CHARACTERIZATION OF NUCLEOPROTEIN GENE SEQUENCE OF AN INDIAN ISOLATE OF RABIES VIRUS

R. JAYAKUMAR*, K.G. TIRUMURUGAAN, G. GANGA, K. KUMANAN, A. MAHALINGA NAINAR

Department of Animal Biotechnology, Madras Veterinary College, Tamil Nadu Veterinary and Animal Sciences University, Chennai-600 0007, India

Received October 17, 2003; accepted February 18, 2004

Summary. – Rabies occurs in all parts of Indian sub-continent except Andaman and Nicobar and Lakshadweep group of islands. The full-length nucleoprotein (N) gene sequence of a rabies virus isolate from India is reported for the first time and the same has been compared with available N gene sequences from the database. A central domain of 230 amino acids (aa) from aa 141 to aa 370 exhibited more than 95% similarity. There were 8 amino acid positions (aa 29, 32, 38, 84, 119, 379, 438, and 439) at which substitution was unique for Indian isolates but common for laboratory strains. In antigenic epitopes, except for a single amino acid difference at the antigenic site IV, the amino acids were conserved. The Indian isolate also possessed two *Bam* HI sites (aa 247 and 278), while the other Asian isolates had only one site at aa 278 or were not digested with *Bam* HI at all. Phylogenetic analysis also demonstrated that the Indian isolate was closely related to the Sri Lankan isolate and grouped in the cluster that comprised of the isolates from other Asian countries namely China and Pakistan.

Key words: Rabies virus; Indian isolate; nucleoprotein; nucleotide sequencing; phylogenetic analysis

Introduction

Rabies virus belongs to the order *Mononegavirales* with non-segmented negative-stranded RNA genomes. Within this order, viruses with a distinct "bullet" shape are classified in the *Rhabdoviridae* family, which includes at least three genera of animal viruses, namely *Lyssavirus*, *Ephemerovirus*, and *Vesiculovirus*. Recent phylogenetic tree constructed from either nucleotide or amino acid sequence data has described the *Lyssavirus* genus as six distinct genetic lineages: genotype 1, classical Rabies virus (Bourhy *et al.,* 1993); genotype 2, Lagos bat virus (Bourhy *et al*., 1993b); genotype 3, Mokola virus (Bourhy *et al*., 1993c); genotype 4, Duvenhage virus (Smith *et al*., 1992); genotype 5, EBL1 virus (Smith *et al.,* 1993); genotype 6, EBL2 virus (Tordo *et al*., 1993).

Rabies occurs in all parts of Indian sub-continent except Andaman, Nicobar and Lakshwadeep group of islands. The increasing human population as well as the increase of the stray dog population in the country had resulted in more human-dog contacts. It has been estimated that 30,000 people die of rabies every year and about 0.7 million people undergo anti-rabies treatment following exposure. Dogs are responsible for 96% of human rabies cases. The mortality in livestock and pet animals is far higher than the number available due to lack of exact reports (Sehgal, 1997).

The nucleoprotein (N) gene is a more convenient target for routine diagnosis, comparison of molecular and antigenic evaluation in taxonomic studies (Kamolvarin *et al*., 1993; Sacramento *et al.*, 1991; Nadin-Davis *et al.*, 1993; Rupprecht and Smith, 1994). It is also highly conserved, strongly expressed (Tordo and Kouknetzoff, 1993) and a target in antigenic classification of Lyssa virus based on the reactivity patterns with anti-N monoclonal antibodies. Current molecular phylogenetic tree for the *Lyssavirus* genus is constructed with the nucleotide or amino acid sequence obtained for the entire coding region of N gene (Bourhy *et*

*E-mail: rjkumar48@yahoo.com; fax: +25362787/+253889445.
Abbreviations: aa = amino acid; FAT = fluorescent antibody test; N = nucleoprotein

al., 1993); however, identical phylogenetic tree has been obtained from partial nucleotide sequence of N gene or genes encoding other viral proteins such as the glycoprotein (Bourhy *et al*., 1993; Ravkov *et al*., 1995; Smith *et al*., 1993; Tordo and Kouknetzoff, 1993). There are a few investigations on rabies virus in Asia, however, molecular epidemiology and the nucleotide or amino acid sequence of the genome of local isolates have not been reported from India.

