Int.J.Curr.Microbiol.App.Sci (2017) *6*(7): 1727-1734

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 6 Number 7 (2017) pp. 1727-1734**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2017.607.208

Glycoprotein Based Phylogenetic Analysis of Rabies Virus Isolates

Neha R. Tomar[1*], Rajiv Kumar[2] and Anil Kumar[3]

[1]Department of Zoology, Kumaun University, S. S. J. Campus, Almora, Uttarakhand, India
[2]Division of Animal Genetics and Breeding, C. S. W. R. I., Avikanagar, Rajasthan, India
[3]Department of Molecular Biology and Genetic Engineering, College of Basic Sciences and Humanities, G. B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India
**Corresponding author*

Keywords

Canids, Rabies, Glycoprotein, Phylogeny, Strain specific PCR.

Article Info

Accepted:
19 June 2017
Available Online:
10 July 2017

A B S T R A C T

Glycoprotein of rabies virus is major contributor to pathogenicity. Glycoprotein based phylogenetic analysis of twenty three rabies and rabies related virus including CVS strain was done by aligning the sequences by clustal W. Glycoprotein sequence based phylogenetic tree was constructed and ten conserved motifs were analyzed among these glycoproteins. The unique sequence of motif 4 was observed only in CVS strain (FJ979833) which might be responsible for discriminating CVS strain from other rabies virus strains. The present study might open the possibility of developing strain specific PCR for identification of rabies virus in canids.

Introduction

Rabies virus (RV) almost always causes a fatal encephalomyelitis in several species of mammals including humans (Dietzschold, 1996). This virus is prototype of the *Lyssavirus* genus of the family *Rhabdoviridae* and is an enveloped virus having non-segmented, negative stranded RNA as genome. The genome of virus is about 12 kb in size and encodes five proteins; the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G), and the RNA-dependent RNA polymerase (L). The viral RNA, which is always encapsidated by N, forms the ribonucleoprotein (RNP), which is the template for viral replication and transcription (Conzelmann, 2004). The RNP together with P and L forms the viral replication complex, which is surrounded by the host cell-derived membrane that also contains glycoprotein (G).

Matrix protein has been proposed to bridge the RNP and the cytoplasmic domain (CD) of RVG to form the bullet-shaped virion (Mebatsion, 1999).

The RVG, which is organized as a trimer, is the major contributor to pathogenicity. The glycoprotein G of the virus is the major antigen responsible for the induction of protective immunity being the sole protein exposed on the surface of the virion. It interacts with cellular receptors (Hanham, 1993), mediates pH-dependent fusion, and

promotes viral entry from a peripheral site into the nervous system (Mazarakis, 2001). Moreover, RVG is involved in the trans-synaptic spread within the central nervous system (Etessami, 2000). Although pathogenicity of RV is a multigenic trait, the G is major contributor to the pathogenicity of a particular RV (Dietzschold, 1985).

Keeping in mind the importance of Rabies virus glycoprotein, *in silico* analysis of glycoprotein gene was done to find out potential sequence motifs to discriminate different strains of rabies virus so that it may be possible to develop strain specific PCR by designing primers targeting the discriminating motif.

Materials and Methods

In silico analysis of glycoprotein gene of rabies virus

Sequence homology search (BLAST alignment and scoring)

The glycoprotein gene used in present study (accession number FJ978833), described earlier (Tomar *et al.,* 2011) was subjected to homology search using Basic Local Alignment Search Tool of NCBI (http://www.ncbi. nlm.nih.gov) (Pruitt *et al.,* 2007). The sequence was translated into six possible reading frames using translate tool (http://ca.expasy.rg/tools/na.html).

Phylogenetic analysis

The sequence of glycoprotein along with known sequences of lyssaviruses was aligned using ClustalW (Thompson *et al.,* 1994) and phylogenetic tree was constructed by N-J branched chain method. A tree was inferred by Bootstrap phylogenetic inference using MEGA3.1. The phylogenetic distance was observed by scoring.

