Venezuelan equine encephalitis (VEE) viruses are mosquito-borne members of the genus Alphavirus, family Togaviridae. The viral genome is a positive-sense, 5′-capped, 3′-polyadenylated RNA genome of about 11.5 kilobases. Four nonstructural proteins are encoded in the 5′ two-thirds of the genome. The structural proteins are processed from a precursor polyprotein (NH₂-capsid-E3-E2-6K-E1-COOH) that is translated from an intracellular, subgenomic 26S mRNA molecule, which is itself transcribed from the 3′ one-third portion of the genomic mRNA. The mature virus is composed of an icosahedral nucleocapsid that is surrounded by a lipid envelope containing the viral E1 and E2 glycoproteins.1

Sporadic epizootics and epidemics of VEE have occurred among equines and humans in South America. The most widespread VEE epizootic-epidemic occurred between 1969 and 1972 in South and Central America and Texas in the United States.2,3 No other VEE epizootics occurred for many years following this outbreak. However, a small epizootic-epidemic occurred in 1992–1993 in Venezuela.4 More recently, the large VEE epizootic-epidemic in more than 20 years occurred in 1995 in the La Guajira region of Venezuela and Colombia. This outbreak caused an estimated 75,000 human cases and 50,000 equine cases,5,6

The viruses of the VEE antigenic complex are currently classified into six antigenic subtypes by cross-neutralization (N) and hemagglutination-inhibition (HI) tests, with subtypes I and III divided into five and three varieties, respectively.7–13 All epizootic equine-virulent strains of VEE virus have been identified as subtype-variety IAB or IC.7,14 The most enigmatic facets of the natural history of VEE viruses are the emergence of epizootic VEE viruses in settings that are ecologically distinct from those of enzootic VEE strains and the failure to isolate IAB or IC viruses during interepizootic periods.14,15

The recent epizootic-epidemic in the Guajira Peninsula, which has again raised questions concerning the origin of epizootic VEE virus, was associated with a prolonged rainy season and a large population of unvaccinated equines.6 Among the hypotheses for the origin of epizootic VEE viruses,7,14 two seem most likely: 1) use of inadequately for-malin-inactivated vaccines using epizootic VEE IAB virus8,17 and 2) evolution from closely related enzootic VEE strains, particularly subtype-variety ID strains.5,18–20

Nucleotide sequences have been determined for the genomes of VEE Trinidad donkey (TRD, IAB),21 71-180 (IAB),17 P676 (IC),19 3880 (ID),19 and 68U201 (IE)22 viruses, as well as the 26S mRNAs of Everglades Fe3-7c (II), and AG80-663 (VI) viruses contradicted previous serologic results. Mutations near the amino terminus of the E2 envelope proteins of Pixuna and AG80-663 viruses probably account for the previously reported low reactivity of the protective monoclonal antibody IA2B-10 with these two viruses. Variations in the distribution of potential glycosylation sites in the E2 glycoprotein are discussed.

**Materials and Methods**

**Viruses.** The virus strains analyzed in this study and their antigenic subtypes or subtype-varieties, places of origin, dates of isolation, references, and Genbank accession numbers for nucleotide sequence data are shown in Table 1. Abbreviations are used in the text for the names of Trinidad donkey (TRD), Mena II (MENA), Everglades (EVE), Mucambo (MUC), Tonate (TON), Pixuna (PIX), and Cabassou (CAB) viruses.35 All viruses, except for 68U201 virus, were previously cloned by picking isolated plaques in Vero or Pekin duck embryo cell monolayers and were used at 12 or fewer passages (including cloning steps) from the original isolate.36 The final passage was in baby hamster kidney-21 cells. These plaque-purified viruses have been used previously to confirm and extend the original antigenic classification of VEE viruses by Young and Johnson,3 and their virulence phenotypes in mice have been determined.9–10,36–38 The 26S mRNA sequence of 68U201 virus is from Oberste and others.22

**Extraction of viral RNA and the reverse transcriptase–polymerase chain reaction (RT-PCR).** Genomic viral RNA was extracted directly from 0.2 ml of virus seed as previously described.39 The precipitated RNA was dissolved in 50
µl of deionized water and stored frozen. The single-tube, extended RT-PCR was performed in a volume of 100 µl in a thin-walled 0.2-ml microtube containing 5 µl of the RNA template, 1 µM each of the upstream, positive, RNA-sense amplimer and downstream, complementary amplimer, 200 µM of each deoxynucleoside triphosphate, 1.8 units of reverse transcriptase (Rous-associated virus-2; Amersham Pharmacia Biotech, Piscataway, NJ), 1.25 units of Taq DNA polymerase (Amplitaq; Perkin-Elmer Corp., Norwalk, CT), and 2.5 units of Taq Extender PCR Additive in Taq Extender buffer (Stratagene Cloning Systems, La Jolla, CA). The RT-PCR was performed in a Model 9600 thermocycler (Perkin-Elmer Corp.) by incubating at 50 °C for 2 min, and then cooling to 18 °C for 1 min, 72 °C for 4 min, and 72 °C for 3 min (with an increase in extension time of 10 sec each cycle), a final extension at 72 °C for 20 min, and then cooling to 18 °C.

