



Received on 04 September, 2015; received in revised form, 06 November, 2015; accepted, 12 December, 2015; published 01 March, 2016

## PHYLOGENETIC ANALYSIS OF RAS SUBFAMILY PROTEINS

Diksha Jain, Ravita Rawat and Vinod Kumar Jatav \*

Department of Biotechnology, MITS Gwalior-474005, Madhya Pradesh, India.

### Keywords:

Cancer, RAS,  
Phylogenetic, ProtParam

### Correspondence to Author: Vinod Kumar Jatav

Assistant Professor, Department of  
Biotechnology, MITS Gwalior-  
474005, Madhya Pradesh, India.

**Email:** vinod.mits@gmail.com


**ABSTRACT:** Cancer is primarily an environmental disease with 90–95% of cases attributed to environmental conditions and 5–10% due to genetics. RAS proteins act as binary molecular switches and play an important role in intracellular signal transduction which regulates processes such as actin cytoskeletal integrity, proliferation, differentiation, cell adhesion, apoptosis, and cell migration. The present bio-computational analysis was performed using web-based tools and servers. Multiple sequence alignment of selected human RAS subfamily proteins with other homologous sequences revealed highly conserved regions. The present work determined the physico-chemical properties of selected RAS proteins such as their hydrophilic nature; alpha-helical structure and close evolutionary relationship with higher vertebrates. Cancerous proteins (gi166706781, gi 4505451) from human having most of homology with the number of species. On the basis of this study we can suggest that these organisms can be taken for the further study and after the successful analysis, these can be implemented on humans.

**INTRODUCTION:** The Ras oncogene proteins are the founding members of this family, which is divided into five major branches on the basis of sequence and functional similarities: Ras, Rho, Rab, Ran and Arf. Small GTPases share a common biochemical mechanism and act as binary molecular switches <sup>1</sup>.

The Ras super family of small guanosine triphosphatases (GTPases) comprise over 150 human members with evolutionarily conserved orthologs found in *Drosophila*, *C. elegans*, *S. cerevisiae*, *S. pombe*, *Dictyostelium* and plants <sup>2</sup>.

High rates of KRAS-activating missense mutations have been detected in non-small cell lung cancer (15 to 20% of tumors) <sup>3</sup>, colon adenomas (40%) <sup>4</sup>, pancreatic adenocarcinomas (95%) <sup>5</sup>, making it the single most common mutationally activated human oncoprotein. KRAS mutations were detected in the primary adenocarcinoma in 40 of 106 tumors (38%) and were significantly more common in smokers compared with non-smokers (43%) <sup>6</sup>. Also there is a clear association between the inactivation of MGMT (O6-Methylguanine DNA methyltransferase) by promoter hypermethylation and the appearance of G to A mutations at KRAS, among 244 colorectal lesions, 88 (36%) had mutant KRAS gene <sup>7</sup>.

In primary colorectal carcinomas, seventy-four of the 160 (46%) presented mutations in KRAS: 54% in codon 12, 42% in codon 13 (particularly G→A transition) and 4% in both <sup>8</sup>. In some tumors, HRAS or NRAS activating mutations were identified in four (8.2%) of 49 well-differentiated

<b>QUICK RESPONSE CODE</b> 	<b>DOI:</b> 10.13040/IJPSR.0975-8232.7(3).1070-80
	<b>Article can be accessed online on:</b> <a href="http://www.ijpsr.com">www.ijpsr.com</a>
<b>DOI link:</b> <a href="http://dx.doi.org/10.13040/IJPSR.0975-8232.7(3).1070-80">http://dx.doi.org/10.13040/IJPSR.0975-8232.7(3).1070-80</a>	

carcinomas WDCs; two (6.7%) of 30 of the tumors were papillary carcinomas, two (10.5%) of 19 of them were follicular carcinomas), in 16 (55.2%) of 29 poorly differentiated carcinomas (PDCs), and in 15 (51.7%) of 29 undifferentiated carcinomas, with a significant association between RAS mutation and poorly or undifferentiated<sup>9</sup>. In addition to mutational activation, RAS genes are amplified or over expressed in some tumors<sup>10</sup>. Other mechanisms leading to RAS over activation in tumor cells include the deletion of genes encoding negative regulators (for example NF1, a GAP for RAS, in neurological tumors)<sup>12-14</sup> and over expression of positive regulators (such as SOS1, a GEF for RAS, in renal cancer cells)<sup>15</sup>. In the present study five members of RAS subfamily: HRAS, NRAS, KRAS, RASSF2, and DIRAS1 proteins were analyzed and find out a correlation among them for finding a drug development relationship by using computational tools and servers.

## MATERIALS & METHODS:

### Retrieval of Protein Sequence:

Five members of RAS subfamily: KRAS, NRAS, HRAS, RASSF2, and DIRAS1 were selected for their characterization, and sequences were retrieved from National Centre for Biotechnology Information (NCBI) database (www.ncbi.nlm.nih.gov/). A total of 174 sequences of KRAS, 140 sequences of NRAS, 185 sequences of HRAS, 125 sequences of RASSF2, 76 sequences of DIRAS1 were selected for this study.

### Phylogenetic Tree Prediction:

Multiple sequence alignments were generated with ClustalW<sup>16</sup>, phylogenetic trees were constructed using the Neighbor-Joining (NJ) method implemented in the software PhyML available at Phylogene.fr<sup>17</sup>. The four steps undertaken were- multiple sequence alignment by MUSCLE software, alignment curation by Gblocks tool, construction of phylogenetic tree by PhyML software, visualization of phylogenetic tree by Tree Dyn tool. In this study, we used more stringent selection option for selecting the blocks in curation and the bootstrapping procedure with 30 bootstraps to construct the phylogenetic tree.

