

# Computational algorithm to assess genetic relationship and functional analysis of non-synonymous substitution of HSP 70 gene of cattle, sheep and goat

## Abstract

This study was carried out using heat shock protein 70 (HSP 70) sequences comprising goat (5), sheep (5) and cattle (5) which was obtained from the GenBank (NCBI). Protein Variant Effect Analyzer (PROVEAN) tool was used to check the non-synonymous single nucleotide polymorphism of the HSP 70 gene of amino acid substitution for cattle, goat and sheep. The study observed that some amino acid substitution returned neutral/beneficial while some deleterious/harmful. Molecular evolutionary genetics analysis (MEGA) 7.0 was used to perform both Tajima's test and phylogenetic relationship for cattle, goat and sheep. The Tajima's test (D) for all the species revealed positive values, an indication of purifying selection. The phylogenetic relationship followed bovidae family speciation. The study concluded that information emanating from this research is significant for further dry and wet laboratory experiment.

**Keywords:** amino acid, gene, hsp 70, nonsynonymous, molecular, speciation, substitution

Volume 5 Issue 6 - 2017

Dauda A,<sup>1</sup> Abbaya HY,<sup>2</sup> Shettima SM,<sup>3</sup> Sinodo S,<sup>3</sup> Adekoya AA,<sup>4</sup> Asiamah S,<sup>4</sup> Malgwi IH<sup>4</sup>  
Hyacinth OU

<sup>1</sup>Department of Animal Science, University of Calabar, Nigeria

<sup>2</sup>Department of Animal Science, Adamawa State University, Nigeria

<sup>3</sup>Department of Animal Science, University of Maiduguri, Nigeria

<sup>4</sup>Faculty of Agricultural and Environmental Sciences, Kaposvar University, Hungary

**Correspondence:** Malgwi IH, Faculty of Agricultural and Environmental Sciences, Kaposvar University, Hungary, Email [zeekofficial@yahoo.com](mailto:zeekofficial@yahoo.com)

**Received:** July 27, 2017 | **Published:** September 11, 2017

## Introduction

Heat shock protein 70 (HSP 70) is significant in providing protection for cells against many stresses. HSP 70 proteins are most available and aware among all the HSPs.<sup>1</sup> Any time cells are under stress stimuli HSP tend to be active.<sup>2</sup> The HSP families consist of HSP 110, HSP 100, HSP 90, HSP70, HSP 60, HSP 40, HSP 10, and small HSP families.<sup>3</sup> Among HSPs, heat shock protein 70 is the most significant because it helps the protein chain to fold correctly into their tertiary shapes, protect and stabilize them. These processes occur in all mammalian cells.<sup>4</sup> With the development of technologies in animal genomics have generated many sequence data of many species. Even with the development these technologies is still challenging to differentiate neutral/beneficial amino acid substitution from the pool of single nucleotide polymorphism.<sup>5</sup> Thus, use of computer to determine harmful non-synonymous amino acid substitution of variant in the sequence exome is important.<sup>6</sup> A large number of computer determinations for amino acid substitutions depend on the basis that protein sequences notice in all living organisms have survived natural selection.<sup>7</sup> Thus, the physical and chemical changes protect amino acid positions across many species are likely to be useful, and amino acid substitutions noticed at determined positions will be deleterious in future on gene functions.<sup>7</sup> In developing countries, such as Nigeria, some quantitative and qualitative measurements have been used for selection and breeding purposes against disease infestation with little or no meaningful improvement in the stocks.<sup>8</sup> This has necessitated the paradigm shift to computational genomics to facilitate the analysis and interpretation of the vast array of molecular data.<sup>8</sup> The reason for this study is to find out non-synonymous single nucleotide polymorphism, evolutionary relationship and test for selection of HSP 70 gene in some selected ruminants animals (cattle, sheep and goats).