Degenerate amplimers were designed to prime cDNA synthesis at two regions of high amino acid conservation in alphaviruses and at the 3’ poly(A) tail (Table 2). Published sequence data indicated that the FKFGAMMKSGMF (amino acid positions nonstructural protein 4 [nsP4]-421-432, using the sequence of the TRD viral genome as reference), WHHGAVQY (capsid-202-209), and MWGGAYCFCD (E1-terminus) C residue and the immediately following 14 A residues of the poly(A) tail of the alphavirus genome.

**Purification of amplicon cDNA and sequencing.** Amplified cDNA was purified by electrophoresis in 0.8% (w/v) agarose gels (Agarose 3:1 High Resolution Blend; Amresco, Solon, OH) in TAE buffer (400 mM Tris, pH 8.3, 200 mM acetic acid, 1 mM disodium EDTA). The DNA bands were stained briefly with ethidium bromide, visualized under UV light, excised from the gel, and extracted from the gel slice by using the GeneClean II kit as recommended by the manufacturer (Bio 101, Inc., La Jolla, CA). Gel-purified amplicon cDNA was sequenced directly without any cloning steps with the Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer Corp.). Sequencing reactions contained 9 µl of template cDNA, 3 µl of 10 µM sequencing primer, and 8 µl of Ready Reaction mixture in a volume of 20 µl. Sequencing was performed by incubation for 35 cycles at 96°C for 15 sec, 50°C for 15 sec, and 60°C for 4 min. The DNA was purified using Centri-sep columns as recommended by the manufacturer (Princeton Separations, Inc., Adelphi, NJ) and analyzed with a Model 373A or 377 automated DNA sequencer (Applied Biosystems, Foster City, CA). Both strands of the cDNA were sequenced.

**Phylogenetic relationships.** Phylogenetic trees were constructed from the aligned nucleotide and amino acid sequences of the 26S translated polyprotein of the VEE virus strains and Sindbis virus by using the PHYLIP software program (Phylogeny Inference Package, version 3.5c, distributed by J. Felsenstein, Department of Genetics, University of Washington, Seattle, WA).50,51 The Sindbis sequence served as the outgroup for the analysis. Statistical analysis of 100 reiterated data sets was performed by the bootstrap algorithm. The data sets were analyzed for phylogenetic relationships by the parsimony method, using the DNAPARS (nucleotide sequence) or PROTPARS (amino acid sequence) algorithm and the CONSENSE algorithm. Trees were plotted with the DRAWGRAM algorithm.

**RESULTS**

**Amplification by RT-PCR and sequencing of cDNA.** The carboxy-terminal 561 nucleotide residues of the gene encoding nsP4 and all of the 26S mRNA region of the VEE virus genomes were amplified with degenerate primers by RT-PCR as three amplicons designated as Region A (1227 basepairs [bp], amplimers ALPHA-6997/cALPHA-8165), Region B (2129 bp, ALPHA-8187/cALPHA-10262), and Region C (approximately 1300 bp, ALPHA-10290/
Table 2
Amplimers/primers used to amplify and sequence Venezuelan equine encephalitis (VEE) cDNA

<table>
<thead>
<tr>
<th>Amplimer*</th>
<th>Sequence (5' -&gt; 3')†</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA-6160</td>
<td>GAATAKGCTATCTTATGTG</td>
</tr>
<tr>
<td>ALPHA-6257</td>
<td>GAAATGCGATCATCCTC</td>
</tr>
<tr>
<td>cALPHA-6165</td>
<td>RTATATGATCTGCCACGT</td>
</tr>
<tr>
<td>ALPHA-6257</td>
<td>GCGATCTCCAGAAATGATT</td>
</tr>
<tr>
<td>cALPHA-10262</td>
<td>RTCCGATACCCATCCTCCCTC</td>
</tr>
<tr>
<td>ALPHA-10290</td>
<td>ATCCTGGGAGCAAATTTTTGG</td>
</tr>
<tr>
<td>cG14T-KEP</td>
<td>CAGAGTGCGGAATGAACAGATGAGG</td>
</tr>
</tbody>
</table>

The standard ambiguity conventions used to designate degenerate nucleotide residues in amplimers/primers are R = A/G; Y = C/T; N = A/C/G/T; W = A/T; K = G/T; M = A/C; B = C/G/T; D = A/G/T; H = A/C/T; V = A/C/G; S = G/C; Y = A/C/G.

