VIROLOGY 240, 100–108 (1998)
ARTICLE NO. VY978907

The Alphavirus 3′-Nontranslated Region: Size Heterogeneity and Arrangement of Repeated Sequence Elements

Martin Pfeffer,*[1] Richard M. Kinney,† and Oskar-Rüger Kaaden*

*Institute for Medical Microbiology, Epidemic and Infectious Diseases, Veterinary Faculty, Ludwig-Maximilians University, Munich, Germany; and †Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention (CDC), Public Health Service, U. S. Department of Health and Human Services, Fort Collins, Colorado

Received July 30, 1997; returned to author for revision August 20, 1997; accepted October 22, 1997

The 3′-nontranslated region (NTR) of representative strains of all known alphavirus species was amplified by reverse transcription–polymerase chain reaction. For 23 of them, the 3′-NTR sequence was determined. Together with previously published data, this allowed an analysis of the 3′-NTR of the viruses in the genus Alphavirus. The length of the 3′-NTRs varied from 77 nt for Pixuna virus to 609 nt for Bebaru virus. The 19-nt conserved sequence element directly adjacent to the poly(A) tract was found in all viruses, supporting the hypothesis that this region is a *cis*-acting sequence element during viral replication and essential for virus growth *in vitro*. Within the 3′-NTR of all alphaviruses, repeated sequence elements of various numbers and lengths were found. Their composition was very consistent in both the Venezuelan equine encephalitis (VEE) and the Sindbis-like viruses, although their number was constant only within the latter group. For the VEE viruses, our data suggested that insertion events rather than deletions from an ancestor with a long 3′-NTR created the various number of repeated sequence elements. Among the remaining viruses, both the number and the composition of repeated sequence elements varied remarkedly.

Key Words: alphavirus; nucleotide sequence data; repeated sequence elements; 3′-nontranslated region (NTR).

INTRODUCTION

The Alphavirus genus (family *Togaviridae*) contains at least 27 arthropod-borne members (Murphy *et al.,* 1995), most of which are important pathogens for livestock animals and humans worldwide. Their virions contain a single-stranded RNA genome of positive sense with a 5′-terminal cap structure and a 3′-terminal poly(A) tract (Strauss and Strauss, 1986). The genome length of those alphaviruses that have been sequenced ranges from 11,442 nt for Semliki Forest virus (SF; Garoff *et al.,* 1980; Takkinen, 1986) to 11,851 nt for Ross River virus (RR; Faragher *et al.,* 1988). The 5′ two-thirds portion of the genome encodes the nonstructural protein genes 1 to 4 followed by the 26S junction region, which promotes transcription of the intracellular subgenomic 26S RNA. This mRNA contains the structural gene region (capsid, E3, E2, 6K, and E1) and represents the remaining 3′ one-third portion of the genome. In addition to the junction region, there are two other untranslated regions in the alphaviral genome: (i) a stretch of 44 to 85 nt at the extreme 5′-end, which is required for plus-strand RNA synthesis (Ou *et al.,* 1983; Strauss and Strauss, 1994), and (ii) the region between the stop codon of the E1 gene and the poly(A) tail. Within the latter region, repeated sequence elements (RSE) 40 to 60 nt in length have been described for all alphaviruses investigated thus far (for review see Strauss and Strauss, 1994). However, the function(s) of these repeats is not fully understood. A Sindbis virus (SIN) mutant missing all three copies of the RSE was viable (Kuhn *et al.,* 1990), showing that they are not essential for virus growth *in vitro.* However, elimination of the RSE resulted in a reduced and delayed release of virus in different cell types, which suggests an effect on the binding of unknown cell-specific proteins that might be involved in virus production (Kuhn *et al.,* 1991). Consistent with this hypothesis are observations that SIN defective interfering RNAs lacking all RSE are defective in translation of viral proteins rather than in replication of the genomic RNA (Leathers *et al.,* 1993).

The length and composition of the 3′-nontranslated region (NTR) and the number and types of RSE have been shown to be very constant for SIN strains of different origin (Shirako *et al.,* 1991) and for Highlands J viruses isolated over a period of more than 40 years (Cilnis *et al.,* 1996). In contrast, both the number of RSE and consequently the nucleotide composition and length of the 3′-NTR can vary within one virus species (Faragher and Dalgarno, 1986). A highly conserved nucleotide sequence 19 nt in length has been found in all alphaviruses

Sequence data from this article have been deposited with the GenBank Data Libraries under Accession Nos. AF023279–AF023301.

[1] To whom correspondence and reprint requests should be addressed at, Institut für Medizinische Mikrobiologie, Infektions- und Seuchenmedizin, Veterinaerstrasse 13, D-80539 Munich, Germany. Fax: + 49-89-2180-2155. E-mail: Martin.Pfeffer@micro.vetmed.uni-muenchen.de.