Proteins by analyzing the phylogenetic trees and MSA results, one sequence each of *Homo sapiens*

from among the several sequences of the five RAS proteins were selected. The selection was done on the basis of maximum similarity found, so that the sequence was a representative of all the other sequences.

### Physico-chemical Properties Prediction:

Physico-chemical properties of the selected RAS proteins were performed using Prot Param<sup>18</sup> tool of the ExPASy proteomics server. The parameters computed by Prot Param include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

### Protein Structure Prediction:

Secondary structure has been predicted using SOPMA<sup>19</sup> and PHYRE2<sup>20</sup> software where the FASTA format of the sequence was given as input. It provides the structural information of the protein sequence in form of coils, helices and strands with template information.

3-D structure prediction was done by using two online tools SWISS-MODEL<sup>21</sup> and PHYRE2. The modeling involves four basic steps, first searching structure showing homology with target, then selecting a best template having maximum identity with the target sequence which follows its alignment with the target and modeling the structure. The modelled structure was then evaluated using Chimera 1.9<sup>22</sup> visualization provided 3D structure of selected PDB ids.

Ramachandran plot of the selected sequences were made for validation using SAVES server (<http://services.mbi.ucla.edu/SAVES/>) here Procheck<sup>23</sup> implemented for the plot formation. The functional domains were then searched using SMART<sup>24</sup> tool. The functional Motifs were then analyzed using MOTIF search (<http://www.genome.jp/tools/motif/MOTIF2.html>).

## RESULT & DISCUSSION:

### Multiple Sequence Analysis:

Multiple Sequence Alignment of the five selected proteins i.e. KRAS, NRAS, HRAS, RASSF2, and DIRAS1 revealed significant conserved regions among them. For KRAS sequences, the patterns

MRDQYMRTGEGFL(67-79), KSFE(88-91), YREQI(96-100), RVKDS(102-106), PMV(110-112), VGNK(127-130), and TSAKTR(157-162) were found in all the 174 sequences. For DIRAS1 sequences, the patterns MPE(94-96), SND(98-100), YRVVVFGA(101-108), GKSSLVLRV(111-121), GTFR (123-126), TYIPT (128-132), EDTYRQVISC DK (134-145), CTL(148-150), ITD TTGSHQFPAMQRLSISKG (152-172), VGNK (212-215), REV(221-223), WKC(235-237), FMETSAK(239-245), and ELFQ (252-255) were present in all the 76 sequences. No conserved regions were found in NRAS, HRAS and RASSF2.

### Phylogenetic Analysis:

**KRAS:** We have constructed the phylogeny tree taking all the 174 sequences of KRAS available on NCBI database (**Fig. 1**). The *Homo sapiens* sequence (gi166706781) shares clad with 76 other sequences. Also, the sequences *Mus musculus* (gi 569017220), *Mus musculus* (gi 569017218), *Mus musculus* (gi 111601545) and *Rattus norvegicus* (gi 564349722) share a different clad. Out of the 174 sequences 172 sequences belong to class Coelomata, one sequence *Clonorchis sinensis* (gi358331595) belong to class Pseudocoelomata and one another sequence *Necator americanus* (gi568268505) belong to class Acoelomata. All other *Homo sapiens* sequences show less homology as compared to this sequence. So after comparing the results of MSA and Phylogenetic tree we have chosen this sequence (gi166706781) for further study.

### NRAS:

Phylogenetic tree constructed 140 sequences of NRAS available on NCBI database (**Fig. 2**). The *Homo sapiens* sequence (gi 4505451) shares clad with 140 other sequences. Out of the 140 sequences, 138 belong to Fungi Metazoa group and other two sequences *Acanthamoeba castellanii* (gi440791055), *Acanthamoeba castellanii* (gi470387771) belong to Acanthamoebidae. Another *Homo sapiens* sequences show less homology as compared to this sequence. So after comparing the results of MSA and Phylogenetic tree we have chosen sequence (gi 4505451) for further study.

**HRAS:** We have constructed the phylogeny tree taking all the 90 sequences of HRAS proteins were used for phylogenetic formation (**Fig. 3**). The *Homo sapiens* sequence (gi 49457536) shares clad with 90 other sequences. All other *Homo sapiens* sequences show less homology as compared to this sequence. So after comparing the results of MSA and Phylogenetic tree we have chosen sequence (gi 49457536) for further study.

### RASSF2:

Constructed phylogeny tree taking all the 20 sequences of RASSF2 available on NCBI database (**Fig. 4**). The *Homo sapiens* sequence (gi 109658938) shares clad with 20 other sequences. Another *Homo sapiens* sequences show less homology as compared to this sequence. So after comparing the results of MSA and Phylogenetic tree we have chosen sequence (gi 109658938) for further study.

### DIRAS1:

31 retrieved sequences of DIRAS1 were used for tree formation (**Fig. 5**). The *Homo sapiens* sequence (gi 21553323) shares clad with 31 other sequences. Out of the 31 sequences, 29 belong to class Sarcopterygii and two other sequences *Danio rerio* (gi41054159), *Danio rerio* (gi 28278625) belong to class Actinopterygii. Another *Homo sapiens* sequences show less homology as compared to this sequence. So after comparing the results of MSA and Phylogenetic tree we have chosen sequence (gi 21553323) for further study.