* c = negative-sense primer; ALPHA = degenerate amplimers designed to amplify cDNA from viruses of the Alphavirus genus; CPX = degenerate primers designed to amplify cDNA from viruses of the VEE complex; numerical designation indicates the genome nucleotide position of the 3' noncoding regions of the viruses. Amplimers were purified by preparative agarose gel electrophoresis and then sequenced. To directly determine the virus-specific nucleotide sequences at the terminal priming sites of Regions A, B, and C, amplimers were derived for each virus by using primer pairs ALPHA-6160/cVEE-7077.CPX, VEE-7698.CPX/cVEE-8506.CPX, and VEE-10035.CPX/cVEE-10541.CPX and then sequenced in the appropriate region (Table 2).

The 26S junction region and 3' nontranslated region. The genes encoding the nonstructural proteins end and those encoding the structural proteins begin at the 26S junction region of the alphavirus genomic RNA. This region is shown for the VEE viruses in Figure 1A. The 3'-terminus of the nsP4 gene, the putative start site of the 26S mRNA, and the AUG initiation codon of the capsid gene are indicated. The two termination codons that are in frame with the nonstructural polyprotein genes are indicated (# in Figure 1A). The underlined sequence indicates the 24-nucleotide conserved element that has been precisely defined by deletion mapping in Sindbis virus as the promoter for transcription of the subgenomic 26S mRNA in alphavirus-infected cells. Nucleotide variation was evident at positions 1, 2, and 13 in the 24-nucleotide conserved sequence element that has been precisely defined by deletion mapping in Sindbis virus as the promoter for transcription of the subgenomic 26S mRNA in alphavirus-infected cells. Variations in the 26S promoter and its surrounding sequences in different alphaviruses may regulate levels of 26S mRNA synthesis or permissive host cell types. The 3'-nontranslated regions (3' NTRs) of alphavirus genomes are highly variable in length. However, the 19 nucleotides immediately preceding the poly(A) tail are highly

cG14T.KEP (Table 2). Amplimers of expected size were obtained from all VEE genomic RNA templates. Since the stringency of amplification was low, non-target bands were also evident in many RT-PCRs. Size heterogeneity of Region C amplimers indicated variability in the nucleotide lengths of the 3' noncoding regions of the viruses. Amplimers were purified by preparative agarose gel electrophoresis and then sequenced. To directly determine the virus-specific nucleotide sequences at the terminal priming sites of Regions A, B, and C, amplimers were derived for each virus by using primer pairs ALPHA-6160/cVEE-7077.CPX, VEE-7698.CPX/cVEE-8506.CPX, and VEE-10035.CPX/cVEE-10541.CPX and then sequenced in the appropriate region (Table 2).

Both strands of each RT-PCR-derived amplicon were sequenced directly. Thus, the weighted sequence of each population of amplified cDNA molecules was determined. The amino acid sequence FKGFMKSGMF (forward priming site for Region A) at nsP4-421-432 was conserved in all but MENA (and 68U201) and AG80-663 viruses, both of which contained a Lys-to-Arg substitution in their FKGFMKSGMF priming site. Since the mismatch occurred at the 5'-end of the long degenerate primer, RT-PCR amplification was not adversely affected. The capsid WHHGAVQY (capsid-202-209, priming site for Region A reverse primer and Region B forward primer) and E1MWGGAYCFCD (E1-88-97, priming site for Region B reverse primer and Region C forward primer) amino acid sequences were conserved in all of the VEE viruses sequenced. However, the nucleotide sequences at these amplimer sites were not conserved. In an earlier study, we did not confirm 26S RNA sequence differences between TRD and MENA viruses on an independent cDNA clone of MENA virus. We repeated the sequence analysis here and detected a single nucleotide error: MENA nucleotide 8813 (using TRD viral genome as reference) was found to be G, not C, and encoded amino acid E2-84-Val instead of Leu.

The 26S junction region and 3' nontranslated region. The genes encoding the nonstructural proteins end and those encoding the structural proteins begin at the 26S junction region of the alphavirus genomic RNA. This region is shown for the VEE viruses in Figure 1A. The 3'-terminus of the nsP4 gene, the putative start site of the 26S mRNA, and the AUG initiation codon of the capsid gene are indicated. The two termination codons that are in frame with the nonstructural polyprotein genes are indicated (# in Figure 1A). The underlined sequence indicates the 24-nucleotide conserved element that has been precisely defined by deletion mapping in Sindbis virus as the promoter for transcription of the subgenomic 26S mRNA in alphavirus-infected cells. Nucleotide variation was evident at positions 1, 2, and 13 in the 24-nucleotide conserved sequence element that has been precisely defined by deletion mapping in Sindbis virus as the promoter for transcription of the subgenomic 26S mRNA in alphavirus-infected cells. Variations in the 26S promoter and its surrounding sequences in different alphaviruses may regulate levels of 26S mRNA synthesis or permissive host cell types. The 3'-nontranslated regions (3' NTRs) of alphavirus genomes are highly variable in length. However, the 19 nucleotides immediately preceding the poly(A) tail are highly
conserved and function as a promoter to initiate synthesis of the negative-sense genomic RNA during replication.\(^5\)\(^4\) The 3' NTRs of the VEE viruses varied from 77 to 195 nucleotides (including the termination codon for the E1 protein) in length (Figure 1B), which correlated well with the size variability of Region C amplicons evident in agarose gels. The 3'-terminal nucleotide sequence immediately preceding the poly(A)-tail of each of the VEE viruses was 5'-ATTTTTGTTTT-TAAATTTTCC-3'. The VEE viruses contain the shortest 3' NTRs of the alphaviruses sequenced to date,\(^5\)\(^9\) and PIX virus contained the shortest 3' NTR (77 nucleotides) of the VEE viruses. The 3' NTRs of the VEE viruses contained 1-4 common repeat sequence elements, which differed in composition from repeat sequence elements identified in the 3' NTRs of other alphaviruses.\(^1\)\(^5\)\(^5\)

