Comparison of the genome sequences and the phylogenetic analyses of the GP78 and the Vellore P20778 isolates of Japanese encephalitis virus from India

SUDHANSHU VRATI

Virology Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, JNU Complex, New Delhi 110 067, India

(Fax, 91-11-6162125; Email, vrati@nii.res.in)

The nucleotide sequence of the complete genomes of two Indian isolates of Japanese encephalitis virus were compared. One of these isolates, GP78 was obtained from northern India in 1978. The other, the Vellore P20778 isolate, was obtained from southern India in 1958. There was 4·40% nucleotide sequence divergence between the two Indian isolates that resulted in a 1·86% amino acid sequence divergence. Phylogenetic analyses showed that in evolutionary terms the north Indian GP78 isolate was close to the SA14 isolate from China whereas the south Indian Vellore P20778 isolate was close to the Beijing-1 isolate, also from China. The two Indian isolates, however, appear to have evolved independently.

1. Introduction

Japanese encephalitis (JE) is an acute viral infection of the central nervous system which is caused by a mosquito-borne flavivirus called the Japanese encephalitis virus (JEV). The virus is active over a vast geographic area that includes India, China, Japan and virtually all of south-east Asia (Umenai *et al* 1985; Ravi and Shankar 1989). Besides this, JEV has been recently isolated from previously non-affected areas of Australia (Hanna *et al* 1999). Approximately 3 billion people live in the JEV endemic area covering much of Asia. Up to 50,000 cases of JE are reported officially each year. Of these, about 10,000 cases result in fatality and a high proportion of survivors have serious neurological and psychiatric sequel.

The presence of JEV in India was first inferred in 1954 through a serological study (Smithburn *et al* 1954). However, the disease was first recognized in India in 1955 at Vellore. Until 1970s the disease was restricted to the southern states. Then in 1973, epidemics of JEV occurred in Bankura and Burdwan districts of West Bengal (Chakravarty *et al* 1975). Between 1977 and 1981 the disease spread to the newer areas that included the northern state of Uttar Pradesh. The virus has since moved to the north-western parts of the country and a number of epidemics have been reported in recent years from these areas (Sharma and Panwar 1991; Sharma *et al* 1991; Prasad *et al* 1993; Kar and Saxena 1998). The JEV is now active in almost all parts of India.

The JEV genome is a plus-sense single-stranded RNA of about 11 kb. It has a single open reading frame that codes for a 3432 amino acids long polyprotein which is subsequently cleaved into a number of structural and non-structural proteins. The 5′ one-third of the genome codes for the structural proteins, namely, capsid, membrane and the envelope while the 3′ two-third of the genome codes for non-structural proteins NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 (Chambers *et al* 1990).

Since the genome of JEV is an RNA molecule, it has high potential for evolution as RNA replication machinery lacks proof reading capability (Holland *et al* 1982). Strain variations have been shown among JEV isolates obtained from different geographical locations around the globe (Hori 1986; Kobayashi *et al* 1984, 1985; Wills *et al* 1992). Within India too, isolates obtained from different parts of the country showed marked differences in their oligonucleotide fingerprints, thus some of them being closer to the Japanese strains of the virus whereas others

Keywords. Flavivirus; RNA evolution; sequence homology

were closer to the Chinese strains (Banerjee and Ranadive 1989). We have recently established the nucleotide sequence of the complete genome of the GP78 isolate of JEV (Vrati *et al* 1999a). This virus was isolated from an epidemic of JE in 1978 in Gorakhpur which is located in the northern state of Uttar Pradesh (Mathur *et al* 1982). Comparison of the genome sequences of various isolates available then demonstrated that the GP78 strain was phylogenetically closer to the Chinese SA14 isolate (Vrati *et al* 1999a). Subsequently, complete genome sequence of another Indian strain Vellore P20778 (GenBank accession No. AF080251) has become available which was isolated from Vellore in 1958. In this paper, we have compared the genome sequences of the two Indian isolates and have carried out phylogenetic analyses to examine the relatedness of the two isolates, if any. In recent months, complete genome sequences of a number of JEV strains from Korea (strain K94P05) and Taiwan (strains HVI, TL and TC) have also become available. This paper presents the phylogenetic relationship based on the complete genome sequences of different JEV isolates obtained from various parts of the world. It is seen that while the north Indian GP78 isolate is phylogenetically closer to the Chinese SA14 strain, the south Indian Vellore P20778 isolate is closer to the Beijing-1 isolate.

