Journal of Genetics and Genetic Engineering
Volume 2, Issue 1, 2018, PP 18-24

Analysis of Thyroglobulin Gene in Some Ruminants

[1]Dauda, A., Gwaza, D. S. and DIM, N. I.

[1]Department of Animal Science University of Calabar P.M.B 1115 Calabar, Nigeria. Department of Animal Breeding and Physiology, University of Agriculture, Markurdi. P.M.B. 2373 Makurdi, Nigeria.

***Corresponding Author:** D.S. Gwaza, Department of Animal Breeding and Physiology, University of Agriculture Makurdi. Nigeria.*

ABSTRACT

A total of thirteen (13) thyroglobulin nucleotide sequences comprising goats (3), sheep (5) and cattle (5) were retrieved from the GenBank (NCBI) (www.ncbi.nlm.nih.gov) where use undertake a computational molecular analysis of thyroglobulin gene in some ruminant (cattle, sheep and goat). The Genbank accession numbers of the sequences are XP_017914074, NP_001301228 and AEE69371 (goat) XP_012038919, XP_011960995, XP_011985627, XP_011980480 and XP_011976781 (sheep) NP_776308, CAA26090, XP_019829570, DAA22816 and NP_001095535 (cattle). The sequence length variation exhibited high level polymorphism. The functional analysis of cattle were performed using PROVEAN (Protein Variant Effect Analyzer) which indicated that all the amino acid substitution are beneficial, while goat and sheep amino acid substitution exhibited both harmful and beneficial which will give hope for future selection. The Tajima's test of neutrality revealed that cattle showed negative (-ve) value which mean selection purification and positive (+ve) value signifying balance selection.The evolutionary distances were computed using the Poisson correction method. The reliability of the trees was calculated by bootstrap confidence values with 1000 bootstrap iterations using MEGA (Multiple Evolutionary Gene Analysis) 7.0 software and the result revealed that phylogenetic tree were in accordance with the well-known evolutionary history of Bovidae subfamily speciation. This study revealed genetic information which may be relevant in selection for early sexual development, reproductive function so as to improve livestock in developing country like Nigeria.

Keywords: *Thyroglobulin, Gene, Sequence, Computational and Molecular.*

INTRODUCTION

Several studies provide evidence to confirm the role of thyroglobulin (TG) in sexual differentiation and gonadal development (Flood et al., 1990; Duarte-Guterman et al., 2002;Mendis-Handagama and SirilAriyaratne,2008 and Jannini et al., 2009). Wagner et al.(2001) discovered that deiodinases [the enzymes responsible for ioding thymoglobulin (TG) to obtain the active forms triiodothyronine (T3) and thyroxine (T4)] as well as thyroid receptors (TRs, encoded by *trα* and *trβ* genes) are present within gonadal tissues, suggesting that THs must have an action on these organs. The presence of TH machinery in testicular tissues implies that TH axis must regulate aspects of testicular functioning. Indeed, it has been shown that hypo- and hyper-thyroid males exhibit testes and sperm dysfunction (Krassaset. al., 1997). In addition, THs are involved in the regulation of androgen receptors(ARs) expression in testicular tissues through thyroid response elements (TREs) located in the promoter of androgen receptor genes(Flood et al., 1990). Furthermore, THs may also regulate other genes involved in androgen biosynthesis and signaling. THs enhance 5α-reductase expression and activity within the testes, increasing 5α-dihydrotestosterone concentrations (Ram and Waxman, 2002; Duarte-Guterman et al., 2002). One of the key issues in biology understands how natural selection drives gene functional diversification across different species and lineages (Toll-Riera et al., 2011). The main goal in animal breeding is to select individuals that have high breeding values for traits of interest as parents to produce the next generation and to do so as quickly as possible. To date, most programs rely on statistical analysis of large data bases with phenotypes on breeding populations by linear mixed model methodology to estimate breeding values on selection candidates. However, there is a long history of research on the use of genetic markers to identify quantitative trait loci and their use in marker-assisted selection but with limited

implementation in practical breeding programs (Dekker, 2012).The objective of this study, therefore, was to investigate using computational method, the molecular genetic variation of thyroglobulin gene of some ruminants (cattle, sheep and goat) with a view to providing relevant genetic information for marker assisted selection in the studied species.