0042-6822/98 $25.00

investigated so far (for review see Strauss and Strauss, 1994). This region is thought to function in *cis* during viral replication and is essential for virus growth *in vitro* (Ou *et al.*, 1982; Kuhn *et al.*, 1990; Hajjou *et al.,* 1996). The aim of our study was to examine the plasticity of the 3′-NTR in representative strains of all classified alphavirus species. Since there appears to be minimal selection pressure in this region (Hajjou *et al.,* 1996), the 3′-NTR sequence data will help to further define genetic relationships among viruses of the Alphavirus genus.

MATERIALS AND METHODS

Virus strains and RNA extraction

The alphavirus strains used in this study are listed in Table 1. Details on their year and location of isolation as well as their passage history have been listed earlier (Pfeffer *et al.,* 1997). The virus passage used in this study was propagated in BHK-21 cells for all viruses except for Pixuna and Ockelbo viruses (Vero cells) and Bebaru and Una viruses (suckling mice). The viral RNA was extracted from a 200-μl aliquot of virus-infected cell culture medium as described earlier (Pfeffer *et al.,* 1997).

Preparation of specific 3′-NTR cDNA

The primers FUSEA.MWG (5′-atg tgg ggn ggn gcn yay tgy tty tgy ga-3′) and cG14T.KEP (5′-cac aga ctg cag cga att cgg tac cttttttttttttttg-3′) were used for the amplification of the 3′-NTRs of all alphaviruses by reverse transcription–polymerase chain reaction (RT-PCR). Ndumu virus (NDU) was the only alphavirus whose viral RNA was not amplified by primers FUSEA.MWG and cG14T.KEP. Primer FUSEA.GVY (5′-ggn gtn tay ccn tty atg tgg gg-3′), which was located 15 nt downstream of FUSEA.MWG, was successfully applied to NDU. The ambiguity conventions used are y = c/t; n = a/c/g/t. Briefly, the RT-PCR mixture consisted of a 100-μmol aliquot of each primer (FUSEA.MWG or FUSEA.GVY and cG14T.KEP), 5 μl of viral RNA template, 50 mM KCl, 10 mM Tris–HCl, pH 8.5, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 μmol dNTP (Boehringer Mannheim GmbH, Mannheim, Germany), 50 mM dithiothreitol, 20 units RNase inhibitor (Boehringer), 1.25 units of *Taq* polymerase (AmpliTaq, Perkin–Elmer, Norwalk, CT), and 2 units of reverse transcriptase (RAV-2, Amersham, Arlington Heights, ILL) in a 100-μl reaction volume. Reactions were incubated at 50°C for 60 min followed by one cycle consisting of 94°C for 2 min, 60°C for 4 min, and 72°C for 5 min. This cycle was followed by 30 cycles of incubation at 94°C for 30 s, 60°C for 90 s, and 72°C for 5 min, with a prolonged extension of 10 s every cycle. A 5-μl aliquot of the RT-PCR was analyzed by electrophoresis in 1% (w/v) agarose gels in Tris–borate–EDTA buffer (Sambrook *et al.,* 1989) stained with ethidium bromide and visualized on a UV transilluminator at 302 nm.

Sequence determination and computer analysis

The preparation of virus-specific amplicons from 1% agarose gels for the sequencing reactions was as described recently (Pfeffer *et al.,* 1997). Sequencing reactions were performed with the *Taq* DyeDeoxy Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) and purified by using Centri-sep columns as recommended by the manufacturer (Princeton Separations, Inc., Adelphia, NJ). The sequences were resolved with an ABI Prism Model 377 automated sequencer (Applied Biosystems). To determine the exact number of T residues at the extreme 3′- end of Buggy Creek virus RNA, the purified amplicon was ligated into pGEM-T vector, as recommended by the manufacturer (Promega, Madison, WI). *Escherichia coli* XL1-Blue bacteria were transformed by electroporation (Dower *et al.,* 1988) and grown in 2X YT medium (Sambrook *et al.,* 1989), and plasmid DNA was purified by using the Qiaprep Spin Plasmid kit (Qiagen, Hameln, Germany). Sequencing primers are available on request. To determine the sequence directly adjacent to the poly(A) tract, primer cKEP (5′-cac aga ctg cag cga att cgg tac c-3′), identical in sequence to the 5′-terminal nucleotides of cG14T.KEP, was used. Sequence alignments and comparison of repeated sequence elements were performed by using the GCG package of the Wisconsin Computer Study Group (Devereux *et al.,* 1984). For the prediction of RNA secondary structure, the DNASIS software was used.