### Sequence Analysis:

Selected Human RAS subfamily proteins: KRAS, NRAS, HRAS, RASSF2 and DIRAS1, consisted of 189, 189, 189, 326, and 198 amino acids respectively. Primary structure analysis provided the physicochemical properties of RAS subfamily proteins. Molecular weight of KRAS, NRAS, HRAS, RASSF2, and DIRAS1 was 21655.8, 21229.1, 21298.1, 37790.2 and 22328.7 respectively. Most abundant amino acid was Valine in HRAS, NRAS, DIRAS1 (8.5%, 9.0%, and 9.1% respectively), Leucine (10.1%) in RASSF2 and Valine and Lysine in KRAS (8.5%). A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. Prot Param server

predicted that HRAS, RASSF2, and DIRAS1 were unstable whereas NRAS and KRAS were stable proteins. The isoelectric point of a protein is an important property, because it is at this point that the protein is least soluble. Computed isoelectric point of KRAS, NRAS, and HRAS was 6.33, 5.01, and 5.03 respectively i.e. below 7 so they are likely to precipitate in acidic buffers, and for RASSF2 and DIRAS1 it was 8.93, and 8.94 respectively, so they are soluble in basic buffers. The aliphatic index of a protein is defined as the relative volume

occupied by aliphatic side chains: alanine, valine, isoleucine, and leucine, which is 85.03, 82.96, 81.96, 82.76 and 81.62 for KRAS, NRAS, HRAS, RASSF2 and DIRAS1 respectively. It may be regarded as a positive factor for the increase of thermo-stability of globular proteins. The Grand Average Hydropathy (GRAVY values) showed that all proteins were hydrophilic, ranging from -0.432 to -0.317, supports the soluble nature of RAS proteins.