## Materials and methods

This study was carried out using heat shock protein 70 (HSP 70) sequences comprising goat (5), sheep (5) and cattle (5) which was obtained from the GenBank (NCBI) ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). The amino acid numbers of the protein are ACR56335, AHX42590, AHX42593, AHZ20721 and AHZ20724 (goat) ABP88261, AEX55799, AOF27447, AOF26313 and AOF25738 (sheep) NP\_776975, CCO75577, NP\_872592, CCO75576 and NP\_001035697 (cattle). Sequences arrangement, changing and comparing of the HSP 70 gene of cattle, goat and sheep were carried out with Clustal W<sup>9</sup> by setting, gap open penalty of 15 and gap extension penalty of 6.66. Computer method was used to determine missense mutations using PROVEAN (Protein Variant Effect Analyzer) with separating value of -2.5. PROVEAN uses a set of similar and distantly related sequences from the NCBI NR protein database using BLASTP (ver.2.2.25) with an E-value separation of 0.1. The sequences were grouped on the basis of sequence similarity of 80% to remove redundancy using the CD-HIT program (ver.4.5.5).<sup>10</sup> PROVEAN score that is smaller than or equal to a given separating value, the polymorphism is predicted as harmful.<sup>7</sup> Evolutionary analyses were conducted in MEGA7. The evolutionary relationship was carried out using the Maximum Likelihood method based on the Equal Input model.<sup>11</sup> The tree revealed the highest log likelihood (-708.1487).

The sequence that are similar are group together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and Bio NJ algorithms to a matrix of pair wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The analysis involved 15 amino acid sequences. The tree is drawn to scale,

with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were deleted. There were a total of 27 positions in the final dataset.<sup>12</sup> The evolutionary distances were computed using the Poisson correction method.<sup>13</sup> The Tajima test statistic<sup>11</sup> was also estimated using MEGA7 was carryout by measuring the difference between two estimators of the population mutation rate,  $q_w$  and  $p$ .<sup>11</sup> Under neutrality, the means of  $q_w$  and  $p$  were approximately equal to one another. Therefore, the expected value of Tajima's D for populations conforming to a standard neutral model is zero. Significant deviations from zero indicate a skew in the allele frequency distribution relative to neutral expectations. Positive values of Tajima's D arise from an excess of intermediate frequency alleles and can result from population bottlenecks, structure and/or balancing selection. Negative values of Tajima's D indicate an excess of low frequency alleles and can result from population expansions or positive selection.<sup>14</sup> The abbreviations used are as follows:  $m$ =number of sites,  $S$ =Number of segregating sites,  $p_s=S/m$ ,  $\Theta=p_s/a_1$ , and  $\pi$ =nucleotide diversity. D is the Tajima test statistic.<sup>11</sup>

## Results

The non-synonymous single nucleotide polymorphism (nsSNP)

of HSP 70 gene cattle, goat and sheep are presented in Table 1-3 respectively. Fifteen (15) amino acid substitutions variant were observed from after arrangement of deduced protein sequence of Cattle. The amino acid substitutions variant (V20L) appeared to be neutral, while the remaining fourteen amino acid substitutions variants were returned as deleterious. Amino acid substitutions in goat revealed that variants (K25E, V26V, E27D, M5Y and A2S) appeared neutral, while variants (D10G, L11G, T13G, Y15S, S16G, C17S, V20G, Q22T and H23I) appeared to be harmful. Amino acid substitutions in sheep showed only two substitutions variants (E23D and C25V) are neutral while the rest of the variants appeared to be harmful. Tajima's test of neutrality for testing selection of cattle, goat and sheep are presented in Table 4. The result revealed that Tajima's test of neutrality (D) is positive for all the species but sheep have the highest nucleotide diversity value (0.923387). The molecular phylogenetic analysis by Maximum Likelihood method is shown in Figure 1. The genetic or evolutionary relationships of the amino acid nucleotides of cattle, goat and sheep as revealed that goat and sheep tend to be closer than cattle. This study concurred with the result of.<sup>9,15</sup>