RESULTS

Amplification of alphaviral RNA and sequence analysis

All alphaviruses except NDU yielded a PCR product by using the primer pair FUSEA.MWG/cG14T.KEP. The fragments usually appeared as single bands after agarose gel electrophoresis and ethidium bromide staining, suggesting that the virus preparations of each individual strain were homogenous with respect to their 3′-NTRs. However, for some of the viruses belonging to the SF serocomplex a second fragment, about 400 to 600 bp smaller, was also visible (Fig. 1). The sizes of the amplicons were estimated by comparison with a molecular weight marker after gel electrophoresis and were found to range from about 1180 bp for Pixuna virus to about 1750 bp for Bebaru virus (BEB; Fig. 1, Table 1). The 3′-end of NDU RNA was amplified by using primers FUSEA.GVY and cG14T.KEP, yielding a DNA fragment of about 1650 bp (Fig. 1, Table 1).

Nucleotide sequence analysis was performed for all of the alphavirus cDNA amplicons except for those where the entire 3′-terminal sequences were published previously (Table 1). For the Highlands J and Middelburg viruses with parts of their 3′-NTR previously published, we found one nucleotide difference in 278 nt and four

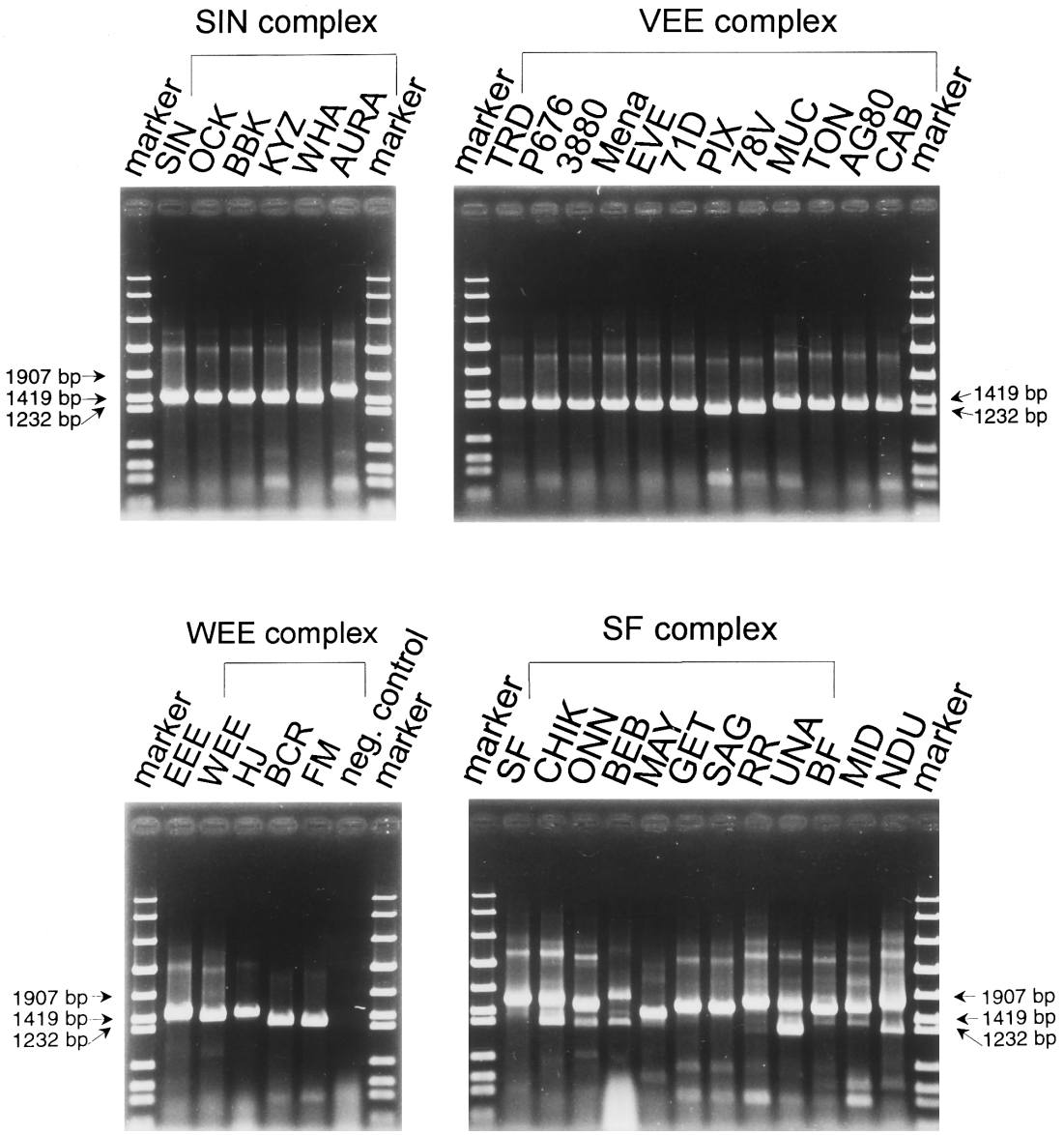

FIG. 1. Ethidium bromide-stained agarose gels showing the DNA fragments resulting from RT-PCR amplification of the 3′-terminus of 35 different alphaviruses. The 3′-terminus of Ndumu virus (NDU) RNA was amplified with primers FUSEA.GVY and cG14T.KEP, the 3′-terminus of all other alphavirus genomes with primers FUSEA.MWG and cG14T.KEP. A 5-μl aliquot of the 100-μl RT-PCR was loaded on the agarose gels. The molecular weights of the DNA marker are shown. The abbreviations are as in Table 1. The negative control contained water instead of RNA in the reaction.

nucleotide changes in 166 nt, respectively (Ou *et al.,* 1982). The 3′-terminus of Chikungunya virus was also sequenced, since it produced two major amplicon bands after RT-PCR (Fig. 1), and because the published sequence data for the polyprotein gene and 3′-NTR did not end with the expected 19-nt conserved sequence element (see below). For Buggy Creek virus, the amplicon was cloned to identify the 19 T residues located 50 nt upstream of the extreme 3′-end of the NTR. Two independent clones were sequenced. A stuttering of the DNA polymerase on the template at this region rather than a mixed DNA population caused the sequencing difficulties. The length heterogeneity, including the