

Molecular characterization of the full-length genome of a rabies virus isolate from India

Tirumuru Nagaraja · Shampur Madhusudana ·
Anita Desai

Received: 31 August 2007 / Accepted: 11 March 2008 / Published online: 25 March 2008
© Springer Science+Business Media, LLC 2008

Abstract Rabies is an important public health problem in South East Asia, with cases in this part of the world contributing to about 70% of the global burden. A large number of rabies cases occur in India, however, there is no organized system of surveillance and hence there is a lack of reliable data. Moreover, comprehensive molecular epidemiological studies have not been performed on Indian virus isolates. In this study, we determined the complete nucleotide and deduced amino acid sequence of a primary isolate of rabies virus obtained from the brain of an infected patient. Comparison of the genomic sequence with those of the ten fully sequenced rabies strains available in GenBank showed nucleotide homology ranging from 97% with AY956319 to 81% with AY705373. Amino acid homology of nucleoprotein ranged from 99.7% with AY352493 to 92% with DQ875051. In case of the glycoprotein gene, the homology ranged from 98.8% with AY956319 to 87.2 % with AY705373. An extensive nucleoprotein, glycoprotein, and full-length genome-based phylogenetic analysis was performed along with sequences available from the GenBank. Phylogenetic analysis of the complete genome sequence indicated that this isolate exhibited close homology with the ex Indian strain AY956319.

Keywords Rabies virus · Full-length genome · Phylogenetic analysis

Primer sequences and the scheme of amplification can be availed from the authors on request.

T. Nagaraja · S. Madhusudana · A. Desai (✉)
Department of Neurovirology, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore 560029, India
e-mail: anita@nimhans.kar.nic.in

Introduction

Rabies is a fatal viral encephalitis, which is caused by a highly neurotropic, single-stranded RNA virus belonging to the genus *Lyssavirus* of the family Rhabdoviridae. The genome of rabies virus comprises a single-stranded, negative-sense RNA molecule of approximately 12 kb that encodes five structural proteins in the order nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G), and large protein (L, RNA-dependent RNA polymerase). Three of the viral proteins are located in the RNP core; they are the N, the non-catalytic polymerase-associated P, and the catalytic L (RNA polymerase). The remaining two structural proteins of the rabies virion, the G and M, are associated with the lipid-bilayer envelope that surrounds the RNP core.

Nucleoprotein is involved in the encapsidation of the genome RNA and protection of the RNA from endogenous ribonucleases activity [1] as well as the modulation of viral RNA transcription and replication [2]. The P protein is capable of oligomerization or binding to the nucleoprotein-RNA template [3]. The cytoplasmic dyenin light chain (LC8) protein has been reported to interact strongly with this protein and this interaction helps in the viral nucleocapsid axoplasmic transport [4]. The P protein acts as a chaperone of soluble nascent N protein [3]. Matrix protein involves in the down regulation of viral RNA transcription, condensation of helical nucleocapsid cores into tight coils, association with membrane bilayers, and involvement in the cytopathogenesis of infected cells [5]. It confers the characteristic bullet shape to the virion. Rabies virus glycoprotein plays an important role in the attachment of virus to the cell surface, pathogenicity [6], and neurovirulence of rabies virus [7]. The L protein or the virus-encoded RNA-dependent RNA polymerase is involved in majority of enzymatic activities of

the polymerase complex in viral RNA transcription and replication [5].

According to the World Health Organization (WHO) estimates, 55,000 deaths due to rabies are reported worldwide every year; the majority of them being in the developing countries of Asia and Africa [8]. Of these 55,000 cases, India alone accounts for approximately 20,000 cases every year [9]. Rabies has been endemic in India since ancient times taking a heavy toll of human lives even in the 21st century. Lack of surveillance and a comprehensive national rabies control programme, increase of stray dog population, lack of proper medical facilities for post-exposure treatment, and public awareness are chiefly responsible for the continued prevalence of rabies in the country. Although dog is the main reservoir and transmitter of the disease to humans and domestic animals, the existence of a sylvatic cycle cannot be ruled out. Along with effective diagnostic facilities, an important component of rabies control is the application of molecular tools to identify the circulating viral variants in host reservoirs. Molecular techniques to evaluate the epidemiology of viral diseases are being increasingly employed to complement conventional methods [10–12]. These techniques can give a clearer understanding of the origin and transmission patterns of the disease. Eventually, data produced from molecular epidemiological studies would lead to a better understanding and more effective strategies to control the spread of rabies. Very few studies have undertaken the molecular characterization of the rabies strains circulating in India. More recently, phylogenetic analyses based on the partial nucleoprotein gene sequences have been reported [13, 14]. However, to date, no complete genome sequence analysis and characterization of an Indian rabies virus isolate has been undertaken. In this study, as a preliminary step to identify and elucidate the genetic diversity, we cloned and sequenced an isolate of rabies virus obtained from the brain of an infected patient and analyzed the phylogenetic relationship of this isolate with other available full-length genomes of rabies virus across the world from the data base.

Materials and methods

Rabies virus isolate

A 50-year-old male patient was admitted to the neurology ward of National Institute of Mental Health and Neurosciences (NIMHANS) with symptoms of progressive flaccid paralysis, weakness in upper and lower limb, behavioral disturbances, and retention of urine. He succumbed to illness after 5 days. He had a history of dog bite five months ago and had taken two doses of antirabies tissue culture vaccine. He had failed to take the complete course of the vaccine. These

two doses of vaccine might not have induced protective levels of antibodies well in time to prevent rabies. Moreover, this patient did not receive immunoglobulins which is an essential part of rabies prophylaxis. Such situations are not infrequent in India and have been reported earlier [9]. At autopsy, the cerebellum tissue of the brain was collected and sent for laboratory confirmation of rabies to the diagnostic unit in the Department of Neurovirology. This department over the years has attempted rabies virus isolation from clinical specimens submitted for confirmation of rabies by fluorescent antibody test (FAT). The brain tissue was positive for rabies antigen by FAT. This was done essentially using the method described by Dean et al. [15]. A small amount of the tissue was homogenized using Dulbecco's minimal essential medium and a 10% homogenate was prepared. For isolation of the virus, 50 μ l of the homogenate was inoculated into 3-week-old Swiss albino mice intracerebrally. After the development of the symptoms in the mice, the brain was harvested, the presence of rabies viral antigen was confirmed by FAT and the tissue was stored in the -70°C for further characterization; the virus isolate was named as NNV-RAB-H.

Primers

The primers were designed based on the full-length genomes of PV strain (GenBank accession number NC_001542) and a strain described from Germany strain (ex-Indian strain, GenBank Ac. No. AY956319) by using freely available primer designing software ("GeneFisher" Interactive PCR primer design software).