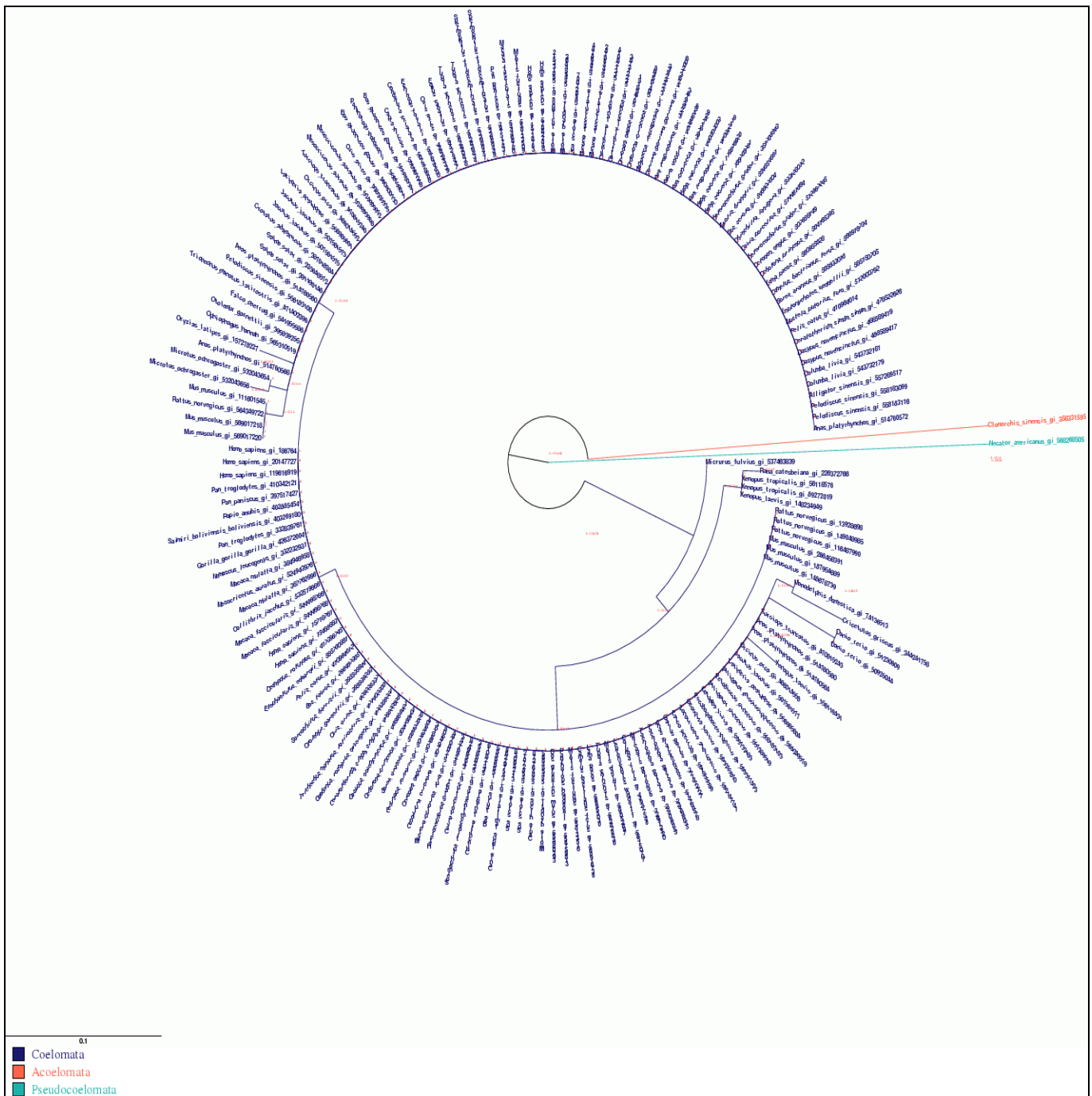


FIG.1: PHYLOGENETIC TREE OF KRAS PROTEINS CONSTRUCTED BY PHYML USING PHYLOGENY.FR



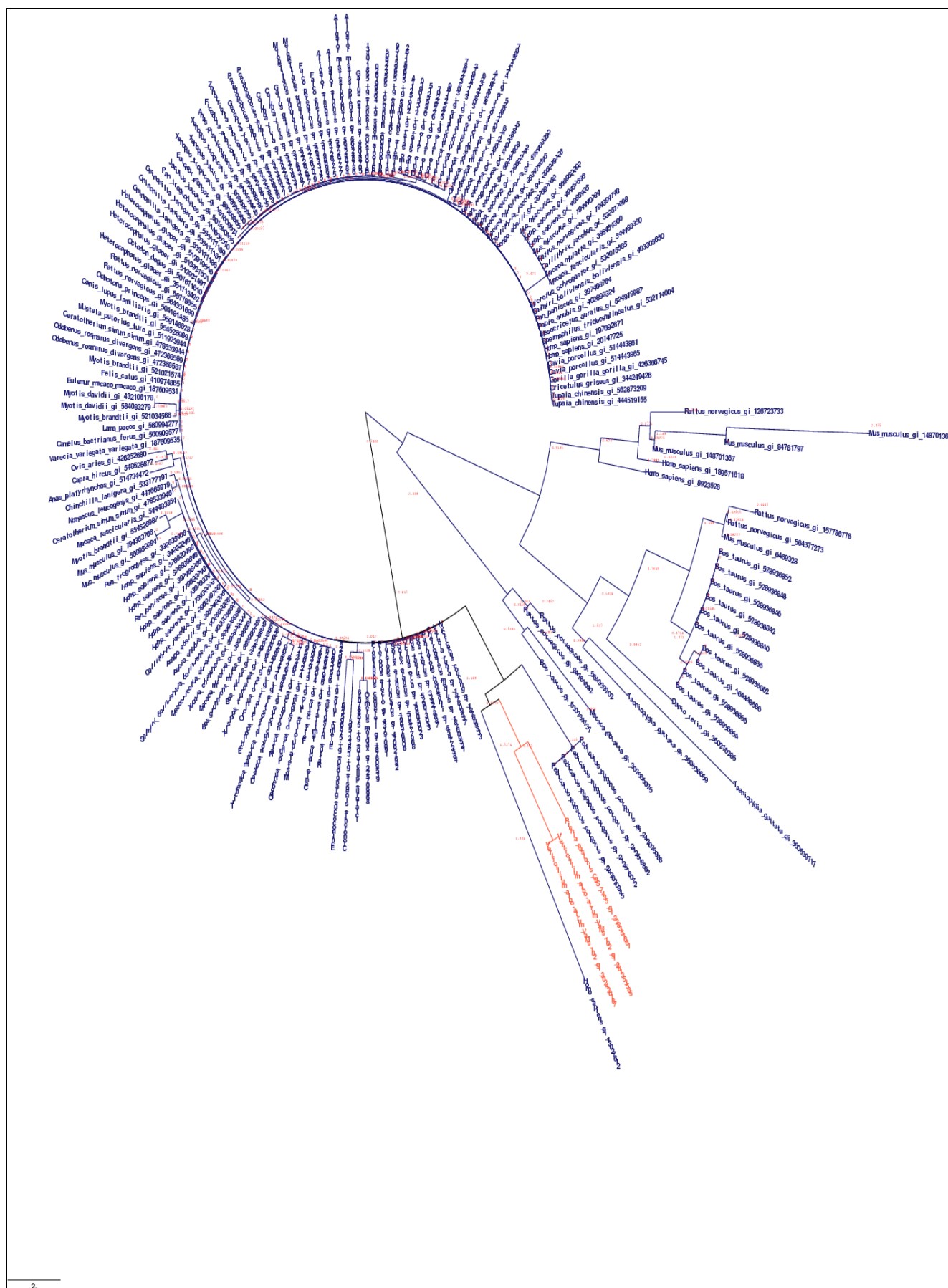