stop codon for the structural polypeptide gene of the alphavirus 3′-NTRs, varied from 77 nt for Pixuna virus to 609 nt for BEB, which reflected the sizes of the respective PCR products. Within some of the virus complexes a remarkable consistency in both nucleotide composition and length of the 3′-NTR was obvious (e.g., SIN), while other serocomplexes consisted of viruses whose 3′-NTRs were heterogenous in length, nucleotide composition, or both (e.g., VEE and SF complexes). Adenine-rich stretches were found between 400 and 600 nt upstream of the poly(A) tract in those viruses that showed two amplicon bands by agarose gel electrophoresis after RT-PCR. This stretch was located 512 nt upstream from the poly(A) tract of BEB and contained 15 adenine residues within 21 bases (A_{15}/N_{21}). This region occurred 470

TABLE 1

List of Alphaviruses Used in This Study

Virus (abbreviation)[a,b]	RT-PCR[c]	Length of 3′-NTR[d]	Type and number of RSE[e]
Barmah Forest (BF)	1500	446	2 RR
Middelburg (MID)	1550	487	2 MID + 3 RR (3)
Ndumu (NDU)	1650	547	4 NDU (1)
Venezuelan equine encephalitis (VEE) (TRD), IAB	1230	121	2 VEE + terminal box
(P676), IC	1230	118	2 VEE + terminal box
(3880), ID	1230	120	2 VEE + terminal box
Menall (MENA), IE	1230	128	2 VEE + terminal box
78V-3531 (78V),[f] IF	1200	82	1 VEE + terminal box
Everglades (EVE), II	1230	118	2 VEE + terminal box
Mucambo (MUC), IIIA	1270	184	4 VEE + terminal box
Tonate (TON), IIIB	1270	184	4 VEE + terminal box
71D-1252 (71D),[f] IIIC	1230	127	2 VEE + terminal box
Pixuna (PIX), IV	1180	77	1 VEE (1) + terminal box
Cabassou (CAB), V	1270	184	3 VEE + terminal box
AG80-663 (AG80),[f] VI	1270	195	3 VEE + terminal box
Eastern equine encephalitis (EEE)	1500	361	5 SIN (4)
Western equine encephalitis (WEE)	1450	304	2 SIN
Buggy Creek (BCR)[f]	1280	214	2 BCR
Fort Morgan (FM)	1250	198	2 BCR
Highlands J (HJ)	1550	379	2 SIN
Aura (AURA)	1580	465	3 SIN
Babanki (BBK)	1420	323	3 SIN
Kyzylagach (KYZ)	1420	323	3 SIN
Ockelbo (OCK)	1420	323	3 SIN
Sindbis (SIN)	1420	323	3 SIN
Whataroa (WHA)	1420	323	3 SIN
Semliki Forest (SF)	1650	264	2 SF (1)
Bebaru (BEB)	1750	609	4 BEB (2)
Chikungunya (CHIK)	1630	503	2 ONN (2) + 3 RR
Getah (GET)	1500	413	3 RR
Mayaro (MAY)	1400	291	2 MAY-1 + 2 MAY-2
O'Nyong-nyong (ONN)	1500	425	5 ONN
Ross River (RR)	1650	524	4 RR
Sagiyama (SAG)	1500	413	3 RR
Una (UNA)	1650	513	3 UNA

[a] Viruses representing the serocomplexes are given in boldface; unless indicated otherwise, abbreviations are formally recognized (Karabatsos, 1985).