RNA extraction

Total RNA from the 10% homogenate of mouse brain-passaged human isolate was extracted using Trizol LS reagent (Invitrogen, Life technologies Pvt Ltd, USA). Trizol reagent (750 μ l) was added to the sample (250 μ l) and vortexed completely until mixed. Chloroform (200 μ l) and glycogen (10 μ g, Sigma-Aldrich, USA), as a RNA carrier, was added, mixed gently, and centrifuged at 13,000 rpm for 10 min. Top aqueous layer was collected and RNA was precipitated by using 500 μ l of isopropyl alcohol. RNA was washed with 1 ml of 70% alcohol, centrifuged, and the pellet was air-dried. The pelleted RNA was dissolved in 50 μ l of diethyl pyrocarbonate (DEPC)-treated water.

Synthesis of cDNA

The extracted RNA was denatured at 65°C for 5 min along with primers designed across the genome followed by 37°C for 15 min. The cDNA was prepared by using Murine Moloney leukemia virus-Reverse transcriptase (200 units)

and dNTPs (20 mM) at 37°C for 2 h in a thermal cycler (Thermal Hybaid, USA) followed by inactivation of the enzyme at 95°C for 5 min.

Polymerase chain reaction (PCR) and cloning

PCR amplification of the defined stretches of the genome (Fig. 1) was carried out by using different sets of primer pairs. The initial denaturation was carried out at 95°C for 2 min followed by 35 cycles using initial denaturation at 95°C for 1 min, annealing at 45°C for 60 s, and extension at 72°C for 60–120 s. The final extension step was done at 72°C for 5 min. The 3' and 5' end of the genome was amplified by rapid amplification of cDNA ends (RACE) method. To achieve the high fidelity, Accu Taq LA DNA Polymerase (Sigma-Aldrich, India) was used in the PCR reactions. The PCR products were detected by electrophoresing in 0.8% agarose gel. The gel containing the PCR products with expected size were excised and the DNA was eluted using gel extraction columns (Auprep, Life Technologies India Pvt Ltd). The DNA was then ligated directly into pGEM-T Easy cloning vector system (Promega, Madison, USA) at 16–19°C, over night. The ligated product was transformed into competent *E. coli* DH5 α cells by heat shock method [16]. The transformed colonies were screened by both ampicillin as resistant marker and blue-white color selection using X-gal, IPTG containing medium. After extracting the plasmid from the transformed bacteria, the insert gene and orientation of the gene was verified by PCR using gene-specific primer and vector-specific primer (T7 or SP6). The plasmid DNA was

purified by using commercially available columns (Auprep, Life Technologies India Pvt Ltd).

Sequencing and phylogenetic analysis

Both strands of the plasmid DNA ranging from 521 to 2,100 bp were sequenced commercially (MWG Biotech Pvt Ltd, Bangalore) using vector- or gene-specific primers. The sequence data was analyzed using Chromas and compiled by using DNASTAR software. The full-length consensus genome was checked for stop codons and submitted to the NCBI GenBank (Ac. No. EF437215). Phylogenetic analysis of the complete genome was carried out using 10 full-length genome sequences of rabies virus strains across the world available in the GenBank. The nucleoprotein gene sequences of 37 rabies virus isolates and glycoprotein sequences of 10 isolates were used for the multiple alignments and to understand the phylogenetic relationships among isolates using the two genes. Similarity calculations between aligned nucleotide and aminoacid sequences were performed using DNASTAR computer software program (DNASTAR Inc., Madison, WI, USA). Phylogenetic analysis was done using aligned sequences by neighbor joining method using Phylip software.

Results

Sequencing and characterization of rabies virus isolate NNV-RAB-H

The complete sequence of NNV-RAB-H was obtained from 13 individual clones (Fig. 1). Sequences were compiled by using DNASTAR SeqMan program and a consensus sequence was obtained was deposited in GenBank (Ac. No. EF437215). The genome length of NNV-RAB-H was 11,928 nucleotides. The ends (3' and 5') of the genome were amplified by rapid amplification of cDNA ends (RACE). These were aligned with full-length genome to ensure that the ends were rabies virus specific. The 3' and 5' end nucleotide sequences matched with the other full-length rabies strains from the GenBank (Fig. 2).

The mRNA and coding sequences of each gene and intergenic regions were characterized. The genome consisted of a 58-nucleotide 3' leader region followed by 1424-nucleotide nucleoprotein mRNA, 990-nucleotide phosphoprotein, 806-nucleotide matrix protein, 2166-nucleotide glycoprotein, 6474-nucleotide large (L) RNA-dependent RNA polymerase followed by 70-nucleotide trailer region. The coding sequences (CDS) for the five mRNAs are as follows: 1353-nucleotide nucleoprotein, 894-nucleotide phosphoprotein, 609-nucleotide matrix

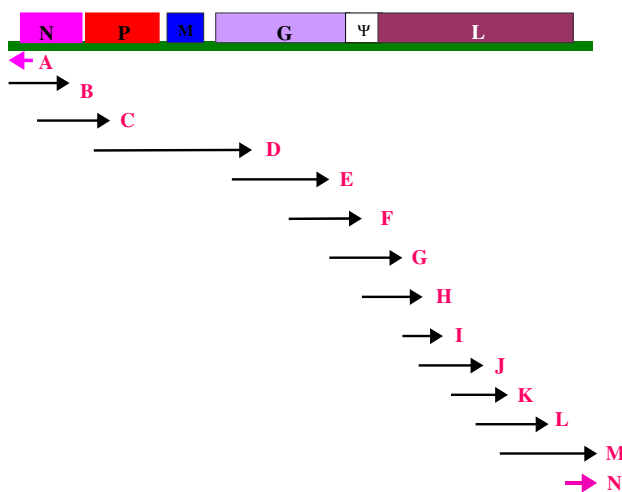
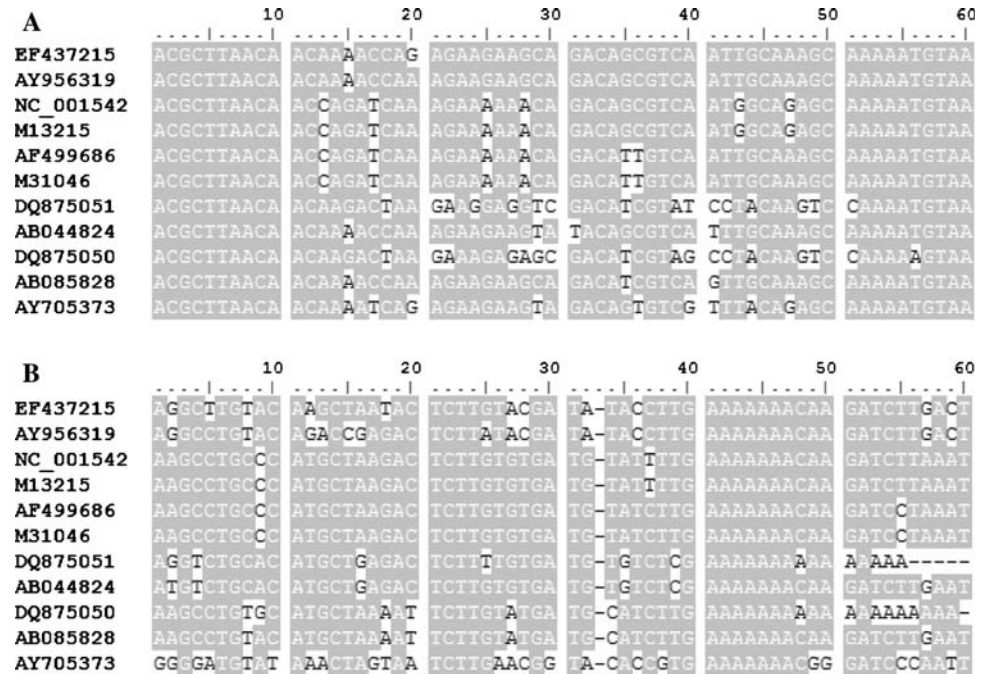