MATERIALS AND METHODS

A total of thirten (13) thyroglobulin nucleotide sequences comprising goat (3), sheep (5) and cattle (5) were retrieved from the Gen Bank (NCBI) (www.ncbi.nlm.nih.gov). The Gen bank accession numbers of the sequences wereXP_017914074, NP_001301228 and AEE69371(goat)XP_012038919,XP_011960995, XP_011985627,XP_011980480andXP_011976781(sheep)NP_776308,CAA26090,XP_019829570,DAA22816andNP_001095535(cattle).Sequences alignment, translation and comparison of the thyroglobulin gene of the various species was done with ClustalW as described by Larkin *et al*. (2007) using IUB substitution matrix, gap open penalty of 15 and gap extension penalty of 6.66**.** *In silico* functional analysis, missense mutations was obtained 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 optimal tree with the sum of branch length = 5.45108648 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are 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 13 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 63 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016).The Tajima test statistic (Tajima, 1989) was estimated using MEGA7. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). 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 π = nucleotide diversity. D is the Tajima test statistic (Tajima, 1993).

RESULTS

Table1.*Accession Number and Sequence length variation of thyroglobulin*

Species	Protein Accession Number	Base Pair Number (bp)	Sequence Length Variation
Cattle	NP_776308	2769	930 – 2769
	CAA26090	930	
	XP_019829570	2770	
	DAA22816	2622	
	NP_001095535	1305	
Goat	XP_017914074	2767	271 – 2767
	NP_001301228	317	
	AEE69371	271	
Sheep	XP_012038919	2767	1241 – 2767
	XP_011960995	1241	
	XP_011985627	2767	
	XP_011980480	1241	
	XP_011976781	1305	

The sequence length variation of cattle, sheep and goat are presented in Table 1. The variation in sequence length in base pair (bp) of cattle protein ranges between 930bp and 2769bp. The variation in sequence length in base pair (bp) of goat protein ranges between 271bp and 2767bp. The variation in sequence length in base pair (bp) of sheep protein ranges between 1241bp and 2767bp. All the species sequence exhibited partial complete coding sequence (<6000bp) from the DNA whose length were less than three thousand base pair (<3000bp). The results of functional analysis of coding ns SNP of thyroglobulin of goat, sheep and cattle are

presented in Table 1,2 and 3 respectively. Thirteen (13) amino acid substitutions of the wild type alleles located in the coding region of goats were obtained from the alignment of deduced amino acid sequences of goat. Out of which, only five amino acid substitutions (A2Q, V7G, G9A, S18L and A41P) were returned neutral indicating that the substitutions did not impair protein function, while the remaining eight amino acid substitutions (N20L, Q25G, P30G, Q37F, Y49A, C52P, D55G and F58A) were returned as deleterious, an indication that the substitutions were harmful. In sheep, nine amino acid substitutions (W6R, F8S, L11A, A14L, S18L and A43D) were returned neutral, an indication that they did not impair protein function while the remaining amino acid substitution(Q29L, C34L, E39P, Y48A, C52L and G56A) was predicted to be harmful. All the fourteen amino acid substitutions in cattle appeared beneficial. The evolutionary relationships are shown in figure 1.The genetic relationship of thyro globulin gene using phylo genetic showed that members of the Bovidae family. Here, sheep is close to goat than cattle. The genetic relationships of the proteins of goat, sheep and cattle as revealed by the phylo genetic tree were in accordance with the well-known evolutionary history of Bovidae subfamily speciation (Floudas, 2007). The Tajima's neutrality test is shown in Table 5. Cattle showed negative value of D while sheep and goat showed positive value of D.

Table2.*Functional analysis of coding nsSNP of the thyroglobulin gene of goat using PROVEAN*

Variant	PROVEAN Score	Prediction
A2Q	-0.861	Neutral
V7G	-1.206	Neutral
G9A	0.257	Neutral
S18L	-1.590	Neutral
N20L	-3.603	Deleterious
Q25G	-2.950	Deleterious
P30G	-4.016	Deleterious
Q37F	-3.300	Deleterious
A41P	-2.416	Neutral
Y48A	-3.860	Deleterious
C52P	-5.800	Deleterious
D55G	-3.183	Deleterious
F58A	-3.748	Deleterious

Default threshold is −2.5, that is; Variants with a PROVEAN score equal to or below −2.5 are considered "deleterious" while Variants withPROVEAN 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 thyroglobulin gene of sheep using PROVEAN*