[b] Classification scheme for the antigenetic subtypes within the VEE serocomplex (in boldface) is according to Calisher *et al.,* (1985) and Kinney *et al.* (1983).

[c] RT-PCR using primers FUSEA.MWG or FUSEA.GVY vs cG14T.KEP; amplicon size estimated from agarose gels.

[d] Sequence data previously reported for TRD (Kinney *et al.,* 1989), P676 and 3880 (Kinney *et al.,* 1992), EVE and MENA (Sneider *et al.,* 1993). EEE (Chang and Trent, 1987), WEE (Hahn *et al.,* 1988), AURA (Rümenapf *et al.,* 1995), OCK (Shirako *et al.,* 1991), SIN (Strauss *et al.,* 1984), SF (Garoff *et al.,* 1980), ONN (Levinson *et al.,* 1990) and RR (Faragher *et al.,* 1988).

[e] Repeated sequence elements (RSE) are named after prototype viruses. The number in parentheses indicates the number of RSE that are truncated. The terminal box of the VEE 3′-NTRs refers to 6 nt within the VEE RSE that are located 5′-adjacent to the 19-nucleotide conserved domain immediately preceding the poly(A) tail.

[f] Not a formally recognized abbreviation.

nt upstream of the poly(A) tail of Una virus (A_{19}/N_{25}), and the cG14T.KEP primer also bound efficiently to an A_{16}/N_{23} stretch located 467 nt upstream of the NDU viral poly(A) tract. An A_{40}/N_{53} region that occurs 294 nt upstream of the poly(A) tail of the O'nyong-nyong viral genome (Levinson *et al.,* 1990) may explain the occurrence of the second lower molecular weight cDNA amplicon following RT-PCR with this virus (Fig. 1). Interestingly, a second A_{18}/N_{23} primer binding site occurred 384 nt upstream of the poly(A) tract of Chikungunya virus as well, thereby extending the reported sequence length of the Chikungunya 3′-NTR by 384 nt (GenBank Accession No. L37661; submitted by Parker, 1994). The alphavirus 3′-NTRs were A/T-rich, ranging from 58.5% for Babanki and Whataroa viruses to 82% for Buggy Creek virus. The average AT content was 67.2% among all 35 viruses investigated (including published sequence data of 12 alphaviruses).

Repeated sequence elements

Inspection of the 3′-NTR sequence data revealed stretches of 18 to 102 bases that occurred at least twice in the viral RNA. A schematic representation of the RSE is given in Fig. 2. The closely related SIN, Ockelbo, Babanki, Whataroa, and Kyzylagach viruses shared three repeats of almost exactly the same length and genomic position and nearly identical nucleotide sequence. In Aura virus these nearly identical 3′-NTR repeats were shifted 27 nt downstream, but the distances between them were similar to those in the other SIN viruses (Rümenapf *et al.,* 1995; Shirako *et al.,* 1991). All members of the VEE serocomplex shared a RSE that consisted of a 6-nt motif flanked by 6 nt upstream and 14 nt downstream. In all VEE 3′-NTRs, this internal motif of 6 nt also appeared directly adjacent to the 19-nt conserved sequence immediately preceding the poly(A) tail. Interestingly, the occurrence of the VEE-type RSE varied from one incomplete RSE plus the internal box at the extreme 3′-end in Pixuna virus to four complete RSE plus the 3′-terminal box in Mucambo and Tonate viruses. Although the latter two viruses had the most RSE copies, their 3′-NTRs were shorter than those of VEE-AG80 and Cabassou viruses, which contained only three copies of the RSE. In summary, the 12 VEE strains shared the same type of RSE, but exhibited five different organizations of the 3′-NTRs (Fig. 2).

Out of the four viruses belonging to the western equine encephalitis (WEE) serocomplex (Calisher *et al.,* 1988), only WEE and Highlands J viruses shared repeated elements with SIN and eastern equine encephalitis viruses. Buggy Creek and Fort Morgan viruses contained a RSE at the same positions of their 3′-NTRs, which differed from those of the other members of the WEE serocomplex.