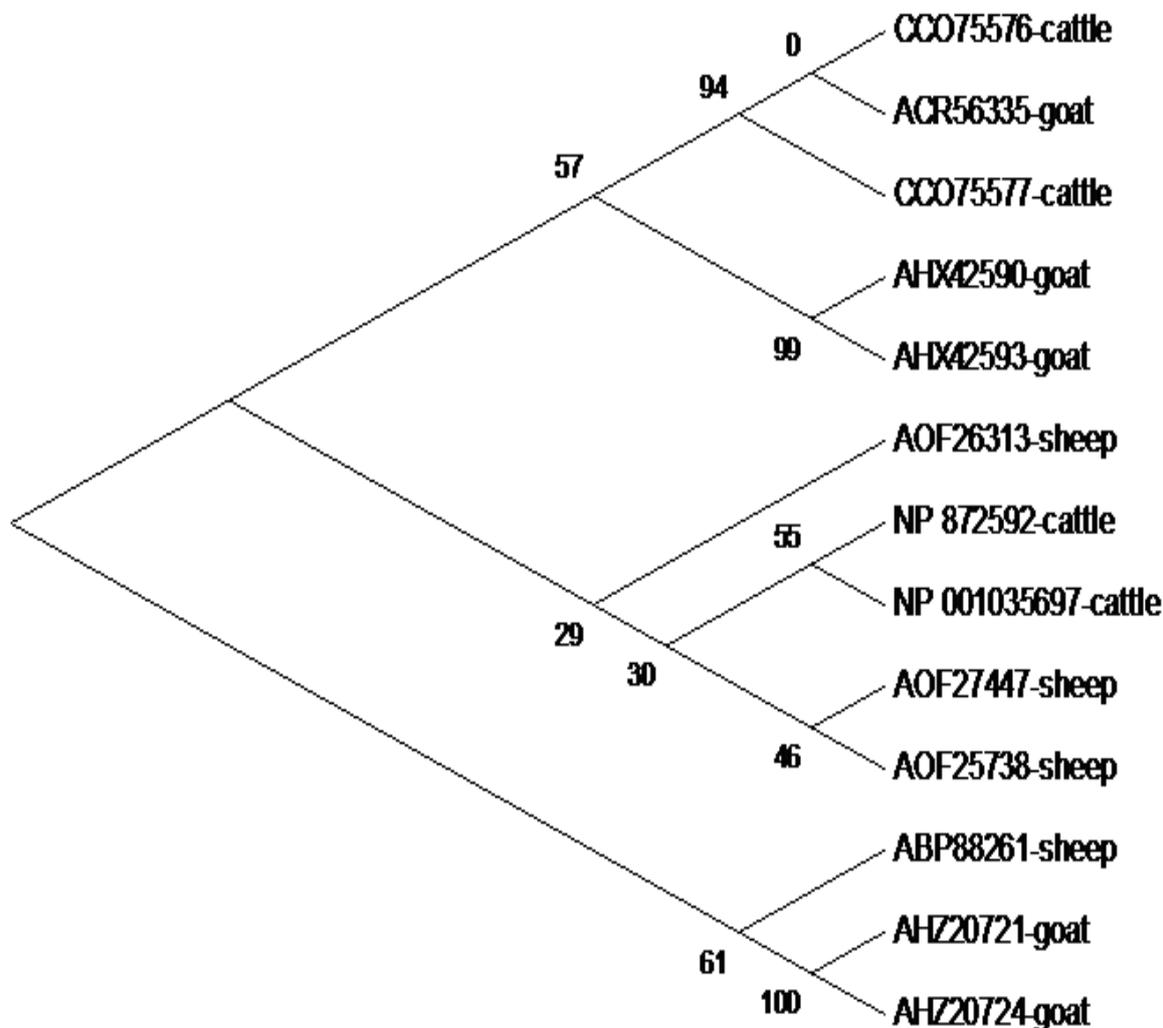


Figure 1 Molecular phylogenetic analysis by maximum likelihood method.

