



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Sequencing of prototype viruses in the Venezuelan equine encephalitis antigenic complex

John D. Meissner ^{a,*}, Claire Y.-H. Huang ^a, Martin Pfeffer ^b,
Richard M. Kinney ^a

^a *Arbovirus Diseases Branch, Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Public Health Service, U.S. Department of Health and Human Services, P.O. Box 2087, Fort Collins, CO 80522, USA*

^b *Institute for Medical Microbiology, Epidemic and Infectious Diseases, Veterinary Faculty, Ludwig-Maximilians University, Munich, Germany*

Received 16 March 1999; received in revised form 4 June 1999; accepted 4 June 1999

Abstract

The 5' nontranslated region (5'NTR) and nonstructural region nucleotide sequences of nine enzootic Venezuelan equine encephalitis (VEE) virus strains were determined, thus completing the genomic RNA sequences of all prototype strains. The full-length genomes, representing VEE virus antigenic subtypes I–VI, range in size from 11.3 to 11.5 kilobases, with 48–53% overall G + C contents. Size disparities result from subtype-related differences in the number and length of direct repeats in the C-terminal nonstructural protein 3 (nsP3) domain coding sequence and the 3'NTR, while G + C content disparities are attributable to strain-specific variations in base composition at the wobble position of the polyprotein codons. Highly-conserved protein components and one nonconserved protein domain constitute the VEE virus replicase polyproteins. Approximately 80% of deduced nsP1 and nsP4 amino acid residues are invariant, compared to less than 20% of C-terminal nsP3 domain residues. In two enzootic strains, C-terminal nsP3 domain sequences degenerate into little more than repetitive serine-rich blocks. Nonstructural region sequence information drawn from a cross-section of VEE virus subtypes clarifies features of alphavirus conserved sequence elements and proteinase recognition signals. As well, whole-genome comparative analysis supports the reclassification of VEE subtype-variety IF and subtype II viruses. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Alphavirus; VEE; Nonstructural proteins; Nucleotide sequence

1. Introduction

Venezuelan equine encephalitis (VEE) viruses are arthropod-borne, Western Hemisphere alphaviruses in the family *Togaviridae* (Walton and Grayson, 1988). Individual VEE virus isolates are

* Corresponding author. Present address: Sequinox Laboratories, 751 Heinz Parkway, Estes Park, CO 80517, USA. Tel: +1-970-5864889; fax: +1-970-5864889.

E-mail address: jdmeissner@usa.net (J.D. Meissner)

designated as either epizootic or enzootic based on origin. Epizootic strains, isolated from sporadic equine epizootics and epidemics, multiply to high titer in non-immune equines and are transmitted indiscriminately by suitable mammalophilic or ornithophilic mosquito vectors to humans, domestic livestock, wild mammals, and birds (Sudia and Newhouse, 1975). The only interepizootic ‘reservoirs’ identified thus far have been in the form of incompletely formalin-inactivated vaccines (Kinney et al., 1992a). In contrast, enzootic strains are readily isolated in discrete geographic foci sharing common elevations, climate, and vegetation (Johnson and Martin, 1974). Elegant early field work implicated *Culex* (primarily subgenus *Melanoconion*) mosquitoes and wild rodent or marsupial hosts in enzootic transmission cycle maintenance (Chamberlain et al., 1964; Galindo et al., 1966; Grayson and Galindo, 1968).

The VEE antigenic complex (Shope et al., 1964), as initially defined by a short-incubation hemagglutination inhibition assay using spiny rat antisera (Young and Johnson, 1969) and later refined and expanded by hemagglutination and neutralization tests (France et al., 1979; Calisher et al., 1982; Kinney et al., 1983; Calisher et al., 1985), currently consists of six subtypes (I–VI). Subtype I is divided into five varieties (IAB, IC, ID, IE, IF) and subtype III is divided into three varieties (IIIA, IIIB, IIIC).

Prior to this decade, VEE viruses isolated from epizootics–epidemics were invariably subtype IAB or IC strains. Not surprisingly, subtype IAB and IC strains are experimentally equine virulent, although virulence varies among strains (Mackenzie et al., 1976). The more divergent enzootic strains occupy remaining antigenic classifications, and are for the most part equine benign. However, intrathecally inoculated subtype ID and IE strains cause a fulminant encephalitis in horses (Dietz et al., 1978), and recent equine outbreaks in southern Mexico with 40–50% case-fatality rates yielded subtype IE strains (Oberste et al., 1998).

This confirmation of the previously suspected involvement of certain enzootic VEE virus strains in outbreaks (Morilla-Gonzalez and Mucha-Ma-

cias, 1969), along with the isolation of long-quiet subtype IC lineages from outbreaks in Venezuela and Columbia (Rico-Hesse et al., 1995; Weaver et al., 1996), raised or renewed questions concerning determinants of equine virulence and sources of equine-virulent strains.

Our laboratory has been engaged in the nucleotide sequencing of prototype strains in the VEE complex to examine genetic relationships and develop improved serologic (Hunt et al., 1990, 1991; Roehrig et al., 1991; Hunt and Roehrig, 1995) and polymerase chain reaction (PCR) reagents (Pfeffer et al., 1997, 1998) for VEE diagnostics. We have previously reported the full-length genomic sequences of subtype IAB, IC, and ID viruses (Kinney et al., 1989, 1992a,b). The complete sequence of a subtype IE virus, strain 68U201, has also been determined (Oberste et al., 1996).

The organization of the VEE virus positive-sense, single-stranded RNA genome is 5'-methylated cap-nontranslated region (5'NTR)-nonstructural protein 1 (nsP1)-nsP2-nsP3-nsP4-26S junction region-capsid-E3-E2-6K-E1-3'NTR-polyadenylated tail. Analogous to other alphaviruses, nonstructural replicase polyproteins are likely translated directly off the plus strand after viral uncoating, orchestrating complementary minus-strand synthesis and (following additional intracellular processing) subgenomic and genomic plus-strand RNA synthesis (de Groot et al., 1990; Strauss and Strauss, 1994; Lemm et al., 1998).

The present study reports the 5'NTR and nonstructural region sequences of nine enzootic VEE virus strains, which, when combined with the 26S mRNA sequences of these strains (Kinney et al., 1998), completes the genomic sequences of all currently identified prototype strains in the VEE virus antigenic complex.

2. Materials and methods

2.1. Viruses

Information concerning strain name, abbreviation, classification, origin, and GenBank accession

number is shown in Table 1. All viruses except 68U201 (Oberste et al., 1996) are plaque-purified stocks that have been used to establish antigenic relationships and virus phenotypes, as summarized previously (Kinney et al., 1998).

2.2. Reverse transcriptase-polymerase chain reaction and sequencing

Extraction of viral genomic RNA, amplification of cDNA by reverse transcriptase-polymerase chain reaction (RT-PCR), and agarose gel purification and sequencing of cDNA were performed as previously described (Kinney et al., 1998). RT-PCR utilized degenerate amplimers designed to

anneal to regions of conserved amino acid sequence in VEE virus and other alphaviruses. Initially, the majority of the nonstructural region was amplified using three overlapping sense/anti-sense amplimer combinations (designations include the 3' terminal nucleotide genomic position in TRD virus, standard ambiguity code is R = A/G, Y = C/T, H = A/C/T, M = A/C, D = A/G/T, K = G/T, V = A/C/G, B = C/G/T, N = A/C/G/T):

VE-72 5'-ATGGAGAARGTTCACGTT
GAYATCGAGG
cVE-2511 5'-TCRTGRTRTRAARTGNACYT
TNAGRCACATCAT

Table 1
Reference information for Venezuelan equine encephalitis (VEE) virus strains

Virus	Abbreviation ^a	Subtype ^b	Origin	Sequence reference/isolation reference ^c	GenBank Accession Number
Trinidad donkey	TRD	IAB	Trinidad, 1943	Kinney et al., 1989	L01442
TC-83	TC-83	^d	^d	Kinney et al., 1989	L01443
71-180	71-180	IAB	Texas, USA, 1971	Kinney et al., 1992a	AF069903
P676	P676	IC	Venezuela, 1963	Kinney et al., 1992b	L04653
3880	3880	ID	Panama, 1961	Kinney et al., 1992b	L00930
Mena II	MENA	IE	Panama, 1962 ^e	Galindo et al., 1966	AF075252
68U201	68U201	IE	Guatemala, 1968	Oberste et al., 1996	U34999
78V-3531	78V-3531	IF	Brazil, 1976	Calisher et al., 1982	AF075257
Everglades Fe3-7c	EVE	II	Florida, USA, 1963	Chamberlain et al., 1964	AF075251
Mucambo BeAn 8	MUC	IIIA	Brazil, 1954	Causey et al., 1961	AF075253
Tonate CaAn 410d	TON	IIIB	French Guiana, 1973	Digoutte and Girault, 1976	AF075254
71D-1252	71D-1252	IIIC	Peru, 1971	Scherer and Anderson, 1975	AF075255
Pixuna-BeAr35645	PIX	IV	Brazil, 1961	Shope et al., 1964	AF075256
Cabassou CaAr 508	CAB	V	French Guiana, 1968	Digoutte and Girault, 1976	AF075259
AG80-663	AG80-663	VI	Argentina, 1980	Calisher et al., 1985	AF075258

^a Strain name abbreviation (Karabatsos, 1985).