FIG. 3: PHYLOGENETIC TREE OF HRAS PROTEINS CONSTRUCTED BY PHYML USING PHYLOGENY.FR

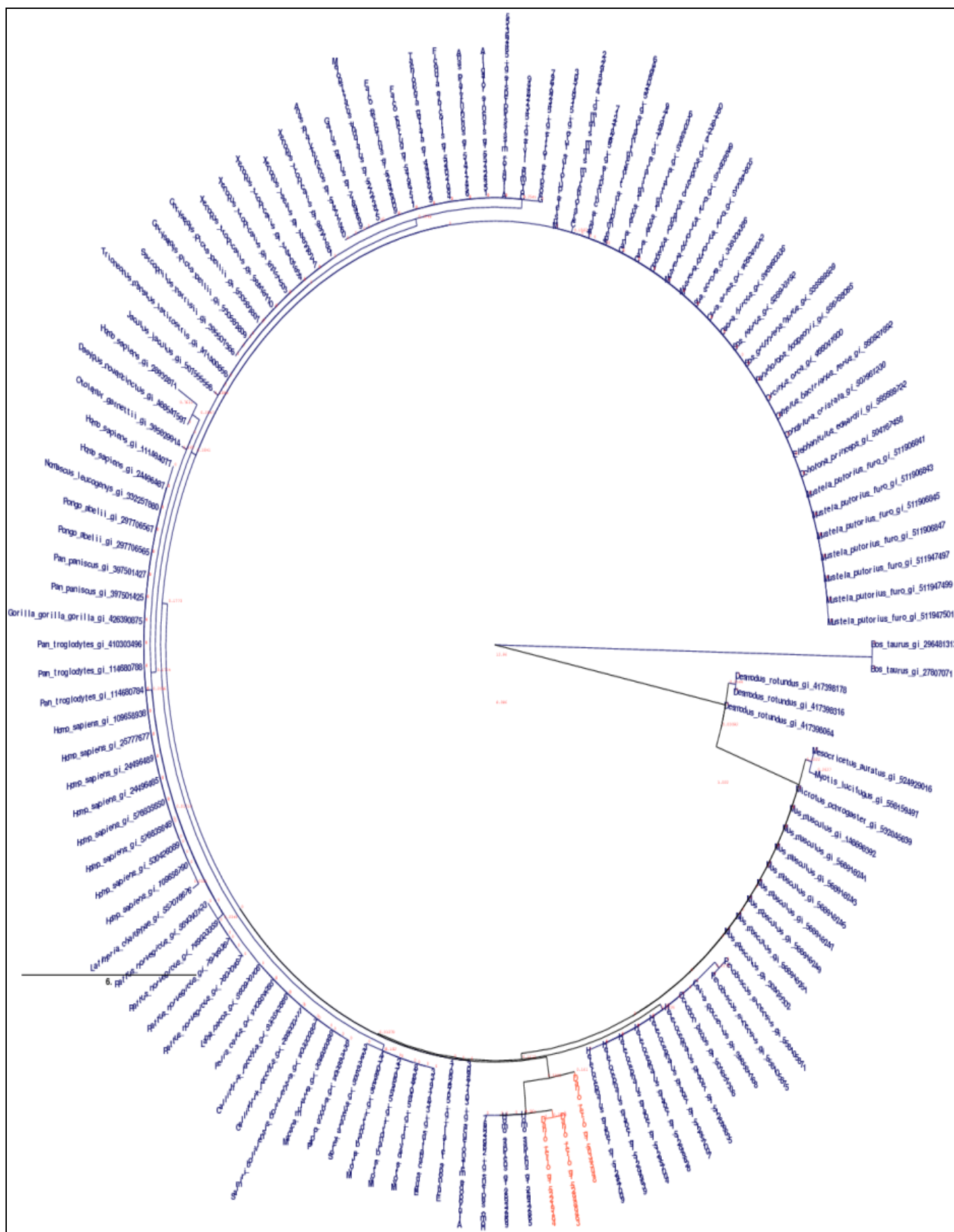
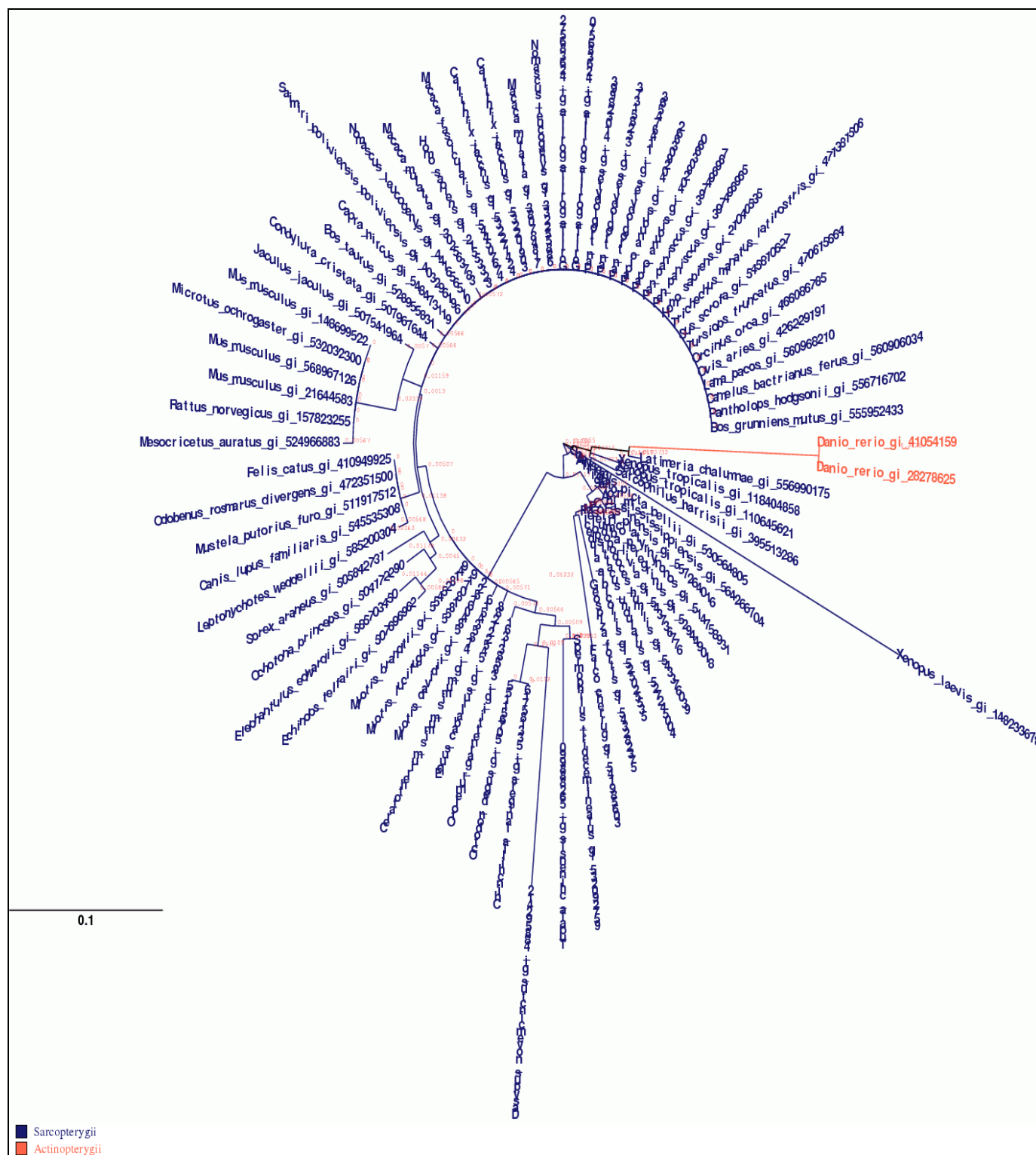


