Journal of Animal & Plant Sciences, 31(6): 2021, Page: 1548-1558
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2021.6.0358

COMPUTATIONAL ANALYSIS OF PHYLOGENETIC RELATIONSHIP OF TAGAP - T CELL ACTIVATION RHO GTPASE ACTIVATING PROTEIN IN SELECTED MAMMALIAN SPECIES

S. H. Abbas[1], A. Mumtaz[1], M. J. ul Hasanain[1], M. E. Babar[2] and M. T. Pervez*[1],

[1]Department of Bioinformatics and Computational Biology, Virtual University of Pakistan
[2]Department of Genetics, Virtual University of Pakistan
*Corresponding author's email: m.tariq@vu.edu.pk

ABSTRACT

The present study was conducted to examine TAGAP gene of selected mammalian species (Cattle, White tufted ear marmoset, Dog, American Beaver, Beluga whale, Horse, Western European Hedgehog, Western lowland Gorilla, Human, Yangtze River Dolphin, Golden Hamster, Mouse, Yangtze finless porpoise, American Mink, Pacific Walrus, Duckbill platypus, Rat, Bolivian Squirrel Monkey, American black bear, Polar Bear and Western Clawed Frog) by evaluating similarity and identity, evolutionary relationship, physicochemical properties, secondary and tertiary structures, and protein-protein interaction. Amino acid and nucleotide sequences were retrieved using UniProt and NCBI. Using MUSCLE integrated in IVisTMSA the MSA (Multiple Sequence Alignment) was produced, while identity and similarity percentage were determined by E-SICT integrated in IVisTMSA. At the end of the experiment, the results showed that there was lowest similarity and identity percentage (40%) among frog and other selected mammalian species. The evolutionary relationship of human and gorilla, mouse and rat evolves from same cluster while all other mammalian species expands cluster according to their related classes. Physicochemical properties of TAGAP protein revealed that Dog TAGAP protein was the least theoretical pI (5.44) while Duckbill Platypus was the highest (8.46) and protein was found to be not only an intra-cytoplasmic protein but also a soluble protein in all selected mammalian species. The secondary structure included the alpha helix, extended strand, and random coil. It can be concluded that TAGAP gene has identical homologue, functional similarity and highly conserved in all selected mammalian species except Western Clawed Frog.

Keywords: TAGAP gene, *in-silico* analysis, evolution, motifs, STRING, Protein-Protein Interaction (PPI)
Published first online March 31, 2021 Published final Nov. 20, 2021.

INTRODUCTION

T Cell Activation Rho GTPase Activating Protein (TAGAP) is located at long (q) arm at 25.3 of chromosome 6 (Garner *et al*. 2009) having 200KB block of linkage disequilibrium (Arshad *et al*. 2018). TAGAP is a protein coding gene, which belongs to the superfamily of Rho GTPase-activator protein and has tremendous role not only in cell cytoskeletal association but also in neuronal development and synaptic capacities (Mao *et al*., 2004); (Amundsen *et al*., 2010); (O'Leary *et al*., 2016); (Berge *et al*., 2016). Genetic mutations in TAGAP are associated with the small-intestine disorder, celiac diseases (CD). TAGAP regulates the Rho GTPase cycle, among active GDP-bound and inactive GDP bound. GEFs catalyze the exchange of GDP-bound active state, GEF catalyze the activation of MLC (Myosin Light Chain) and LIMK (LIM Domain Kinase) which improves the focal adhesion and rearrange actin cytoskeleton. CD involves the alteration of actin cytoskeleton and cell shape. When TAGAP inversely regulates the Rho GTPase cycle, actin cytoskeleton become unstable, alters the function and promote CD (Huang *et al*., 2017). Polymorphism in TAGAP induces many autoimmune diseases. Moreover, risk loci of TAGAP responsible for several diseases, including rheumatoid arthritis (O'Leary *et al*., 2016), crohn's disease (Festen *et al*., 2011), celiac disease (Festen *et al*., 2011), and multiple sclerosis (Berge *et al*., 2016). The recent study suggested that targeting of TAGAP is more promising option for future diagnosis against crohn's disease (Festen *et al*., 2011), Type I Diabetes (Xu *et al*., 2013) and multiple sclerosis (Berge *et al*., 2016). Rheumatoid arthritis (RA) is an autoimmune disease chronic inflammation of bone which leads to bone damage and finally loss of function. TAGAP is a shared risk factor between Rheumatoid arthritis, Type I Diabetes and celiac disease (Arshad *et al*., 2018). Several transcripts downregulated in Down Syndrome (DS) code for proteins involved in T-cell and B-cell receptor signaling (e.g., TAGAP, CD46, LY6E, FOD) (Xu *et al*., 2013). The recent study suggested that targeting of TAGAP is more promising option for future diagnosis against celiac disease, Rheumatoid arthritis and Type I Diabetes.

TAGAP gene reveals high level of divergence within mammalian species that may influence its

structural composition and functionality in various organisms. With this in view, in the current study, we analyzed, TAGAP gene of selected mammalian species (Asia specific) with respect to physiochemical properties, identity and similarity percent of amino acid sequences, phylogenetic relationship in addition, domain prediction and their protein-protein interaction. Therefore, an attempt was made in the current work to understand the gene's functions, structure, and phylogenetic relationship.

MATERIALS AND METHODS

This study was conducted from June 2019 to October 2019 at Department of Bioinformatics and Computational Biology, Virtual University of Pakistan with an objective of characterization of TAGAP in 21 selected mammals under *in-silico* platform.

Retrieval of TAGAP nucleotides and amino acids sequences: Amino acid and nucleotide sequences of *Bos Taurus* - cattle (Q08DB1), *Callithrix jacchus* - White tufted ear marmoset (F6PUL3), *Canis lupus* - Dog (J9NY68), *Castor canadensis* - American Beaver (A0A250YL04), *Delphinapterus leucas* - Beluga whale (A0A2Y9M287), *Equuas caballus* – Horse (A0A3Q2I277), *Erinacells european* - Western European Hedgehog (A0A1S2ZAC0), *Gorilla-gorilla* - Western lowland Gorilla (G3QPM6), *Homo sapiens* - Human (Q8N103), *Lipotes vexillifer* - Yangtze River Dolphin (A0A340XR51), *Mesocrietus auratus* - Golden Hamster (A0A1U7QEL1), *Mus musculus* - Mouse (B2RWW0), *Neophocaena asiaeorirntails* - Yangtze finless porpoise (A0A341BM46), *Neovison vison* - American Mink (U6CTG1), *Odobenus rosmarus divergens* - Pacific Walrus (A0A2U3X459), *Ornithorhynchus anatinus* - Duckbill platypus (F7BJJ5), *Rattus norvegicus* - Rat (A0A128DYU5), *Saimiri boliviensis* - Bolivian Squirrel Monkey (A0A2K6U8B8), *Ursus americanus* - American black bear (A0A452SAW4), *Urus maritimus* - Polar Bear (A0A384BVW0) and *Xenopus tropical* - Western Clawed Frog (F7AKF6) were retrieved in FASTA format from National Center for Biotechnology Information - NCBI (Acland *et al*. 2014) and UniProt (Bateman *et al*. 2017).