Analysis of motifs in glycoprotein and their correlation with phylogeny

Motifs present in rabies CVS glycoprotein sequence and other glycoprotein sequences from lyssaviruses were analyzed using MEME software version 4.1.0 (Bailey *et al.,* 1994).

Results and Discussion

Sequence homology search (BLAST alignment and scoring)

The nucleotide sequence obtained after sequencing was translated into protein and analyzed *in silico* for characterization. The sequence was subjected to Basic Local Alignment Search Tool (BLAST) to reveal the similarity at protein level with other existing glycoprotein of lyssaviruses. It showed the highest score for sequence similarity with other rabies viruses. The query sequence matches solely with the rabies virus glycoprotein for which it showed the score > 200. This score indicated high level of homology of our rabies CVS glycoprotein sequence with other rabies virus glycoprotein sequences.

Sequence alignment and their phylogenetic study (Clustal W and MEGA)

The query sequence (Accession no. FJ979833) along with known glycoprotein sequences from lyssaviruses when subjected to multiple sequence alignment by clustalW server, revealed some interesting features of similarity between rabies virus glycoprotein and lyssavirus glycoprotein.

The sequences of glycoprotein along with the query sequence (Rabies CVS) showed some completely conserved regions while some stretches showed the differences occasionally at certain locations across the length of sequences only. The stretch from aa 635-654

was found conserved among all lysaviruses except that at position 635 in clustal W alignment, T was replaced by A in the rabies CVS (FJ979833), while it was found conserved in all other lyssaviruses including rabies viruses. The stretch from 736-742 was uniformly observed in all lyssaviruses. The stretch from 772-784, 870-882 was also found conserved in most lyssaviruses except in lagos bat viruses at Position 775 V was replaced by T, at position 780 the replacement of L with F and at position 877 replacement of F with Y was observed. In the stretch 938-947 in all the DUV and EBLV at position 945 usual replacement of S with A, and in some EBLV at position 946 the replacement of L with M was observed. In the clustal W alignment it was observed that most differences were found in lagos bat viruses which supported the position of lagos bat viruses in the phylogenetic tree as separate cluster.

Phylogenetic analysis

The coding region of the G gene of rabies CVS isolate analyzed in this study consisted of 1590 nt coding for a 524 aa protein. A set of 23 complete G gene sequences consisting of seven rabies viruses (genotype 1), eight European bat lyssaviruses; five of which were from genotype 6 and three were from genotype 5, two duvenhage viruses (genotype 4), four Lagos bat viruses (genotype 2) and two Australian bat lyssaviruses (genotype 7) isolates.

The available full length glycoprotein sequences for genotypes 1, 2, 4, 5, 6, 7 and CVS isolate (FJ979833) were assembled for phylogenetic analysis and phylogenetic tree construction using N-J branched chain method revealed two major clusters (Figure 1). All seven RVG sequences from genotype 1 (FJ979833, AY009097, RVU03767, U11752, AF325475, AF325465, AF325460) were found to be clustered close to ABLV from genotype 7 (AF426297, AF426292) while other viruses were lying in different clusters. The N-J method indicated high boot strap support for all the major clusters representing different lyssavirus genotypes. Bootstrap similarity was above > 45% in 1000 cycles of sampling.

The rabies viruses along with rabies CVS isolate (FJ979833) were clearly distinguished with high bootstrap support (Figure 1).

Calculation of P distances

The intrinsic variation of genotype 1 was 88.4-92.4% on the amino acid level. The closest identity of genotype 1 isolate to another lyssaviruses genotype was the identity to genotype 7. Genotype 1 isolate had a 72.9-76 % aa identity to genotype 7 whereas it had 71.9-72.9% aa identity to genotype 6, 69.7-69.8% aa identity to genotype 5. All the genotype 1 isolates were found to have a higher sequence identity to each other than to genotype 7 isolates (Figure 2).