The composition, length, and number of the RSE varied considerably among the viruses of the SF virus serocomplex and the sole members of the Barmah Forest, Middelburg, and NDU serocomplexes (Fig. 2). Only Sagiyama and Getah viruses showed an almost identical organization of their 3′-NTR, which contained three RSE of a type found in RR (Faragher and Dalgarno, 1986). The two RSE found in Barmah Forest virus were of the RR-type as well. In addition to the tandemly arranged RSE in Middelburg virus, described by Ou *et al.* (1982), we found three copies of other RSE that represent part of the RR-type RSE but which are not tandemly arranged. The organization of the 3′-NTR in Chikungunya virus revealed two copies of an incomplete RSE found in O'nyong nyong virus (Levinson *et al.,* 1990) and three copies of a RSE that contained the entire RR-type RSE and repetitive sequences specific for Chikungunya virus. In five of the SF-related viruses, RSE are in part or in total (Sagiyama and Getah viruses) tandemly arranged. In general, the RSE of members of the SF serocomplex were longer than those of any of the other alphaviruses, reaching up to 102 nt for BEB (Fig. 2).

The 19-nt conserved sequence element

The 19-nt conserved sequence element (19nt-CSE) showed a high degree of identity between all alphaviruses. The 40 nt at the 3′-end of the genomic RNA of each virus is aligned in Fig. 3.

Previously published data on deletion and substitution mutants within the 19nt-CSE of SIN indicated certain positions where modifications are tolerated. The sequence data presented here reflect many of the substitution mutants described by Kuhn *et al.,* (1990). At position 2, referring to the first nucleotide upstream of the poly(A) tract as position -1, we found a C residue in Middelburg virus instead of the T residue occurring in all other viruses investigated. In four SIN and two WEE group viruses, the T residue at position 6 was substituted by a C (SIN) or an A residue (WEE). Other spots of nucleotide changes were positions -18 and -19. At -18, 9 of the 12 SF-related viruses showed an A, while at -19 two WEEs and two SF-related viruses (BEB, NDU, SF, UNA) had either a G or a C instead of the A. At position -15, four of the SF-related viruses had a G residue. Interestingly, these latter four viruses contained an additional C at the extreme 3′-end, corresponding to position +1 relative to the other alphaviruses. Another interesting observation was that in addition to RR, three more viruses isolated in far east Asia (SAG, GET, BF) had identical insertions of TT and TA dimers within this highly conserved region (Fig. 3). Taken together, the 19nt-CSE was highly conserved within the Alphavirus genus, further supporting its importance for viral replication.

DISCUSSION

We investigated the plasticity of the 3′-NTR of virus strains representing all known alphavirus species. This work extends previous analyses of alphavirus 3′-NTRs (for review see Strauss and Strauss, 1994) by including new nucleotide sequence data for 22 alphaviruses. The lengths of alphavirus 3′-NTRs varied from 77 to 609 nt, which extended the previously identified lower and upper nucleotide length limits for this region by 44 and 11 nt, respectively. This length range was in good agreement with recently described SIN recombinants derived from transfected nonreplicative RNAs with 3′-NTRs, ranging from 150 to 600 nt (Hajjou *et al.,* 1996), and recently described avirulent SF variant containing a 3′-NTR of 598 nt (Santagati *et al.,* 1994). Although packaging of up to 4 kb has been demonstrated in different alphavirus expression systems (Bredenbeck and Rice, 1992), it seems likely that BEB virus with a 3′-NTR of 609 bases represents the upper limit of naturally occurring alphaviruses.