^b Subtype-variety classification scheme for the VEE antigenic complex (Young and Johnson, 1969; France et al., 1979; Calisher et al., 1982; Kinney et al., 1983; Calisher et al., 1985; Roehrig and Mathews, 1985; Roehrig et al., 1991).

^c Previously sequenced VEE viruses in **bold**, isolation references included in sequencing reports referenced.

^d Vaccine strain derived from TRD virus by passage in tissue culture (Berge et al., 1961).

^e Two subtype IE strains were isolated from the same patient in consecutive years, the latter in 1962 (M.A. Grayson, personal communication).

VE-2008	5'-CAYGGHGGRGCNYTGAAY ACNGAYGARGACTACTA or
VE-2518	5'-GGDGAYCCKAARCARTGYG GYTYTYTYAAYATGATGTG
cVE-4506	5'-TCHTCVGAGATRCATAT CTCYTCBRYNGCYTCYCT
VE-4477	5'-GCRGATGTRGCYATMTAYT GYMGRGAYAARAARTGGGA
cVE-7155	5'-ACYACBGCRTCDATRATYT TNACYTCCATRTTYARCCA

Additional degenerate and virus-specific primers (sequences available on request) were designed to fill in sequence gaps. The infrequency of invariant neighboring wobble position nucleotides in the nonstructural region required synthesis of more degenerate primers than were used for the 26S mRNA region (Kinney et al., 1998). All initial priming sites in the genome were derived as internal sequences in other amplicons, and both strands of the amplified cDNA were sequenced.

2.3. 5' nontranslated region sequence determination

5'NTR sequences were obtained using the 5' RACE System kit (GIBCO BRL, Bethesda, MD) according to supplied protocol. The anti-sense oligomer cVE-963 (5'-GTBACYTTGCARCACAAGAATWCCCTCGCGRTGCAT-3', where W = A/T) primed synthesis of the first strand cDNA. Following RNA degradation, the purified cDNA from each strain was divided into two aliquots, and the aliquots tailed with either dATP or dCTP. The dATP-tailed cDNA was amplified using a synthesized poly-T primer (5'-CACAG-ACTGCAGCGAATTCGGTACCT-TTTTTTTT-TTTTT-3') and cVE-256 (5'-GGR-CARAYRCARTGRTAYTTRTKHTTHGART-ACATTC-3'), a VEE virus-specific antisense primer. The dCTP-tailed cDNA was amplified using the supplied anchor primer and cVE-256. Comparison of sequences determined from the alternatively-tailed cDNAs assured the identity of the 5' terminal nucleotide.

2.4. Published partial or complete Venezuelan equine encephalitis virus nonstructural region sequences

We previously reported partial N-terminal nsP1 and C-terminal nsP4 coding sequences of enzootic strains included in this study (Pfeffer et al., 1997; Kinney et al., 1998). Full-length nucleotide sequences have been determined for five naturally-isolated VEE virus subtype I strains and for TC-83, the TRD-derived vaccine strain (Kinney et al., 1989, 1992a,b; Oberste et al., 1996). In the course of this study, an A → T artifact in the TRD virus sequence at nt 4809 was discovered. As originally reported (Kinney et al., 1989), the nucleotide at this position using plaque-purified TRD virus as starting template for sequence determination was T. Resequencing using cDNA amplified from low-passage (equine-1/mouse-1, 1961 seed) virus that had not been plaque-purified revealed an A at this position, as is found in TC-83 and every VEE virus strain. A partial MENA virus nsP3-nsP4 coding sequence, GenBank accession no. U34978 (Oberste et al., 1996), is identical to sequence obtained using our laboratory stock of MENA virus. Conflicts with a previous partial C-terminal nsP4 coding sequence determination, ranging from one nucleotide difference in MUC virus to 11 differences in 78V-3531 virus, are likely due to sequencing methods employed (Weaver et al., 1992).

2.5. Sequence analysis

Nucleotide and amino acid sequence analysis was performed using programs available in the GCG package, version 8.0 (Devereux et al., 1984), with default settings used in all cases. Phylogenetic analysis, including bootstrap resampling (Felsenstein, 1985), was performed using maximum parsimony algorithms implemented by PAUP V3.1.1 (Swofford, 1993). Characters were either unweighted or were unweighted in codon first and second positions when codon third position characters were given zero weight. Throughout this study, all comparative or phylogenetic analyses involving aligned VEE virus deduced polyprotein P1234 sequences exclude the C-termi-

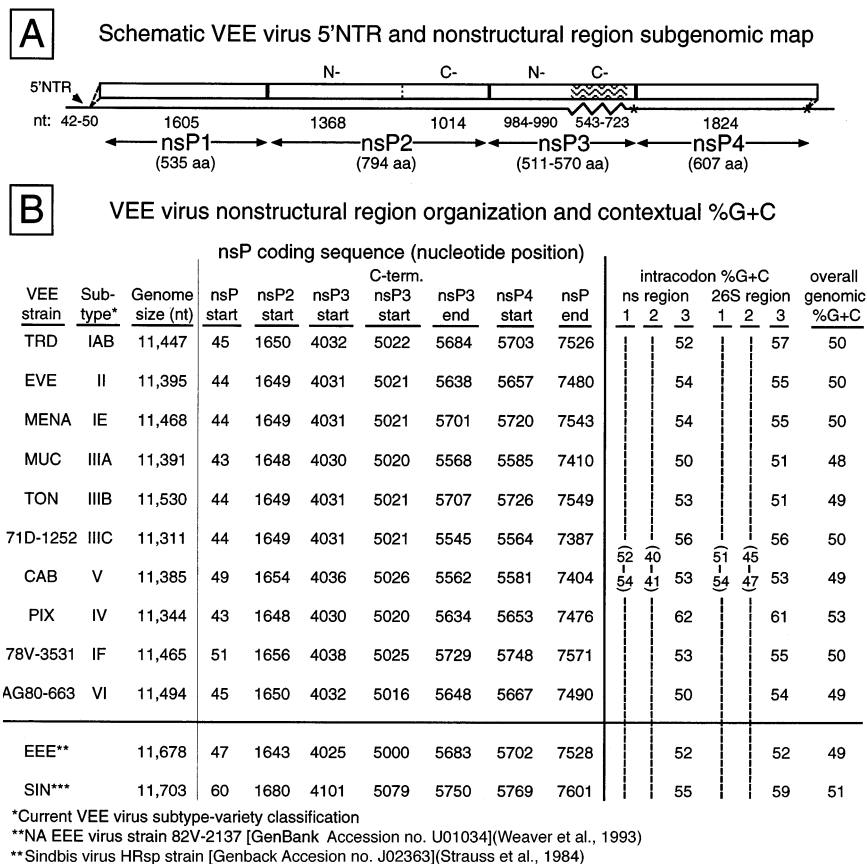


Fig. 1. (A) Schematic VEE virus 5'NTR and nonstructural region subgenomic map (not drawn to scale), with ranges of 5'NTR and nsP coding sequence lengths above and deduced amino acid sequence lengths below individual nsP designations. The C-terminal nsP3 domain is shaded. Asterisks indicate stop codons. (B) Total genome lengths and nonstructural region landmarks of prototype VEE viruses and representative alphaviruses. Nucleotide positions for the ends of the nsP3 and nonstructural polyprotein coding sequences include respective stop codons. Polyprotein codon position and total genomic %G + C content are also listed (right). The C-terminal nsP3 domain coding sequence is excluded from nonstructural region %G + C content calculations. For strain abbreviations, see Table 1.

nal nsP3 domain up to and including the putative nsP3/nsP4 cleavage site.