**Table 1** Functional analysis of coding nsSNP of the HSP 70 gene of cattle using PROVEAN

Variant	Provean score	Prediction
V26R	-4.699	Deleterious
G24V	-4.329	Deleterious
F21P	-5.866	Deleterious
V20L	-1.834	Neutral
V18F	-3.427	Deleterious
S16R	-3.155	Deleterious
Y15S	-5.591	Deleterious
T13L	-3.708	Deleterious
G12A	-3.734	Deleterious
DO10V	-5.57	Deleterious
L11P	-4.261	Deleterious
H23D	-3.369	Deleterious
C17A	-5.659	Deleterious
I9R	-4.295	Deleterious
I7P	-3.538	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 HSP 70 gene of goat using PROVEAN

Variant	Provean score	Prediction
D10G	-4.313	Deleterious
L11G	-4.928	Deleterious
T13G	-4.333	Deleterious
Y15S	-5.591	Deleterious
S16G	-2.526	Deleterious
C17S	-6.287	Deleterious
V20G	-4.515	Deleterious
Q22T	-3.288	Deleterious
H23I	-5.308	Deleterious
K25E	-1.872	Neutral
V26V	0	Neutral
E27D	-1.668	Neutral
M5Y	-0.66	Neutral
A2S	-0.033	Neutral

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 HSP 70 gene of sheep using PROVEAN

Variant	Provean score	Prediction
T6E	-5.541	Deleterious
A7N	-4.258	Deleterious
G8Q	-7.454	Deleterious
T10L	-5.566	Deleterious
H11P	-9.255	Deleterious
G13S	-5.721	Deleterious
G14K	-7.553	Deleterious
D16P	-6.718	Deleterious
D18P	-6.721	Deleterious
L21Q	-4.09	Deleterious
E23D	-0.201	Neutral
C25V	-1.142	Neutral
Q27G	-4.339	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 4** Results from Tajima’s Neutrality test

Species	M	S	P <sub>s</sub>	θ	Π	D
Cattle	4	191	0.989637	0.539802	0.784111	4.746763
Sheep	5	27	1.00000	0.480000	0.740741	4.051175
Goat	4	124	1.00000	0.545455	0.923387	7.256215

M, number of sites; S, number of segregating sites; p<sub>s</sub>, S/m, θ, p<sub>s</sub>/a<sub>1</sub>, and π, nucleotide diversity  
D is the Tajima test statistic

## 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).<sup>16</sup> Variation at these sites may affect the antigen binding groove and antigenic-peptide binding ability, and hence peptide specificity.<sup>17</sup> HSP70 genes are located within the major histocompatibility complex class I.<sup>18</sup> The present finding revealed that the HSP 70 gene of bovine, caprine and ovine is highly polymorphic in nature. The amino acid substitution in showed both variants that are neutral and deleterious for all the three species. The neutral substitution is a clear indication that the substitution did not impair the protein structure in terms of structure and function which will help in prevention of susceptibility to disease and also give hope for future selection while the deleterious dose the opposite.<sup>15</sup> Since the amino acid substitution in this study revealed both neutral and deleterious variants, this implies that any attempt to increase the number of beneficial allele there is high frequency of also increasing deleterious allele. The modern genetic analysis of determining singly nucleotide polymorphism offer great hope for future selection stock as it enable to chosen molecular

markers for selection. A positive Tajima's D observed for all the species under study signifies low levels of low and high frequency polymorphisms, an indications of a decrease in population size and/or balancing selection.<sup>19</sup> This might aid in purifying selection i.e. selection against deleterious amino acids substitution variants. This is also an indication that the breeds have not undergone any recent bottleneck or any recent reduction in the effective population size and are at mutation drift equilibrium. The presence of many alleles at a certain Major histocompatibility complex (MHC) locus is prove of long term physical and chemical changes that occur at locus.<sup>8</sup> This is suggesting by frequency with which some alleles in one species are more closely related to the alleles in a closely related species than to the alleles in the same species.<sup>20</sup> The closeness of a gene among ruminants may be as a result of threshold in physical and chemical changes due to the same selection pressure which the ruminants undergo during physical and chemical changes. Information emanating from this research may pave a way for subsequent dry and wet laboratory experiment in order to increase cattle, goat and sheep production in developing country like Nigeria.<sup>21-23</sup>

## Conclusion

The level of molecular diversity showed at the HSP 70 locus of cattle, goat and sheep, the beneficial amino acid substitution, the Tajima test of neutrality for selection and phylogenetic relationship revealed in this study is significant in selection for thermoregulation especially in tropical country like Nigeria.