**Genetic relationships among viruses of the VEE complex.** The aligned, deduced amino acid sequences of the precursor polyproteins translated from the 26S mRNAs of the VEE viruses are shown in Figure 2. The percent identities among the nucleotide (lower left) and amino acid (upper right) sequences of the translated VEE 26S domains are shown in Table 3. Boxes are drawn around the closely related VEE subtype I and II viruses, the subtype III viruses, and the more distantly related 78V-3531 (IF) and AG80-663 (VI) viruses (Table 3). When the relationships among the genetically nearly identical IAB strains (TRD and its vaccine derivative TC-83, and 71-180) were excluded, percent identities among the polyproteins of the VEE viruses varied from a low of 77.5% for PIX (subtype IV) versus AG80-663 (VI) virus and CAB (V) versus 78V-3531 (IF) virus to a high of 98.4% for P676 (IC) versus 3880 (ID) virus (Table 3). The 26S mRNAs of the two subtype-variety IE viruses, Mena II\(^2\)\(^3\) and 68U201\(^1\)\(^2\), had an overall amino acid identity of 98%. The 26S polyprotein of Sindbis virus showed only 45.2-46.1% identity with the polyprotein sequences of the VEE viruses. The nucleotide sequences of the coding region of the 26S mRNAs of the VEE viruses differed to a greater extent because of the degeneracy of the genetic code. Nucleotide sequence identities ranged from a low of 69.6% for 78V-3531 versus P676 to a high of 96.5% for P676 versus TRD virus. As observed by Powers and others,\(^2\) the genetic relationships among the VEE viruses corresponded well with previous classification by serologic or other molecular methods, except for EVE (II) and 78V-3531 (IF) viruses. The EVE virus was closely related to subtype I viruses. The nucleotide and translated amino acid sequences of each of the individual capsid, E3, E2, 6K, and E1 genes of EVE virus were more closely related to those of the cognate genes of TRD, P676, and 3880 viruses than were the cognate genes of MENA virus. The PIX, CAB, 78V-3531, and AG80-663 viruses were distinct, although the latter two viruses were more similar to each other than to the other VEE viruses analyzed.

The aligned nucleotide and amino acid sequences of the 26S mRNA region of the genome encoding the entire polyproteins of the VEE viruses and Sindbis virus,\(^4\) which served as the outgroup virus, were analyzed with the PHYLP program.\(^5\)\(^0\)\(^5\)\(^1\) The IAB viruses and P676 (IC), 3880 (ID) and EVE (II) viruses formed a highly consistent (all 100 bootstrap iterations) monophyletic group when comparing aligned nucleotide or deduced amino acid data for the 3'-terminal 561 nucleotide residues of nsP4, the entire 26S translated region (amino acid phylogenetic tree shown in Figure 3), and the individual E1 and E2 genes, as well as for the nucleotide data for the capsid gene. The 26S and E1 amino acid sequences of MENA (IE) virus also formed a consistent monophyletic group with the preceding viruses. The 78V-3531 virus, which is currently classified as VEE subtype-variety IF, did not group with the subtype I viruses by this analysis. Instead, it consistently (100 iterations) grouped with AG80-663 virus after analysis of the nucleotide or deduced amino acid sequence of the translated 26S