3. Results

3.1. Genome organization

A schematic VEE virus subgenome and accompanying chart (Fig. 1) provide nonstructural region landmarks for TRD virus and the nine enzootic strains sequenced in this study — Mena II (antigenic subtype-variety IE), 78V-3531 (IF),

Everglades Fe3-7c (II), Mucambo BeAn 8 (IIIA), Tonate CaAn 410d (IIIB), 71D-1252 (IIIC), Pixuna BeAr 35645 (IV), Cabassou CaAr 508 (V) and AG80-663 (VI) — as well as for North American Eastern equine encephalitis (EEE) virus strain 82V-2136 (Weaver et al., 1993) and Sindbis virus (SIN) strain HR_{sp} (Strauss et al., 1984).

The actual N-termini of the VEE virus nonstructural polyproteins or processed proteins have not been directly confirmed by protein sequencing. The favorable context for the putative initiation codon (Kozak, 1981, 1987) and shared amino acid identity with the cognate alphavirus nsP ter-

mini along with the known specificity of the papain-like proteinase (Strauss and Strauss, 1994), suggest these are the most likely termini. The VEE virus nsP2 is divided (dotted line) into N- and C-terminal domains after the isoleucine residue at position 456 (numbering based on TRD virus), corresponding to the final residue in the alignment of SIN nsP2 with certain single-stranded RNA plant viral helicases (Ahlquist et al., 1985) and the region where the N-terminus of the SIN nsP2 proteinase domain was mapped using infectious clone deletion mutants (Hardy and Strauss, 1989). Comparative analysis of VEE virus sequences (see below) indicates that selection pressures are differ-

ent in N- and C-terminal portions of this protein. The VEE virus N-terminal nsP3 domain ends at residue 330, a conserved aliphatic residue following an invariant tyrosine residue. Only strains 78V-3531 and AG80-663 contribute to the variation in N-terminal nsP3 domain size.

3.2. 5' nontranslated region

Individual VEE virus 5'NTR sequences, shortest among the alphaviruses, vary in length by 0–8 nucleotides (Fig. 2A). Despite an overall lack of sequence conservation, alphavirus 5'NTRs are predicted to form stable secondary structures (Ou

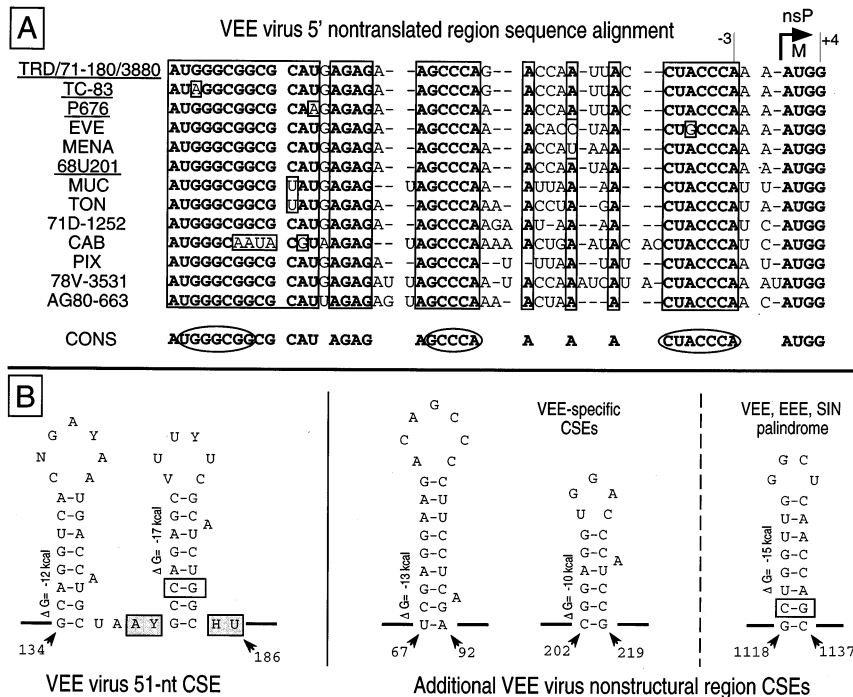


Fig. 2. VEE virus conserved sequence elements. (A) 5'NTR sequence alignment, with both conserved nucleotides (bold) and consensus sequence (CONS) derived from nucleotides present in at least 5 of 6 VEE virus subtypes. A dash indicates a gap used to improve alignment. 5'NTR sequences of 71–180 and 3880 viruses are identical to TRD virus (Kinney et al., 1992a,b). Previously reported sequences are underlined. See Table 1 for strain abbreviations. (B) Proposed secondary structures for the VEE virus 51-nt CSE (left) and additional VEE virus nsP1 coding region CSEs (right). The 51-nt CSE nucleotide positions and putative stem sequence are based on TRD virus. Remaining non-stem 51-nt CSE sequence is a composite of all VEE virus strains (Y = A/U, V = A/G/C, H = A/C/U, N = A/C/G/U). The C-G pairs replaced by U-A pairs in certain strains are boxed. The shaded boxes indicate additional pairs which could hydrogen bond in subtype IE MENA virus. For the putative hairpin beginning at TRD nt-67, the sequence is that of 71D-1252 virus, although a similar structure with slightly lower stability can be modeled for other VEE virus strains. TRD virus sequence is used for the other two hairpins. To facilitate comparison with earlier studies (Ou et al., 1983; Niesters and Strauss, 1990a), an older method of calculating the free energy of duplex formation is used (Tinoco et al., 1973), relying on a strict interpretation of Table 1 in this reference without modifications proposed in the text. ΔG = free energy at 25°C.

Table 2

Individual nonstructural protein (nsP) coding sequence identities, adding increasingly more divergent VEE virus (and alphavirus) strain sequences

	nsP1	N-term. nsP2	C-term. nsP2	N-term. nsP3	nsP4
	%nt(%aa)	%nt(%aa)	%nt(%aa)	%nt(%aa)	%nt(%aa)
TRD, TC-83, 71–180, P676, 3880, EVE	89 (98)	84 (96)	84 (96)	85 (98)	86 (98)
+MENA, 68U201	77 (93)	70 (93)	68 (88)	70 (92)	70 (93)
+MUC, TON, 71D-1252	67 (87)	61 (86)	56 (78)	59 (82)	60 (87)
+CAB, PIX, 78V-3531, AG80-663	60 (79)	55 (79)	48 (66)	55 (75)	55 (81)
+EEE ^a	49 (60)	52 (72)	38 (48)	44 (56)	50 (72)
+SIN ^a	41 (48)	44 (56)	27 (30)	35 (40)	45 (63)

^a NA EEE virus strain 82V-2136 sequence, SIN strain HR_{sp} sequence.

et al., 1983) potentially required for ribosome orientation or (via the 3' complement on the minus strand) promotion of plus-strand RNA synthesis (Niesters and Strauss, 1990b; Pardigon and Strauss, 1996). Putative VEE virus 5'NTR hairpins are less stable and structurally less complex than stem–loop structures modeled for other alphavirus 5'NTR sequences. A secondary structure previously modeled for TRD virus (Dubuisson et al., 1997) has a calculated free energy of –4.4 kcal by the method used here (Tinoco et al., 1973). A 5' terminal hairpin formed from the conserved UGGGCGG heptamer (circled in the consensus sequence) starting at TRD virus nt-2 and either the conserved GCCCA (nt-21) or the conserved CUACCCA (nt-36) has a calculated free energy of –11.2 kcal using the former combination and –7.2 kcal (–8.2 kcal on the minus strand) using the latter combination.

In contrast to conserved sequence element (CSE) putative stem–loop structures noted below, the VEE virus 5'NTR stems are not supported by compensatory pairs of nucleotide changes (Noller and Woese, 1981; Pace et al., 1989; Niesters and Strauss, 1990b) in sequences of other VEE virus strains.

