilies indexed by function. Genome Research, *13*: 2129-2141.

THOMAS, P.D. and A. KEJARIWAL (2004): Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. Proc. Natl. Acad. Sci. USA, *101*: 15398-15403.

THOMAS, P.D., A. KEJARIWAL, N. GUO, H. MI, M.J. CAMPBELL, A. MURUGANUJAN, B. LAZAREVA-ULITSKY (2006): Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res., *34:* 645-650.

TRAN, N.T., I. JAKOVILIC, W-M. WANG (2015): In silico characterisation, homology modelling and structure-based functional annotation of blunt snout bream (Megalobrama amblycephala) Hsp70 and Hsc70 proteins. Journal of Animal Science and Technology, *57*: 44 DOI 10.1186/s40781-015-0077-x.

UGBO, S.B., A. YAKUBU, J.N. OMEJE, B.S. BIBINU, I.S. MUSA, J.O. EGAHI, N.I. DIM (2015): 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, *5*: 145-158.

YAKUBU, A., E.A. ALADE, N.I. DIM (2014): Molecular analysis of solute carrier family 11 member a1 (Slc11a1) gene in ruminants and non-ruminants using computational method. Genetika, *46*: 925-934

YAKUBU, A. (2014): Comparative evaluation of the functionality of non-synonymous substitutions in MHC DRB gene of Nigerian goats. *Proceedings of Tropentag Conference on "Bridging the gap between increasing knowledge and*

decreasing resources", September 17-19, 2014, Czech University of Life Sciences, Prague, Czech Republic (Abstract). P. 198.

YAKUBU, A., A.E. SALAKO, M. DE DONATO, M.I. TAKEET, S.O. PETERS, M.A. ADEFENWA, M. OKPEKU M. WHETO, B.O. AGAVIEZOR, T.M. SANNI, O.O. AJAYI, G.O. ONASANYA, O.J. EKUNDAYO, B.M. ILORI, S.A. AMUSAN, I.G. IMUMORIN (2013): Genetic diversity in Exon 2 at the major histocompatibility complex DQB1 locus in Nigerian indigenous goats. Biochemical Genetics, *51*: 954-966.

ZHAO, Y., H. XU L. SHI, J. ZHANG (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.

PRIMENA KOMPJUTERSKIH ALGORITAMA U PROUČAVANJU FUNKCIONALNOSTI NE-SINONIMNIH SUBSTITUCIJA U MHC DRB GENU NIGERIJSKIH KOZA

Abdulmojeed YAKUBU[1,2,3], Adebowale E. SALAKO[2], Marcos DE DONATO[3.4] i Ikhide G. IMUMORIN[3]

[1]Departman za stočarstvo, Nasarawa državni Univerzitet, Keffi, Shabu-Lafia Campus, P.M.B., Lafia, Nigerija

[2]Laboratorija za oplemenjivanje i genetiku životinja, Departman za stočarstvo, Univerzitet u Ibadanu, Ibadan, Nigerija

[3]Laboratorija za genetiku i genomiku životinja, Internacionalni Pogrami, Koledž za poljoprivredu i prirodne nauke, Kornel Univerzitet, Itaka, Njujork 14853, SAD

[4]Laboratorija za molekularnu genetiku, IBB, Univerzitet Oriente, Kumana, Venecuela

Izvod

Glavni histokompatibilni kompleks (MHC) sadrži visoko varijabilne multi-genske familije, koje imaju ključnu ulogu u adaptivno imunom odgovoru kičmenjaka. Unutar klase II Capra MHC gena, DRB lokus je visoko polimorfan, posebno u egzonu 2, koji kodira mesto vezivanja antigena. Pretpostavlja se da mnogi ne-sinonimni SNP (nsSNP) DRB lokusa kod koza utiču na funkcionalnost proteina. Cilj ovog rada bio je da se uporedi efikasnost šest komjuterskih pristupa u proceni pojedinačnih ne-sinonimnih kodirajućih SNP-ova koji utiču na funkcionalnost proteina. Korišćeni su PANTHER, SNAP, SIFT, PolyPhen-2, PROVEAN i nsSNP Analyzer bioinformatičke tehnike u otkrivanju štetnih i korisnih efekata supstitucije H57G, Y89R, V104D i Y112I na peptidne veze DRB gena kod Nigerijskih koza. Rezultati analize PANTHER-om ukazaju da su supstitucije H57G, Y89R i Y112I ($P_{deleterious}$= 0.113, 0.204 i 0.472) bile korisne, dok su V104D bile štetne ($P_{deleterious}$= 0.756), odnosno nisu bile neutralne. Prema SNAP-u, supstitucije H57G i Y89R, bile su neutralne, sa očekivanom preciznošću od 53 odnosno 69%, dok su V104D i Y112I bile štetne. Supstitucije H57G i Y89R su takođe bile bezopasne, prema SIFT analizi. Samo su H57G (PROVEAN) i V104D (nsSNPAnalyzer) supstitucije aminokiselina označene kao korisne. Interesantno je da je i predviđena 3D struktura nativnog i mutant DRB proteina bila slična dobijenoj Ramachandran plotom. Saglasnost dobijena korišćenjem PANTHER, SNAP, SIFT i PolyPhen-2 pristupa, ukazuje na njihovu moguću primenu u identifikovanju genotipova koza otpornih na bolesti.

Primljeno 06.IV. 2016.
Odobreno 25. XII. 2016.