FIG.4: PHYLOGENETIC TREE OF RASSF2 PROTEINS CONSTRUCTED BY PHYML USING PHYLOGENY.FR



The functional domains of these proteins were analyzed using SMART tool. The results were shown in the **Table 1**.

**TABLE 1: FUNCTIONAL DOMAIN OF KRAS & NRAS**

Protein Family	Domain	Start	End	E-Value
KRAS (gi166706781)	RAS	1	166	9.11e-123
NRAS(gi 4505451)	RAS	1	166	1.09e-120



The functional motifs of these proteins were analyzed using MOTIF-Search tool (Table 2).

TABLE 2: FUNCTIONAL MOTIFS OF KRAS & NRAS

Pfam	KRAS (gi 166706781)	NRAS(gi 4505451)
Ras	5..164	5..164
Miro	5..119	5..119
Arf	3..160	3..162
GTP_EFTU	47..161	47..163
MMR_HSR1	6..117	6..116
Gtr1_RagA	5..123	5..108
FeoB_N	108..154	108..154
ATP_bind_1	55..133	Not Found
DUF258	6..44	6..27
AAA_22	5..117	5..117
AAA_25	6..25	6..25
Septin	3..50	3..49
DUF2075	6..65	Not Found
AAA_14	6..35	6..28
CbiA	7..25	Not Found
DUF1413	145..175	Not Found
Ldh_1_N	Not Found	5..49

**Structure Analysis:**

The secondary structure studies gave the following alpha helix predicted values 44.97%, 37.57% for

KRAS & NRAS respectively. Secondary structure studies gave the following  $\beta$ -turn predicted values: 4.76%, 7.41% for KRAS, NRAS respectively.