Fig. 1 Order of clones constructed and sequenced. Varied length of the genome (from 0.5–2 kb, B–M) was cloned as seen in the figure. A, N are 3' and 5' Rapid Amplification of cDNA clones (RACE). The viral DNA stretches were sequenced to obtain the consensus sequence of NNV-RAB-H

Fig. 2 Untranslated terminal regions 3' (a) and 5' (b) of the isolate NNV-RAB-H (EF437215) in the multiple sequence alignment with 10 isolates from the GenBank indicated that the sequences of the terminal regions were specific to the rabies virus genome



protein, 1575-nucleotide glycoprotein, and 6384-nucleotide large protein (RNA-dependent RNA polymerase, Table 1). The intergenic nucleotides between mRNAs are as follows: two nucleotides between nucleoprotein and phosphoprotein gene, five nucleotides between phosphoprotein and matrix protein; as well as matrix protein and glycoprotein genes and 423-nucleotide Ψ gene between glycoprotein and RNA polymerase gene (Table 1).

The comparison of complete NNV-RAB-H genomic sequence with complete genome sequences available in the GenBank (10 sequences) were used for molecular

characterization and to understand the relative similarity of NNV-RAB-H to other rabies strains. Similarly, the nucleoprotein and glycoprotein gene sequences of NNV-RAB-H were compared with 37 nucleoprotein gene sequences and 10 glycoprotein sequences available in the GenBank. The percentage homology of NNV-RAB-H at nucleotide level with other full-length genomes ranged from 97% with AY956319 to 81% with AY705373. The percentage homology of NNV-RAB-H nucleoprotein gene with other isolates ranged from 97% with AY956319 to 84.1% with AY705373. In the case of phosphoprotein gene with other

Table 1 Transcriptional start and stop signals of Indian isolate NNV-RAB-H (EF437215)

Gene	mRNA		Coding sequence (CDS)					
	Start position	Stop position	Start position		Stop position			
N	AACACCCCT	59–67	ATG(A) ₇	1473–1482	ATG	71	TAA	1423
P	AACACCCCT	1485–1493	GTG (A) ₇	2467–2475	ATG	1485	TAA	2407
M	AACACCACT	2481–2489	GTG(A) ₇	3278–3287	ATG	2496	TAA	3104
G	AACATCCCT	3293–3301	AAC(A) ₇	3190–3199	ATG	3320	TGA	4894
			A(G) ₅ (A) ₃ TC	4940–4950				
L	AACACTTCT	5384–5392	GAG(A) ₇	5350–5359	ATG	5414	TGA	11797
			TTG(A) ₇	11849–11858				
Intergenic signals			Length			Coordination		
Leader RNA			58 ntds			1–58		
N–P			2 ntds TC			1483–1484		
P–M			5 ntds CAGGC			2476–2480		
M–G			5 ntds CTATT			3288–3292		
G–L (ψ)			423 ntds			4961–5383		

isolates ranged from 97.5% with AY956319 to 80.1% with AY705373. The percentage homology of NNV-RAB-H matrix protein gene with other isolates ranged from 98.5% with AY956319 to 81.6% with AF360857. The percentage homology of NNV-RAB-H of glycoprotein gene ranged from 97% with AY956319 to 79.4% with AY705373. In the Large RNA-dependent RNA polymerase the percentage homology was 98% with AY956319 to 82% with AY705373. At amino acid level (Table 2), the percentage homology of different genes of NNV-RAB-H were as follows: nucleoprotein gene ranged from 99.7% with AY352493 to 92% DQ875051, phosphoprotein gene ranged from 98.3% with AY956319 to 86.1% AY854611, matrix protein ranged from 98.5% with AY956319 to 91.6% with AF499686, while the percentage homology of glycoprotein gene ranged from 98.8% with AY956319 to 87.2% with AY705373 & DQ875051 and the Large RNA-dependent RNA polymerase ranged from 99.2% with AY956319 to 95.1% with AY705373.

The mRNA transcription start site of five genes of NNV-RAB-H with different rabies virus isolates available in GenBank (10 isolates) was also analyzed (Fig. 3). In nucleoprotein transcription start site (59–67) two changes were observed. At nucleotide position 65 all isolates including NNV-RAB-H have cytosine, whereas 3 isolates such as PV, RAVMMGN (M13215), and Nishigahara strain (AB044824) have thymine (C→T). At position 66, strains described from China (DQ875050 and DQ875051) have guanine, whereas the remaining nine including NNV-RAB-H have cytosine. Two changes were also observed in the phosphoprotein transcription start site (1485–1493) of different isolates. At position 1490, all the isolates analyzed including NNV-RAB-H have cytosine except isolates DQ875050 and AB085828 (C→T). At position 1491, Nishigahara strain (AB044824) has thymine, whereas all the remaining has cytosine. In the matrix and glycoprotein gene, transcription start sites were conserved in all sequences. Sequences of the transcription start sites of large protein in all isolates were conserved except in isolates DQ875050 and AB044824 at position 5389 (T→C).

The amino acid differences in the five rabies proteins were analyzed. In the glycoprotein gene, eight antigenic sites (I, II, III, IV, V, VI, a, G) were recognized, out of eight, five sites were mapped; these are antigenic site I (231), II (34–42 and 198–200), III (330–338), VI (264), and 'a' (342–343). Amino acids in antigenic site I (amino acid E), antigenic site 'a' (amino acids M and E), at antigenic site III (F, G, K, A, Y, T, I, F, and N), and antigenic site VI (amino acid T) in all the isolates including NNV-RAB-H were conserved. Antigenic site II was discontinuous and conserved in all isolates including NNV-RAB-H except AB085828 (I→L at amino acid position 38) and AY705373 (E→V at amino acid position 200).