Multiple sequence alignment: All sequences (Amino acid) of the 21 selected mammals were aligned/mapped with MUSCLE - Multiple Sequence Comparison by Log-Expectation integrated in IVisTMSA (Pervez *et al*., 2015).

Similarity and identity percent determination: The similarity and identity percent identification of TAGAP sequences (amino acid) in all 21 selected mammals was determined by performing a pairwise sequence similarity using E-SICT integrated in IVisTMSA (Pervez *et al*., 2015).

Phylogenetic analysis: The phylogenetic (Evolutionary) relationship utilizing the sequences (amino acid) of all selected mammals obtained from UniProt was performed according Tamura technique. By using Molecular Evolutionary Genetic Analysis (MEGA - X) (Kumar *et al*., 2018), the evolutionary phylogenetic relationship was inferred. The phylogenetic relationship between 21 selected mammals was evaluated using statistical method of maximum probability (Likelihood) estimation based on WAG model (Whelan and Goldman 2001). With bootstrap method of 1000 replication, the accuracy of the deduced phylogenetic tree was assessed.

Determination of physicochemical properties of TAGAP protein: The physicochemical characteristics of all sequences (amino acid) of TAGAP were calculated utilizing ProtParam embedded in ExPASy - Expert Protein Analysis System tool. AA: Amino Acid, MW: Molecular weight, Th pI: Theoretical pI, NCR: negatively charged residues (Aspartic acid + Glutamic acid), PCR: positively charged residues – (Arginine + Lysine), Extinction coefficients, abs: absorbance, Insl: instability index, AI: aliphatic index and GRAVY: Grand average of hydropathicity of TAGAP among the 21 selected mammals were estimated on the ExPASy server. Furthermore, SOSUI (Hirokawa *et al*., 1998) and WoLF PSORT (Horton *et al*., 2007) software predicated the solubility and subcellular location of TAGAP protein.

Structural characterization of TAGAP: The 21 selected mammals' secondary structure "Alpha helix, extended strands and Random coils" of TAGAP protein was evaluated using GORIV (Prediction method for secondary structure) (Garnier *et al*., 1996); (Singh al., 2013).Furthermore, the TAGAP gene 3D structure (Tertiary Structure) was predicted on the principle of the canonical a.a (amino acid) sequence acquired from phyre2 according to Kelley and Stemberg (Kelley *et al*., 2015).

Motif scanning and its functional analysis: PROSITE (Gasteiger *et al*., 2003) software was used to predict motifs in the sequences (amino acid) of TAGAP protein of 21 selected mammals; all sequences were scanned for matches individually.

Prediction of Protein - Protein Interaction (PPI): To determine the particular interaction of the TAGAP protein during speciation with other molecules due to biochemical occurrences, we used the protein sequences of TAGAP from each mammalian species. We used STRING (Szklarczyk *et al*., 2015) in order to generate various networks and predict protein association with other group of protein. This is essential to examine the diversity formed by evolution in TAGAP gene. In addition, Venn diagram were designed to compare and view overlapping PPI (Protein-Protein Interaction) by using two online tools Venny 2.1

(http://bioinfogp.cnb.csic.es/tools/venny/) (Oliveros, 2007) and (http:// bioinformatics.psb. ugent.be/software /details/Venn-Diagrams).

RESULTS

Retrieved nucleotide and amino acids sequences of TAGAP gene: The sequences showed variations in amino acid sequences length. TAGAP amino acid sequences showed a large variation in sequences from *Bos Taurus* to *Xenopus tropicalis* in bps (base pairs), the sequences of amino acid length varied from 701-914 residues of amino acid. The residue length of amino acid for the 21 selected mammals; *Homo sapiens, Gorilla* and *Saimiri boliviensis*; *Canis lupus*, *Odobenus rosmarus divergens*, *Ornithorhynchus anatinus, Ursus americanus* and *Ursus maritimus*; *Delphinapterus leucas* and *Neophocaena asiaeorientalis*; *Lipotes vexillifer* and *Neovison vison*; *Mus musculus* and *Rattus norvegicus* were the same 731; 723; 722; 721; and 714, respectively as shown in Table 1.

Table 1 Detailed information of obtained amino acid sequences of 21 selected mammals.

Mammals	UniProtKB Accession No.	Amino Acid Sequence Length	Taxonomic Identifier [NCBI]	Proteomic database (PaxDb ID)
Cattle *(Bos taurus)*	Q08DB1	716	9913	UP000009136
White tufted ear marmoset *(Callithrix jacchus)*	F6PUL3	730	9483	UP000008225
Dog *(Canis lupus)*	J9NY68	723	9615	UP000002254
American Beaver *(Castor Canadensis)*	A0A250YL04	715	51338	-
Beluga Whale *(Delphinapterus leucas)*	A0A2Y9M287	722	9749	UP000248483
Horse *(Equus caballus)*	A0A3Q2I277	914	9796	UP000002281
European Hedgehog *(Erinaceus europaeus)*	A0A1S2ZAC0	701	9365	UP000079721
Western Lowland Gorilla- (Gorilla gorilla gorilla)	G3QPM6	731	9595	UP000001519
Human *(Homo sapiens)*	Q8N103	731	9606	UP000005640
Yangtze River Dolphin *(Lipotes vexillifer)*	A0A340XR51	731	118797	UP000265300
Golden Hamster *(Mesocricetus auratus)*	A0A1U7QEL1	707	10036	UP000189706
Mouse *(Mus musculus)*	B2RWW0	714	10090	UP000000589
Yangtze Finless Porpoise *(Neophocaena asiaeorientalis)*	A0A341BM46	722	1706337	UP000252040
American Mink *(Neovison vison)*	U6CTG1	721	452646	-
Pacific Walrus *(Odobenus rosmarus divergens)*	A0A2U3X459	723	9708	UP000245340
Duckbill Platypus *(Ornithorhynchus anatinus)*	F7BJJ5	723	9258	UP000002279
Rat *(Rattus norvegicus)*	A0A128DYU5	714	10116	-
Bolivian Squirrel Monkey *(Saimiri boliviensis)*	A0A2K6U8B8	731	39432	UP000233220
American Black Bear *(Ursus americanus)*	A0A452SAW4	723	9643	UP000291022
Polar Bear *(Ursus maritimus)*	A0A384BVW0	723	29073	UP000261680
Western Clawed Frog *(Xenopus tropicalis)*	F7AKF6	780	8364	UP000008143

Similarity and identity percent determination among amino acid sequences of TAGAP gene: The identity and similarity percentage of amino acid sequences depict significant differences (**Fig. 1.**). Percentage identity among amino acid sequences of Polar Bear, American Black Bear; Finless Porpoise, Beluga Whale; Human and Gorilla revealed over 90% identity with each other while percent similarity showed more than 80% (**Fig. 1.**). Similarity and Identity matrix displayed that Frog (*Xenopus tropicalis*) and Horse (*Equus caballus*) had less than 40% and 50% respectively similarity and identity with all other TAGAP protein sequences of selected mammals. However, the human had a higher percentage with Gorilla, Squirrel Monkey as well as Tufted ear marmoset (Fig. 1.).