We observed 82.7% sequence identity in the 3′-NTRs

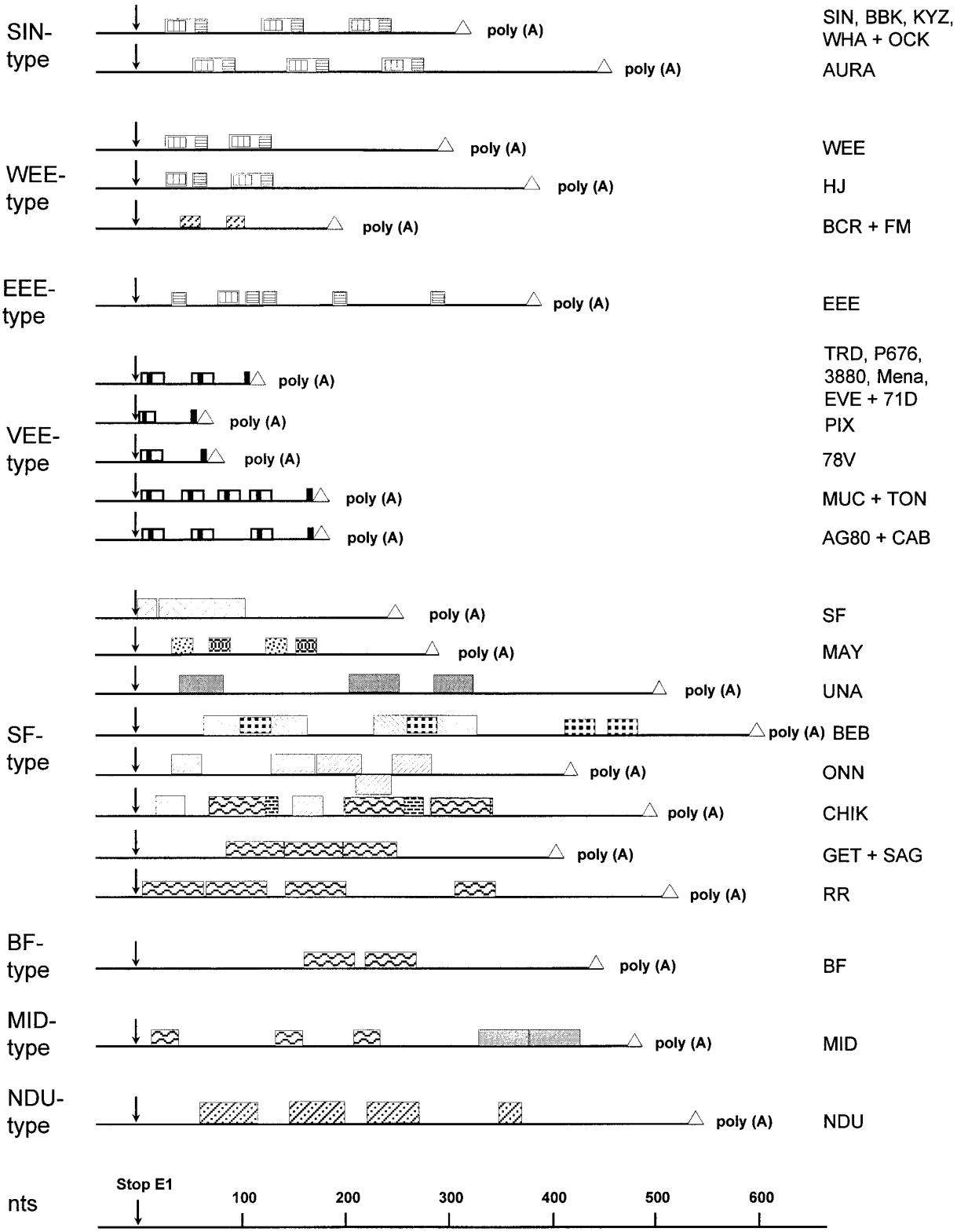

FIG. 2. Schematic representation of repeated sequence elements (RSE) and lengths of 3′-NTRs of 35 alphaviruses (adapted from Strauss and Strauss, 1994). Both the lengths of the repeats and the overall lengths of the 3′-NTRs have been drawn to scale. The downward-pointing arrow indicates the stop codon of the structural protein gene region. Different types of shading indicate repeats of unique sequence. Repeats consisting of two or three distinct motifs are indicated by different types of shading. The open triangles preceding the poly(A) tail indicate the highly conserved 19-nt sequence.

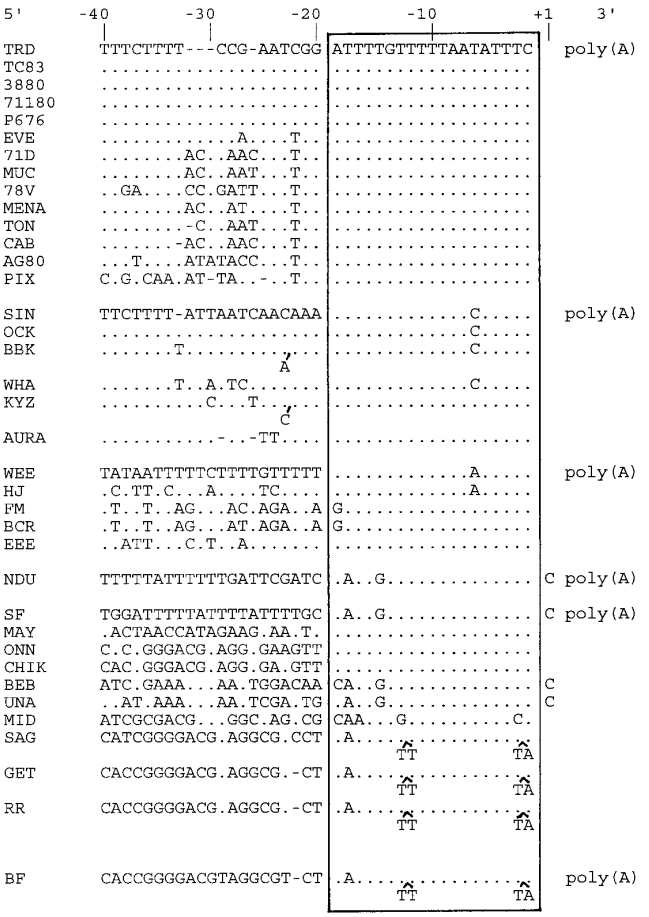

FIG. 3. Alignment of the nucleotides preceding the poly(A) tail at the extreme 3′-end of the alphavirus genome. The 19-nt conserved sequence element of the 35 alphaviruses is boxed. Dashes were introduced to optimize the alignment. Dots indicate sequence identity with the sequence of VEE-TRD virus. Abbreviations are as in Table 1.