Analysis of conserved motifs (MEME, MAST)

To find out the conserved motifs in the glycoprotein sequence of different lyssavirus, these sequences were further subjected to MEME program for motif analysis. A total of 10 conserved motifs were analyzed (Figure 3). Presence of high density of nucleotide motifs and signature sub sequences were an indication of high level of homology existing among the analyzed sequences. Motif 1, 5, 6, 8 and 9 were found to be conserved in all seven rabies virus glycoproteins while it showed some amino acid substitutions in other viruses. Among these, motif 8 was peculiar to RVG only. This motifs 8 might play important role for putting them together in one subcluster. Motif 9 is universally present in all lyssavirus

glycoproteins. Among the 10 motifs, Rabies CVS strain differed from all the rabies related strains and rabies strains in motif 4 at position 3 in that it had Alanine (A), while other lyssaviruses including rabies viruses had Threonine (T) at this position (Figure 3). The unique sequence of motif 4 was observed only in CVS strain (FJ979833). This motif might be responsible for separating the Rabies CVS strain from other strains of rabies viruses. Motif 7 and 10 were found in all strains of rabies viruses except AY009097. These motifs might play a role in discriminating this virus from the glycoproteins of other viruses. The sequences of different motifs have been shown in figure 4. MEME motifs are represented by position-specific probability matrices that specify the probability of each possible letter appearing at each possible position in an occurrence of the motif (Figure 4). These are displayed as "sequence LOGOS", containing stacks of letters at each position in the motif. The total height of the stack is the "information content" of that position in the motif in bits. The height of the individual letters in a stack is the probability of the letter at that position multiplied by the total information content of the stack.

Fig.1 Phylogenetic tree showing evolutionary relationship among various strains of lyssa Viruses including rabies CVS (Accession No. FJ979833)