2. Materials and methods

2.1 *JEV isolates*

The GP78 strain (GenBank accession No. AF075723) came from isolate 78668A which was obtained from postmortem brain of a patient from Gorakhpur (northern India) in 1978 (Mathur *et al* 1982; Vrati *et al* 1999a). The Vellore P20778 strain (GenBank accession No. AF080251) was obtained from a patient from Vellore (south India) during 1958 (Chen *et al* 1990). The Beijing-1 strain (GenBank accession No. L48961) was isolated from human brain in 1949 from Beijing, China (Hashimoto *et al* 1988). The SA14 strain (GenBank accession No. U14163) was isolated from *Culex tritaeniorhynchus* mosquito in 1954 from Xian, China (Nitayaphan *et al* 1990). RP-9 (GenBank accession No. AF014161) and RP2ms (GenBank accession No. AF014160) originated from isolate NT109 obtained from *C. tritaeniorhynchus* mosquito in 1985 from Taiwan. The JaOArS982 (GenBank accession No. M18370) strain was isolated from a pool of mosquitoes from Osaka (Japan) in 1982 (Sumiyoshi *et al* 1987). The P3 strain (GenBank accession No. U47032) was obtained in 1949 from mosquitoes in Beijing, China (Ni and Barrette 1996). JEV strains HVI (GenBank accession No. AF098735), TL (GenBank accession No. AF098737) and TC (GenBank accession No. AF098736) were isolated in Taiwan and K94P05 (GenBank accession No. AF045551) was from Korea. No other details regarding the year of isolation and the source of these viruses are known.

2.2 *Sequence comparison and phylogenetic analyses*

Alignment of nucleotide or the amino acid sequences was carried out as per the Jotun Hein method (Hein 1990) using the Lasergene software (DNASTAR, Inc., Madison, WI, USA). The PHYLIP (Felsenstein 1995) and the MEGA packages (Kumar *et al* 1993) were used for constructing the phylogenetic tree. Reliability of the tree was evaluated statistically by using 500 bootstrap replicates.

3. Results and discussion

Recently we published the complete nucleotide sequence of the GP78 isolate from Gorakhpur in Uttar Pradesh (Vrati *et al* 1999a). This constituted the first complete genome sequence of an Indian isolate of JEV. At this point of time, complete genome sequences were available only for the JaOArS982 strain from Japan (Sumiyoshi *et al* 1987), Beijing-1 (Hashimoto *et al* 1988), SA14 (Nitayphan *et al* 1990) and P3 strains from China (Ni and Barrette 1996), and the RP-9 and RP2ms strains from Taiwan (Lin *et al* 1996). Analyses based on the available complete genome sequences showed that the Indian GP78 strain was phylogenetically closer to the Chinese SA14 strain (Vrati *et al* 1999a). Subsequent to the publication of the sequence for the GP78 isolate, complete genome sequence for another isolate (Vellore P20778 isolate) from India has become available. The Vellore P20778 isolate was obtained from an epidemic of JE from Vellore in south India during 1958. In order to examine the relatedness of the two Indian isolates we have compared their nucleotide and predicted amino acid sequences (table 1). The two isolates had genome organization similar to that of the prototype JaOArS982 strain. While the GP78 genome was 10976 bases long, the Vellore P20778 isolate was 10977 bases long. Thus the Vellore P20778 isolate had an additional base compared to the prototype virus. However, since this additional base was part of the 3′-non-coding region (NCR), this did not affect the length of the viral polyprotein and hence the predicted size of all the viral proteins was similar between the GP78, Vellore P20778 and the JaOArS982 strains. When the genomes of the GP78 and the Vellore P20778 isolates were compared, 484 nucleotide substitutions were noted between the two, resulting in a rate of about 4·4% nucleotide substitution. Nucleotide substitution rate was highest for the NS3 coding sequence (5·7%) whereas it was lowest for the 5′-NCR (1%). Substitution rate for the 3′-NCR was 3·7%. The

Table 1. Comparison of the nucleotide and the amino acid sequences of the GP78 and the Vellore P20778 isolates of Japanese encephalitis virus.