3.3. Nonstructural region

Overall VEE virus nonstructural region deduced amino acid sequence identity is 77%, compared to 60% in the structural region (Kinney et al., 1998). Table 2 details the stepwise approach

to the nonstructural region consensus sequence, starting with a consensus determined for selected subtype I and II viruses and adding increasingly more divergent strains. EEE and SIN genotypes (Strauss et al., 1984; Weaver et al., 1993) are added to the overall VEE virus consensus sequence to emphasize certain trends. For any particular VEE virus nsP coding sequence, the majority of nucleotide changes are silent, occurring initially at the codon wobble positions and continuing until the wobble positions are essentially saturated, as noted previously in other alphavirus sequence comparisons (Strauss and Strauss, 1994). For example, the nsP4 coding sequences of subtype IAB, IC, ID, and II strains share only 86% nucleotide identity yet are virtually identical at the deduced amino acid level. It is clear from every generated consensus sequence that individual nsPs (and protein domains) diverge at different rates. The VEE virus N-terminal nsP2 domain, for example, is much more conserved than the C-terminal nsP2 domain. This divergence is even more striking when SIN and EEE virus nsP2 sequences are included in the comparison.

A ranking of invariant nsP amino acid residues, arranged by total numbers of invariant alanine, arginine, etc. residues in the deduced polyprotein P1234 sequence, was generated, and ratios of total invariant residues/total residues for each amino acid were calculated for every sequenced strain (data not shown). In general, four-codon amino acid families are more prevalent in the VEE con-

sensus sequence than two-codon families, and amino acids with specific, irreplaceable roles in protein structure or enzymatic activity, i.e. glycine, cysteine, and tryptophan, are well-conserved and seldom extraneous.

Previously identified alphavirus nsP or cognate ‘Sindbis-like supergroup’ protein functions and features are listed in Table 3, along with invariant residue correlates in VEE virus sequences. Only one Sindbis-like supergroup invariant residue in the replicase complex is not strictly conserved by all VEE virus strains. This nsP2 helicase domain valine residue (Gorbalenya et al., 1989) is an isoleucine in strain 71D-1252.

3.4. Conserved sequence elements

The VEE virus nsP1 deduced amino acid consensus sequence is less conserved than the nsP4 sequence (Table 2, line 4), while the corresponding nsP1 coding sequence is more conserved. One

reason for this is the presence of CSEs, most notably the alphavirus 51-nt CSE, in the nsP1 coding region. The 51-nt CSE has been well-characterized in SIN (Niesters and Strauss, 1990a) and may represent a promoter for RNA minus-strand synthesis. The VEE virus 51-nt CSE is shown in Fig. 2B, along with three other nsP1 coding region CSEs.

Unlike the putative VEE virus 5’NTR stem–loop structure, proposed RNA secondary structures for these CSEs are supported by pairs of compensatory nucleotide changes (Noller and Woese, 1981; Pace et al., 1989) in different VEE virus strains. The boxed C-G pair in the stem of the second hairpin of the 51-nt CSE (Fig. 2) is a U-A pair in all three of the subtype III strains and in strains 78V-3531 and AG80-663. Similarly, the boxed C-G pair in the stem of the 20-nt near-perfect palindrome beginning at TRD virus nt-1118 is a U-A pair in EVE and MUC viruses. Compensatory changes are lacking in the proposed stems

Table 3

Essential alphavirus nonstructural protein (or cognate ‘Sindbis-like supergroup’ protein) features — VEE virus nonstructural protein sequence correlates

Nonstructural protein or protein domain	(putative) Function(s)	Conserved residues/motifs/domains ^a	Comments
nsP1 ^b	membrane-associated methyltransferase, guanylyltransferase	H-37, D-89, R-92, Y-248 [VEE virus specific: N-terminal 55 amino acid (aa) residues]	VEE virus C-terminal residues vary — C-terminal truncations well-tolerated in SIN and SF virus constructs
N-terminal nsP2 domain	RNA helicase, ATPase, GTPase	6–7 conserved segments, including GVPGSGKS and DEAF NTP-binding signature motifs, 14 ‘invariant’ aa residues	‘Invariant’ nsP2 V-389 is isoleucine in strain 71D-1252
C-terminal nsP2 domain	Papain-like proteinase	Catalytic dyad residues C-477 and H-546 (C-481 and H-558 in SIN), W-547	Other plus strand RNA viral papain-like proteinases have flexible catalytic dyad spacing
N-terminal nsP3 domain	Phosphoprotein necessary for minus-strand, 26S mRNA synthesis		Only aa similarity is with rubella virus and coronavirus proteins of undetermined functions
nsP4	RNA-dependent RNA polymerase	N-terminal Y residue, GDD motif, 8 conserved domains	

^a Invariant VEE virus residues in **bold**, numbering based on TRD virus nonstructural protein sequences.

^b References: nsP1 (Mi et al., 1989; Mi and Stollar, 1991; Rozanov et al., 1992; Laakkonen et al., 1994; Peranen et al., 1995; Wang et al., 1996; Ahola et al., 1997; Pfeffer et al., 1997); nsP2 (Gorbalenya et al., 1988, 1989, 1991; Hodgman, 1988; Ding and Schlesinger, 1989; Hardy and Strauss, 1989; Strauss et al., 1992; Rikkonen et al., 1994); nsP3 (Gorbalenya et al., 1991; LaStarza et al., 1994a,b); nsP4 (Kamer and Argos, 1984; Koonin, 1991; Shirako and Strauss, 1998).

of the two additional VEE virus-specific CSEs. However, equivalent calculated free energies and proximity to the 51-nt CSE double hairpins support the involvement of these CSEs in a large 5' terminal secondary structure, as has been modeled for SIN (Niesters and Strauss, 1990b).

3.5. Proteinase sites

Proteinase recognition sites have been identified in other alphaviruses by N-terminal sequencing of processed nsPs and by mutations which abolish *in vitro* cleavage (Strauss and Strauss, 1994). By alignment, the putative substrates for the VEE virus papain-like proteinase are predicted to be (residues given as P4P3P2P1/P1'P2'P3'P4', variable residues in brackets): nsP1/nsP2 — EAGA/G[S,T]VE; nsP2/nsP3 — EAG[C,S,T]/APSY; and nsP3/nsP4 — [D,E]AGA/YIFS. The P3, P2, and P1 residues constitute the major recognition signal for alphavirus proteinase cleavage (de Groot et al., 1990; Strauss and Strauss, 1994), with the P3 residue generally alanine, the P2 residue invariably glycine, and the P1 residue generally alanine or glycine, although any non-bulky residue may be tolerated, as indicated by the P1 residues allowed at the putative VEE virus nsP2/nsP3 cleavage site. The P4 residue is acidic in all VEE virus strains. The VEE virus proteinase may require an acidic P4 residue or a particular 3-dimensional conformation for cleavage, if for no other reason than the presence of additional AGA motifs in the VEE virus polyprotein P1234 consensus sequence besides those at putative cleavage sites.

3.6. C-terminal nsP3 domain

Because of large duplications and amplification of small serine-rich blocks in different strains, the deduced VEE virus C-terminal nsP3 domain ranges in size from 174–234 amino acids, with only 27 amino acid residues invariant (Fig. 3). Of the four nsP3 carboxyl region domains previously identified by comparison of five subtype I strains (Oberste et al., 1996), only domain 4 remains inviolate in the overall VEE virus alignment. Although divergence in this region is so great that any proposed alignment likely contains errors, the

majority of VEE virus C-terminal nsP3 domain sequences coalesce around two motifs corresponding to truncated domains 2 and 3. The domain 2 SXWSXPXASDF motif (where 'X' indicates a variable amino acid residue) is conserved by all strains except PIX. Only strains 78V-3531 and AG80-663 do not preserve at least one copy of the imperfect PXPAPRT repeat in domain 3. One copy of this repeat is also found in EEE virus strain 82V-2137 (Weaver et al., 1993).

Duplication events appear to be common in the alphavirus C-terminal nsP3 domain (Strauss et al., 1988), although the variety and size of direct repeats in VEE virus sequences have not been previously noted. A 34-amino acid duplication is present in all VEE virus subtype IAB, IC, ID, and II strains. One copy of this duplication was deleted during propagation of a TRD virus infectious clone, with no demonstrable effect on viability (Davis et al., 1989). Both subtype IE strains possess a larger upstream duplication obscured by a subsequent deletion event. TON virus has two duplications of 34 and 20 amino acids, neither of which is found in other subtype III strains. The region corresponding to domain 2 has been duplicated at least once (Fig. 3) in subtype IV PIX virus. Beyond domain 2, sequences of strains 78V-3531 and AG80-663 comprise little more than short serine, alanine, and/or arginine-containing repeats. Translation in alternative reading frames rules out the possibility that sequencing errors or recent indels mask motif conservation in these strains.