Table 2 Homology comparison of nucleotide and amino acid sequences of N, G and complete genome of NNV-RAB-H (EF437215) with reference strains

Strain	Nucleotide			Amino acid	
	N (%)	G (%)	Full length (%)	N (%)	G (%)
NC_001542*	89.4	86.9	87	97.3	92.3
AY956319*	97.3	97	97	99.7	98.8
AF374721	86.1	–	–	96.6	–
AY352493	97	–	–	99.5	–
DQ875051*	87.2	84.7	86	92.4	87.2
DQ875050*	88.9	86.9	87	97.3	90.6
AB044824*	87.8	86.3	86	95.5	91.2
AY705373*	84.1	79.4	81	94.6	87.2
AF499686*	89.4	86.3	87	97.7	91.9
AB085828*	89.2	87.3	87	98.2	93.3
M13215*	89.4	86.9	87	97.3	92.3
M31046*	89.2	86.8	87	97.7	92.3
AY352495	96.6	–	–	99.5	–
AY352494	92.7	–	–	99.3	–
DQ076131	92.6	–	–	99.1	–
RAVN783	92.9	–	–	98.8	–
DQ521212	94.5	–	–	98.4	–
DQ099525	89.0	87	–	97.7	91.2
U42607	89.3	–	–	97.3	–
U22475	89.3	–	–	96.6	–
U22481	88.9	–	–	97.7	–
U42704	89.8	–	–	98.0	–
AY352491	89.2	–	–	97.3	–
AY352475	88.6	–	–	97.7	–
U43025	89.3	–	–	97.3	–
AF406696	89.5	–	–	97.7	–
AF406695	89.0	–	–	96.4	–
U22633	89.0	–	–	98.2	–
AY500827	89.4	–	–	98.6	–
DQ515993	87.1	–	–	98.2	–
AB041964	86.6	–	–	97.7	–
AY705373	84.1	–	–	94.6	–
AF070450	84.8	–	–	96.2	–
AF394888	85.9	–	–	96.2	–
AF045166	84.7	–	–	96.4	–
U22478	84.4	–	–	96.4	–
AY854595	84.3	–	–	95.7	–

Full genome sequence of these isolates marked with * were available in the GenBank

In the nucleoprotein gene, three antigenic sites (I, II, III) have been recognized. Out of the three, two (I and III) are characterized. Antigenic site I and III in all the isolates analyzed including NNV-RAB-H are conserved. At position 332, four isolates (EF437215, AY956319, AY352493, and AY352495) have G while other isolates have A or T.

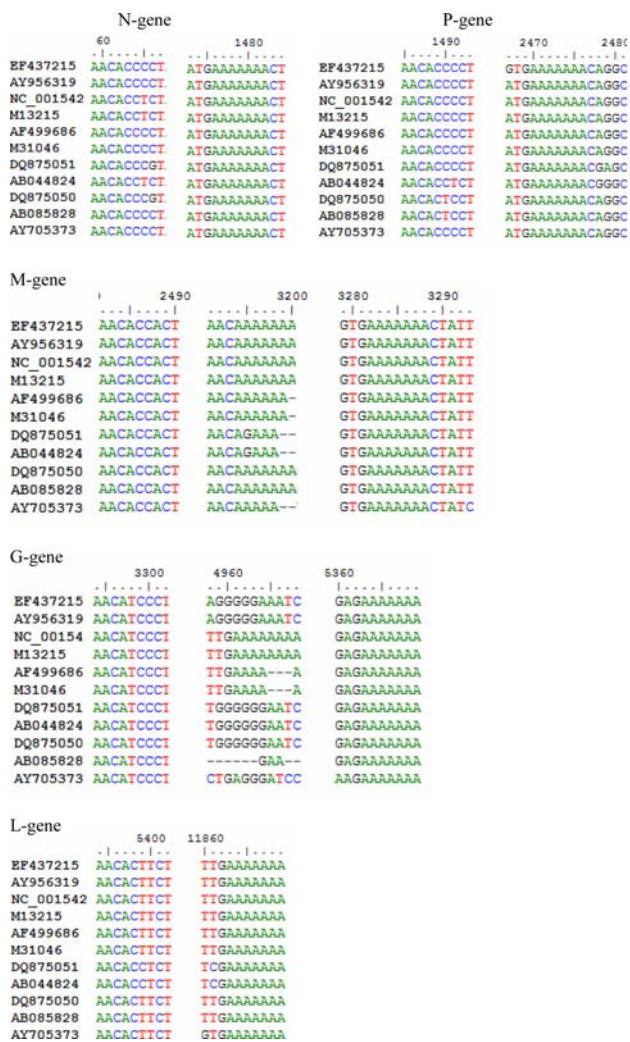


Fig. 3 Comparison of transcription start sites for five genes in different isolates

The rabies virus P protein has been shown to interact with LC8 (cytoplasmic dynein light chain) at residues aa 138–172 [17]. This region is highly hydrophobic and contains many phosphorylation sites [18]. The motif (K/R) STQT—residues 145–149—is important to interact with LC8, a protein that contributes to the axonal transport of rabies virus within neurons [4]. All the isolates analyzed in this study have (K/R) STQT motif except from five sequences such as AY705373 (S→A at aa 146); AY854611 (S→A at aa 146); AY049119 (S→Q at aa 146); EF157976 (Q→R at aa 148), and NC_006429 (K/R→I at aa 145; S→Q at aa 146, and T→I at aa 147). The short lysine-rich motif (FSKKYKF) corresponding to 214–220 amino acids was critical for the binding to the RNP [19]. All the isolates analyzed in this study were conserved for this motif, with the exception of NNV-RAB-H (Y→H at aa 218) and AF049119 (K→R at aa 217).

Two late domain motifs involved in the budding of the virus were identified in rabies virus matrix protein; these

are PPXY (aa 35–38) and PX (T/S) AP motifs [20]. The PPXY motif is conserved in all isolates including NNV-RAB-H and is reported to bind class I WW-domain-containing E3-ubiquitin ligases [21]. The PX (T/S) AP motif is conserved in all isolates including NNV-RAB-H and apparently not required for rabies virus budding [22].

The large protein is the most hydrophobic of the five lyssavirus proteins. Comparison of large proteins from members of the *Mononegavirales*, the conserved residues is not distributed randomly, but cluster into six conserved domains with invariant motifs [23]. Domain one (aa 233–424) contains an invariant GHP motif (aa 373–375) and is present in all isolates analyzed and probably plays an important functional role by part of turn structure with exposed H residue [23]. Domain two (aa 505–608) contains charged residues with a central conserved stretch containing a KEKE motif. This motif has been shown to be involved in the positioning and binding of the RNA template [24]. In this study, it was noticed that all the isolates have serine at position 526 while NNV-RAB-H has Tyrosine (S→T at aa 526).