Genome segment	Size		Nucleotide substitution		Amino acid substitution		NRAC* (%)
	Nucleotides	Amino acids	No. of substitutions	Per cent substitution	No· of substitutions	Per cent substitutions	
5′-NCR	95	0	1	1·05	0	0·00	0·00
Capsid	381	127	18	4·72	5	3·93	2·77
Membrane	501	167	25	4·99	4	2·39	1·60
Envelope	1494	500	66	4·41	8	1·60	12·12
NS1	1236	412	34	2·75	6	1·45	17·64
NS2A	501	167	21	4·19	2	1·19	9·52
NS2B	393	131	19	4·83	2	1·52	10·52
NS3	1857	619	106	5·70	14	2·26	13·20
NS4A	866	289	42	4·84	4	1·38	9·52
NS4B	345	115	15	4·34	4	3·47	26·66
NS5	2715	905	113	4·16	13	1·43	11·50
3′-NCR	586	0	22	3·75	0	0·00	0·00
Complete	10976	3432	484	4·40	62	1·86	12·80

*NRAC, Nucleotide substitution resulting in an amino acid change.

higher level of nucleotide sequence conservation in the 5′-NCR may be due to the functional significance of this sequence in virus replication. It may be noted that both the 3′-NCR and the 5′-NCR contain sequences that may be involved in the synthesis of viral RNA during replication (Chambers *et al* 1990; Ta and Vrati 2000). In addition, the 5′-NCR acts as a leader sequence that may be involved in control of viral protein synthesis. The 5′-NCR is also predicted to be involved in packaging the genomic RNA into the viral capsid (Zhao *et al* 1986; Khromykh *et al* 1998). These important functions of 5′-NCR would ensure a tight conservation of its sequence.

Only about 13% of all the nucleotide substitutions resulted in an amino acid change as most others were located at the third position of the codon (table 1). Thus, while the rate of the nucleotide substitution between the GP78 and the Vellore P20778 isolates was 4·40%, at the amino acid level it was only 1·86%. The amino acid substitution rate varied between different genes; highest amino acid substitution (3·93%) was recorded for the capsid protein while NS2A protein was most conserved (only 1·19% substitution). Of the three structural proteins, the envelope was most conserved. This is commensurate with the key biological functions related to the virus replication, such as virus attachment, penetration and fusion that are associated with the envelope protein (McMinn 1997; Vrati *et al* 1999b).

RNA viruses have the capacity to evolve rather fast due to the lack of the proof reading mechanism during RNA replication (Holland *et al* 1982). Methods based on oligonucleotide fingerprinting have been used to demonstrate the nucleotide sequence variation among various isolates of JEV (Banerjee and Ranadive 1989). However these methods are crude and cannot be as accurate as those using the nucleotide sequence to study the phylogenetic relatedness and evolution of different JEV strains. Using methods based on complete genome sequence, we recently demonstrated that the GP78 isolate from India was phylogenetically closer to the SA14 strain of JEV from China (Vrati *et al* 1999a). Since then, complete genome sequences of three more isolates from Taiwan, one from Korea and one from south India have become available. In order to establish the phylogenetic relationships, we have carried out pair-wise alignment of the complete genome sequences of the 12 JEV isolates for which the complete genome sequence is available. Results in figure 1 show that the HVI and RP9 strains from Taiwan were least diverged (1·2%) while the Vellore P20778 strain from south India and the K94P05 strain from Korea were most diverged (16·1%). To establish evolutionary relationship among different isolates, a phylogenetic tree was constructed based upon the complete genome sequences of the 12 JEV isolates. The dendogram (figure 2) shows that the 12 JEV isolates fall basically in four clusters. The first of these clusters contained all the 5 Taiwanese isolates of JEV, namely, TC, TL, HVI, RP2ms and the RP9 isolates. The second cluster had the two Chinese isolates, SA14 and the P3, and the Japanese JaOArS982 isolate together with the north Indian GP78 isolate. The third cluster had the Beijing-1 isolate from China and the Vellore P20778 isolate from the south India. The fourth cluster had only the Korean K94P05 isolate which appears to be highly diverged from all other isolates. An interesting finding from this analysis is that the two Indian isolates, GP78 and Vellore