The relative ages of these direct repeats can be determined by nucleotide alignment. The two copies of the 20 amino acid duplication in TON virus, for example, share 59 of 60 nucleotides. Using an estimate of 10^{-4} substitutions/nucleotide (Strauss and Strauss, 1986), this duplication occurred < 100 generations prior to isolation.

The opal termination codon is preserved by all VEE virus strains, as are 12 of 14 amino acid residues immediately upstream (Fig. 3). Amino acid sequence conservation upstream of the opal codon is atypical for alphaviruses (Strauss and Strauss, 1994).

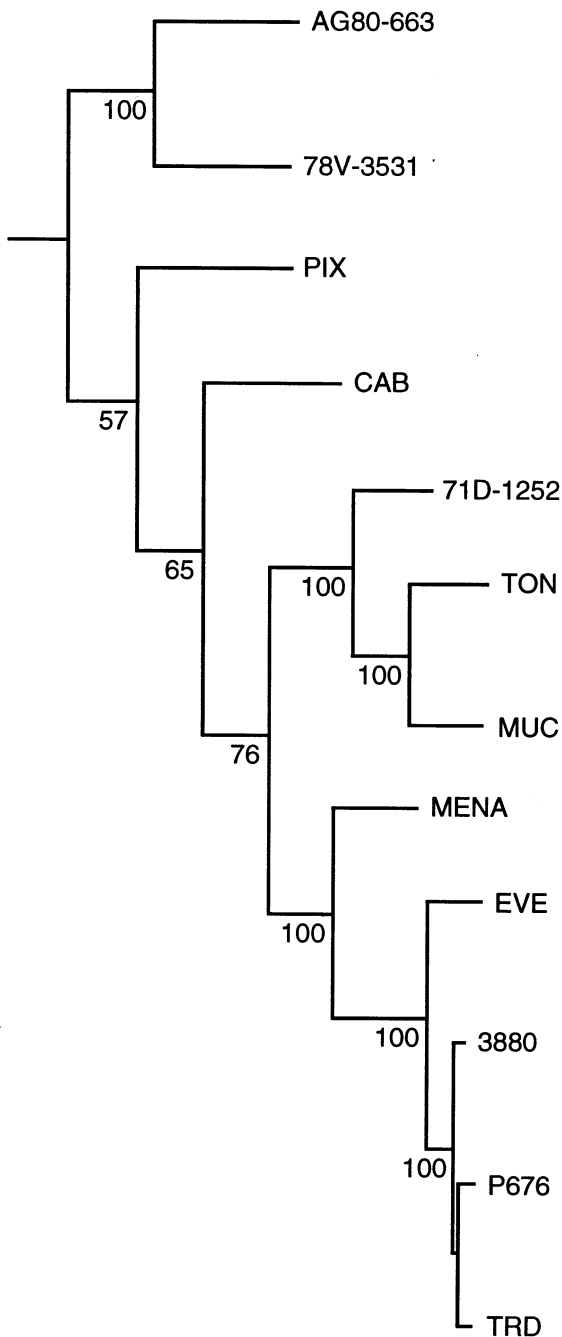


Fig. 4. (Continued)

are deduced amino acid sequences or codon first and second position nucleotides. Maximum parsimony analysis (Swofford, 1993) using polyprotein

P1234 sequences or corresponding codon first and second position nucleotides produce the same topology as that shown in Fig. 4. The tree in Fig. 4 is derived from combined nonstructural and structural region codon first and second position nucleotides, with EEE virus strain 82V-2136 (Weaver et al., 1993) serving as the outgroup. This tree is well-supported by bootstrap resampling (Felsenstein, 1985), as only two partitions are present in fewer than 70% of resampled trees. An identical branching pattern is again reproduced when combined deduced nonstructural and structural region amino acid sequences are used, although bootstrap proportions in this case are somewhat lower (data not shown).

4. Discussion

Comparative analysis is often the best initial experimental method for determining secondary structures of RNA molecules (Noller and Woese, 1981; Pace et al., 1989), and is especially appropriate for examining alphavirus genomic elements potentially involved in formation of secondary structures required for host or viral protein interactions. Essential features of the alphavirus 51-nt CSE structure, for example, are confirmed or clarified by 51-nt CSE sequences reported here. Substitutions in putative stem structures of enzootic VEE virus strains represent natural experiments identical to SIN mutants constructed in vitro (Niesters and Strauss, 1990a). The requirement for strict maintenance of stem length in the second stem-loop of the 51-nt CSE as demonstrated by anti-pairing changes in O'Nyong-nyong virus (Levinson et al., 1990) is apparently violated

Fig. 4. Maximum-parsimony phylogenetic tree for VEE virus prototype strains derived from branch-and-bound search in PAUP V3.1.1 (Swofford, 1993) using codon first and second position nucleotides from combined nonstructural (excluding the C-terminal nsP3 domain coding sequence) and structural regions, rooted with EEE virus strain 82V-2136 (Weaver et al., 1993) as the outgroup (7056 total characters, 1255 parsimony-informative characters). Percentages to the left of internal nodes indicate bootstrap support for 1000 pseudoreplicates (Felsenstein, 1985), using 10 random-addition heuristic searches per pseudoreplicate.

by MENA virus, which could form a stem lengthened by two base pairs (Fig. 2B). An equivalent calculated free energy for the MENA virus hairpin, due to the presence of a non-canonical G-U pair within the putative helix (Tinoco et al., 1973), may indicate that overall stability, rather than absolute stem length, determines possible nucleotide changes.

Secondary structure estimations using cross-species comparisons of homologous RNA molecules define a covariation that preserves pairing (such as a U-A pair mutating to a C-G pair) in a putative helical region as supportive of a stem, with two independent covariations taken as 'proof' of that stem (Pace et al., 1989). While this definition is more appropriate for purely structural RNA molecules, the alphavirus 51-nt CSE second stem is 'proven' and the VEE virus 20-nt palindrome stem supported by available sequences. In VEE virus, both CSEs are located in the only regions of the genome characterized by high concentrations of invariant wobble position nucleotides, and both may be part of more extensive secondary structures. The proposed involvement of the 51-nt CSE in a large 5' terminal secondary structure (Niesters and Strauss, 1990a) has been mentioned. The 20-nt palindrome is within the region (nt 735-1255) corresponding to the putative SIN packaging signal (Weiss et al., 1989). Experiments using SIN RNA transcribed from progressively truncated cDNA clones rule out the palindrome as the exclusive packaging signal (Weiss et al., 1989), but capsid binding may require a secondary structure which includes this palindrome.

The proposed VEE virus 5'NTR secondary structures are less stable than those of other alphaviruses (Ou et al., 1983), and are not supported by covariation of potential stem-forming pairs. Because it would be helix-disruptive, the G-to-A mutation at nt 3 contributing to attenuation of the TC-83 strain (Fig. 2) indirectly supports a putative VEE virus 5'NTR stem (Dubuisson et al., 1997). However, while the contribution of this mutation to TC-83 attenuation is well-established (Kinney et al., 1993), the contribution of this mutation to attenuation as a result of 5'NTR secondary structure disruption awaits

demonstration. As evidence from SIN 5'NTR and 51-nt CSE mutants indicates (Niesters and Strauss, 1990a,b), nucleotide changes can influence a secondary structure model and a viral genome disproportionately. Preservation of 'proximal' nucleotides in the putative SIN 5'NTR stem (Niesters and Strauss, 1990b), or preservation of the SIN 51-nt CSE linear nucleotide sequence over and above preservation of amino acid sequence and stem-loop free energy (Niesters and Strauss, 1990a), may be more important determinants of viral fitness than secondary structure preservation. For another member of the *Togaviridae*, preservation of secondary structure may not even be required. The rubella virus 51-nt CSE homolog conserves many alphavirus codon wobble position nucleotides and deduced amino acid residues despite lacking an apparent alphavirus-like secondary structure (Dominguez et al., 1990).

Comparative analysis is the foundation for much of our current understanding of the less tractable, cellularly less plentiful alphavirus non-structural proteins. Many of the experiments which helped define functions or functional residues for these proteins were prompted by or directed by sequence comparisons (Ahlgquist et al., 1985; Hardy and Strauss, 1989; Rozanov et al., 1992; Strauss et al., 1992; Wang et al., 1996; Ahola et al., 1997; Shirako and Strauss, 1998), and support for additional proposed nsP properties relies on shared identity with viral or nonviral proteins of known function (Gorbalenya et al., 1988; Rikkinen et al., 1994). Potential VEE virus epizootic residues have been identified in structural proteins by sequence comparison (Powers et al., 1997; Kinney et al., 1998; Oberste et al., 1998). Functional attributes of 'epizootic nsPs' must be defined for similar comparisons in the nonstructural region to have meaning (LaStarza et al., 1994a; Oberste et al., 1996).