---

**Figure 1.** A. Alignment of the nucleotide sequences at the 26S junction regions of viruses of the Venezuelan equine encephalitis (VEE) antigenic complex. The 3' terminus of the nonstructural protein 4 (nsP4) gene, the putative start site of the 26S mRNA, and the AUG initiation codon of the capsid gene are indicated. The underlined sequence indicates the 24-nucleotide conserved domain that functions as promoter for the transcription of the subgenomic 26S mRNA in alphavirus-infected cells.\(^5\)\(^2\)\(^5\) # = termination codon in frame with the nsP4 gene. Dots indicate sequence identity with that of TRD virus. The 26S junction of PIX virus was three nucleotides shorter (indicated by dashes) than those of the other viruses. B. Nucleotide lengths, including the termination codon for the structural genes and excluding the poly(A) tail, of the 3' nontranslated regions of VEE viruses. Virus abbreviations: TRD = Trinidad donkey; EVE = Everglades Fe3-7c; MENA = Mena II; MUC = Mucambo BeAn 8; TON = Tonate CaAn 410d; PIX = BeAr 35645; CAB = Cabassou CaAr 508.
region or the amino acid sequence of the E1 gene, as well as the amino acid sequences of the capsid and E2 genes (98 of 100 iterations). The 78V-3531 virus grouped with AG80-663 virus for 87–93 of 100 iterations following phylogenetic analysis of the nucleotide data for the individual capsid, E1, and E2 genes. These two viruses were both unique by analysis of the nucleotide or deduced amino acid sequence of the 3'-terminal 561 nucleotides of the nsP4 gene. The subtype III viruses (MUC, TON, 71D-1252) formed a monophyletic group in 96–100 of 100 trees calculated for the nucleotide and amino acid data of the 26S translated region and the individual capsid and E2 genes, as well as the amino acid data of E1. This grouping was less consistent (77–88 of 100 iterations) by analyses of the E1 nucleotide sequences and the 3'-terminal region of the nsP4 gene. The PIX (IV) and CAB (V) viruses were unique and did not group with any other VEE viruses studied (Figure 3).

**Capsid protein.** The capsid proteins of EVE, MENA (and 68U201)22, 78V-3531, and AG80-663 viruses contained 274, 274, 279, and 278 amino acid residues, respectively; those of the remaining VEE viruses were 275 residues in length. Amino acid positions 56-109 were highly variable (Figure 2), and this variable domain accounted for the majority of amino acid sequence differences among the capsid proteins of these viruses. The variable, positively charged amino acids were highly conserved sequences of the amino-terminal portion of the capsid proteins of alphaviruses interact with the genomic RNA molecule, while the highly conserved carboxyl portion of the protein probably functions in protein-protein interactions during virus maturation and assembly.1 Conserved residues His-156, Asp-178, and Ser-230 (positions include gaps introduced into the aligned sequences in Figure 2) in the capsid protein of the VEE viruses correspond to the autolytic triad of the capsid protease identified for Sindbis virus.1 The highly conserved VEE virus sequence KPGKRQRMVKL齐鲁 at capsid positions 115-129 (including gaps) is nearly identical to the alphavirus-conserved sequence that has been shown to bind to ribosomes and to lie within a region that binds genomic RNA.1,56 The conserved WHHGAVQY locus at capsid-206-213 (including gaps) served as an amplimer binding site for degenerate amplimers designed in this study.

**Envelope E2 glycoprotein.** The precursor protein PE2 is cleaved by a cellular protease after the conserved dibasic cleavage site R-X-(K/R)-R, where X = any amino acid, in alphaviruses to produce E3 and E2.1 This cleavage motif was conserved in all of the VEE viruses, except for 71D-1252 virus, which contained a conservative Lys substitution at the first position of this motif (Figure 2). The potential site of Asn-linked glycosylation (underlined and identified with an asterisk in Figure 2) in E3, which is conserved in all alphaviruses sequenced so far, was present in the E3 protein of all of the VEE viruses (underlined N-X-S/T motifs in Figure 2, where X = any of a number of amino acid residues). The IAB viruses TRD, TC-83, and 71-180 contained three sites of potential Asn-linked glycosylation at E2 positions 212, 291, and 318. Glycosylation of the TRD and TC-83 E2 proteins has been demonstrated.57 The EVE virus was the only other VEE virus that contained the 212 site, but it lacked the 291 site as did PIX, 78V-3531, and AG80-663 viruses. The E2-318 glycosylation site was conserved in all of the VEE viruses and may play an important role in the biology of VEE virus. Viruses MUC, TON, and 71D-1252 all contained a subtype III-specific potential glycosylation site at E2-305 that was not shared by any of the other VEE strains. Differences in the degree of glycosylation probably account for the variable electrophoretic mobilities of the E2 proteins of VEE viruses whose structural proteins have been analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE).5,9

The E2 region containing the most amino acid sequence variation occurred between amino acid positions 178-224. This E2 region has been shown to be a major N domain for VEE56,59 and Sindbis1,20,62 viruses. The E2-204 Gly residue was substituted with Ser in 78V-3531 and AG80-663 viruses; this position was deleted in MUC, TON, 71D-1252, PIX, and CAB viruses (Figure 2). Sequence differences in this domain may affect monoclonal antibody (MAb) reactivities, virus neutralization, or host cell binding of VEE viruses.13,63