Beijing-1	**1**										
GP78	**2**	4.1									
HVI	**3**	2.8	3.2								
JaOArS982	**4**	3.1	2.9	2.3							
K94P05	**5**	12.3	12.2	11.9	11.7						
P3	**6**	2.7	2.9	2.0	2.3	11.8					
RP2ms	**7**	6.3	6.7	4.6	5.8	15.3	5.6				
RP9	**8**	2.9	3.2	1.2	2.3	11.9	2.0	3.5			
SA14	**9**	2.5	2.3	1.5	1.5	11.7	1.3	4.9	1.5		
TC	**10**	6.3	6.8	4.8	6.1	15.5	5.7	8.7	5.2	5.2	
TL	**11**	3.1	3.6	1.6	2.9	12.2	2.6	5.2	1.9	2.0	4.4
Vellore P20778	**12**	7.1	8.4	7.3	7.5	16.1	7.1	10.8	7.2	6.8	11.1	7.8
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**

Figure 1. Nucleotide sequence divergence among different JEV isolates. Alignment of the complete nucleotide sequences of different JEV isolates was carried out using the Jotun Hein method (Hein 1990). Per cent nucleotide sequence divergence was calculated by comparing sequence pairs in relation to the phylogeny reconstructed by MEGALIGN using Lasergene software (DNASTAR Inc., Madison, WI).

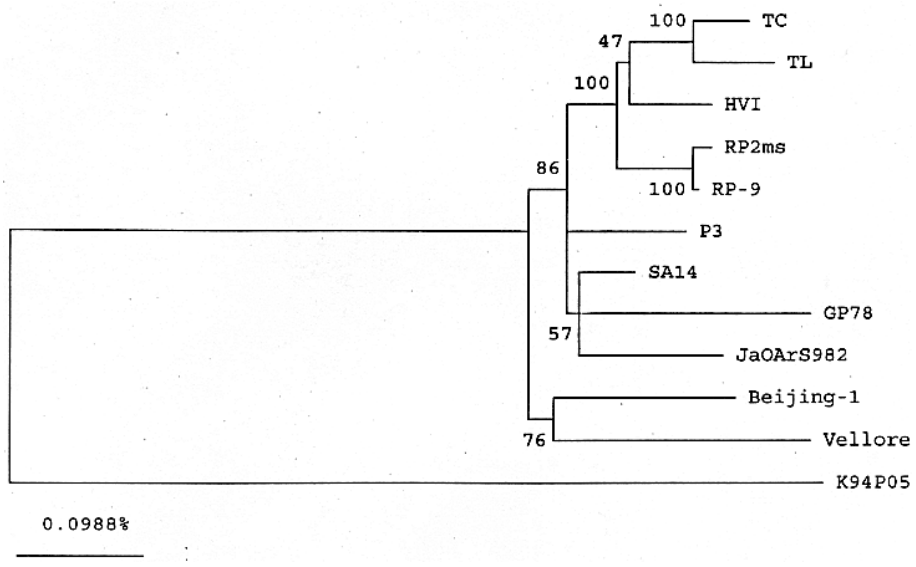

Figure 2. Phylogenetic relationship among different isolates of Japanese encephalitis virus. Complete nucleotide sequences of different JEV isolates were aligned using CLUSTAL W (Higgins *et al* 1996). The dendogram was constructed using the MEGA program (Kumar *et al* 1993) that used 100 bootstrap replicates of the sequence data. The numbers at the forks indicate the number of times the group consisting of the species to the right of the fork occurred among 100 trees. Shown is an unrooted tree.