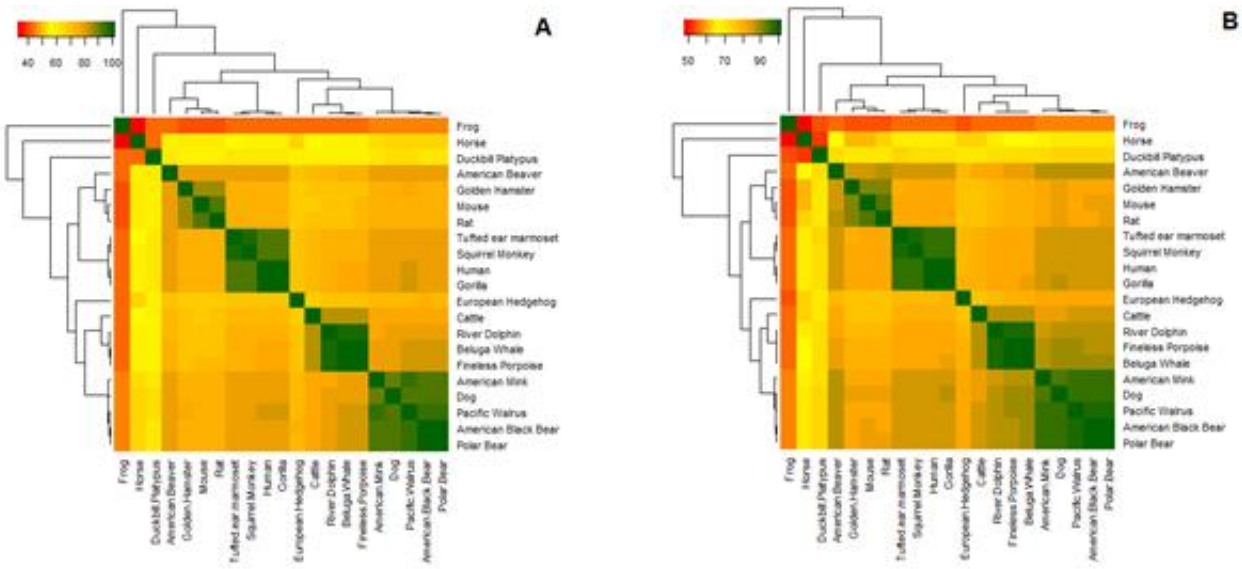

Figure 1 Heatmap of Identity (A) and Similarity (B) of all 21 selected mammalians species

Phylogenetic relationship of TAGAP gene among the 21 selected mammals: The phylogenetic analysis was performed to reveal the evolutionary relationship among the different mammalian species. In this study of phylogenetic analysis, MEGA X describes many models for selected mammals but model with low score of BIC according to JTT + G Jones Taylor Thornton with Gamma distribution was perceived to be the MEGA X finest model to describe the pattern of substitution. The evolutionary relationship analysis of the selected mammals revealed that the TAGAP originated from common ancestral root but diverged into two significant clades. Clade A subdivided into two sub clades in the course of evolution. The evolutionary analysis of TAGAP amino acid sequences disclose that the Human and Gorilla, Mouse and Rat were much closer to each other as shown in (**Fig. 2.)**, So it can be reported that this protein is conserved in two species having similar function (Sahoo *et al*., 2018). Although human and all other selected mammalian species when compare with European hedgehog showed different behaviors and more time of divergence which indicate more adaptive changes of protein (P. S. Soltis and Soltis 2003); (D. E. Soltis and Soltis 2003). Overall, Phylogenetic tree indicated that both cluster (Clade A and Clade B) are functionally conserved during evolution (Christin *et al*. 2012).

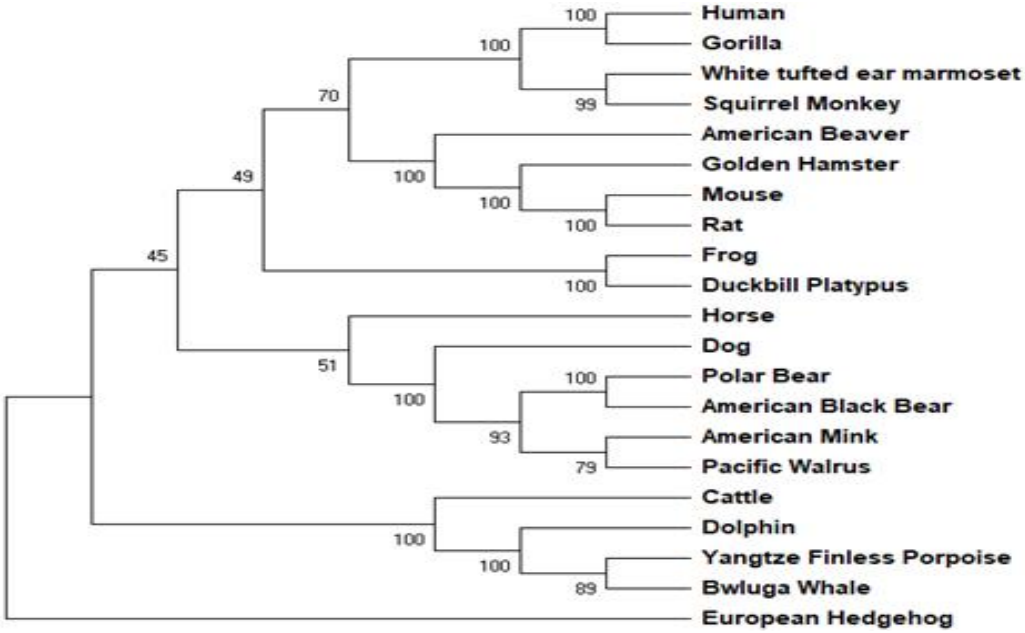

Figure 2 Phylogenetic tree showing evolutionary relationship among mammalians species of TAGAP