Hence, in this study, we determined the full-length N gene nucleotide sequence of an Indian isolate of a street rabies virus and compared it with the available sequences of N genes from the database to find the relationship among rabies virus isolates circulating worldwide.

Materials and Methods

Brain specimens were collected during necropsy from dogs suspected for rabies. Rabies diagnosis was done by a direct fluorescent antibody test (FAT) using anti-rabies N polyclonal antibody conjugate (Sanofi Diagnostic Pasteur, France). The tested specimens were stored at -86^{0}C for future use. One specimen found positive by FAT was used for RNA extraction.

Total RNA extraction was performed as described earlier (Chomczynski and Sacchi, 1987). Briefly, total RNA was extracted from 1 g of brain material by a standard acid guanidium thiocyanate-phenol-chloroform method.

Primers. Oligonucleotide primers NF28 (5'-GCG GAT CCC ACC TCT ACA ATG GAT GCC G-3') and NR32 (5'-TTA TGA GTC ACT CGA ATA TGT CT-3') was designed to amplify the full-length N gene of the local isolate of rabies virus.

RT-PCR. cDNA was prepared using 1 µg of total RNA using the THERMOSCRIPT™ Reverse Transcriptase (Gibco-BRL, USA) and random hexamers according to the manufacturer's instructions. PCR was performed in MJ Research Thermal Cycler (PTC-200) programmed for 5 cycles with denaturation at 94^{o}C for 2 mins; annealing at 45^{o}C for 1min, and elongation at 72^{o}C for 2 mins. Subsequently, the annealing temperature was increased to 50^{o}C for 5 cycles and to 55^{o}C for 20 cycles. The ultimate elongation was at 72^{o}C for 10 mins.

Nucleotide sequencing and phylogenetic analysis. The amplicon was purified and sequenced using the forward and reverse PCR primers in an ABI Prism 3700 DNA Sequencer (PE Applied Biosystems). The nucleotide sequences thus obtained were aligned to get the full-length N gene sequence of the isolate and had been deposited in Genbank under Acc. No. AF374721. Multiple sequence alignment of the local isolate and fixed viruses was generated by MEGA (Kumar *et al.*, 1993) to find the varying amino acids. The 5'-400 bp N gene sequence data was analyzed using the DNAdist (maximum likelihood option) and the Neighbour programs of the PHYLIP package (Felsenstein, 1989) using estimated transition transversion ratios derived from the PUZZEL Program. The phylogenetic tree was produced using the Drawtree Program and the TREEVIEW Program visualized bootstrap values.

Results and Discussion

RT-PCR amplified a product of 1,380 bp corresponding to the full-length coding region of the N gene as expected.

The complete nucleotide sequence of the positive sense strand revealed one long ORF from the first ATG codon at position 19 to the stop codon TAA at position 1,371 that was able to code for a polypeptide of 450 aa long, of which 306 were conserved in all the isolates (Fig. 1). The putative phosphorylation site mapped to the serine at position 389. A central domain of 230 amino acids (aa 141–370) exhibited more than 95% similarity. There were 8 amino acid positions (29, 32, 38, 84, 119, 379, 438, and 439) at which substitution was unique for the Indian isolates but common for the laboratory strains. Except for a single amino acid difference, there was no difference in the amino acids that constituted the antigenic epitopes (Lafon and Wiktor, 1985). The Indian isolate sequence possessed two *Bam*HI restriction sites (247 and 278), while the Asian isolates had a single restriction site (278) or did not have a restriction site at all.

The unrooted phylogenetic tree (Fig. 2) separated the compared rabies isolates into 8 clusters: (1) Africa 1a, (2) Africa 1b, (3) Africa 2, (4) Africa 3, (5) Africa 4, (6) Vaccine 1–3, (7) Europe 1–6, and (8) Asia. The comparison also showed that the Indian isolate was closely related to the Sri Lanka isolate and grouped in the cluster that comprised of the isolates from other Asian countries like China and Pakistan. Besides, the Indian isolate and the Sri Lanka isolate formed a cluster that was distinct from other rabies virus isolates worldwide (Arai *et al.*, 2001).