Variant	PROVEAN Score	Prediction
W6R	-1.297	Neutral
F8S	-0.699	Neutral
L11A	-1.476	Neutral
A14L	0.200	Neutral
S18L	-1.590	Neutral
Q29L	-3.283	Deleterious
C34L	-4.833	Deleterious
E39P	-2.733	Deleterious
A43D	-1.317	Neutral
Y48A	-3.860	Deleterious
C52L	-4.833	Deleterious
G56A	-2.900	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. *Functional analysis of coding nsSNP of the thyroglobulin gene of cattle using PROVEAN*

Variant	PROVEAN Score	Prediction
M1M	0.00	Neutral
W6W	0.00	Neutral
E46E	0.00	Neutral
Q62Q	0.00	Neutral
C34C	0.00	Neutral
D27D	0.00	Neutral
I14I	0.00	Neutral
G9G	0.00	Neutral
V61V	0.00	Neutral
F58F	0.00	Neutral
A28A	0.00	Neutral
N20N	0.00	Neutral
K44K	0.00	Neutral
L43L	0.00	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.

Table5. *Results from Tajima's Neutrality test*

Species	M	S	Ps	Θ	Π	D
Cattle	5	179	0.890547	0.427463	0.357711	-1.243992
Sheep	5	1234	0.994359	0.477293	0.752619	4.412808
Goat	5	109	0.990909	0.475636	0.889091	6.610065

m = number of sites, S = Number of segregating sites, p_s = S/m, $\Theta = p_s/ a_1$, and π = nucleotide diversity. D is the Tajima test statistic

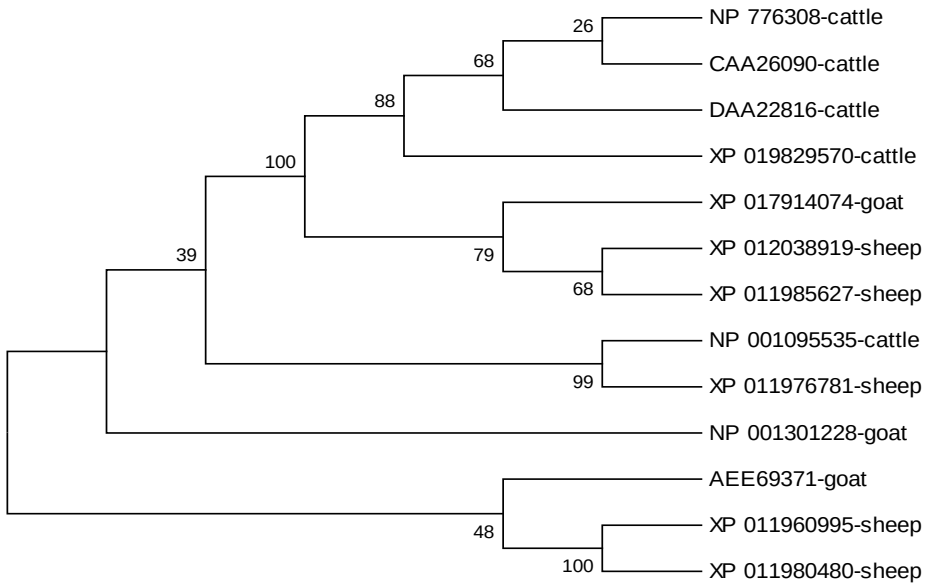

Figure1. *Evolutionary relationships of goat, sheep and calttle nucleotide sequence..*

DISCUSSION

Puberty is a stage of sexual development determined by the interaction of many loci and environmental factors. Identification of genes contributing to genetic variation in this character can assist with selection for early puberty, improving genetic progress in livestock breeding (Fernández, 2014). The variation in sequence length within and among species might results from evolution and differentiation (Yakubu *et al*., 2014). There are cases where variability might results from DNA duplication, DNA rearrangement, short tandem repeat (STR), insertions or deletion of sequences (Vincent *et al*., 2014). The length variation observed within and across species might also attributed to differences in the genomic

region where the sequences were obtained from and differences due to complete coding or partial coding (Dauda *et al*., 2016). In this study the protein sequences are partial coding sequences (CDS) from DNA and had sequence length that are less than six thousand base pair (<6000bp)(Dauda *et al*., 2016).This variability might initiate unique structures between individual members in conferring different biological activities (Dauda *et al*., 2016). This frame shift mutation often results in a completely different translation from the original protein and is also likely to cause a stop codon to be read which truncates further synthesis of protein (Williams and Eerne green, 2013).