Physicochemical properties of TAGAP protein: The ProtParam tool was utilized to determine physical and chemical characteristics of TAGAP sequences (amino acid) (**Table 2**). The physiochemical analysis of TAGAP protein of 21 selected mammalians displayed different physical and chemical characteristics. The outcome revealed that *Canis lupus* (Dog) TAGAP protein was the least ThpI - theoretical pI (5.44) while *Ornithortynchus anatinus* (Duckbil Platypus) TAGAP protein was the highest (8.46). The molecular weights of TAGAP proteins from the 21 selected mammalian species revealed that human TAGAP protein weigh 8703.01 kDa while *Bos Taurus* (Cattle), *Mesocricetus auratus* (Golden Hamster) and *Mus musculus* (Mouse) TAGAP Protein weigh 78410.32, 78575.93, 78920.09 kDa respectively. *Erinaceus europaeus* (European Hedgehog) TAGAP protein had a weight of 76773.61 kDa, which was the least as shown in **Table 2**. Furthermore, the number of amino acids in TAGAP protein of the mammalian species revealed that Human, Western Lowland Gorilla and Bolivian Squirrel Monkey (**Table 2**). Moreover, this protein was found to not only an intra-cytoplasmic protein but also soluble protein in all selected mammalians species showed in **Table 2**.

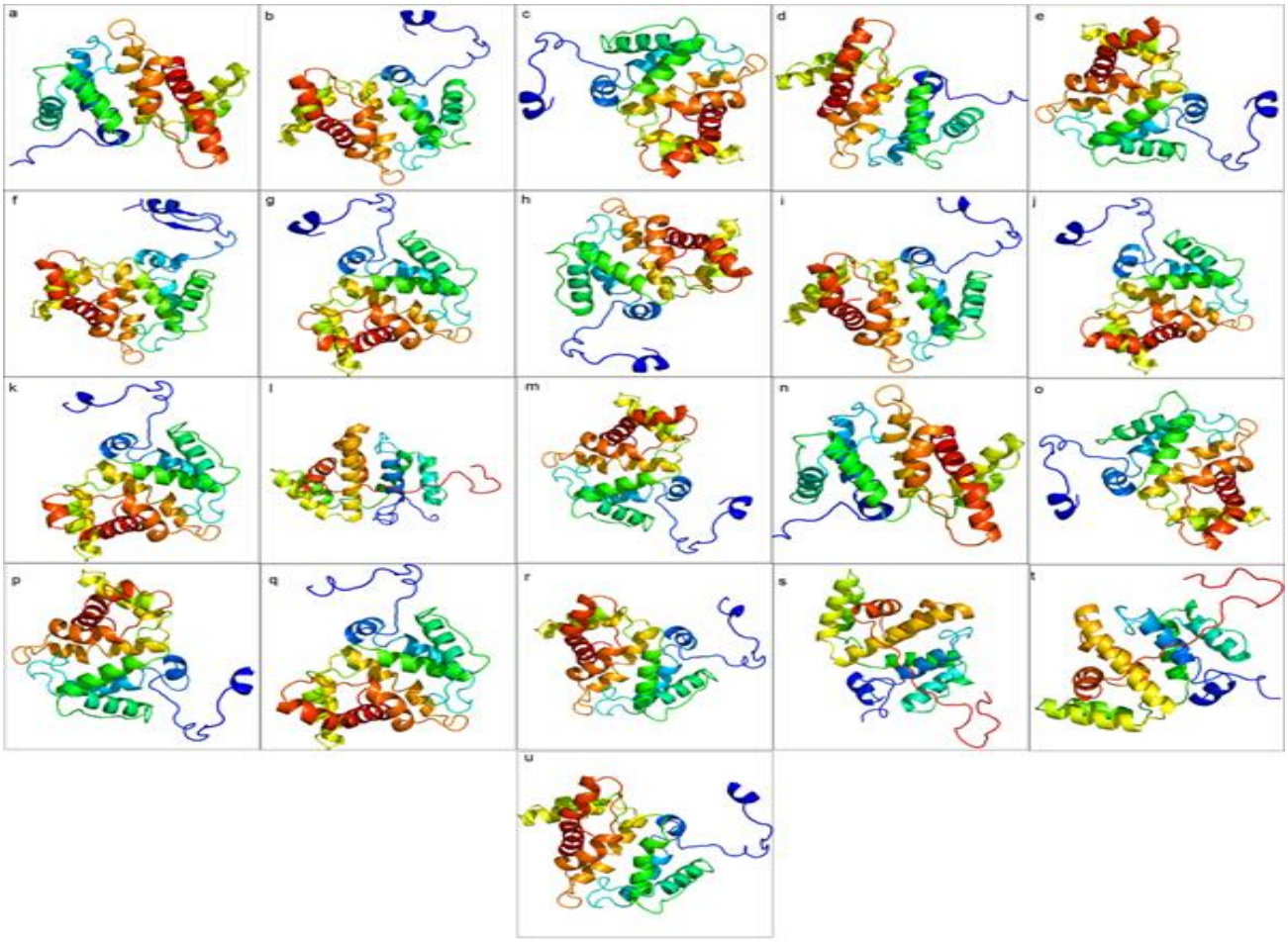

Figure 3 Tertiary structures of TAGAP proteins as calculated by Phyre2

Protein secondary structures, structural characterization of TAGAP gene: The TAGAP protein secondary structure for 21 selected mammals was anticipated using GOR4. It is observed that TAGAP protein contains essentially alpha helix, extended strand and random coil. The human TAGAP protein alpha helix was 27.22% while it was 32.96% for cattle as shown in **Table 3**. The least was for *Saimiri boliviensis* (Monkey) whose occupies alpha helix 24.21% of the structure. Conversely, found that *Xenopus tropicalis* (Frog) has the longest extended strand of the TAGAP protein (15.09%) with that of the *Mesocricetus auratus* (Golden Hamster) being the shortest (8.71%). There were variations existed on the random coil of TAGAP secondary structure of the 21 selected mammals as shown in **Table 3**.

Characterization of functional motifs: In this study, sequences of selected mammalian species were separately screened for matches using InterPro. One domain (Rho GTPase-activating protein (RHOGAP), PS50238) with

varying frequency across 21 species was observed (**Fig. 4.**). Comparative study of the predicted intra-domain features indicates only one RHOGAP domain in all selected mammalians species with different features and predicted score of domains. The homology of domain shows that most amino acid in this region are significantly conserved as shown in **Fig. 4**.