This report presents, for the first time, the full-length coding sequence of a non-passaged rabies virus street isolate. The predicted protein product of this isolate was highly conserved, and differred marginally from the N of previously characterized rabies strains (Mannen *et al.*, 1991). The variations were found to cluster at the termini of the protein. The sequence reported had a unique *Bam*HI restriction site that was absent in other Asian isolates. However, this restriction site needs further study and confirmation in other isolates. The results of this study also confirm the earlier studies (Smith *et al.*, 1992; Arai *et al*., 1997) and the importance of nucleotide sequence analysis of selected regions of rabies virus N gene as an excellent tool for determining the areas of the world where exposure to rabies virus had occurred.

Acknowledgement. The work was supported by the Indian Council of Agricultural Research, Governement of India. We thank Drs. K. Thangaraj and L. Singh, both of the Centre for Cellular and Molecular Biology, Hyderabad, India, for their help in nucleotide sequencing. We also thank Drs. L.M. McElhinney and S.A. Nadin Davis, Centre of Expertise for Rabies, Ottawa Laboratory, Canadfa, for their contribution during the initial stages of this study.

```
AF374721  MDADKIVFKV NNQVVSLKPE IIVDQYEYRY PRIKDLKEPS ITLGKAPDLN KAYKSVLSGM   NAAKLDPDDV CSYLAAAMQF FEGSCPEDWT SYGILIARKG DKITPDSLVD IKRTDVEGSW  120
3AG       ........RA .......R.. ..A.....K. .A.....K.C .......... ..........   .......... .......... ...T...... .........R .....N.R.E ...N....N.
CTN       ........R. .......... ........K. .A.....K.. .......... ..........   .......... .......... ...T.....N .......... .......... ........N.
CVS       .......... .......... ........K. .A.....K.C .......... ..........   .......... .......... ...T...... .......... .R...N...E ........N.
PV        .......... .......... ........K. .A.....K.C .......... ........C.   S......... .......... ...T...... ....V..... .....G...E ........N.
RCHL      ....R...RA .......R.. ..A.....K. .A.....K.C .......... ..........   .......... .......... ...T...... .......... .....N...E ...N....N.
SAD       .......... .......... ........K. .A.....K.C .......... ..........   S....N.... .......... ...T...... ....V..... .....G...E ........N.

AF374721   ALTGGMELTR DPTVSEHASL VGLLLSLYRL SKISGQNTGN YKTNIADRIE QIFETAPFVK   IVEHHTLMTT HKMCANWSTI PNFRFLAGTY DMFFSRIEHL YSAIRVGTVV TAYEDCSGLV  240
3AG       .......M.. .......S.. .......... .......... .......... ..........   .......... .......N.. .......... .L........ .......... ..........
CTN       .......... .......... .......... ......S... .......... ..........   .......... .......... .......... .......... .......... ..........
CVS       .......... .......... .......... .......... .......... ..........   .......... .......... .......... .......... .......... ..........
PV        .......... ....P..... .......... ......S... .......... ..........   .......... .......... .......... .......... .......... ..........
RCHL      .......... .......... .......... .......... .......... ..........   .......... .......... .......... .......... .......... ..........
SAD       .......... ....P..... .......... .......... .......... ..........   .......... .......... .......... .......... .......... ..........

AF374721   SFTGFIKQIN LTAREAILYF FHKNFEEEIR RMFEPGQETA VPHSYFIHFR SLGLSGKSPY   SSNAVGHVFN LIHFVGCYMG QVRSLNATVI AACAPHEMSV LGGYLGEEFF GKGTFERRFF  360
3AG       .......... .......... .......... .......... .......... ..........   .......... ..........I....S... .......... .......... .R........
CTN       .......... ...K...... .......... .......... .......... ..........   .......... .......... .......... .......... .......... ..........
CVS       .......... .......... .......... .......... .......... ..........   .......... .......... .......... .......... .......... ..........
PV        .......... .......... .......... .......... .......... ..........   .......... .......... .......... .......... .......... ..........
RCHL      .......... ......T... .......... .......... .......... ..........   .......... ..........I......... .......... .......... .R........
SAD       .......... .......... .......... .......... .......... ..........   .......... .......... .......... .......... .......... ..........

                                      *
AF374721  RDEKELQEYE AAELTKTDIA LADDGTVNSD DEDYFSGETR SPEAVYTRIM MNGGRLKRSH   IRRYVSVSSN HQARPNSSPE FLNKTYSSDS  450
3AG       .......... ........V. .......... .......... G.....A... ..........   .......... .......FA. ..........
CTN       .......... ........L. .......... .......... .......... ..........   .......... .......FA. ..........
CVS       .......... ......S.V. .......... .......... .......... ..........   .......... .......FA. .......N..
PV        .......... ........V. .......... .......... .........I ..........   .......... .......FA. ..........
RCHL      .......... ........V. .......D.. ...H.P..A. G.....A... ..........   .......... .......FA. ..........
SAD       .......... ........V. .......... .......... .......... ..........   .......... .......FA. ..........
```