The variety of alphavirus 5'NTR and nonstructural region mutations leading to attenuation for cell lines or laboratory animals (Niesters and Strauss, 1990b; Kuhn et al., 1992; Kinney et al., 1993; Rikkinen, 1996; Dryga et al., 1997) make it unwise to dismiss identified VEE virus nucleotide or amino acid differences as irrelevant in the

absence of functional studies. This is especially true for the 5'NTR, which is emerging as a major determinant of viral replication and pathogenesis (Kinney et al., 1993; Dubuisson et al., 1997).

On the other hand, the overall 77% amino acid sequence identity in the VEE virus polyprotein P1234 (excluding the C-terminal nsP3 domain) does not include conservative amino acid changes or changes near the ends of processed proteins, where structure is almost certainly less constrained. Engineered into a TRD virus genetic background, many enzootic strain nsP conservative amino acid substitutions would likely have negligible attenuating effect. An example of the probable lesser (or independently insufficient) role of naturally-occurring VEE virus nonstructural region mutations in attenuation or epizootic strain emergence is provided by the recent equine outbreak in Chiapas, Mexico.

Isolation of a 68U201-like subtype IE strain from this outbreak (Oberste et al., 1998) indicates the epizootic phenotype can be maintained over a range of nonstructural region sequences. More than 120 amino acid residues in the polyprotein P1234 sequence differ between epizootic IAB or IC strains and strain 68U201. Provided the Mexican VEE virus isolate is not a recombinant with a IAB or IC strain nonstructural region and a 68U201-like structural region [the near-identity of partial nsP3 coding sequences of the Mexican isolate and strain 68U201 (Oberste et al., 1998) make this unlikely], replicase modules at least 5% dissimilar are capable of equivalent equine virulence.

VEE virus sequence comparisons provide insight into the C-terminal nsP3 domain coding sequence and the polyprotein codon wobble position nucleotides, two of the least conserved 'regions' of the genome. In vitro, the alphavirus nsP3 tolerates C-terminal domain deletions, duplications, or linker insertions (Davis et al., 1989; LaStarza et al., 1994a) provided the reading frame is preserved. Given the plasticity of alphavirus C-terminal nsP3 domain size (Strauss et al., 1988; LaStarza et al., 1994a), the infrequency of deletion events and predominance of duplication events in VEE virus C-terminal nsP3 domains is noteworthy. A TRD virus C-terminal nsP3 do-

main direct repeat secondary structure has been proposed (Davis et al., 1989), and related structures can be drawn for the duplications found in enzootic VEE virus strains. The mechanism for the generation of duplications may be related to the sequences themselves, as G + C content analysis or secondary structure modeling of C-terminal nsP3 coding sequences stripped of direct repeats fail to reveal sequence qualities peculiar to this region that would favor polymerase slippage or template switching.

Conservation of certain motifs by most VEE virus strains and a high serine concentration by all strains (Fig. 3) suggests a C-terminal region function beyond that of separating proteins in the replication complex. The hypothesis that nonconserved portions of the C-terminal nsP3 domain may determine host protein interactions or vector competence (Oberste et al., 1996) is not disproved by the additional sequences reported here, since there are nonconserved portions of the C-terminal domain specific to each VEE virus subtype and each strain. However, such a model would have to explain the finding of widely divergent deduced C-terminal nsP3 domain sequences (e.g. CAB and 78V-3531 virus sequences) in strains with the same identified mosquito vector (Digoutte and Girault, 1976; Calisher et al., 1982), or of essentially the same sequence (e.g. P676 and EVE virus sequences) in strains with different vectors and hosts (Chamberlain et al., 1964; Mackenzie et al., 1976).

While the lack of nucleotide sequence conservation renders the C-terminal nsP3 coding region unreliable for RT-PCR diagnosis of VEE virus infection using nondegenerate primers (Brightwell et al., 1998), we had no difficulty amplifying the C-terminal regions of all strains using a single degenerate primer pair hybridizing to neighboring conserved regions. Because nucleotide sequences of the C-terminal nsP3 region distinguish between even closely-related strains, the diagnostic utility of this region is obvious.

Few non-methionine, non-tryptophan codon third position nucleotides are invariant in the VEE virus nonstructural region. Provided changes are silent and preserve RNA secondary structure, codon third position nucleotides may simply drift,

or may be selected for based on prevailing vector or host genome codon biases. The high wobble position G + C content in PIX virus, for example, could represent a quasispecies minority population founder effect or a specific adaptation to indigenous fauna.

The use of data sets uncorrected for superimposed changes at codon third positions as more divergent VEE virus sequences are added can produce misleading phylogenies. Even without this correction, the monophyly of VEE virus subtype I and subtype II strains (excluding subtype IF 78V-3531 virus) and the monophyly of subtype III strains are absolute, regardless of the region analyzed or algorithm used (Powers et al., 1997; Kinney et al., 1998). The phylogenetic clustering of EVE virus with subtype IAB, IC, and ID strains has been previously noted (Powers et al., 1997; Kinney et al., 1998), as has the clustering of 78V-3531 virus with subtype VI AG80-663 virus. A reclassification of VEE virus subtype IF and subtype II strains has been proposed (Kinney et al., 1998), to which the complete nucleotide sequences reported here provide further support.

Acknowledgements

We would like to thank Ryan J. Barrett and Kevin M. Myles for technical assistance. All synthetic oligonucleotide primers used in this study were supplied by the Biotechnology Core Facility Branch, Centers for Disease Control and Prevention, Atlanta, Georgia.

References

- Ahlquist, P., Strauss, E.G., Rice, C.M., Strauss, J.H., Haseloff, J., Zimmern, D., 1985. Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. *J. Virol.* 53, 536–542.
- Ahola, T., Laakkonen, P., Vihinen, H., Kaariainen, L., 1997. Critical residues of Semliki Forest virus RNA capping enzyme involved in methyltransferase and guanylyltransferase-like activities. *J. Virol.* 71, 392–397.
- Berge, T.O., Banks, I.S., Tigertt, W.D., 1961. Attenuation of Venezuelan equine encephalomyelitis virus by in vitro cultivation in guinea-pig heart cells. *Am. J. Hyg.* 73, 209–218.
- Brightwell, G., Brown, J.M., Coates, D.M., 1998. Genetic targets for the detection and identification of Venezuelan equine encephalitis viruses. *Arch. Virol.* 143, 731–742.
- Calisher, C.H., Kinney, R.M., de Souza Lopes, O., Trent, D.W., Monath, T.P., Francly, D.B., 1982. Identification of a new Venezuelan equine encephalitis virus from Brazil. *Am. J. Trop. Med. Hyg.* 31, 1260–1272.
- Calisher, C.H., Monath, T.P., Mitchell, C.J., Sabbatini, M.S., Cropp, C.B., Kerschner, J., Hunt, A.R., Lazuick, J.S., 1985. Arbovirus investigation in Argentina, 1977–1980. III. Identification and characterization of viruses isolated, including new subtypes of Western and Venezuelan equine encephalitis viruses and four new Bunyaviruses (Las Maloyas, Resistencia, Barranqueras, and Antequera). *Am. J. Trop. Med. Hyg.* 34, 956–965.
- Causey, O.R., Causey, C.E., Maroja, O.M., Macedo, D.G., 1961. The isolation of arthropod-borne viruses, including members of two hitherto undescribed serological groups, in the Amazon region of Brazil. *Am. J. Trop. Med. Hyg.* 10, 227–249.
- Chamberlain, R.W., Sudia, W.D., Coleman, P.H., Work, T.H., 1964. Venezuelan equine encephalitis virus from south Florida. *Science* 145, 272–274.
- Davis, N.L., Willis, L.V., Smith, J.F., Johnston, R.E., 1989. In vitro synthesis of infectious Venezuelan equine encephalitis virus RNA from a cDNA clone: Analysis of a viable deletion mutant. *Virology* 171, 189–204.
- de Groot, R.J., Hardy, W.R., Shirako, Y., Strauss, J.H., 1990. Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase: Evidence for temporal regulation of polyprotein processing in vivo. *EMBO J.* 9, 2631–2638.
- Devereux, J., Haeblerli, P., Smithies, O., 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* 12, 387–395.
- Dietz, W.H., Jr., Alvarez, O., Jr., Martin, D.H., Walton, T.E., Ackerman, L.J., Johnson, K.M., 1978. Enzootic and epizootic Venezuelan equine encephalomyelitis virus in horses infected by peripheral and intrathecal routes. *J. Infect. Dis.* 137, 227–237.
- Digoutte, J.-P., Girault, G., 1976. The protective properties in mice of Tonate virus and two strains of Cabassou virus against neurovirulent Everglades Venezuelan encephalitis virus. *Ann. Microbiol. (Institute Pasteur)* 127B, 429–437.
- Ding, M., Schlesinger, M.J., 1989. Evidence that Sindbis virus nsP2 is an autoprotease which processes the virus non-structural polyprotein. *Virology* 171, 280–284.
- Dominguez, G., Wang, C.-Y., Frey, T.K., 1990. Sequence of the genome RNA of rubella virus: Evidence for genetic rearrangement during Togavirus evolution. *Virology* 177, 225–238.
- Dubuisson, J., Lustig, S., Ruggli, N., Akov, Y., Rice, C.M., 1997. Genetic determinants of Sindbis virus neurovirulence. *J. Virol.* 71, 2636–2646.
- Dryga, S.A., Dryga, O.A., Schlesinger, S., 1997. Identification of mutations in a Sindbis virus variant able to establish persistent infection in BHK cells: The importance of a mutation in the nsP2 gene. *Virology* 228, 74–83.

- Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39, 783–791.
- France, J.K., Wyrick, B.C., Trent, D.W., 1979. Biochemical and antigenic comparisons of the envelope glycoproteins of Venezuelan equine encephalomyelitis strains. *J. Gen. Virol.* 44, 725–740.
- Galindo, P., Srihongse, S., de Rodaniche, E., Grayson, M.A., 1966. An ecological survey for arboviruses in Almirante, Panama, 1959–1962. *Am. J. Trop. Med. Hyg.* 15, 385–400.
- Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P., Koonin, E.V., 1989. An NTP-binding motif is the most conserved sequence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication. *J. Mol. Evol.* 28, 256–268.
- Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., Blinov, V.M., 1988. A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett.* 235, 16–24.
- Gorbalenya, A.E., Koonin, E.V., Lai, M.M.-C., 1991. Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, α - and coronaviruses. *FEBS Lett.* 288, 201–205.
- Grayson, M.A., Galindo, P., 1968. Epidemiologic studies of Venezuelan equine encephalitis virus in Almirante, Panama. *Am. J. Epidemiol.* 88, 80–96.
- Hardy, W.R., Strauss, J.H., 1989. Processing of the nonstructural polyproteins of Sindbis virus: Nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans. *J. Virol.* 63, 4653–4664.
- Hodgman, T.C., 1988. A new superfamily of replicative proteins. *Nature* 333, 22–23.
- Hunt, A.R., Johnson, A.J., Roehrig, J.T., 1990. Synthetic peptides of Venezuelan equine encephalomyelitis virus E2 glycoprotein. *Virology* 179, 701–711.
- Hunt, A.R., Roehrig, J.T., 1995. Localization of a protective epitope on a Venezuelan equine encephalomyelitis (VEE) virus peptide that protects mice from both epizootic and enzootic VEE virus challenge and is immunogenic in horses. *Vaccine* 13, 281–288.
- Hunt, A.R., Short, W.A., Johnson, A.J., Bolin, R.A., Roehrig, J.T., 1991. Synthetic peptides of the E2 glycoprotein of Venezuelan equine encephalomyelitis virus. II. Antibody to amino terminus protects animals by limiting viral replication. *Virology* 185, 281–290.
- Johnson, K.M., Martin, D.H., 1974. Venezuelan equine encephalitis. *Adv. Vet. Sc. Comp. Med.* 18, 79–116.
- Kamer, G., Argos, P., 1984. Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucl. Acids Res.* 12, 7269–7282.
- Karabatsos, N. (Ed.), 1985. International Catalogue of Arboviruses, Including Certain Other Viruses of Vertebrates, Third ed. The Subcommittee on Information Exchange of the American Committee on Arthropod-Borne Viruses, American Society of Tropical Medicine and Hygiene, San Antonio, Texas, USA.
- Kinney, R.M., Chang, G.-J., Tsuchiya, K.R., Sneider, J.M., Roehrig, J.T., Woodward, T.M., Trent, D.W., 1993. Attenuation of Venezuelan equine encephalitis virus strain TC-83 is encoded by the 5'-noncoding region and the E2 envelope glycoprotein. *J. Virol.* 67, 1269–1277.
- Kinney, R.M., Johnson, B.J.B., Welch, J.B., Tsuchiya, K.R., Trent, D.W., 1989. The full-length nucleotide sequences of the virulent Trinidad donkey strain of Venezuelan equine encephalitis virus and its attenuated vaccine derivative, strain TC-83. *Virology* 170, 19–30.
- Kinney, R.M., Pfeffer, M., Tsuchiya, K.R., Chang, G.-J.J., Roehrig, J.T., 1998. Nucleotide sequences of the 26S mRNAs of the viruses defining the Venezuelan equine encephalitis antigenic complex. *Am. J. Trop. Med. Hyg.* 59, 952–964.
- Kinney, R.M., Trent, D.W., France, J.K., 1983. Comparative immunological and biochemical analyses of viruses in the Venezuelan equine encephalitis complex. *J. Gen. Virol.* 64, 135–147.
- Kinney, R.M., Tsuchiya, K.R., Sneider, J.M., Trent, D.W., 1992a. Molecular evidence for the origin of the widespread Venezuelan equine encephalitis epizootic of 1969 to 1972. *J. Gen. Virol.* 73, 3301–3305.
- Kinney, R.M., Tsuchiya, K.R., Sneider, J.M., Trent, D.W., 1992b. Genetic evidence that epizootic Venezuelan equine encephalitis (VEE) viruses may have evolved from enzootic VEE subtype I-D virus. *Virology* 191, 569–580.
- Koonin, E.V., 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* 72, 2197–2206.
- Kozak, M., 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucl. Acids Res.* 9, 5233–5252.
- Kozak, M., 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucl. Acids Res.* 15, 8125–8148.
- Kuhn, R.J., Griffin, D.E., Zhang, H., Niesters, H.G.M., Strauss, J.H., 1992. Attenuation of Sindbis virus neurovirulence by using defined mutations in nontranslated regions of the genome RNA. *J. Virol.* 66, 7121–7127.
- Laakkonen, P., Hyvonen, M., Peranen, J., Kaariainen, L., 1994. Expression of Semliki Forest virus nsP1-specific methyltransferase in insect cells and in *Escherichia coli*. *J. Virol.* 68, 7418–7425.
- LaStarza, M.W., Grakoui, A., Rice, C.M., 1994a. Deletion and duplication mutations in the C-terminal nonconserved region of Sindbis virus nsP3: Effects on phosphorylation and on virus replication in vertebrate and invertebrate cells. *Virology* 202, 224–232.
- LaStarza, M.W., Lemm, J.A., Rice, C.M., 1994b. Genetic analysis of the nsP3 region of Sindbis virus: Evidence for roles in minus-strand and subgenomic RNA synthesis. *J. Virol.* 68, 5781–5791.
- Lemm, J.A., Bergqvist, A., Read, C.M., Rice, C.M., 1998. Template-dependent initiation of Sindbis virus RNA replication in vitro. *J. Virol.* 72, 6546–6553.