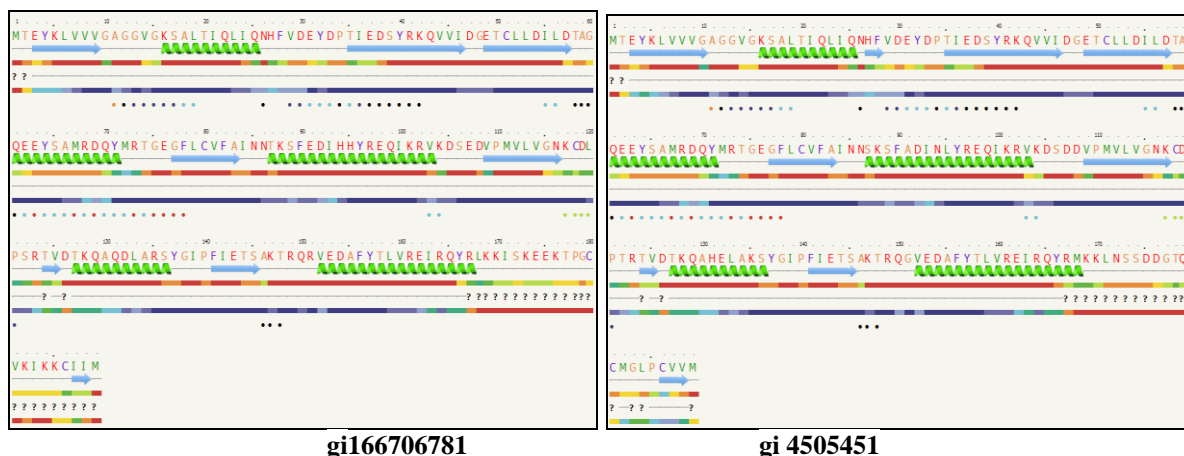


FIG. 6: SECONDARY STRUCTURE OF KRAS & NRAS PROTEIN PREPARED BY PHYRE2

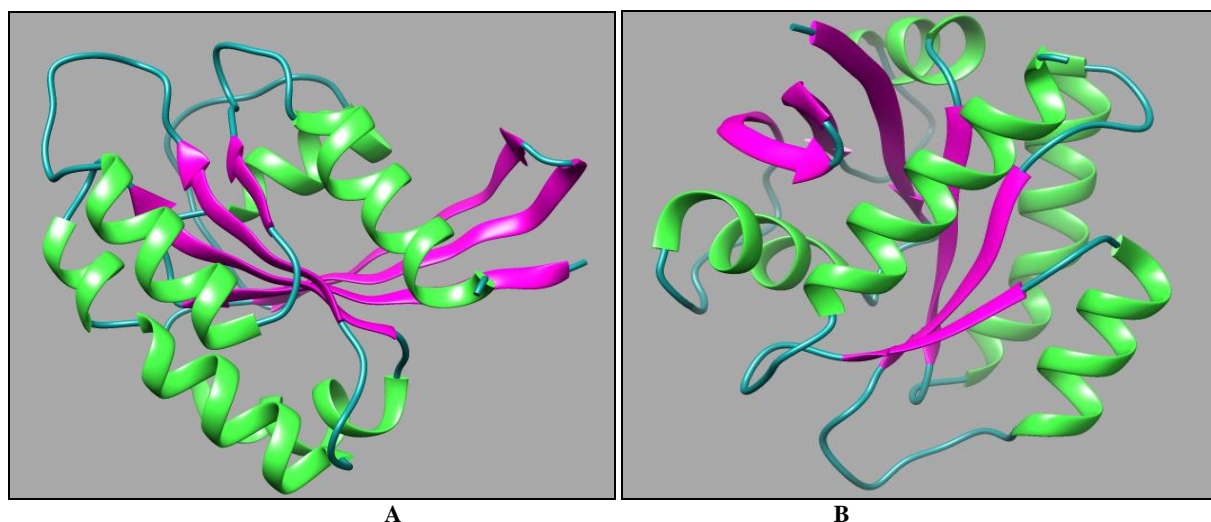


FIG.7: 3D STRUCTURE OF KRAS DEVELOPED BY SWISS-MODEL (A) & PHYRE2 (B)

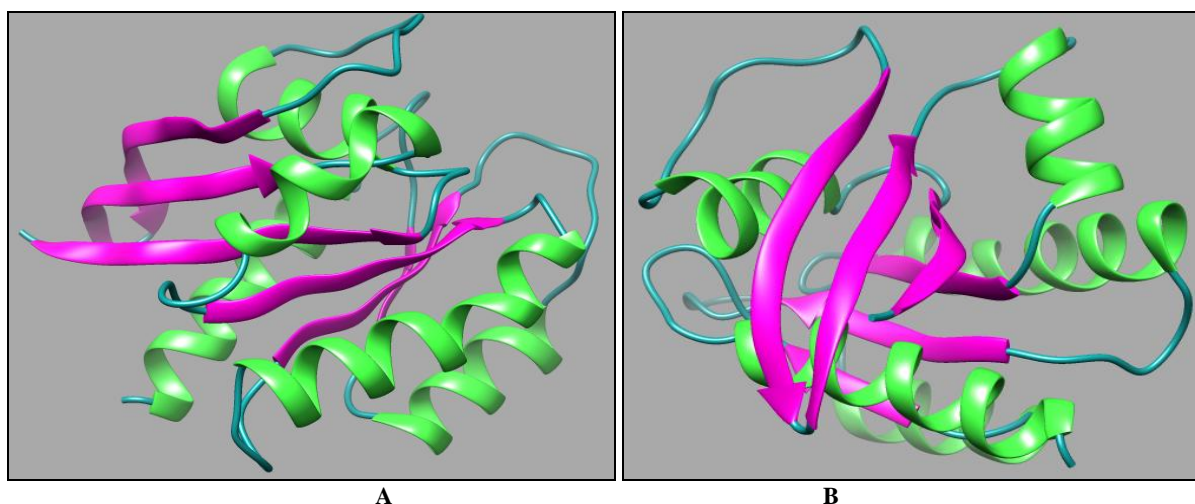


FIG. 8: 3D STRUCTURE OF NRAS DEVELOPED BY SWISS-MODEL (A) & PHYRE2 (B)

For KRAS, out of all the templates given by PHYRE2 server, the one with highest % i.d. (95 %) was used to predict the 3D structure. In case of SWISS-MODEL out of several templates, the one with highest percentage of residues in the favored regions (98.2%) was used to predict the 3D structure. For NRAS, out of all the templates given by PHYRE2 server, the one with highest % i.d. (99 %) was used to predict the 3D structure. In case of

SWISS-MODEL out of several templates, the one with highest percentage of residues in the favored regions (98.2%) was used to predict the 3D structure. Models show the structure of proteins. The description of the colors is as follows-

1. Helices: green
2. Sheets: magent
3. Turns: Light sea green

TABLE 3: SHOWS THE RAMACHANDRAN PLOT & OVERALL QUALITY ANALYSIS

Parameters	KRAS by SWISS-MODEL	KRAS by PHYRE2	NRAS by SWISS-MODEL	NRAS by PHYRE2
Total no. of residues	165	165	165	164
No. of residues in favored region	98.2	98.2	98.2	98.2
No. of residues in allowed region	0.6	0.6	0.6	1.8
No. of residues in outlier region	1.2	1.2	1.2	0
Overall quality factor	94.375	97.351	93.125	90.566

**CONCLUSION:** In the present study, RAS protein sequences were taken from NCBI and result shows the cancerous proteins (gi166706781, gi 4505451) having most of homology with the number of species. On the basis of our study we can suggest that for cancer development and treatment, we can choose these organisms for the further study and after the successful result, these can be implemented on humans. This analysis revealed the importance of computational approaches for drug designing and discovery. This study proposes to put forward a constructive conception to designing RAS protein inhibitors.