Protein protein interaction (PPI) of TAGAP gene in selected mammalians species: To infer Protein-protein interaction that formed due to genetic interaction, co-localization and any physical connection with TAGAP in the selected mammalian species being studied, STRING was utilized to construct the network of protein by collecting experimental laboratory outcomes from the database (**Fig. 5.**) and based on phylogenetic findings construct Venn diagram for each cluster clade (**Fig. 6.**). STRING couldn't detect any network of protein for American Beaver (*Castor Canadensis*), Beluga Whale (*Delphinapterus leucas*), Horse (*Equus caballus*), Yangtze Finless Porpoise (*Neophocaena asiaeorientalis*), American Mink (*Neovison vison*), Pacific Walrus (*Odobenus rosmarus divergens*), and American Black Bear (*Ursus americanus)* and so in this assessment therefore, excluded. Our result showed that the protein interactome of the TAGAP protein varied significantly across species **Fig. 5**. Generally, a distinct protein was observed in all species that clustered with TAGAP. All species in their cluster had 10 proteins. Venn diagram revealed that, had the highest TAGAP Protein – Protein Interaction (PPI), in total 6 protein sets (RHOB RHOA RAC1 RAC2 RAC3 RHOG) were specific to this group as shown in **Fig. 6 (A)**. Similarly, cattle and dog had highest PPI with each other (**Fig. 6 (B)**).

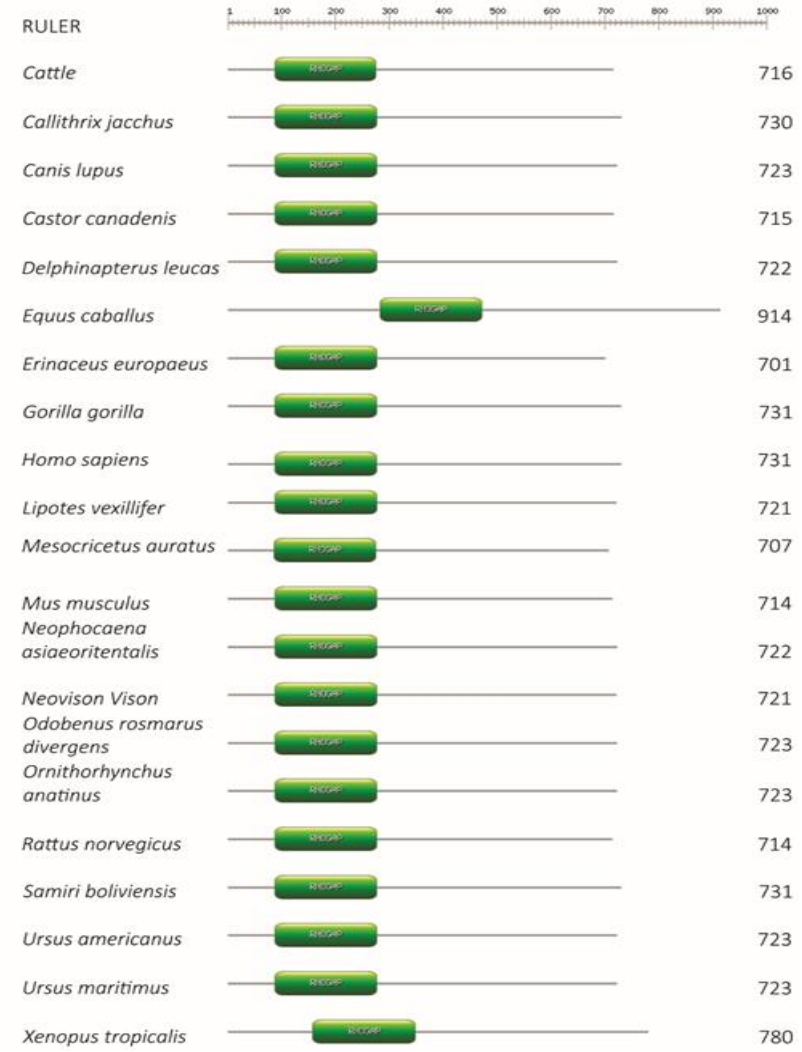

Figure 4. Comparison of predicted intra-domain features of TAGAP protein. This comparison shows RHOGAP, which provide additional information about the structure and function of critical amino acids in the 21 selected mammalians species