Fig. 1

Comparative alignment of the predicted amino acid sequence of the Indian Isolate of Rabies N gene sequence

The sequence was aligned with the 3AG (AF155039), CTN (AF367863), CVS (AB069972), PV (M13215), RCHL (AB009663), and SAD (M31046). The single-letter amino acid code is used. Only amino acid changes relative to the sequence of AF374721 are indicated. The putative phosphorylation site mapping serine at position 389 is marked with an asterisk. The epitope 816 (underlined) and the epitope 377, 802 (double underlined) as reported by Lafon and Wiktor, (1985) on the ERA strain of Rabies virus are also shown. The amino acids unique to the Indian isolate (aa 29, 32, 38, 74, 119, 379, 338, and 339) are in bold.

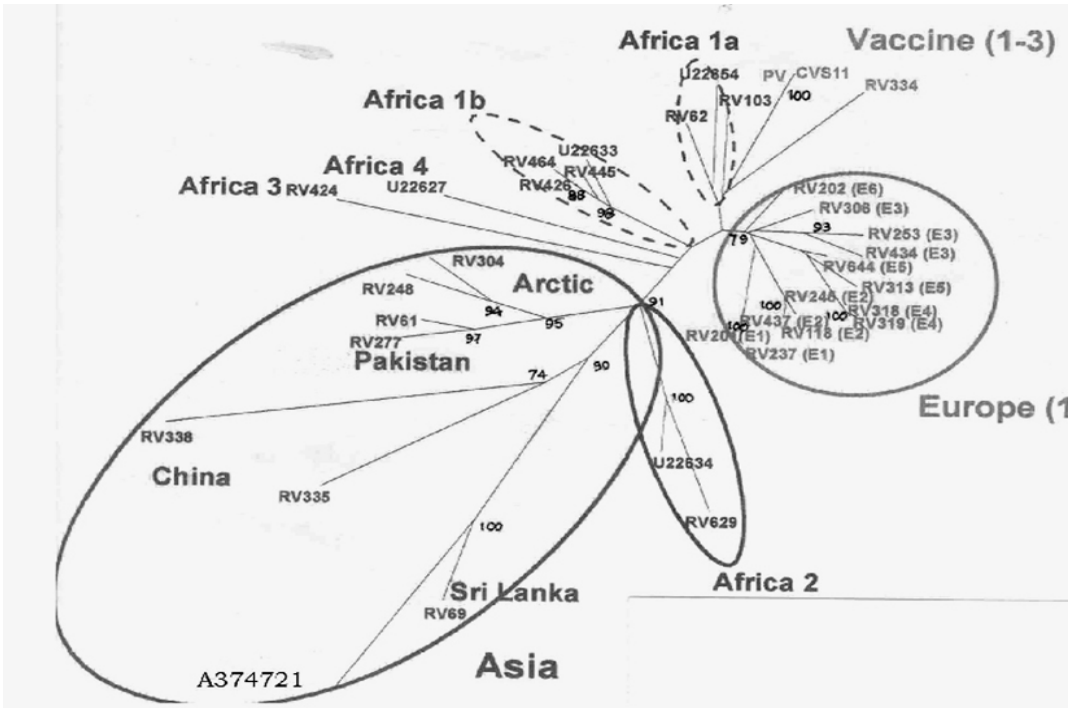

Fig. 2

An unrooted phylogenetic tree indicating genetic relationships among Rabies virus isolates from various countries

The phylogenetic tree was constructed by comparing the 5'-400 bp of the Indian street virus N gene sequence (AF374721) with available N gene sequences by the neighbor joining method.