P20778, were not closer to each other. Rather, the GP78 isolate was closer to the SA14 strain while the Vellore P20778 strain was closer to the Beijing-1 strain. Thus both the Indian strains for which complete genome sequence is available are phylogenetically close to the Chinese isolates although the GP78 and the Vellore P20778 strains appear to have evolved independently. This observation is not completely surprising since these two strains were isolated from geographically distant locations at a time gap of about 20 years. However, this reinforces the need for carrying out the molecular epidemiology of JEV in India that may have implications for JEV vaccine design and use.

Acknowledgments

Our work is supported by financial grants from the Department of Biotechnology, New Delhi. I thank Sudhir Sahdev for help with computer analyses.

References

Banerjee K and Ranadive S N 1989 Oligonucleotide fingerprint analysis of Japanese encephalitis virus strains of different geographical origin; *Indian J. Med. Res.* **89** 201–216

Chakravarty S K, Sarkar J K, Chakravarty M S, Mukherjee M K, Mukherjee K K, Das B C and Hati A K 1975 The first epidemic of Japanese encephalitis studied in India-virological studies; *Indian J. Med. Res.* **63** 77–82

Chambers T J, Hahn C S, Galler R and Rice C M 1990 Flavivirus genome organization, expression, and replication; *Annu. Rev. Microbiol.* **44** 649–688

Chen W R, Tesh R B and Rico-Hesse R 1990 Genetic variation of Japanese encephalitis virus in nature; *J. Gen. Virol.* **71** 2915–2922

Felsenstein J 1995 PHYLIP (Phylogeny Inference Package, version 3.572c) (Seattle, WA: Department of Genetics, University of Washington)

Hanna J N, Ritchie S A, Phillips D A, Lee J M, Hills S L, van den Hurk A F, Pyke A T, Johansen C A and Mackenzie J S 1999 Japanese encephalitis in north Queensland, Australia, 1998; *Med. J. Aust.* **170** 533–536

Hashimoto H, Nomoto A, Watanabe K, Mori T, Takezawa T, Aizawa C, Takegami T and Hiramatsu K 1988 Molecular cloning and complete nucleotide sequence of the genome of Japanese encephalitis virus Beijing-1 strain; *Virus Genes* **1** 305–317

Hein J 1990 Unified approach to alignments and phylogenies. *Methods Enzymol.* **183** 625–645

Higgins D G, Thompson J D and Gibson T J 1996 Using CLUSTAL for multiple sequence alignments; *Methods Enzymol.* **266** 383–402

Holland J, Spindler K, Horodyski F, Grabau E, Nichol S and VandePol S 1982 Rapid evolution of RNA genomes; *Science* **215** 1577–1585

Hori H 1986 Oligonucleotide fingerprint analysis pf Japanese encephalitis (JE) virus strains of different geographic origins; *Trop. Med.* **28** 179–190

Kar N J and Saxena V K 1998 Some epidemiological characteristics of Japanese encephalitis in Haryana state of northern India; *J. Commun. Dis.* **30** 129–131

Khromykh A A, Varnavski A N and Westaway E G 1998 Encapsidation of the flavivirus kunjin replicon RNA by using a complementation system providing Kunjin virus structural proteins in trans; *J. Virol.* **72** 5967–5977

Kobayashi Y, Hasegawa H, Oyama T, Tamai T and Kusaba T 1984 Antigenic analysis of Japanese encephalitis virus by using monoclonal antibodies; *Infect. Immun.* **44** 117–123

Kobayashi Y, Hasegawa H and Yamauchi T 1985 Studies on the antigenic structure of Japanese encephalitis virus using monoclonal antibodies; *Microbiol. Immunol.* **29** 1069–1082

Kumar S, Tamura K and Masatoshi N 1993 MEGA: Molecular Evolutionary Genetics Analysis, version 1·01 (University Park, PA 16802: The Pennsylvania State University)