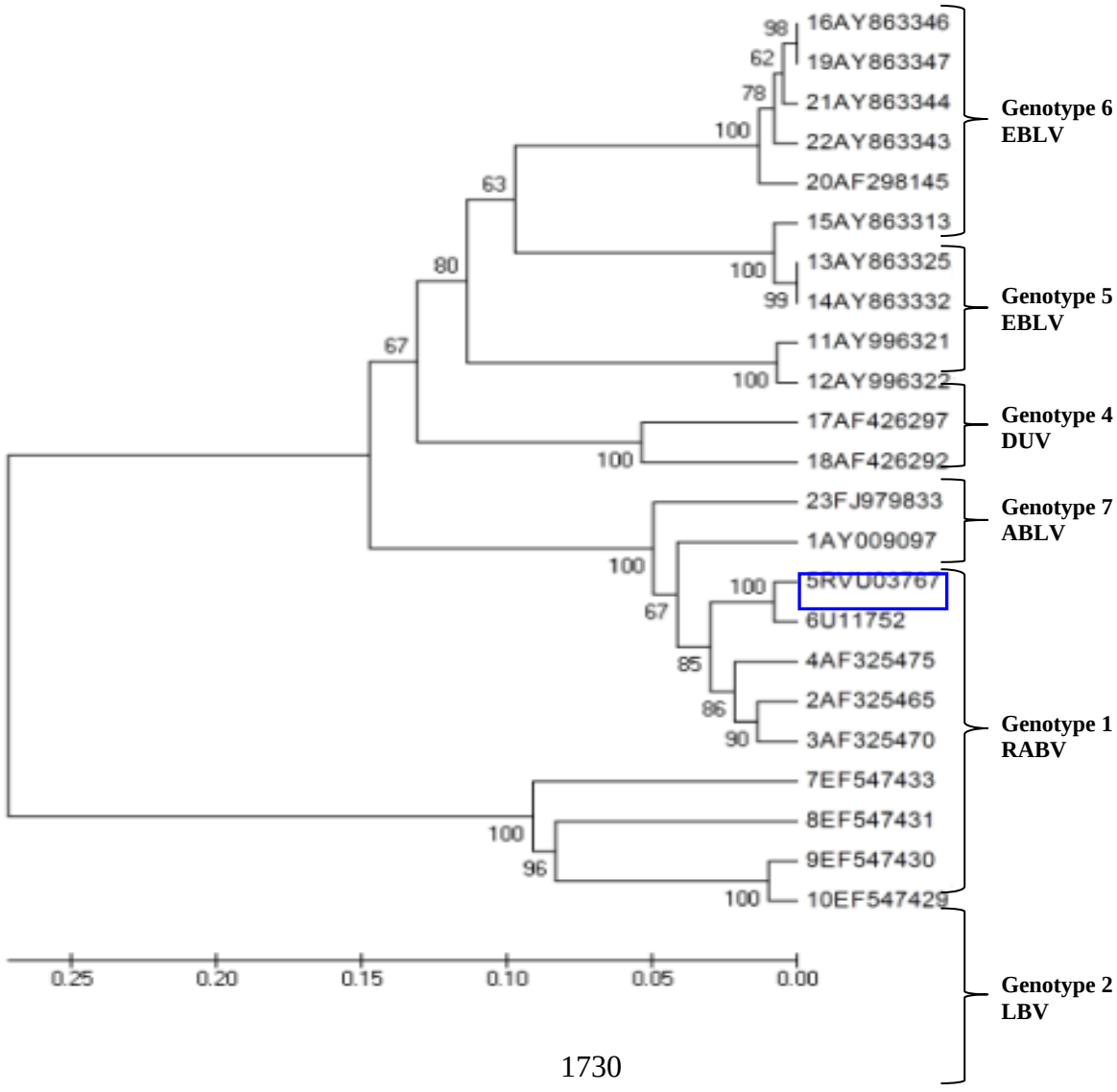


Fig.2 Different lyssaviruses showing the Paired distances to each other

Percent Identity

Divergence

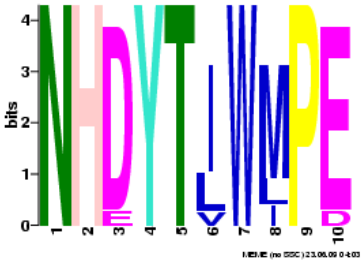
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22		
1		77.1	74.0	91.6	70.0	70.2	70.2	74.0	74.2	73.9	68.1	68.5	57.1	57.3	57.9	56.9	90.5	92.7	91.0	72.9	94.8	96.2	1	AF325475RABV
2	27.4		90.3	75.6	76.5	76.3	76.3	80.2	80.2	79.8	72.9	73.7	58.2	58.2	58.4	58.8	76.0	76.5	74.1	78.8	77.3	78.1	2	AF426292ABLV
3	31.9	10.4		73.3	73.3	73.5	73.5	76.9	76.9	76.5	71.2	71.6	57.9	57.7	58.0	58.2	72.9	74.2	72.4	75.6	74.0	75.0	3	AF426297ABLV
4	8.9	29.6	33.1		69.5	69.7	69.7	72.7	72.5	72.1	68.3	68.5	56.1	56.3	57.3	56.7	88.4	89.9	95.4	71.2	92.0	93.7	4	AY009097RABV
5	38.2	28.2	33.1	39.2		98.9	98.9	80.3	80.5	80.0	78.6	78.6	58.2	58.2	57.7	58.6	69.7	71.0	68.3	79.0	71.8	71.6	5	AY863313EBLV
6	37.9	28.5	32.8	38.8	1.2		100.0	80.7	80.9	80.3	78.8	78.8	58.4	58.4	57.3	58.8	69.8	71.2	68.3	79.4	71.9	71.8	6	AY863325EBLV
7	37.9	28.5	32.8	38.8	1.2	0.0		80.7	80.9	80.3	78.8	78.8	58.4	58.4	57.3	58.8	69.8	71.2	68.3	79.4	71.9	71.8	7	AY863332EBLV
8	31.9	23.1	27.7	33.9	22.9	22.3	22.3		99.0	98.7	78.2	79.0	57.3	57.3	57.1	55.7	73.1	74.0	71.8	97.5	75.0	75.4	8	AY863343EBLV
9	31.6	23.1	27.7	34.2	22.6	22.1	22.1	1.0		99.2	78.6	79.4	57.1	57.1	56.9	55.6	73.3	74.2	71.9	98.1	74.4	75.6	9	AY863344EBLV
10	32.2	23.6	28.2	34.8	23.4	22.9	22.9	1.3	0.8		78.1	78.8	56.5	56.5	56.3	55.0	72.9	73.9	71.6	98.3	74.0	75.2	10	AY863347EBLV
11	41.4	33.6	36.4	41.1	25.2	24.9	24.9	25.8	25.2	26.0		97.7	56.1	56.1	54.8	55.9	68.5	68.5	67.2	77.1	68.7	69.7	11	AY996321DUV
12	40.8	32.5	35.7	40.8	25.2	24.9	24.9	24.7	24.2	24.9	2.3		56.7	56.5	55.4	56.3	68.9	69.1	67.4	77.9	69.1	70.0	12	AY996322DUV