**Envelope E1 glycoprotein.** The single site of potential glycosylation at amino acid position E1-134 was conserved in all of the VEE viruses. The VEE virus E1 protein has been shown to be glycosylated,57 and the E1 glycoproteins of VEE virus strains have nearly identical apparent molecular weights by SDS-PAGE analysis.8,9 The 17-amino acid VFTGVYPFMWGGAYCFC putative fusion domain at E1-80-96 (stippled bar in Figure 2) is highly conserved among alphaviruses and is thought to be involved in fusion of the viral envelope with cellular membranes to release the nucleocapsid into the cytoplasm of the infected cell.1,40,64 Only a single conservative amino acid substitution occurred at E1-82 in this domain of EVE, MENA, and 68U201 viruses (Figure 2). The MWGGAYCFCF portion of this domain served as an amplimer binding site for degenerate amplimers designed in this study. Cytosolic Cys residues located at the carboxyl terminus of the E2 and E1 glycoproteins are usually present in alphaviruses and serve as sites for acetylation of the protein.1 Although numerous conserved Cys residues were present at the carboxyl terminus of the E2 protein of the VEE viruses (Figure 2), the carboxyl terminus of the E1 protein of these viruses, like that of eastern equine encephalitis virus, lacked Cys residues and is probably not palmitoylated.1

**DISCUSSION**

This study reports new sequence information for VEE 78V-3531, AG80-663, MUC, TON, 71D-1252, PIX, and CAB viruses. Genetic relationships among 14 members of the VEE antigenic complex, including all of the prototype viruses for the various subtypes and varieties of VEE virus, were analyzed for the carboxyl-terminal 561 nucleotides of nsP4 and the entire 26S mRNA region of the genome. The phylogenetic tree, showing relationships among these viruses based on the amino acid sequences of the entire translated region of the 26S RNA region of the genome (Figure 3), was very similar to the E3/E2 sequence-based tree reported recently.20 Although detailed genetic analyses have largely supported the current classification scheme of the VEE serocomplex, which was established by Young and Johnson1 and expanded by other investigators,6–11 some modification of the scheme is warranted. The VEE subtype II virus (EVE) is as closely related to subtype-variety IAB and IC viruses as is
Figure 2. Alignment of the deduced amino acid sequences of the polyproteins translated from the 26S mRNAs of Venezuelan equine encephalitis (VEE) virus strains. Sequence gaps (−) were introduced to optimize the alignment. Solid dots indicate amino acid sequence identity with VEE TRD virus. Underlining and asterisks (*) indicate potential glycosylation sites (N-X-S or T, where X is any amino acid) in the E3, E2, and E1 glycoproteins. Solid bars over E2, 6K, and E1 domains indicate transmembrane domains. The stippled bar in E1 indicates the putative conserved fusion domain of alphaviruses. The 26S amino acid sequences have been published previously for TRD, TC83, P676, 3880, EVE, MENA, MUC, TON, P1X, 7BV, AG80, and CAB.
958 KINNEY AND OTHERS

and EVE 23 viruses, as well as for the E3 and E2 gene sequences for the remaining viruses. The E3 and E2 sequences were determined independently in this study. Virus abbreviations: TRD = Trinidad donkey; TC83 = TC-83; 71180 = 71-180; EVE = Everglades Fe3-7c; MENA = Mena II; MUC = Mucambo BeAn 8; TON = Tonate CaAn 410d; 71D = 71D-1252; PIX = BeAr 35645; 78V = 78V-3531; AG80 = AG80-663; CAB = Cabassou CaAr 508.

<table>
<thead>
<tr>
<th>Virus Abbreviation</th>
<th>Sequence Details</th>
<th>E2-159</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P676</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3880</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MENA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68U201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TON</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6K

**and EVE**

**viruses, as well as for the E3 and E2 gene sequences for the remaining viruses.**

**The E3 and E2 sequences were determined independently in this study.**

**Virus abbreviations:**

- TRD = Trinidad donkey
- TC83 = TC-83
- 71180 = 71-180
- EVE = Everglades Fe3-7c
- MENA = Mena II
- MUC = Mucambo BeAn 8
- TON = Tonate CaAn 410d
- 71D = 71D-1252
- PIX = BeAr 35645
- 78V = 78V-3531
- AG80 = AG80-663
- CAB = Cabassou CaAr 508.

**and EVE**

**viruses, as well as for the E3 and E2 gene sequences for the remaining viruses.**

**The E3 and E2 sequences were determined independently in this study.**

**Virus abbreviations:**

- TRD = Trinidad donkey
- TC83 = TC-83
- 71180 = 71-180
- EVE = Everglades Fe3-7c
- MENA = Mena II
- MUC = Mucambo BeAn 8
- TON = Tonate CaAn 410d
- 71D = 71D-1252
- PIX = BeAr 35645
- 78V = 78V-3531
- AG80 = AG80-663
- CAB = Cabassou CaAr 508.
FIGURE 2. Continued.