Domain three (aa 609–832) has four long conserved regions. In this study it was noticed that region A (615–632) was conserved in all isolates including NNV-RAB-H analyzed except in two isolates such as DQ875051 (631 T→P) and AY735373 (616 H→R). Conserved region B (687–714), C (726–735), and D (791–811) were conserved in all isolates including NNV-RAB-H. The pentapeptide (QGDNQ; aa 728–732) in region C is proposed to be the active site for template recognition and/or phosphodiester bond formation [25]. In this study also it was noticed that this pentapeptide was conserved in all isolates analyzed.

Domain four (890–1061) is rich in proline residues and is involved in nucleotide binding [23]. Domain five (1091–1327) contain numerous cysteine and histidine invariant residues, while domain six (1674–1749) is less conserved with a GXGXX motif (1705–1710) preceded by a lysine 19–22 residues upstream, which could play the role of polyadenylation or protein kinase activity [23]. In this study also it was noticed that GXGXX motif was conserved in all isolates analyzed.

Phylogenetic analysis

To understand the genetic relationships and evolution of rabies virus strains, phylogenetic analysis of the 11 complete genome sequences available in the GenBank (Fig. 4), 11 complete glycoprotein sequences (Fig. 5), and 38 nucleoprotein sequences (Fig. 6) including those of NNV-RAB-H strain were analyzed. With respect to the complete genome, glycoprotein gene and the nucleoprotein gene, NNV-RAB-H strain (EF437215) grouped with AY956319 with high bootstrap value of 100, 100, and 72, respectively.

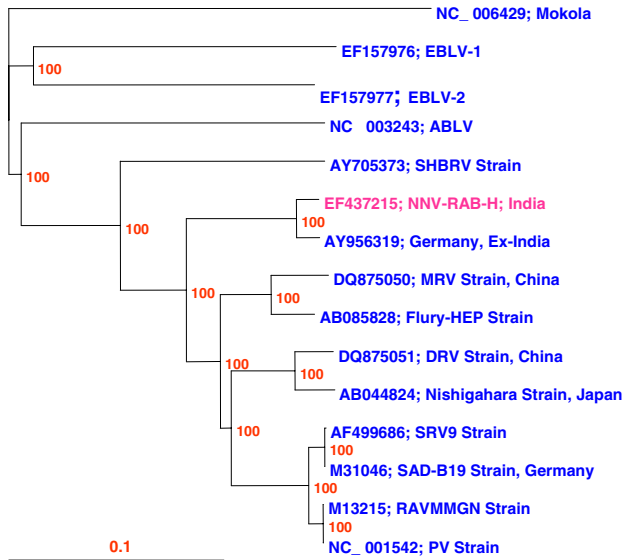


Fig. 4 Phylogenetic analysis of rabies virus isolate (NNV-RAB-H) from India. The phylogenetic tree was constructed with full-length genome sequence of NNV-RAB-H (pink in color) along with full-length rabies virus isolates and other genotype sequences available in GenBank by using Neighbor-Joining program in PHYLIP software. The numbers below the branches are bootstrap values for 100 replicates. Note that Indian isolate (NNV-RAB-H) is clustered with Ex-Indian isolate (AY956319). Mokola virus was taken as outgroup

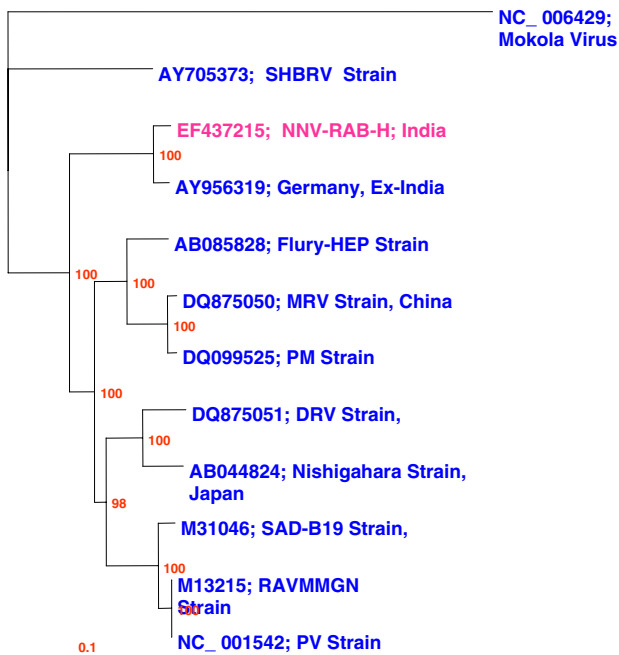


Fig. 5 Phylogenetic analysis of glycoprotein sequence of NNV-RAB-H (pink in color) along with glycoprotein sequences available in GenBank by using Neighbor-Joining program in PHYLIP software, from India. The numbers below the branches are bootstrap values for 100 replicates. Note that Indian isolate (NNV-RAB-H) is clustered with Ex-Indian isolate (AY956319). Mokola virus was taken as outgroup

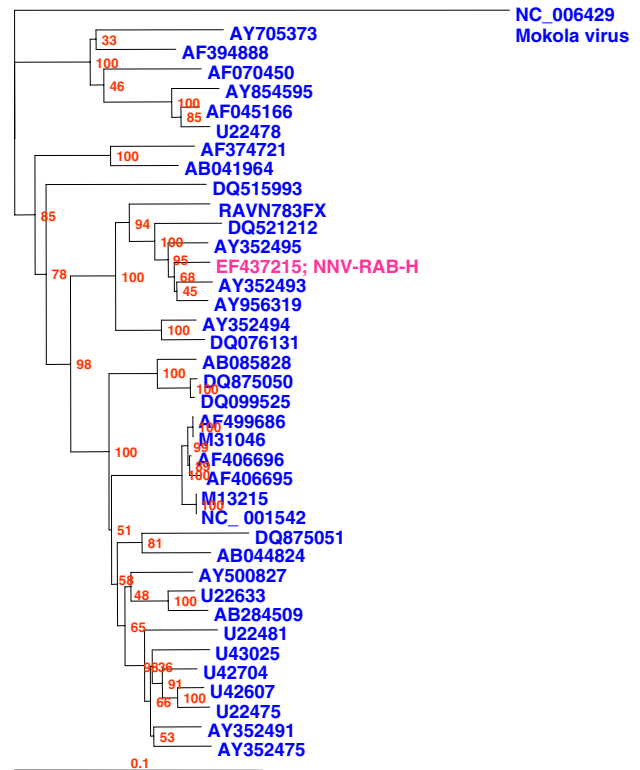


Fig. 6 Phylogenetic analysis of nucleoprotein sequence of NNV-RAB-H (pink in color) along with nucleoprotein sequences available in GenBank by using Neighbor-Joining program in PHYLIP software, from India. The numbers below the branches are bootstrap values for 100 replicates. Note that Indian isolate (NNV-RAB-H) is clustered with both, Ex-Indian isolates (AY352495 and AY956319), whereas the Indian isolate AF374721 grouped more closely with Sri Lankan isolate AB041964. Mokola virus was taken as outgroup

In the phylogenetic analysis of the nucleoprotein gene (Fig. 6), one of south Indian isolate AF374721 described earlier [26] was grouped with the Sri Lankan isolate AB041964 [14]. On the other hand, NNV-RAB-H (EF437215) clustered with all ex-Indian strains AY956319, AY352493, and AY352495.