- Levinson, R., Strauss, J.H., Strauss, E.G., 1990. Complete sequence of the genomic RNA of O'Nyong-nyong virus and its use in the construction of alphavirus phylogenetic trees. *Virology* 175, 110–123.
- Mackenzie, R.M., de Siger, J., Parra, D.V., 1976. Venezuelan equine encephalitis virus: Comparison of infectivity and virulence of strains V-38 and P676 in donkeys. *Am. J. Trop. Med. Hyg.* 25, 494–499.
- Mi, S., Durbin, R., Huang, H.V., Rice, C.M., Stollar, V., 1989. Association of the Sindbis virus RNA methyltransferase activity with the nonstructural protein nsP1. *Virology* 170, 385–391.
- Mi, S., Stollar, V., 1991. Expression of Sindbis virus nsP1 methyltransferase activity in *Escherichia coli*. *Virology* 184, 423–427.
- Morilla-Gonzalez, A., Mucha-Macias, J., 1969. Estudio de una epizootia de encefalitis equina de Venezuela ocurrida en Tamaulipas. *Mex. Rev. Invest. Salud Publica. (Mex.)* 29, 3–20.
- Niesters, H.G.M., Strauss, J.H., 1990a. Mutagenesis of the conserved 51 nucleotide region of Sindbis virus. *J. Virol.* 64, 1639–1647.
- Niesters, H.G.M., Strauss, J.H., 1990b. Defined mutations in the 5' nontranslated sequence of Sindbis virus RNA. *J. Virol.* 64, 4162–4168.
- Noller, H.F., Woese, C.R., 1981. Secondary structure of 16S ribosomal RNA. *Science* 212, 403–411.
- Oberste, M.S., Fraire, M., Navarro, R., Zepeda, C., Zarate, M.L., Ludwig, G.V., Kondig, J.F., et al., 1998. Association of Venezuelan equine encephalitis virus subtype IE with two equine epizootics in Mexico. *Am. J. Trop. Med. Hyg.* 59, 100–107.
- Oberste, M.S., Parker, M.D., Smith, J.F., 1996. Complete sequence of Venezuelan equine encephalitis virus subtype IE reveals conserved and hypervariable domains within the C terminus of nsP3. *Virology* 219, 314–320.
- Ou, J.-H., Strauss, E.G., Strauss, J.H., 1983. The 5'-terminal sequences of the genomic RNAs of several alphaviruses. *J. Mol. Biol.* 168, 1–15.
- Pace, N.R., Smith, D.K., Olsen, G.J., James, B.D., 1989. Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA — A review. *Gene* 82, 65–75.
- Pardigon, N., Strauss, J.H., 1996. Mosquito homolog of the La autoantigen binds to Sindbis virus RNA. *J. Virol.* 70, 1173–1181.
- Peranen, J., Laakkonen, P., Hyvonen, M., Kaariainen, L., 1995. The alphavirus replicase protein nsP1 is membrane-associated and has affinity to endocytic organelles. *Virology* 208, 610–620.
- Pfeffer, M., Kinney, R.M., Kaaden, O.-R., 1998. The alphavirus 3'-nontranslated region: size heterogeneity and arrangement of repeated sequence elements. *Virology* 240, 100–108.
- Pfeffer, M., Proebster, B., Kinney, R.M., Kaaden, O.-R., 1997. Genus-specific detection of alphaviruses by a semi-nested reverse transcription-polymerase chain reaction. *Am. J. Trop. Med. Hyg.* 57, 709–718.
- Powers, A.M., Oberste, M.S., Brault, A.C., Rico-Hesse, R., Schmura, S.M., Smith, J.F., Kang, W., Sweeney, W.P., Weaver, S.C., 1997. Repeated emergence of epidemic/epizootic Venezuelan equine encephalitis from a single genotype of enzootic subtype ID virus. *J. Virol.* 71, 6697–6705.
- Rico-Hesse, R., Weaver, S.C., de Siger, J., Medina, G., Salas, R.A., 1995. Emergence of a new epidemic/epizootic Venezuelan equine encephalitis virus in South America. *Proc. Natl. Acad. Sci. USA* 92, 5278–5281.
- Rikkinen, M., 1996. Functional significance of the nuclear-targeting and NTP-binding motifs of Semliki Forest virus nonstructural protein nsP2. *Virology* 218, 352–361.
- Rikkinen, M., Peranen, J., Kaariainen, L., 1994. ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2. *J. Virol.* 68, 5804–5810.
- Roehrig, J.T., Bolin, R.A., Hunt, A.R., Woodward, T.M., 1991. Use of a new synthetic peptide-derived monoclonal antibody to differentiate between vaccine and wild-type Venezuelan equine encephalomyelitis viruses. *J. Clin. Microbiol.* 29, 630–631.
- Roehrig, J.T., Mathews, J.H., 1985. The neutralization site on the E2 glycoprotein of Venezuelan equine encephalomyelitis (TC-83) virus is composed of multiple conformationally stable epitopes. *Virology* 192, 347–356.
- Rozanov, M.N., Koonin, E.V., Gorbalenya, A.E., 1992. Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. *J. Gen. Virol.* 73, 2129–2134.
- Scherer, W.F., Anderson, K., 1975. Antigenic and biologic characteristics of Venezuelan encephalitis virus strains including a possible new subtype, isolated from the Amazon region of Peru in 1971. *Am. J. Epidemiol.* 101, 356–361.
- Shirako, Y., Strauss, J.H., 1998. Requirement for an aromatic amino acid or histidine at the N terminus of Sindbis virus RNA polymerase. *J. Virol.* 72, 2310–2315.
- Shope, R.E., Causey, O.R., de Andrade, A.H.P., Theiler, M., 1964. The Venezuelan equine encephalomyelitis complex of group A arthropod-borne viruses, including Mucambo and Pixuna from the Amazon region of Brazil. *Am. J. Trop. Med. Hyg.* 13, 723–727.
- Strauss, E.G., de Groot, R.J., Levinson, R., Strauss, J.H., 1992. Identification of the active site residues in the nsP2 proteinase of Sindbis virus. *Virology* 191, 932–940.
- Strauss, E.G., Levinson, R., Rice, C.M., Dalrymple, J., Strauss, J.H., 1988. Nonstructural proteins nsP3 and nsP4 of Ross River and O'Nyong-nyong viruses: Sequence and comparison with those of other alphaviruses. *Virology* 164, 265–274.
- Strauss, E.G., Rice, C.M., Strauss, J.H., 1984. Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* 133, 92–110.
- Strauss, E.G., Strauss, J.H., 1986. Structure and replication of the alphavirus genome. In: Schlesinger, S., Schlesinger, M.J. (Eds.), *The Togaviridae and Flaviviridae*. Plenum Publishing Corporation, New York, pp. 35–90.
- Strauss, J.H., Strauss, E.G., 1994. The alphaviruses: Gene expression, replication, and evolution. *Microbiol. Rev.* 58, 491–562.

- Sudia, W.D., Newhouse, V.F., 1975. Epidemic Venezuelan equine encephalitis in North America: A summary of virus-vector–host relationships. *Am. J. Epidemiol.* 101, 1–13.
- Swofford, D.L., 1993. PAUP: Phylogenetic analysis using parsimony, version 3.1.1. Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.
- Tinoco, I., Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M., Gralla, J., 1973. Improved estimation of secondary structure in ribonucleic acids. *Nature New Biol.* 246, 40–41.
- Walton, T.E., Grayson, M.A., 1988. Venezuelan equine encephalomyelitis. In: Monath, T.P. (Ed.), *The arboviruses: Epidemiology and Ecology*, vol. IV. CRC Press, Boca Raton, Florida, pp. 203–233.
- Wang, H.-L., O’Rear, J., Stollar, V., 1996. Mutagenesis of the Sindbis virus nsP1 protein: effects on methyltransferase activity and viral infectivity. *Virology* 217, 527–531.
- Weaver, S.C., Bellew, L.A., Rico-Hesse, R., 1992. Phylogenetic analysis of alphaviruses in the Venezuelan equine encephalitis complex and identification of the source of epizootic viruses. *Virology* 191, 282–290.
- Weaver, S.C., Hagenbaugh, A., Bellew, L.A., et al., 1993. A comparison of the nucleotide sequences of eastern and western equine encephalomyelitis viruses with those of other alphaviruses and related RNA viruses. *Virology* 197, 375–390.
- Weaver, S.C., Salas, R., Rico-Hesse, R., Ludwig, G.B., Oberste, M.S., Boshell, J., Tesh, R.B., et al., 1996. Re-emergence of epidemic Venezuelan equine encephalomyelitis in South America. *Lancet* 348, 436–440.
- Weiss, B., Nitschko, H., Ghattas, I., Wright, R., Schlesinger, S., 1989. Evidence for specificity in the encapsidation of Sindbis virus RNAs. *J. Virol.* 63, 5310–5318.
- Young, N.A., Johnson, K.M., 1969. Antigenic variants of Venezuelan equine encephalitis virus: Their geographic distribution and epidemiologic significance. *Am. J. Epidemiol.* 89, 286–307.