TABLE 3

Percent nucleotide (lower left) and translated amino acid (upper right) identities among the aligned sequences of the structural polyprotein genes of Venezuelan equine encephalitis viruses*

<table>
<thead>
<tr>
<th>TRD</th>
<th>TC-83</th>
<th>71-180</th>
<th>P676</th>
<th>3880</th>
<th>EVE</th>
<th>MENA</th>
<th>TON</th>
<th>710-1252</th>
<th>PIX</th>
<th>78V-3531</th>
<th>AG80-663</th>
<th>CAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>96.0</td>
<td>96.0</td>
<td>95.6</td>
<td>94.8</td>
<td>94.1</td>
<td>84.0</td>
<td>83.6</td>
<td>84.0</td>
<td>83.4</td>
<td>83.4</td>
<td>83.4</td>
<td>85.7</td>
<td>87.6</td>
</tr>
<tr>
<td>96.0</td>
<td>96.0</td>
<td>95.6</td>
<td>94.8</td>
<td>94.1</td>
<td>84.0</td>
<td>83.6</td>
<td>84.0</td>
<td>83.4</td>
<td>83.4</td>
<td>83.4</td>
<td>85.7</td>
<td>87.6</td>
</tr>
<tr>
<td>96.0</td>
<td>96.0</td>
<td>95.6</td>
<td>94.8</td>
<td>94.1</td>
<td>84.0</td>
<td>83.6</td>
<td>84.0</td>
<td>83.4</td>
<td>83.4</td>
<td>83.4</td>
<td>85.7</td>
<td>87.6</td>
</tr>
<tr>
<td>96.0</td>
<td>96.0</td>
<td>95.6</td>
<td>94.8</td>
<td>94.1</td>
<td>84.0</td>
<td>83.6</td>
<td>84.0</td>
<td>83.4</td>
<td>83.4</td>
<td>83.4</td>
<td>85.7</td>
<td>87.6</td>
</tr>
<tr>
<td>96.0</td>
<td>96.0</td>
<td>95.6</td>
<td>94.8</td>
<td>94.1</td>
<td>84.0</td>
<td>83.6</td>
<td>84.0</td>
<td>83.4</td>
<td>83.4</td>
<td>83.4</td>
<td>85.7</td>
<td>87.6</td>
</tr>
</tbody>
</table>

* Comparison of 3,777 aligned nucleotides and 1,259 aligned amino acids. Virus abbreviations: TRD = Trinidad donkey; EVE = Everglades Fe3-7c; MENA = Mena II; MUC = Mucambo BeAn 8; TON = Tonate CaAn 410d; PIX = Pixuna ReAr 35645; CAB = CaAr 508. Boxes indicate closely related viruses.

† Stippled bar indicates homologous, 100% nucleotide or amino acid sequence identity.

the IE virus (MENA) by N and HI analyses using polyclonal sera6-10 and MAbs12,13 and by molecular analysis.65 This and other genetic analyses support reclassification of EVE virus as another variety of subtype I.18,20,23

Although the 78V-3531 virus was originally identified as VEE subtype-variety IF9,10 genetic analyses of the E3/E2 genes10 and the entire structural polyprotein translated from the 26S mRNA (this study) indicated that 78V-3531 virus is more closely related to subtype VI AG80-663 virus. Among the 45 amino acid positions that were uniquely shared between the 26S polyproteins of the 78V-3531 and AG80-663 viruses were a unique 3-amino acid insert (Ser-Lys-Asn or Ala-Lys-Asn) after position 111 in the capsid protein, the presence of Phe rather than Tyr at the amino terminus of the E1 protein, and an Ala-to-Pro substitution at E1-22 (Figure 2). Although HI tests using anti-VEE E2 subunit antisera indicated a close one-way relationship with subtype I viruses,9,10 cross-N tests with these subunit antisera,9 as well as HI tests using immune sera from short-haired guinea pigs immunized with VEE viruses,10 clearly differentiated 78V-3531 virus from viruses belonging to VEE subtypes I-V. The genetic relationship between 78V-3531 and AG80-663 viruses was robust when based on the sequence of the entire structural polyprotein gene (Figure 3), but was reportedly not robust when the shorter E3/E2 gene region was analyzed.20 These two viruses were clearly differentiated from each other by reciprocal HI tests using anti-E2 subunit sera,11 and their 3'-NTRs differed considerably in length (Figure 1B). It now appears reasonable to exclude 78V-3531 virus from subtype I of the VEE serocomplex. However, inclusion of 78V-3531 virus in subtype VI with AG80-663 virus is probably premature at this time, pending more extensive genetic and antigenic characterization of the two viruses.

The proposed changes in the classification of EVE and...
78V-3531 viruses are supported by the virulence characteristics of these viruses. The EVE virus and all of the virus varieties of VEE subtype I, except for 78V-3531 (IF) virus, cause fatal infections in weanling and adult mice after peripheral challenge, although in outbred older mice EVE or MENA virus may kill mice nonuniformly or only at high challenge dose, respectively. The 78V-3531 virus and the other enzootic viruses of subtypes III, IV, V and VI do not kill mice even at very high peripheral challenge doses ($\geq 10^6$ plaque-forming units). Given this proposed reclassification, lethal susceptibility of weanling or adult mice to peripheral challenge with virus may constitute a biological marker for viruses belonging to VEE subtype I.