Discussion

As a preliminary step to understand the molecular characteristics of circulating rabies virus strains in India, we undertook the complete genome sequence analysis of a rabies virus isolate obtained from an infected human brain. The genome was cloned to define the viral genes and to characterize the transcriptional and replication regulatory sequences. The strategy employed was to obtain the clones by walking along the genome using different sets of primers. The consensus sequence obtained was characterized and submitted to the GenBank (Ac. No. EF437215). It is interesting to note that the lyssavirus genomes that have been

sequenced to date, all have been found to have an even number of nucleotides. This was also observed with our isolate NNV-RAB-H which has a total genome length of 11,928 nucleotides. This may be a requirement for efficient replication and is equivalent to the Paramyxovirus “rule of six” [27].

Comparison of nucleotide sequences of rabies virus isolates across the world have indicated extensive divergence except for the transcription regulatory signals and limited stretches of the nucleoprotein [28]. The presence of a remnant protein gene between the G and L cistrons in rabies virus suggests a rapid evolution of this region among Rhabdoviridae [29]. All non-segmented negative-strand RNA viruses use the same basic mechanism for the replication: the genomic RNA (–) is first transcribed into monocistronic RNAs (+, leader and mRNAs), then it is replicated into a complete antigenome (+) which will serve as template for the synthesis of novel genomic RNAs (–). Comparative studies between the rabies genomes can allow defining more precisely the sequences involved in the transcription and replication [28].

The leader RNA is a 3′-transcription product of the genomic RNA, which has been known for a long time in Vesicular Stomatitis Virus (VSV) and different CVS strain of the rabies virus [30, 31]. The 3′ location of the leader region in genome implies that essential events such as initiation of transcription, replication [32], the switching between these two functions [33, 34], and the initiation of encapsidation take place here [35]. Kurilla et al. [31] have shown some common features between VSV and rabies leader RNA, they are of similar length (about 50–58 nucleotides) with a high content of adenine residues (50%), they all initiate with the same trinucleotide (ACG) including the isolate (NNV-RAB-H) used in this study, which seems to be conserved throughout the Rhabdoviridae family and have a very conserved 3′ end [36]. Comparison of the PV and CVS strains indicates that certain sequences are less important than had previously been thought. For example, the hexanucleotide 3′₋₁₄UUUGGU₁₉-5′ (genomic sense) is present in the CVS strain as throughout the *Vesiculovirus* genus (one nucleotide upstream) where it is thought to be implicated in the RNA synthesis initiation [36]. Since this hexanucleotide is absent from the PV strain due to one U/C change at position 15 and one G/A change at position 17, it is uncertain that it plays the same role in the rabies genome. In this study, with the strain NNV-RAB-H, this hexanucleotide sequence 3′₋₁₄UUUGGU₁₉-5′ was changed like other strains (NC_001542, M13215, AF499686, M31046, DQ875051, DQ875050, and AY705373). Most of the changes were observed at position 15 (U→G), 17, and 18 (C→T). When 11 isolates including NNV-RAB-H, were analyzed for nucleotide changes within the leader region (58 nucleotides), 30 nucleotides

out of 58 nucleotides (52%) showed nucleotide differences. Consequently, the leader region appears to be an area of rapid evolution in the rabies virus genome.

To compare the nucleoprotein transcription initiation site, 10 sequences from GenBank were collected and analyzed along with NNV-RAB-H (Fig. 3). It was noticed that fewer changes were observed in transcription start site of NNV-RAB-H isolate. Rabies nucleoprotein central region corresponding to 298–352 amino acids involved in the interaction of RNA is highly conserved [28]. Analysis of this stretch of amino acids from 38 isolates including NNV-RAB-H showed that this RNA binding region was highly conserved.

The transmembrane glycoprotein (G) is responsible for the induction of virus neutralizing antibodies [37] and consists of 524 amino acids encoded between 3318 and 4890 nucleotides of PV strain. In this study, it was observed that all antigenic sites of glycoprotein of NNV-RAB-H were highly conserved like the rabies virus PV strain. Analysis of the nucleotide sequence of NNV-RAB-H helped defining sequences involved in the transcription of the structural genes and to outline fundamental aspects of the evolutionary relationship with other rabies virus isolates. Two consensus sequences located at the boundary of each structural gene were 5′ start and the 3′ end of the mRNA sequence. These sequences most probably represent transcriptional initiation and termination signals. Tordo et al. [28] showed that these sequences were similar to those of another member of the family Rhabdoviridae, viz., the VSV and appear to retain a fair degree of homology with Sendai virus, a Paramyxoviridae. This homology between transcription signals strongly contrasts with the divergence found in other parts of the genome, in particular most of the structural genes, with the exception of the nucleoprotein gene. This shows that among unsegmented negative-strand RNA viruses, in spite of their high rate of evolution, many viruses conserve the same basal genomic organization and the signal sequences essential for transcription [28].

Tordo et al. [28] characterized the transcription initiation, termination, intergenic, and coding sequences for PV strain. NNV-RAB-H nucleoprotein mRNA transcription initiates around position 59 (5′-AACACCCCT-3′) which was similar to PV strain but it differs slightly from the (5′-AACAG-3′) VSV 5′ mRNA initiation site. Downstream from the nucleoprotein coding sequence, this nine nucleotide sequence also appears in position 1482 and it was assumed that it represents the transcription initiation site of the phosphoprotein mRNA. This assumption is strengthened by the finding of the sequence 5′-ATG(A)₇-3′ three nucleotides upstream (position 1473–1482) of the phosphoprotein transcription gene, closely similar to the consensus 3′ end of VSV mRNAs. It was also observed that a consensus mRNA start sequence (nine nucleotides) consists of 12–30 residues upstream of the translation initiation codon.