The functional analysis of coding non synonymous polymorphism (nsSNP) of the thyroglobulin gene of goat, sheep and cattle presented in Table 2-4 is important to consider the neutral or beneficial amino acid substitutions. These substitutions 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. 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). The beneficial nsSNPs obtained in the present study, therefore, offer hope for future genetic improvement of goats, sheep and cattle at the thyroglobulin locus. This is due to the fact that nsSNPs have been reported to be linked to economically important traits and disease development (Vincent, 2014). Some SNP benefits are; have direct effects on sexual development, reproductive function and associated molecular mechanisms and pathways (Fernández, 2014). Tariq *et al*. (2013) noted that the prediction of SNPs status is promising in modern genetics analysis and breeding programmes as they have been used to identify that animalswith higher breeding value.Genetic improvement and prediction using genetic markers is a goal for livestock breeding. The identification of new genes and/or mutations contributing to genetic variation can assist selection by reducing the generation interval and increasing fertility and genetic progress (Johnston *et al*., 2009: Fortes *et al*., 2012). The phylo genetic tree in figure 1.Revealedclustering and some level of intermingling between sheep cattle and goat sheep. The implication of the similarities in the proteins is that any selection programmes design for early puberty and improving genetic progress of sheep might give inside for making genetic progress in goat and cattle. This is an evidence of trans-species evolution which might be attributed to the coding nature of the sequences. A negative Tajima's D signifies an excess of low frequency polymorphisms relative to expectation, indicating population size expansion (bottleneck or a selective sweep) and/or purifying selection. A positive Tajima's D signifies low levels of low and high frequency polymorphisms, indicating a decrease in population size and/or balancing selection (Fu and Li, 1993). This might aid in purifying selection i.e. selection against non-synonymous mutations because of their deleterious effect.

CONCLUSION

This study revealed that sequence length variation exist between and across specie for thyroglobulin gene. The thyroglobulin gene is polymorphic that has many mutation which some can be harmful or beneficial. The Tajima's D for cattle is positive signifying purifying selection and for sheep and goat is negative indicating balance selection. The phylo genetic analysis revealed they are closely related. This study revealed genetic information which may be relevant in selection for early sexual development, reproductive function so as to improve livestock in developing country especially Nigeria.

REFERENCE

[1] 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

[2] Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. and Chan, A.P. (2012) Predicting the Functional Effect of Amino AcidSubstitutions and Indels. *PLoS ONE*, **7**, e46688

[3] Dauda, A., Abbaya, H. Y., Malgwi, I. H. and Abare, E. A. (2016).Sequence Length Variation and Transmembrane Analysis of Contagious Bovine Pleuropneumonia (CBPP) and Contagious CaprinePleuropneumonia (CCPP) Proteins.*Journal of Dairy, Veterinary & Animal Research.*4(3): 00121

[4] Dekkers, J.C.M. (2012) Application of Genomics Tools to Animal Breeding.*Current Genomics*, 13:207-212.

[5] Duarte-Guterman, P., Navarro-Martín, L. and Trudeau, V. L. (2014) Mechanisms of crosstalk between endocrine systems: Regulation of sex steroid hormone synthesis and action by

thyroid hormones. *Gen Comp Endocrinol*. doi:10.1016/j.ygcen.2014.03.015

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

[7] Fernández, M.E., Goszczynski, D.E., Prando, A.J., Garcia, P. P., Baldo, A., Giovambattista, G.And Liron, J. P. (2014).Assessing the association of single nucleotide polymorphisms in thyroglobulin gene with age of puberty in bulls, J AnimSciTechnol, 56: 17

[8] Flood, D. E., Fernandino, J. I. andLanglois, V. S. (2013). Thyroid hormones in male reproductive development: evidence for direct crosstalk between the androgen and thyroid hormone axes. *Gen. Comp.Endocrinol*. 192: 2-14.

[9] Floudas, C. A. (2007) Computational methods in protein structure prediction. *Biotechnology and Bioengineering* 97(2): 207-213.

[10] Fortes, M. R. S., Lehnert, S. A., Bolormaa, S., Reich, C., Fordyce, G., Corbet, N. J., Whan, V., Hawken, R. J. andReverter, A. (2012).Finding genes for economically important traits: Brahman cattle puberty.Anim Prod Sci., 52: 143-150.

[11] Fu, Y. X. and Li, W. H. (1993).Statistical tests of neutrality of mutations. *Genetics.*133 (3): 693–709.