## REFERENCES:

1. Vetter IR, A Wittinghofer. The guanine nucleotide-binding switch in three dimensions. *Science* 2001; 294:1299-1304.

2. Colicelli J. Human RAS superfamily proteins and related GTPases. *Sci. STKE* 2004, RE13.
3. Mitsuuchi Y, Testa JR. Cytogenetics and molecular genetics of lung cancer. *Am. J. Med. Genet.* 2002; 115:183-188.
4. Grady WM, Markowitz SD. Genetic and epigenetic alterations in colon cancer. *Annu. Rev. Genomics Hum. Genet.* 2002; 3:101-128.
5. Jaffee EM, Hruban RH, Canto M, Kern SE. Focus on pancreas cancer. *Cancer Cell.* 2002; 2:25-28.
6. Ahrendt SA, Decker PA, Alawi EA, Zhu Y, Sanchez-Cespedes M, Yang SC, Haasler GB, Kajdacsy-Balla A, Demeure MJ, Sidransky D. Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. *Cancer* 2001; 92: 1525-1530.
7. Esteller M, Toyota M, Sanchez-Cespedes M, Capella G, Peinado MA, Watkins DN, Issa JPI, Sidransky D, Baylin SB, Herman JG. Inactivation of the DNA Repair Gene O<sup>6</sup>-Methylguanine-DNA Methyltransferase by Promoter Hypermethylation Is Associated with G to A Mutations

- in *K-ras* in Colorectal Tumorigenesis. *Cancer Res.* 2000; 60: 2368
8. Bazan V, Migliavacca M, Zanna I, Tubiolo C, Grassi N, Latteri MA, La Farina M, Albanese I, Dardanoni G, Salerno S, Tomasino RM, Labianca R, Gebbia N, Russo A. Specific codon 13 *K-ras* mutations are predictive of clinical outcome in colorectal cancer patients, whereas codon 12 *K-ras* mutations are associated with mucinous histotype. *Ann. Oncol.* 2002; 13(9): 1438-1446.
  9. Garcia-Rostan G, Zhao H, Camp RL, Pollan M, Herrero A, Pardo J, Wu R, Carcangiu ML, Costa J, Tallini G. *ras* mutations are associated with aggressive tumor phenotypes and poor prognosis in thyroid cancer. *J. Clin. Oncol.* 2003; 21:3226–3235.
  10. Vageli D, Kiaris H, Delakas D, Anezinis P, Cranidis A, Spandidos DA. Transcriptional activation of H-ras, K-ras and N-ras proto-oncogenes in human bladder tumors. *Cancer Lett.* 1996; 107:241–247.
  11. Xu GF, O'Connell P, Viskochil D, Cawthon R, Robertson M, Culver M, Dunn D, Stevens J, Gesteland R, White R, et al. The neurofibromatosis type 1 gene encodes a protein related to GAP. 1990; 62:599–608.
  12. Xu GF, Lin B, Tanaka K, Dunn D, Wood D, Gesteland R, White R, Weiss R, Tamanai F. The catalytic domain of the neurofibromatosis type 1 gene product stimulates ras GTPase and complements *ira* mutants of *S. cerevisiae*. *Cell.* 1990; 63:835–841.
  13. Martin GA, Viskochil D, Bollag G, McCabe PC, Crosier WJ, Haubruck H, et al. The GAP-related domain of the neurofibromatosis type 1 gene product interacts with ras p21. *Cell.* 1990; 63:843–849.
  14. Ballester R, Marchuk D, Boguski M, Saulino A, Letcher R, Wigler M, Collins F. The NF1 locus encodes a protein functionally related to mammalian GAP and yeast IRA proteins. *Cell.* 1990; 63:851–859.
  15. Shinohara N, Ogiso Y, Tanaka M, Sazawa A, Harabayashi T, Koyanagi T. The significance of Ras guanine nucleotide exchange factor, son of sevenless protein, in renal cell carcinoma cell lines. *J. Urol.* 1997; 158:908–911.
  16. Li W, Cowley A, Uludag M, Gur T, Mc William H, Squizzato S, Park YM, Buso N, Lopez R. European Bioinformatics Institute, EMBL Outstation, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, Cambridge, UK. *Nucleic Acids Research* 2015.
  17. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 2008 Jul 1; 36.
  18. Gasteiger E. Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, Humana Press 571-607.
  19. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci.* 1995; 11: 681-684.
  20. Lawrence Kelley, Benjamin Jefferys. *Phyre2: Protein Homology/analog Y Recognition Engine V 2.0*. Structural Bioinformatics Group, Imperial College, London 2011; Retrieved 22 April.
  21. Marco Biasini, Stefan Bienert, Andrew Waterhouse, Konstantin Arnold, Gabriel Studer, Tobias Schmidt, Florian Kiefer, Tiziano Gallo Cassarino, Martino Bertoni, Lorenza Bordoli, Torsten Schwede. *SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information*. *Nucleic Acids Research*; (1 July 2014) 42 (W1): W252-W258; doi: 10.1093/nar/gku340.
  22. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera – A visualization system for exploratory research and analysis. *J Comput Chem* 2004; 25(13):1605-12.
  23. Laskowski RA, MacArthur MW, Thornton JM. PROCHECK: validation of protein structure coordinates, in *International Tables of Crystallography, Volume F. Crystallography of Biological Macromolecules*, eds. Rossmann M G & Arnold E, Dordrecht, Kluwer Academic Publishers 2001, The Netherlands, pp. 722-725.
  24. Letunic I, Doerks T, Bork P, SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res* 2014; doi:10.1093/nar/gku949.

**How to cite this article:**

Jain D, Rawat R and Jatav VK: Phylogenetic Analysis of RAS Subfamily Proteins. *Int J Pharm Sci Res* 2016; 7(3): 1070-80.doi: 10.13040/IJPSR.0975-8232.7(3).1070-80.

All © 2013 are reserved by International Journal of Pharmaceutical Sciences and Research. This Journal licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

This article can be downloaded to **ANDROID OS** based mobile. Scan QR Code using Code/Bar Scanner from your mobile. (Scanners are available on Google Playstore)