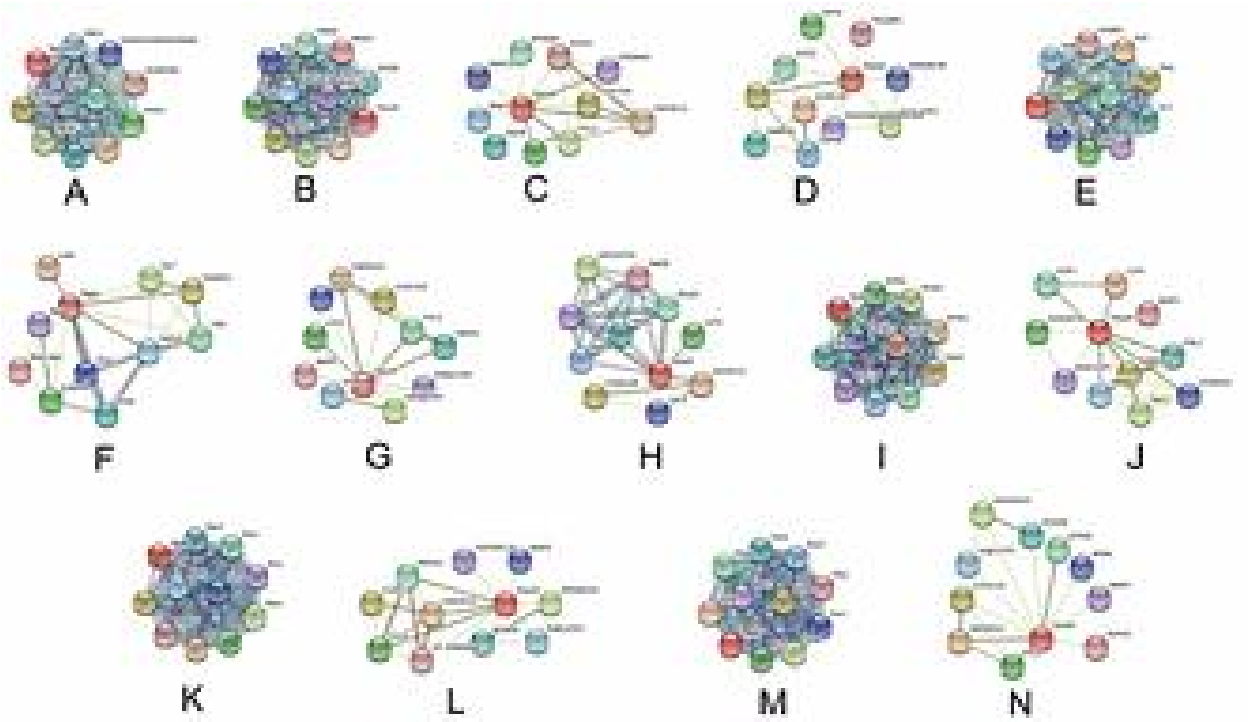

Figure 5. Network view of predicted associations for group of protein in TAGAP. The newtwork nodes are protein. The edges represent the functional associations. The thickness of the line indicates the degree of confidence prediction for the protein. Red line, indicates presence of fusion evidence; Green line, niehhborhood evidence; Blue line, co-occurrence evidence; Purple line, experimental evidence; Yellow line, text mining evidence; Light blue line, database evidence. (A) Cattle, (B) Dog, (C) Yangtze River Dolphin, (D) Duckbill Platypus, (E) Western Clawed Frog, (F) Golden Hamster, (G) Western Lowland Gorilla, (H) European Hedgehog, (I) Human, (J) Bolivian Squirrel Monkey, (K) Mouse,(L) Polar Bear, (M) Rat, (N) White tufted ear marmoset.

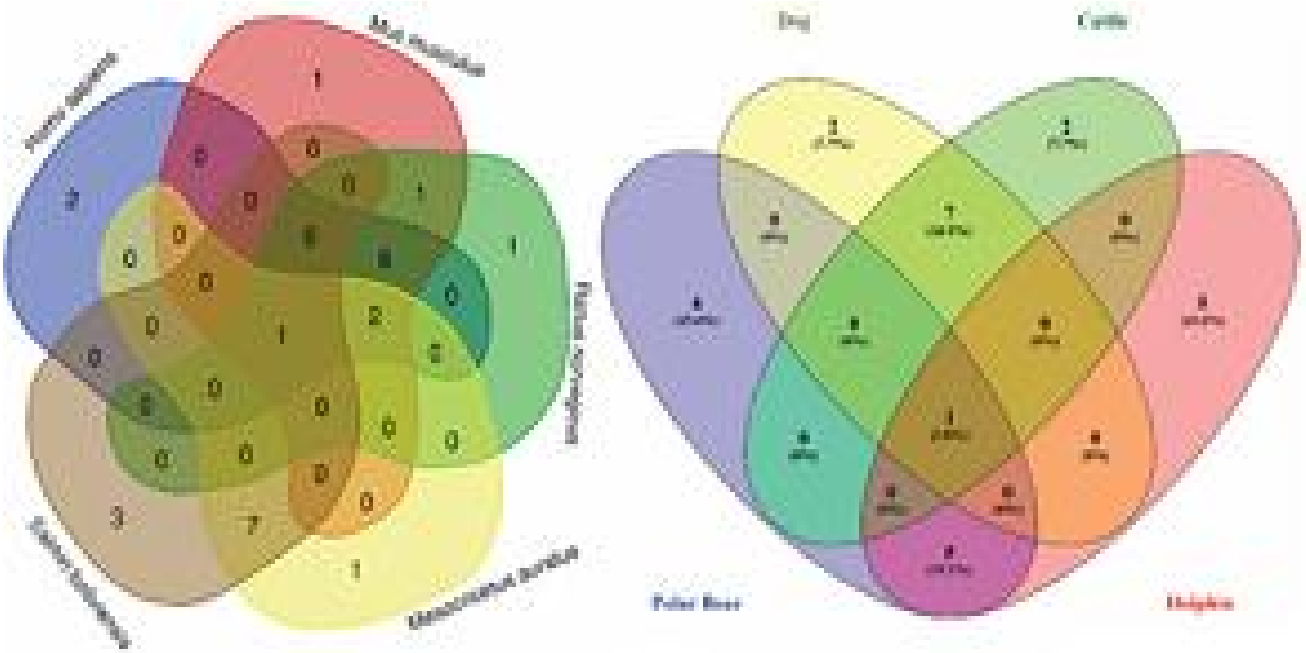

Figure 6. Venn diagram showing the proportion of intersection and unique genes depicting evolutionary diversity of TAGAP molecule.

Table 2 Physiochemical properties of TAGAP protein in 21 selected mammalian species

Species	AA	MW	Th.pI	NCR	PCR	Ext. Co	Abs	InsI	AI	GRAVY	Solubility and Localization	
Cattle *(Bos taurus)*	716	78410.32	6.3	98	92	34950	0.446	66.08	75.94	-0.554	Soluble	Cytoplasmic
White tufted ear marmoset *(Callithrix jacchus)*	730	80583.57	5.9	101	87	33055	0.41	58.92	76.66	-0.565	Soluble	Cytoplasmic
Dog *(Canis lupus)*	723	80624.85	5.44	104	85	47160	0.585	63.45	79.2	-0.571	Soluble	Cytoplasmic
American Beaver *(Castor Canadensis)*	715	80046.67	6.29	98	91	38680	0.483	62.66	78.15	-0.578	Soluble	Cytoplasmic
Beluga Whale *(Delphinapterus leucas)*	722	79450.29	6.33	97	91	42940	0.54	60.04	73.95	-0.592	Soluble	Cytoplasmic
Horse *(Equus caballus)*	914	101368.89	6.99	119	118	65650	0.648	67.71	78.32	-0.528	Soluble	Cytoplasmic
European Hedgehog *(Erinaceus europaeus)*	701	76773.61	7.57	88	89	34670	0.452	57.56	77.09	-0.541	Soluble	Cytoplasmic
Western Lowland Gorilla *(Gorilla gorilla gorilla)*	731	80847.15	5.84	103	89	33180	32430	63.39	77.65	-0.55	Soluble	Cytoplasmic
Human *(Homo sapiens)*	731	80703.01	6.04	101	90	31690	0.393	62.74	77.77	-0.542	Soluble	Cytoplasmic
Yangtze River Dolphin *(Lipotes vexillifer)*	721	79605.61	6.38	97	91	44430	0.558	58.75	73.91	-0.599	Soluble	Cytoplasmic
Golden Hamster *(Mesocricetus auratus)*	707	78575.93	6.33	94	89	44430	0.565	60.58	77.65	-0.554	Soluble	Cytoplasmic
Mouse *(Mus musculus)*	714	78920.09	6.23	92	85	38930	0.493	57.94	76.58	-0.531	Soluble	Cytoplasmic
Yangtze Finless Porpoise *(Neophocaena asiaeorientalis)*	722	79336.14	6.24	96	89	42940	0.541	58.85	73.95	-0.587	Soluble	Cytoplasmic
American Mink *(Neovison vison)*	721	80702.94	5.7	104	90	52785	0.654	58.9	77.28	-0.608	Soluble	Cytoplasmic
Pacific Walrus *(Odobenus rosmarus divergens)*	723	80815.23	5.75	102	89	52785	0.653	59.01	78.92	-0.571	Soluble	Cytoplasmic
Duckbill Platypus *(Ornithorhynchus anatinus)*	723	79745.24	8.46	88	95	32525	0.408	57.53	73.9	-0.617	Soluble	Cytoplasmic
Rat *(Rattus norvegicus)*	714	79249.65	6.09	95	87	34795	0.439	56.47	79.45	-0.511	Soluble	Cytoplasmic
Bolivian Squirrel Monkey *(Saimiri boliviensis)*	731	80525.63	5.86	100	86	33180	0.412	59.95	75.1	-0.554	Soluble	Cytoplasmic
American Black Bear *(Ursus americanus)*	723	80398.46	5.86	102	89	45670	0.568	62.54	75.95	-0.607	Soluble	Cytoplasmic
Polar Bear *(Ursus maritimus)*	723	80471.47	5.81	103	89	45670	0.568	63.37	75.27	-0.624	Soluble	Cytoplasmic
Western Clawed Frog *(Xenopus tropicalis)*	780	88386.6	6.5	99	94	59955	0.678	54.5	79.83	-0.48	Soluble	Cytoplasmic