## Acknowledgements

None.

## Conflict of interest

Author declares that there is no conflict of interest.

## References

1. Gade N, Mahapatra RK, Sonawane A, et al. Molecular Characterization of Heat Shock Protein 70-1 Gene of Goat (*Capra hircus*). *Molecular Biology International*. 2010;2010:108429.
2. Christians ES, Zhou Q, Renard J, et al. Heat shock proteins in mammalian development. *Seminars in Cell and Developmental Biology*. 2003;14(5):283-290.
3. Feder ME, Hofmann GE. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annual Review of Physiology*. 1999;61:243-282.
4. Kiang JG, Tsokos GC. Heat shock protein 70 kDa: molecular biology, biochemistry, and physiology. *Pharmacol Ther*. 1998;80(2):183-201.
5. George Priya Doss C, Rajasekaran R, Sudandiradoss C, et al. A Novel Computational and Structural Analysis of nsSNPs in *CFTR* Gene. *Genomic Med*. 2008;2(1-2):23-32.
6. Liu L, Kumar S. Evolutionary Balancing Is Critical for Correctly Forecasting Disease Associated Amino Acid Variants. *Mol Biol Evol*. 2013;30(6):1252-1257.
7. Choi Y, Sims GE, Murphy S, et al. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*. 2012;7(10):e46688.
8. Ugbo SB, Yakubu A, Omeje JN, 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(4):145-158.
9. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X Version 2.0. *Bioinformatics*. 2007;23(21):2947-2948.
10. Li W, Godzik A. Cd-Hit: A Fast Program for Clustering and Comparing Large Protein or Nucleotide Sequences. *Bioinformatics*. 2006;22(13):1658-1659.
11. Tajima F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123(3):585-595.
12. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33(7):1870-1874.
13. Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. In: V Bryson, HJ Vogel, editors. *Evolving Genes and Proteins*. New York: Academic Press; 1965. p. 97-166.
14. Biswas S, Akey JM. Genomic insights into positive selection. A review. *Trends Genet*. 2006;22(8):437-446.
15. Bibinu BS, Yakubu A, Ugbo SB, et al. Computational Molecular Analysis of the Sequences of BMP15 Gene of Ruminants and Non-Ruminants. *Open Journal of Genetics*. 2016;6(2):39-50.
16. Zhao Y, Xu H, Shi L, et al. Polymorphisms in Exon 2 of MHC Class II DRB3 Gene of 10 Domestic Goats in Southwest China. *Asian-Aust J Anim Sci*. 2011;24(6):752-756.
17. Zhou H, Hickford JG, Fang Q. Polymorphism of the DQA2 Gene in Goats. *J Anim Sci*. 2005;83(5):963-968.
18. Cameron PU, Tabarias HA, Pulendran B, et al. Conservation of the central MHC genome: PFGE mapping and RFLP analysis of complement, HSP70, and TNF genes in the goat. *Immunogenetics*. 1990;31(4):253-264.
19. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993;133(3):693-709.
20. Takeshima S, Chen S, Miki M, et al. Distribution and Origin of Bovine Major Histocompatibility Complex Class II DQA1 Genes in Japan. *Tissue Antigens*. 2008;72(3):195-205.
21. Sreedhar AS, Csermely P. Heat shock proteins in the regulation of apoptosis: new strategies in tumor therapy—a comprehensive review. *Pharmacol Ther*. 2004;101(3):227-257.
22. Wang X, Su L, Pan X, et al. Isolation and cDNA Characteristics of MHC-DRA Genes from Gayal (*Bos frontalis*) and Gaytle (*Bos frontalis* × *Bos taurus*). *Biotechnol Biotechnol Equip*. 2015;29(2):33-39.
23. Zemla A, Kostova T, Gorchakov R, et al. Genesv—An Approach to Help Characterize Possible Variations in Genomic and Protein Sequences. *Bioinform Biol Insights*. 2014;8:1-16.