The sporadic occurrence of epizootic-epidemics of equine-virulent VEE IAB and IC viruses may be due to re-emergence from unidentified enzootic foci or mutation of enzootic strains of subtype-variety ID, which show close genetic relationship with the epizootic strains. The data presented here support this hypothesis because the other enzootic VEE viruses are too genetically disparate to give rise easily to equine-virulent, epizootic varieties. Historically, epizootics of VEE IAB virus may have been caused by inadequate formalin-inactivated vaccines derived from virulent IAB viruses, including TRD virus. Near genetic identity between the genomes of TRD virus, which was isolated in 1943 on the island of Trinidad and has been used in formalin-inactivated VEE vaccines, and 71-180 virus, which was isolated in 1971 in Texas, has been established. The E3 and E2 genes of other VEE IAB viruses isolated between 1938 and 1973 are also genetically similar to those of TRD virus. No epizootics of VEE IAB virus have occurred since formalized vaccines were replaced by live, attenuated VEE virus vaccines during and following the extensive 1969–1972 outbreak.

Selection of N escape variants with MAbs identified a major N domain between E2 residues 182-207 in VEE virus and 170-220 in Sindbis virus. The sequence variation observed in this E2 N domain of VEE virus would be expected for a domain that is subject to immune selection pressures. Competition assays between VEE anti-peptide antibodies and MAbs have shown that the amino-terminal part of the E2 protein overlaps the N domain. A protective epitope, consisting of the nine amino-terminal amino acids of the E2 glycoprotein of VEE virus, has been identified using MAb 1A2B-10. This MAb recognizes all VEE complex viruses.
except the TC-83 vaccine strain and AG80-663 virus in ELISA tests.\textsuperscript{68} The failure of MAb 1A2B-10 to recognize TC-83 virus is probably due to the E2-7 Lys-to-Asn (TRD-to-TC-83) mutation.\textsuperscript{68} The recent sequence information indicates that the lack of reactivity of MAB 1A2B-10 with AG80-663 virus\textsuperscript{68} is probably due to the E2-2-4 Thr-Glu-Glu (TRD) to Leu-Asp-Asp (AG80-663) substitutions. The low reactivity of MAb 1A2B-10 with PIX virus\textsuperscript{68} may be due to the Glu-to-Asp substitution at position E2-3 in PIX virus. The 1A2B-10 MAb passively protects mice from challenge with both epizootic and enzootic VEE viruses, and peptide E2-1-19 elicits virus-reactive antibodies in horses.\textsuperscript{69}

The attenuated phenotype of the TC-83 vaccine strain of VEE virus in mice is due to the synergistic effects of mutations in the 5' nontranslated region and E2 protein of the parent TRD virus.\textsuperscript{71} Single amino acid mutations in E2 are capable of attenuating TRD virus.\textsuperscript{71±73} It is possible that the E2 protein is involved in the determination of virulent phenotype in epizootic VEE viruses. The ectodomain of the E2 proteins of two subtype II viruses contain 19 or 21 amino acid substitutions, three of which are shared with IC and ID viruses, relative to that of TRD virus.\textsuperscript{20,23} Six strains of subtype-variety ID virus contain only 5–8 amino acid substitutions in the ectodomain of E2 relative to that of TRD virus, and four of these substitutions are shared with epizootic IC strains.\textsuperscript{19,20}

The E2 glycoproteins of the VEE viruses demonstrated considerable variability in the number of potential sites for Asn-linked glycosylation. Only the glycosylation site located at E2-318 was conserved in all of the VEE viruses studied here. Glycosylation at this location, as well as the single conserved glycosylation site in E1, may be important in the antigenicity or biological function of these viruses. The number of potential glycosylation sites used or the extent of glycosylation at certain sites must vary for the different VEE viruses, since the variable molecular weights of their E2 proteins do not correlate uniformly with the number of potential glycosylation sites that are present in the amino acid sequence of E2.\textsuperscript{8,9}

Acknowledgments: We thank the Biotechnology Core Facility Branch, Centers for Disease Control and Prevention (Atlanta, GA) for supplying the synthetic oligonucleotide primers used in this study.

Authors’ addresses: Richard M. Kinney, Kyotaka R. Tsuchiya, Gwong-Jen J. Chang, and John T. Roehrig, Arbovirus Diseases Branch, Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, PO Box 2087, Fort Collins, CO 80522. Martin Pfeffer, Institute for Medical Microbiology, Infectious and Epidemic Diseases, Ludwig-Maximilians University, Veterinary Faculty, Veterinaestr. 13, D-80539, Munich, Germany.

Reprint requests: Richard M. Kinney, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, PO Box 2087, Fort Collins, CO 80522.

REFERENCES


60. Davis NL, Pence DF, Meyer WJ, Schmaljohn AL, Johnston RE,