AA: No. of Amino Acid, MW: Molecular Weight, Th.pI: Theoretical pI, NCR: Total No. of negatively charged residue, PCR: Total No. of positively charged residue, Ext. Co: Extinction Coefficient, Abs: Absorbance, Insl: Instability Index, AI, Aliphatic Index, GRAVY: Grand average hydropathicity.

Table 3 Protein secondary structure of TAGAP protein as calculated by GORIV

Species	Alpha helix	%	Extended strand	%	Random coil	%
Cattle *(Bos taurus)*	236	32.96	66	9.22	414	57.82
White tufted ear marmoset *(Callithrix jacchus)*	183	25.07	108	14.79	439	60.14
Dog *(Canis lupus)*	281	38.87	63	8.71	379	52.42
American Beaver *(Castor Canadensis)*	217	30.35	88	12.31	410	57.34
Beluga Whale *(Delphinapterus leucas)*	240	33.24	89	12.33	393	54.43
Horse *(Equus caballus)*	283	30.96	131	14.33	500	54.7
European Hedgehog *(Erinaceus europaeus)*	236	33.67	86	12.27	379	54.07
Western Lowland Gorilla *(Gorilla gorilla gorilla)*	214	29.27	72	9.85	445	60.88
Human *(Homo sapiens)*	199	27.22	82	11.22	450	61.56
Yangtze River Dolphin *(Lipotes vexillifer)*	218	30.24	96	13.31	407	56.45
Golden Hamster *(Mesocricetus auratus)*	227	32.11	62	8.77	418	59.12
Mouse *(Mus musculus)*	183	25.63	104	14.57	427	59.8
Yangtze Finless Porpoise *(Neophocaena asiaeorientalis)*	228	31.58	90	12.47	404	55.96
American Mink *(Neovison vison)*	243	33.7	76	10.54	402	55.76
Pacific Walrus *(Odobenus rosmarus divergens)*	275	38.04	76	10.51	372	51.45
Duckbill Platypus *(Ornithorhynchus anatinus)*	210	29.05	91	12.59	422	58.37
Rat *(Rattus norvegicus)*	209	29.27	89	12.46	416	58.26
Bolivian Squirrel Monkey *(Saimiri boliviensis)*	177	24.21	104	14.23	450	61.56
American Black Bear *(Ursus americanus)*	232	32.09	95	13.14	396	54.77
Polar Bear *(Ursus maritimus)*	232	32.09	92	12.72	399	55.19
Western Clawed Frog *(Xenopus tropicalis)*	222	28.46	124	15.09	434	55.64

DISCUSSION

Changes in the DNA and proteins caused by mutations are essential to phylogeny. The building blocks of these biological macromolecules, nucleotide bases, and amino acids sequence assessed the primary structure of the molecules (Teng *et al*., 2008). The sequence length variation observed in TAGAP protein of the 21 selected mammalian species compared with each other might have resulted from evolution, mutation, and differentiation. According to (Kedersha *et al*., 1999), sequence length variation is caused by insertion and deletion.

The presence of similar length of amino acid sequences between species is an indication of evolution. Following the aforementioned, the protein sequence length of *Homo sapiens, Gorilla* and *Saimiri boliviensis*; *Canis lupus*, *Odobenus rosmarus divergens*, *Ornithorhynchus anatinus, Ursus americanus* and *Ursus maritimus* was found to be the same; 731 amino acids (a.a). The protein sequence length was also observed to be similar between *Delphinapterus leucas* and *Neophocaena asiaeorientalis*; 722 amino acids (a.a); *Lipotes vexillifer* and *Neovison vison*; 721 amio acid (a.a) and *Mus musculus* and *Rattus norvegicus;* 714 amino acid (a.a). However, the percentage identity and similarity of TAGAP revealed the highest values of 100 between Human and Gorilla; American Black bear and Polar bear, A 90% similarity was observed between Mouse and Rat; Squirrel Monkey and tufted ear marmoset and American mink, Dog and Pacific Walrus. The interpretation is that all these organisms were related and had conserved protein sequence. Based on the result of obtained, the phylogenetic relationship of TAGAP gene among these organisms showed a close relationship among the species investigated with the exception of European Hedgehog which could be considered as an out group and thus distantly related with other species studied. Human and Gorilla; Mouse and Rat were closely related and shared some ancestry with other mammalians species.

Evolutionarily, domain is more conserved than other protein region and tends to develop as units, which are gained, lost or shuffled as one module. The identification of motifs and domain in protein is an important aspect of the classification and functional annotation of protein sequences (Jiang, *et al*., 1997). The domain architecture revealed that these proteins contain mainly one domain (Rho GTPase-activating protein (RHOGAP), PS50238) with varying frequency across 21 species. Furthermore, the STRING database was used to annotate TAGAP protein network with other protein molecules that may have evolved together during speciation.

Conclusion: The present study concluded that due to their structural and functional homologies, the TAGAP gene is more conserved among the 21 selected mammalians species. Phylogenetic relationship of TAGAP showed high relatedness among mammalian

species from the same clade. Though TAGAP gene has been indicated to cause Diabetes Mellitus, Insulin-Dependent and Rheumatoid arthritis in human but also involve in major functions such as GTPase activator activity, guanyl-nucleotide exchange factor activity, signal transduction and play important roles during T-cell activation. So, this study will provide a better understanding of TAGAP gene at biochemical and molecular level to other researcher for further study.