Lin Y L, Liao C L, Yeh C T, Chang C H, Huang Y L, Huang Y Y, Jan J T, Chin C and Chen L K 1996 A highly attenuated strain of Japanese encephalitis virus induces a protective immune response in mice; *Virus Res.* **44** 45–56

Mathur A, Chaturvedi U C, Tandon H O, Agarwal A K, Mathur G P, Nag D, Prasad A and Mittal V P 1982 Japanese encephalitis epidemic in Uttar Pradesh, India during 1978; *Indian J. Med. Res.* **75** 161–169

McMinn P C 1997 The molecular basis of virulence of the encephalitogenic flaviviruses; *J. Gen. Virol.* **78** 2711–2722

Ni H and Barrett A D 1996 Molecular differences between wild–type Japanese encephalitis virus strains of high and low mouse neuroinvasiveness; *J. Gen. Virol.* **77** 1449–1455

Nitayaphan S, Grant J A, Chang G J and Trent D W 1990 Nucleotide sequence of the virulent SA-14 strain of Japanese encephalitis virus and its attenuated vaccine derivative, SA-14-14-2; *Virology* **177** 541–552

Prasad S R, Kumar V, Marwaha R K, Batra K L, Rath R K and Pal S R 1993 An epidemic of encephalitis in Haryana: serological evidence of Japanese encephalitis in a few patients; *Indian Pediatr.* **30** 905–910

Ravi V and Shankar S K 1989 Japanese encephalitis: the Indian scene; in *Progress in clinical neurosciences* (eds) K K Sinha and P Chandra (Ranchi: Neurological Society of India) pp 69–84

Sharma R C, Saxena V K, Bharadwaj M, Sharma R S, Verghese T and Datta K K 1991 An outbreak of Japanese encephalitis in Haryana-1990; *J. Commun. Dis.* **23** 168–169

Sharma S N and Panwar B S 1991 An epidemic of Japanese encephalitis in Haryana in the year 1990; *J. Commun. Dis.* **23** 204–205

Smithburn K C, Kerr J A and Gatne P B 1954 Neutralising antibodies against certain viruses in the sera of residents of India; *J. Immunol.* **72** 248–251

Sumiyoshi H, Mori C, Fuke I, Morita K, Kuhara S, Kondou J, Kikuchi Y, Nagamatu H and Igarashi A 1987 Complete nucleotide sequence of the Japanese encephalitis virus genome RNA; *Virology* **161** 497–510

Ta M and Vrati S 2000 Mov34 protein from mouse brain interacts with the 3′ noncoding region of Japanese Encephalitis virus; *J. Virol.* **74** 5108–5115

Umenai T, Krzysko R, Bektimivov T A and Assad F A 1985 Japanese encephalitis: current world wide status; *Bull. W.H.O.* **63** 625–631

Vrati S, Giri R K, Razdan A and Malik P 1999a Complete nucleotide sequence of an Indian strain of Japanese encephalitis virus: sequence comparison with other strains and phylogenetic analysis; *Am. J. Trop. Med. Hyg.* **61** 677–680

Vrati S, Agarwal V, Malik P, Wani S A and Saini M 1999b

Molecular characterization of an Indian isolate of Japanese encephalitis virus that shows an extended lag phase during growth; *J. Gen. Virol.* **80** 1665–1671

Wills M R, Sil B K, Cao J X, Yu Y X and Barrett A D 1992 Antigenic characterization of the live attenuated Japanese encephalitis vaccine virus SA14-14-2: a comparison with isolates of the virus covering a wide geographic area; *Vaccine* **10** 861–872

Zhao B, Mackow E, Buckler-White A, Markoff L, Chanock R M, Lai C J and Makino Y 1986 Cloning full-length dengue type 4 viral DNA sequences: analysis of genes coding for structural proteins; *Virology* **155** 77–88

MS received 24 April 2000; accepted 17 July 2000

Corresponding editor: SEYED E HASNAIN