In NNV-RAB-H, nucleoprotein–phosphoprotein and phosphoprotein–matrix protein intergenic regions contain a single polyadenylation signal and the matrix protein–glycoprotein and glycoprotein–RNA-dependent RNA polymerase intergenic regions have two putative polyadenylation signals at positions 3190/3278 and 4940/5350, respectively like the PV strain. Interestingly, based on the size of the mRNA for the matrix protein, the second polyadenylation signal (position 3278) is more likely to be the polyadenylation site and the second poly (A) sequence (position 5350) is used as polyadenylation signal for glycoprotein. Furthermore, as a result of the identical multiplication mechanism of these non-segmented negative-strand RNA viruses, a high conservation of transcriptional start and stop signals is observed [29]. Interestingly it was observed that the matrix and glycoprotein transcription start sites were more conserved than nucleoprotein, phosphoprotein, and RNA-dependent RNA polymerase gene mRNA start sites (Table 1).

Tordo et al. [28] suggested that the existence of mRNA start and stop consensus sequences near its extremities of Ψ gene indicates that this gene perhaps was also transcribed. In NNV-RAB-H, 423-nucleotide Ψ gene is presented in the intergenic region located between the G and L cistrons. Assuming that this region is transcribed, the sizes of the NNV-RAB-H genome intergenic regions become 2, 5, 5, 9, and 24 residues, respectively. Considering that the flanking sequences probably result from the degeneration of consensus transcription signals and this G–L intergenic region is the sole large genomic region extensively blocked in all reading frames, it represents a remnant protein gene [28]. This was supported by the identification of a sixth gene, similar in length to the rabies G–L intergenic region, between the G and L cistrons of a fish Rhabdoviridae, infectious hematopoietic necrosis virus [28]. This gene encodes the M_r 12 kDa nonvirion structural protein (NV), whose role remains unknown. The rabies G–L intergenic region, similar in length to those of infectious hematopoietic necrosis virus but with multiple stop codons, probably represents an intermediate stage of rhabdoviridae evolution. In this context, it is interesting to note that, unlike Rhabdoviridae, most Paramyxoviridae have two glycoprotein genes. The additional one is located in a region corresponding to the rabies G–L intergenic region. This emphasizes the plasticity of this region in unsegmented negative-strand RNA viruses [28]. The rabies G–L intergenic region is a pseudogene in this group of viruses.

Rabies virus contain a long (approximately 425 nucleotides) region of non-coding sequence between the G and L genes that was suggested to represent a remnant or pseudogene (Ψ) which was once functional, but had become vestigial. Two transcriptional termination and

polyadenylation (TTP) motifs were found in this region in the Pasteur strain (PV) of rabies virus isolate [28]. The first was 70 nucleotides downstream from the translation stop codon of the G gene and the other was located at the end of the proposed Ψ region, 24 nucleotides upstream of the L gene start. This arrangement suggested that the upstream TTP motif was the TTP signal for G mRNA transcription and downstream motif was presumably related to the previous function of the pseudogene. The presence of both upstream and downstream TTP sites in a second rabies strain (SAD-B19) suggested that this may be a common gene arrangement in rabies viruses [38]. Morimoto et al. [39] noticed two species of G mRNA (1.9 and 2.3 kb in length) in cells infected with the ERA rabies strain. Sacramento et al. [40] and Sakamoto et al. [41] noticed the lack of upstream TTP signal and synthesis of 2.3 kb mRNA by using downstream TTP in Flury-HEP, CVS, and Nishigahara strains.

To elucidate which of the TTP signals was ubiquitous in the synthesis of the G mRNA in wild-type rabies virus isolates, Ravkov et al. [42] sequenced the G–L intergenic region (from 4665 to 5543) in laboratory rabies virus strains and a large number of diverse wild rabies viruses. Only one distinct lineage of the laboratory strains and none of the wild-type rabies viruses contained the upstream TTP-like signal, indicating that only the downstream TTP motif is the authentic G mRNA transcription termination and polyadenylation signal. These data indicated that this region of the rabies virus genome encodes a G mRNA with long 3' non-coding region with no evidence of a pseudogene. In this study also it was noticed that PV (NC_001542) and M13215 isolate have the first downstream TTP signal compared to the all isolates analyzed (15 isolates excluding PV and M13215: EF437215, AY956319, AF499686, M31046, DQ875051, AB044824, DQ875050, AB085828, AY705373, NC_003243, EF157976, EF157977, and NC_006429). In case of upstream TTP signal all isolates (15 isolates) analyzed have the conserved TTP signal.

In the phylogenetic analysis of the complete genome sequence of 11 rabies virus isolates, NNV-RAB-H clustered with ex-Indian strain AY956319 and this was supported by a high bootstrap value (Fig. 4). Phylogenetic analysis based on glycoprotein gene sequence also revealed that NNV-RAB-H was closely related to the isolate AY956319 (Fig. 5). The isolate AY956319 described from Germany, was obtained from the saliva of a symptomatic organ recipient and the transmission was attributed to solid organ transplantation imported from India. Based on the nucleoprotein gene sequence analysis alone, 97.3% homology was noticed between EF437215 (NNV-RAB-H) and AY956319 (Germany). Phylogenetic analysis of nucleoprotein sequences revealed that NNV-RAB-H grouped with isolates from Asian countries such as Pakistan (AY352494), ex-Indian strain (AY352493) and the more recent Indian strain

(AY956319; Fig. 6). On the other hand, the isolate NNV-RAB-H exhibited low homology (86%) at nucleotide level with a full-length nucleoprotein gene sequence of AF374721 described earlier from Chennai, India [10]. This was also reflected in the phylogenetic placement in Fig. 6. Furthermore, it was also noted that AF374721 clustered with the Sri Lankan isolate AB041964 with a bootstrap value of 100 (Fig. 6). This indicates the existence of genetic diversity among rabies virus isolates circulating in the country. This has been explained by Denduangboripant et al. [43] who suggested that transmission of rabies virus may be related to availability of transportation routes, human activity particularly human migration as well as human and animal population density. Genomic heterogeneity of nucleoprotein and glycoprotein among various rabies viruses isolated from different geographical regions and several sources including human, dog, cat, cattle, and sheep has been reported previously [13, 14, 26].

Conclusion

For the first time in the country, a complete genome sequence analysis of a rabies virus isolate NNV-RAB-H from an infected human brain was achieved. The characterization and phylogenetic analysis of the complete genome emphasizes the underlying genetic diversity amongst the circulating strains in the country. A more detailed molecular analysis of a large number of rabies virus strains obtained from different parts of the country would provide a better understanding of the complexity of the strains circulating in India.

Acknowledgements T. Nagaraja is a senior Research fellow and is supported by the Council of Scientific and Industrial Research, Government of India.