Acknowledgements: The author would like to thank the Department of Bioinformatics and Computational Biology, Virtual University of Pakistan.

REFRENCES

Acland, Abigail, R. Agarwala, T. Barrett, J. Beck, D. A. Benson, C. Bollin and E. Bolton (2014). Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 42(D1): 8–20.

Amundsen, S. S., J. Rundberg, S Adamovic, A. H. Gudjonsdottir, H. Ascher, J. Ek, S. Nilsson, B. A. Lie, A. T. Naluai and L. M. Sollid (2010). Four Novel Coeliac Disease Regions Replicated in an Association Study of a Swedish-Norwegian Family Cohort. Genes and Immunity, 11(1): 79–86.

Arshad, Maria, A. Bhatti, P. John, F. Jalil, F. Borghese, J. Z. Kawalkowaska, R. O. Williams and F. I. L . Clanchy (2018). T Cell Activation Rho GTPase Activating Protein (TAGAP) Is Upregulated in Clinical and Experimental Arthritis. Cytokine, 104(October): 130–35.

Bateman, Alex, M. J. Martin, C. O'Donovan, M. Magrane (2017). UniProt: The Universal Protein Knowledgebase. Nucleic Acids Res. 45(D1): D158–69.

Berge, T., I. S. Leikfoss, I. S. Brorson, S. D. Bos, C. M. Page, M. W. Gustavsen, A. Bjolgerud, T. Holmoy, E. G. Celius, J. Damoiseaux, J. Smolders, H. F. Harbo and A. purkland (2016). The Multiple Sclerosis Susceptibility Genes TAGAP and IL2RA Are Regulated by Vitamin D in CD4+ T Cells. Genes and Immunity, 17(2): 118–27.

Christin, Pascal-Antoine, Guillaume Besnard, Erika J. Edwards, and Nicolas Salamin. 2012. Effect of Genetic Convergence on Phylogenetic Inference. Mol. Phylo. and Evol. 62(3): 921–27.

Festen, Eleonora A.M., P. Goyette, T. Green, G. Boucher, C. Beauchamp, *et al*. (2011). A Meta-Analysis of Genome-Wide Association Scans Identifies IL18RAP, PTPN2, TAGAP, and PUS10 as Shared Risk Loci for Crohn's Disease and Celiac Disease. PLoS Genetics, 7(1): 3–8.

Garner, C. P., J.A. Murray, Y.C. Ding, Z. Tien, D.A. van Heel and S. L. Neuhausen (2009). Replication of Celiac Disease UK Genome-Wide Association Study Results in a US Population. Human Mol. Gen. 18(21): 4219–25.

Garnier, Jean, Jean François Gibrat, and Barry Robson. (1996). GOR Method for Predicting Protein Secondary Structure from Amino Acid Sequence. Methods in Enz. 266(1995): 540–53.

Gasteiger, Elisabeth, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch (2003). ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis. Nucleic Acids Res. 31(13): 3784–88.

Hirokawa, Takatsugu, Seah Boon-Chieng, and Shigeki Mitaku (1998). SOSUI: Classification and Secondary Structure Prediction System for Membrane Proteins. Bioinformatics, 14(4): 378–79.

Horton, Paul, K. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier and K. Nakai (2007). WoLF PSORT: Protein Localization Predictor. Nucleic Acids Res. 35(SUPPL.2): 585–87.

Huang, Shi Qi, N. Zhang, Z. Xing. Zhou, C.C. Huang, C-Li. Zeng (2017). Association of LPP and TAGAP Polymorphisms with Celiac Disease Risk: A Meta-Analysis. Int. J. of Env. Res. and Pub. Heal. 14(2).

Jiang, Weining, Yan Hou, and Masayori Inouye (1997). CspA, the Major Cold-Shock Protein of Escherichia Coli, Is an RNA Chaperone. J. of Bio. Chem. 272(1): 196–202.

Kedersha, Nancy, M. Gupta, W. Li and P. Anderson (1999). EIF-2 to the Assembly of Mammalian Stress Granules. The J. of Cell bio. 147(7): 1431–41.

Kelley, Lawrence A., S. Mezulis, C. M. Yates, M. N. Wass and M. J.E. Sternberg (2015). The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. Nature Protocols.

Kumar, Sudhir, G. Stecher, M. Li, C. Knyaza and K. Tamura (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biology and Evo. 35(6): 1547–49.

Mao, Mao, M.C. Biery, S.V. Kobayashi, T. Ward, G. Schimmack (2004). T Lymphocyte Activation Gene Identification by Coregulated Expression on DNA Microarrays. Genomics, 83(6): 989–99.

O'Leary, Nuala A., M.W. Wright, J. R. Brister, S. Ciufo, D. Haddad (2016). Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. Nucleic Acids Res. 44(D1): D733–45.

Oliveros, J.C. 2007. VENNY. An Interactive Tool for

Comparing Lists with Venn Diagrams.

Pervez, Muhammad Tariq, M. E. Babar, A. Nadeem, N. Aslam (2015). IvisTMSA: Interactive Visual Tools for Multiple Sequence Alignments. Evolutionary Bioinf. 11: 35–42.

Sahoo, Pravas Ranjan, Tapas Kumar Patabandha, and Anjan Kumar Sahoo (2018). Phylogenetic Analysis of P53 Tumor Suppressor Gene of Bos Taurus through in Sillico Platform. The Pharma Inn. J. 7(3): 309–11.

Singh, Rajbir, Neha Jain, and Dheeraj Pal Kaur (2013). GOR Method for Protein Structure Prediction Using Cluster Analysis. Int. J. of Computer Appl. 73(1): 1–6.

Soltis, Douglas E, and Pamela S Soltis (2003). Update on Molecular Systematics The Role of Phylogenetics in Comparative Genetics Society 132(August): 1790–1800.

Soltis, Pamela S., and Douglas E. Soltis (2003). Applying the Bootstrap in Phylogeny Reconstruction. Statistical Sci. 18(2): 256–67.

Szklarczyk, Damian, A. Franceschini, S. Wyder, K. Forslund, D. Heller (2015). STRING V10: Protein-Protein Interaction Networks, Integrated over the Tree of Life. Nucleic Acids Res.

Teng. S, M. Alexova, E. Alexov (2008). Curr. Pharma. Biot. 9(2): 123-133.

Whelan, Simon, and Nick Goldman (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. Mol. Bio. and Evol. 18(5): 691–99.

Xu, Yong, W. Li, X. Liu, H. Chen, K. Tan (2013). Identification of Dysregulated MicroRNAs in Lymphocytes from Children with Down Syndrome. Gene, 530(2): 278–86.