References

Arai YT, Takahashi H, Kameoka Y, Shiino T, Wimalaratne O, Lodmell DL (2001): Characterization of Sri Lanka rabies virus isolates using nucleotide sequence analysis of nucleoprotein gene. *Acta Virol.* **45**, 327–331.

Arai YT, Yamada K, Kameoka Y, Horimoto T, Yamamoto K, Yabe S, Nakayama M, and Tashiro M (1997): Nucleoprotein gene analysis of fixed and street rabies virus variants using RT-PCR. *Arch. Virol.* **142**, 1787–1796.

Bourhy H, Kissi B, Lafon M, Sacramento D, Tordo N (1993): Molecular diversity of the Lyssavirus genus. *Virology* **194**,70–81.

Bourhy H, Kissi B, Badrane H, Tordo N (1993b): Analyse de la variabilite genetique des lyssavirus. *Med. Malad. Infect.* **23**, 533–536.

Bourhy H, Kissi B, Tordo N (1993c): Taxonomy and evolutionary studies on lyssaviruses with special reference to Africa. Proceedings of the research workshop on rabies, 3–4 May 1993, Pretoria. *Onderstepoort J. Vet. Res.* **60**, 277–282.

Chomczynski P, Sacchi P (1987): Single-step method of RNA isolation by Acid Guanidium Thiocyanate-Phenol Chloroform extraction. *Anal. Biochem*. **162**, 156–159.

Felsenstein J (1989): Distance methods for inferring phylogenies: a justification. *Evolution* **38**, 16–24.

Kamolvarin N, Tirawatnpong T, Rattanasiwamoke R, Tirawatnpong S, Panpanich T Hemachudha T (1993): Diagnosis of rabies by polymerase chain reaction with nested primers. *J. Infect. Dis*. **167**, 207–210.

Kumar S, Koichir Tamura, Masatoshi Nei (1993): MEGA: Molecular Evolutionary Genetic Analysis, Version 1.0, The Pennysylvania State University, University Park, PA 16802.

Mannen K, Hiramatsu K, Mifune K, Sakamoto S (1991): Conserved nucleotide sequence of rabies virus cDNA encoding the nucleoprotein. *Virus Genes* **5**, 69–73.

Nadin-Davis SA, Casey GA, Wandler A (1993): Identification of regional variants of the rabies virus within the Canadian province of Ontario. *J. Gen. Virol.* **75**, 829–837.

Ravkov EV, Smith JS, Nichol ST (1995): Rabies virus glycoprotein gene contains along 3'non-coding region which lacks pseudogene properties. *Virology* **206**,718–723.

Rupprecht CE, Smith JS (1994): Raccoon rabies: the re-emergence of an epizootic in densely populated area. *Semin. Virol*. **5,** 155–164.

Sacramento D, Bourhy H, Tordo N (1991): PCR technique as an alternative method for diagnosis and molecular epidemiology of rabies virus. *Mol. Cel. Probes* **5**, 229–240.

Sehgal S (1997): In Dodet B, Meslin FX (Eds): *Rabies Control in Asia*. Meslin Elsevier, Paris, pp. 140–145.

Smith JS, Orciari LA, Yager PA, Seidel HD, Warner CK (1992): Epidemiological and historical relationships among 87 rabies virus isolates as determined by limited sequence analysis. *J. Infect. Dis.* **166**, 296–307.

Smith JS, Yager PA, Orciari LA (1993): Rabies in wild and domestic carnivores of Africa: Epidemiological and historical associations determined by limited sequence analysis. *Onderstepoort J. Vet. Res*. **60**, 307–314.

Tordo N, Kouknetzoff A (1993): The rabies virus genome: an overview. *Onderstepoort J. Vet. Res*. **60,** 263–269.

Tordo N, Badrane H, Bourhy H, Sacramento D (1993): Molecular epidemiology of lyssaviruses: focus on the glycoprotein and pseudogenes. *Onderstepoort J. Vet. Res*. **60**, 315–323.

Lafon M, Wiktor TJ (1985): Antigenic sites on the ERA rabies virus nucleoprotein and non-structural protein. *J. Gen. Virol.* **66,** 2125–2133.