13	62.7	60.2	61.1	64.9	60.2	59.8	59.8	62.3	62.7	64.0	64.9	63.6		96.9	85.2	81.8	56.3	57.3	56.1	55.6	57.1	57.5	13	EF547429LBV
14	62.3	60.2	61.5	64.4	60.2	59.8	59.8	62.3	62.7	64.0	64.9	64.0	3.1		84.1	81.2	56.7	57.3	56.5	55.6	57.1	57.7	14	EF547430LBV
15	61.1	59.8	60.6	62.3	61.5	62.3	62.3	62.7	63.1	64.4	67.9	66.6	16.5	17.9		80.3	56.5	57.9	57.5	55.4	57.7	58.2	15	EF547431
16	63.1	59.0	60.2	63.6	59.4	59.0	59.0	65.7	66.2	67.5	65.3	64.4	20.9	21.7	23.0		56.7	57.5	55.7	54.0	57.5	57.9	16	EF547433LBV
17	10.2	29.0	33.6	12.7	38.8	38.5	38.5	33.3	33.1	33.6	40.8	40.1	64.4	63.6	64.0	63.6		89.9	88.0	71.9	91.0	92.4	17	FJ979833.1RABV
18	7.6	28.2	31.6	10.9	36.7	36.4	36.4	31.9	31.6	32.2	40.8	39.8	62.3	62.3	61.1	61.9	10.9		89.3	72.9	93.3	95.0	18	RVU03767RABV
19	9.6	31.3	33.9	4.7	41.1	41.1	41.1	35.4	35.1	35.7	43.0	42.7	64.9	64.0	61.9	65.7	13.1	11.6		70.6	91.4	92.9	19	AB009663RABV
20	33.6	24.9	29.6	36.4	24.7	24.2	24.2	2.5	1.9	1.7	27.4	26.3	66.2	66.2	66.6	69.7	35.1	33.6	37.3		73.1	74.2	20	AF298145EBLV
21	5.3	27.1	31.9	8.5	35.4	35.1	35.1	30.4	31.3	31.9	40.4	39.8	62.7	62.7	61.5	61.9	9.6	7.0	9.1	33.3		97.5	21	AF325465RABV'
22	3.9	26.0	30.4	6.6	35.7	35.4	35.4	29.9	29.6	30.2	38.8	38.2	61.9	61.5	60.2	61.1	8.1	5.1	7.4	31.6	2.5		22	AF325470
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22		

Fig.3 Distribution of conserved motifs among Lyssaviruses

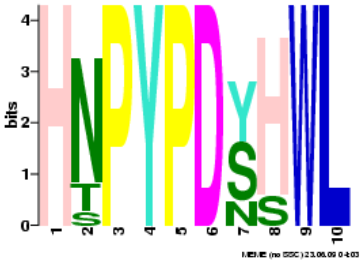
Isolate	Conserved Motifs
AY863346 EBLV	
AY863347 EBLV	
AY863344 EBLV	
AY863343 EBLV	
AF298145 EBLV	
AY863313 EBLV	
AY863325 EBLV	
AY863332 EBLV	
AY996321 DUV	
AY996322 DUV	
AF426297 ABLV	
AF426292ABLV	
FJ979833 CVS	
AY009097 RABV	
RVU03767 RABV	
U11752 RABV	
AF325475 RABV	
AF325465 RABV	
AF325470 RABV	
EF547430 LBV	
EF547429 LBV	
EF547431 LBV	
EF547433 LBV	