References

1. W.H. Wunner, in *The Natural History of Rabies*, ed. by G.M. Baer (CRC Press, Boca Raton, FL, 1991), pp. 66–67
2. J. Yang, H. Koprowski, B. Dietzschold, Z.F. Fu, *J. Virol.* **73**(2), 1661–1664 (1999)
3. B. Gigant, F. Iseni, Y. Gaudin, M. Knossow, D. Blondel, *J. Gen. Virol.* **81**(Pt 7), 1757–1761 (2000)
4. H. Raux, A. Flamand, D. Blondel, *J. Virol.* **74**(21), 10212–10216 (2000)
5. W.H. Wunner, in *Rabies*, ed. by A.C. Jackson, W.H. Wunner (Academic Press, USA, 2002), pp. 23–61
6. K. Morimoto, D.C. Hooper, S. Spitsin, H. Koprowski, B. Dietzschold, *J. Virol.* **73**(1), 510–518 (1999)
7. C. Tuffereau, H. Leblois, J. Bénéjean, P. Coulon, F. Lafay, A. Flamand, *Virology* **172**(1), 206–212 (1989)
8. World Health Organization (WHO) expert consultation on rabies. Introduction. In: *WHO Technical Report Series 931* (World Health Organization, Geneva, 2005), p. 3
9. M.K. Sudarshan, S.N. Madhusudana, B.J. Mahendra, N.S. Rao, D.H. Ashwath Narayana, S. Abdul Rahman, F.X. Meslin, D. Lobo, K. Ravikumar, Gangaboraiah, *Int. J. Infect. Dis.* **11**, 29–35 (2007)
10. B.F.V. Von Teichman, G.R. Thomson, C.D. Meredith, L.H. Nel, *J. Gen. Virol.* **76**, 73–82 (1995)
11. L.K. Crawford-Miksza, D.A. Wadford, P.D. Schunurr, *J. Clin. Virol.* **14**, 207–219 (1999)
12. D. David, B. Jakobson, J.S. Smith, Y. Stram, *J. Clin. Microbiol.* **38**, 755–762 (2000)
13. T. Nagarajan, B. Mohanasubramanian, E.V. Seshagiri, S.B. Nagendrakumar, M.R. Saseendranath, M.L. Satyanarayana, D. Thiagarajan, P.N. Rangarajan, V.A. Srinivasan, *J. Clin. Microbiol.* **44**, 3218–3224 (2006)
14. S.A. Nadin-Davis, G. Turner, J.P. Paul, S.N. Madhusudana, A.I. Wandeler, *Emerg. Infect. Dis.* **13**, 111–116 (2007)
15. D.J. Dean, M.K. Abelseh, P. Atanasiu, in *Laboratory Techniques in Rabies*, 4th edn., ed. by F.X. Meslin, M.M. Kaplan, H. Koprowski (WHO, Geneva, 1996)
16. J. Sambrook, D.W. Russell (eds.), *Molecular Cloning*, 3rd edn., vol. 3, (Cold Spring Harbor Laboratory Press, New York, 2001)
17. T. Mebatsion, *J. Virol.* **75**(23), 11496–11502 (2001)
18. N. Tordo, O. Poch, A. Ermine, G. Keith, *Nucleic Acids Res.* **14**(6), 2671–2683 (1986)
19. Y. Jacob, E. Real, N. Tordo, *J. Virol.* **75**(20), 9613–9622 (2001)
20. H.R. Jayakar, K.G. Murti, M.A. Whitt, *J. Virol.* **74**(21), 9818–9827 (2000)
21. R.N. Harty, J. Paragas, M. Sudol, P. Palese, *J. Virol.* **73**(4), 2921–2929 (1999)
22. T. Irie, J.M. Licata, J.P. McGettigan, M.J. Schnell, R.N. Harty, *J. Virol.* **78**(6), 2657–2665 (2004)
23. O. Poch, B.M. Blumberg, L. Bougueleret, N. Tordo, *J. Gen. Virol.* **71**(Pt 5), 1153–1162 (1990)
24. R. Müller, O. Poch, M. Delarue, D.H. Bishop, M. Bouloy, *J. Gen. Virol.* **75**(Pt 6), 1345–1352 (1994)
25. O. Poch, N. Tordo, G. Keith, *Biochimie* **70**(8), 1019–1029 (1988)
26. R. Jayakumar, K.G. Tirumurugan, G. Ganga, K. Kumanan, A. Mahalinga Nainar, *Acta Virol.* **48**, 47–50 (2004)
27. D. Warrilow, I. Smith, B. Harrower, G.A. Smith, *Virology* **297**, 109–119 (2002)
28. N. Tordo, O. Poch, A. Ermine, G. Keith, F. Rougeon, *Proc. Natl. Acad. Sci.* **83**, 3914–3918 (1986)
29. N. Tordo, P. De Haan, R. Goldbach, R. Poch, *Semin. Virol.* **3**, 341–357 (1992)
30. R.J. Colonno, A.K. Banerjee, *Cell* **8**, 197–204 (1976)
31. M.G. Kurilla, C.D. Cabridilla, B.P. Holloway, J.D. Keene, *J. Virol.* **50**, 773–778 (1984)
32. S.U. Emerson, *Cell* **31**, 635–642 (1982)
33. B.M. Blumberg, M. Leppert, D. Kolakofsky, *Cell* **23**, 837–845 (1981)
34. H. Arnheiter, N.L. Davis, G. Wertz, M. Schubert, R.A. Lazzarini, *Cell* **41**, 259–267 (1985)
35. B.M. Blumberg, C. Giorgi, D. Kolakofsky, *Cell* **32**, 559–567 (1983)
36. A. Kouznetzoff, M. Buckle, N. Tordo, *J. Gen. Virol.* **79**, 1005–1013 (1998)
37. B. Dietzschold, H.H. Wang, C.E. Rupprecht, E. Celis, M. Tollis, H. Ertl, E. Heber-Katz, H. Koprowski, *Proc. Natl. Acad. Sci. USA* **84**, 9165–9169 (1987)
38. K.K. Conzelmann, J.H. Cox, L.G. Schneider, H.J. Thiel, *Virology* **175**(2), 485–499 (1990)
39. K. Morimoto, A. Ohkubo, A. Kawai, *Virology* **173**(2), 465–477 (1989)
40. D. Sacramento, H. Badrane, H. Bourhy, N. Tordo, *J. Gen. Virol.* **73**(Pt 5), 1149–1158 (1992)

41. S. Sakamoto, T. Ide, H. Nakatake, S. Tokiyoshi, M. Yamamoto, A. Kawai, J.S. Smith, *Virus Genes* **8**(1), 35–46 (1994)
42. E.V. Ravkov, J.S. Smith, S.T. Nichol, *Virology* **206**(1), 718–723 (1995)
43. J. Denduangboripant, S. Wacharapluesadee, B. Lumlertdacha, N. Ruankaew, W. Hoonsuwan, A. Puanghat, T. Hemachudha, *BMC Infect. Dis.* **29**(5), 52 (2005)