[12] Jannini, E. A., Crescenzi, A., Rucci, N., Screponi, E., Carosa, E., De Matteis, A., Macchia, E, D'Amati, G. and D'Armiento, M. (2000).Ontogenetic pattern of thyroid hormone receptor expression in the human testis.*J.Clin.Endocrinol.Metab,* 85: 3453-3457.

[13] Johnston, D. J., Barwick, S.A., Corbet, N.J., Fordyce, G., Holroyd, R.G., Williams, P.J. and Burrow, H. M. (2009).Genetics of heifer puberty in two tropical beef genotypes in northern Australia and associations with heifer and steer-production traits. Anim. Prod. Sci., 49: 399-412.

[14] Krassas, G. E., Papadopoulou, F., Tziomalos, K., Zeginiadou, T.Pontikides, N. (2008). Hypothyroidism has an adverse effect on human spermatogenesis: A prospective, controlled study. *Thyroid.* 18:1255-1259.

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

[16] Larkin, M.A., Blackshields G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace,I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007).Clustal W and Clustal X Version2.0.*Bioinformatics*, **23**, 2947-2948

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

[18] Mendis-Handagama, S. M. andSirilAriyaratne, H. B.(2005).Leydig cells, thyroid hormones and steroidogenesis. *Indian J. Exp. Biol.* 43 (Suppl11): 939-962

[19] Mullur, R., Liu, Y.Y. and Brent, G. A. (2014) Thyroid hormone regulation of metabolism.Physiol Rev., 94 (Suppl2): 355-382.CrossRefPubMedPubMedCentralGoogle Scholar

[20] Nei, M. and Kumar, S. (2000).Molecular Evolution and Phylogenetics. Oxford University Press, New York.

[21] Ram, P.A. and Waxman, D. J. (1990)Pretranslational control by thyroid hormone of rat liver steroid 5 alpha-reductase and comparison to the thyroid dependence of two growth hormone-regulated CYP2C mRNAs. *J. Biol. Chem.* 265 (31):19223-19229.

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

[23] Sugawara, T., Kiriakidou, M., McAllister, J. M., Holt, J. A., Arakane, F. and Strauss, J. F. (1997). Regulation of expression of the steroidogenic acute regulatory protein (StAR) gene: a central role for steroidogenic factor 1. Steroids. 62: 5-9.

[24] Tajima, F. (1989).Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

[25] Tariq, A.M., Al-Shammari, A.S., Al-Muammar, M.N. and Alhamdan, A.A. (2013) Evaluation and Identification ofDamaged SNPs in COL1A1 Gene Involved in Osteoporosis. *Archives of Medical Science*, 9:899-905.

[26] Toll-Riera, M., Laurie, S. and Alba, M.M. (2011) Lineage-Specific Variation in Intensity of Natural Selection inMammals. *Molecular Biology and Evolution*, 28:383-398

[27] Vincent ST, Momoh OM, Yakubu A (2014) Bioinformatics analysis of beta-casein gene in some selected mammalian species. Research Opinions in Animal and Veterinary Sciences 4(10): 564-570.

[28] Wagner, M. S., Wajner, S. M., Maia, A. L. (2009). Is there a role for thyroid hormone on spermatogenesis?.*Microsc Res. Tech.*72: 796-808.

[29] Williams, L.E. and Eernegreen, J.J. (2013) Sequence Context of Indel Mutations and Their Effect on Protein Evolutionin a Bacterial Endosymbiont. *Genome Biology and Evolution*, 5:599-605.

[30] Yakubu, A., Alade, E. D. and Dim, N. I. (2014.)Molecular analysis of Solute Carrier Family 11 A1 (SLC11A1) gene in ruminants and non-ruminants using computational method.*Genetika* 46(3): 925-934.

[31] Zhao, L., Bakke, M., Krimkevich, Y., Cushman, L. J., Parlow, A. F., Camper, S. A. and Parker, K. L. (2001)Steroidogenic factor 1 (SF1) is essential for pituitary gonadotrope function. *Development.*128 (2):147-154.

[32] Zhou, R. J., Bonneaud, N., Yua, C. X., Barbara, P. D., Boizet, B., Tibor, S., Scherer, G., Roeder, R. G., Poulat, F. and Berta, P. (2002). SOX9 interacts with a component of the human thyroid hormone receptor-associated protein complex. *Nucleic Acids Res.*30 (14):3245-3252.

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