Fig.4 The sequence logos of analyzed motifs

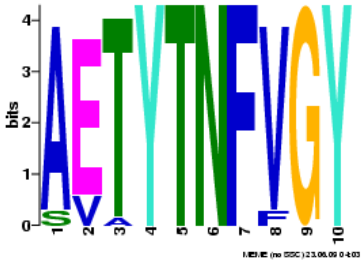
Motif-1



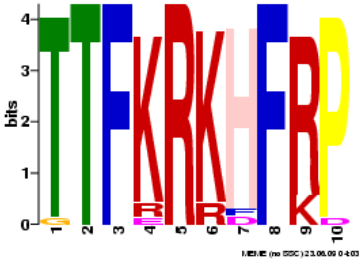
Motif-2



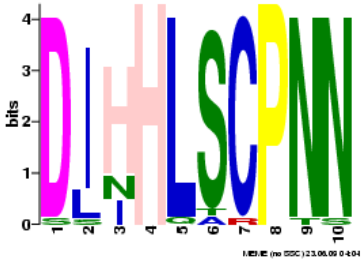
Motif-3



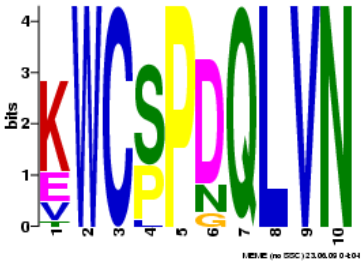
Motif-4



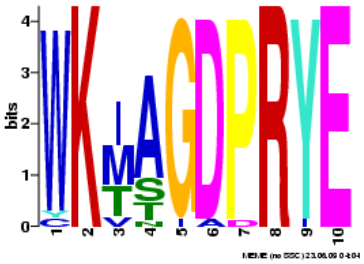
Motif-5



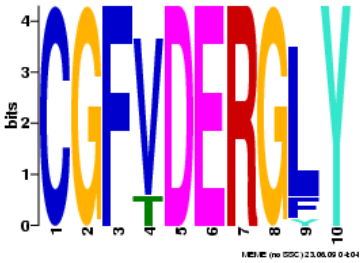
Motif-6



Motif-7



Motif-8



Motif-9

Motif-10

All the 23 sequences when subjected to BLAST revealed more than 85% homology with all rabies virus glycoprotein. All the 23 sequence along with Rabies CVS glycoprotein sequences subjected to multiple sequence alignment and phylogenetic tree construction using N-J branched chain method revealed two major clusters. One cluster containing all the lyssaviruses except lagos bat virus and the other cluster was having Lagos bat viruses.

In order to derive phylogenetic relationship between CVS rabies virus isolate and other lyssaviruses, the glycoprotein gene was sequenced. A high level of protein sequence conservation indicated a close phylogenetic relationship between virus isolates from different regions. These isolates appeared to be distinct from but closely related to European strains of rabies virus. The Rabies CVS (FJ979833) was found to be clustered close to the PG strain of rabies virus (AY009097) showing 89 % similarity in glycoprotein sequence and all the rabies viruses were found to be clustered close to Australian bat lyssaviruses.

In the phylogenetic analysis performed in this study, genotype 1 (Rabies viruses accession numbers FJ979833, AY009097, RVU03767, U11752, AF325475, AF325470, AF325465) was observed as a separate cluster with high bootstrap support regarding to the glycoprotein used. P-distance analysis of G protein's amino acid sequences indicated that overlaps between the intragenotypic and intergenotypic identities of lyssavirus genotypes occurred. The occurrence of different motifs also supported the position of our strain (FJ979833) in the phylogenetic tree. The unique sequence of motif 4 was observed in the CVS strain only (FJ979833) which might be responsible for differentiating this strain from other rabies viruses. Discrimination of certain virus strain on the basis of peculiar sequences can prove a key in development of strain specific PCR. Further the study can also open the possibility to design novel molecules against conserved regions of glycoprotein which can block viral pathogenesis since it is related to glycoprotein itself (Tomar *et al.*, 2010). The conserved motifs analyzed among rabies viruses can also be used to design epitope based peptide vaccines against rabies virus (Tomar *et al.*, 2011).

In this study we analysed phylogenetic distances for different rabies strains on the basis of glycoprotein gene and we analysed ten different kinds of motifs present in glycoprotein. Among these, motif 8 was found in RVG only. This motif might play important role for putting them together in one subcluster. Out of these 10 motifs our strain differs from rest of strain in motif 4 which was peculiar to our strain only. This motif can justify the position of our strain (accession no. FJ979833) in phylogenetic tree. This study will help in designing novel molecules targeting conserved motifs, designing of epitope based peptide vaccines and development of strain specific PCR targeting the novel motif to identify different strains of rabies virus in canids.

Acknowledgement

This work was supported by Centre of Bioinformatics, G. B. Pant University of Agriculture and Technology, Pantnagar.

References

Bailey, T. L., and Elkan, C. 1994. "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, AAAI Press, Menlo Park,*

California. pp. 28-36.

Conzelmann, K.K., Cox, J.H., Schneider, L.G., Thiel, H.I., 1990. Molecular cloning and complete nucleotide sequence of attenuated rabies virus SAD BI9. *Virol.* 175: 485-499.

Dietzschold, B., Rupprecht, C. E., Fu, Z.F. and Koprowski, H. 1996. Rhabdoviruses: In B.N., Knipe D.M. and Howley P.M. Fields *Virol.* 3rd Ed. Lippincott Raven Publishers Philadelphia. pp. 1137-1159.

Etessami, R., Conzelmann, K.K., Fadai-Ghotbi, B., Natelson, H., Tsiang, and Ceccaldi, P.E. 2000. Spread and pathogenic characteristics of a G deficient rabies virus recombinant: an in vitro and in vivo study. *J. Gen. Virol.* 81:2147–2153.

Hanham, C. A., Zhao, F. and Tignor, G.H. 1993. Evidence from the antiidiotypic network that the acetylcholine receptor is a rabies virus receptor. *J. Virol*. 67:530–542.

Mebatsion, T., Weiland, F. And Conzelmann, K.K. 1999. Matrix protein of rabies virus is responsible for the assembly and budding of bullet shaped particles and interacts with the transmembrane spike glycoprotein G. *J. Virol.* 73: 242-250.

Mazarakis, N. D., Azzouz, M., Rohll, J.B., Ellard, F.M., Wilkes, F.J., Olsen, A.L., Carter, E.E., Barber, R.D., Baban, D.F., Kingsman, S.M., Kingsman, A.J., O'Malley, K. and Mitrophanous, K.A. 2001. Rabies virus glycoprotein pseudotyping of lentiviral vectors enables retrograde axonal transport and access to the nervous system after peripheral delivery. *Hum. Mol. Genet*. 10:2109–2121.

Pruitt KD., Tatusova T., Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res, 35, D62-D6.

Thompson JD., Higgins DG., Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. Nov 11; 22(22):4673-80.

Tomar NR, Chandra R, Kumar R, Tiwari A K and Kumar A 2011. Expression of rabies virus glycoprotein gene into eukaryotic system and determination of potential T-cell epitopes. Indian J. Experimental Biology. 49(8):594-599.

Tomar NR**,** Singh V, Marla SS, Chandra R, Kumar R and Kumar A. 2010. Molecular Docking Studies with Rabies Virus Glycoprotein to Design Viral Therapeutics, Indian J. Pharm. Sci. 72 (4): 486-490.

Tomar N R, Marla S S' Chandra R, Kumar R and Kumar A. 2011. *In silico* epitope (T-228) rabies virus vaccine candidate. Online Journal of Bioinformatics, Volume 12(1):46-56.

How to cite this article:

Neha R. Tomar, Rajiv Kumar and Anil Kumar. 2017. Glycoprotein Based Phylogenetic Analysis of Rabies Virus Isolates. *Int.J.Curr.Microbiol.App.Sci.* 6(7): 1727-1734.
doi: https://doi.org/10.20546/ijcmas.2017.607.208