of Old World virus isolates showing a SIN virus-like organization of their 3′-NTR. A very close relationship between SIN and SIN-like viruses isolated from South Africa, Australia, and Europe over a period of 30 years has been described (Shirako *et al.,* 1991). The virus strains in our study were also isolated from these three continents over a period of 30 years. These SIN-like viruses should be considered as varieties of SIN virus as supported by recent sequence data from the amino-terminal third of the nsP1 gene (Pfeffer *et al.,* 1997). Aura virus, the only SIN-like virus described in the Americas, differs in its 3′-NTR and other regions of the genome from the other SIN viruses (Rümenapf *et al.,* 1995). In contrast to the SIN viruses, different numbers of RSE have been described for RR and SF viruses (Faragher and Dalgarno, 1986; Santagati *et al.,* 1994). There was no correlation between the variable organization of the 3′-NTR and geographic location or year of isolation for the VEE viruses.

The generation of alphavirus 3′-NTR of different lengths containing various numbers of an identical sequence element could, in principle, be explained either by sequence insertions into a smaller ancestral 3′-NTR or by deletions from a large ancestral 3′-NTR. For the tick-borne encephalitis viruses (*Flaviviridae*), the latter possibility was preferred because none of their 3′-NTRs contained any sequence element that was not present also in one of the longest 3′-NTRs, and it is unlikely that the same type of RSE (including the variable nucleotides surrounding it) was created by independent insertion events (Wallner *et al.,* 1995). In contrast, pestiviruses (*Flaviviridae*) have been shown to acquire even nonviral sequences (Meyers *et al.,* 1989). Although independently occurring deletion events from an ancestor with a long 3′-NTR might be involved in generating variable-length 3′-NTRs (Wallner *et al.,* 1995), such a mechanism does not answer the question of how the RSE originated. Ou *et al.* (1982) first suggested that a template-switching mechanism could generate RSE. Recent studies by Hajjou *et al.* (1996) with nonreplicative SIN RNAs showed that recombination occurs via template switching of the viral replicase. The authors defined two hot spots of recombination, which had no sequence identity to each other but showed similar predicted secondary structures that might dislodge the viral polymerase. The 3′-NTRs of the VEE viruses studied here showed that each of the repeats, including the truncated one of Pixuna virus, formed a predicted stem–loop structure with the 6-nt internal motif being part of the loop (data not shown). Secondary structures have been claimed to be the reason for polymerase jumping in tombus viruses (White and Morris, 1995). Another reason supporting insertion rather than deletion to create different numbers of RSE is that AT-rich sequences are known to be involved in recombination and duplication events leading to homologous recombinants in positive-stranded RNA viruses (Nagy and Bujarski, 1996). Hahn *et al.* (1988) first suggested that WEE is a recombinant virus with most of its structural protein genes derived from a SIN-like ancestor while the nonstructural genes and the extreme 3′-end of its 3′-NTR are derived from an EEE-like virus. By examining the 3′-NTR of HJ we found a very similar organization to the 3′-NTR of WEE. Both 3′-NTRs contained the first two SIN-type RSE although the distance between them was shortened compared to SIN. Thus, the second crossover site should be 3′ of the second RSE, which corresponds to nt position 130 in WEE and HJ and to nt position 161 in SIN. Nucleotide sequence alignments showed a higher level of sequence identity of WEE and HJ to SIN than to EEE up to the region that corresponds to the 3′-end of the third SIN RSE. Since the nucleotides further downstream did align better to EEE than to SIN, the recombination point should be somewhere around the third SIN RSE. These findings are in contrast to Hahn *et al.* (1988), who stated that only the first 60 nt of the WEE 3′-NTR is very similar to SIN.

Based on the variable growth behavior in different cell

lines of SIN mutants lacking all three copies of their RSE, Kuhn *et al.* (1990) concluded that binding of the RSE to a yet unknown host cell protein might explain these phenotypic differences. With respect to the variety of RSE described in this paper, this implies that different cellular proteins might bind to the different RSE types. Further, it suggests that viruses sharing the same type of RSE also share the same vector or host. The large number of vectors and hosts known to cycle, e.g., SIN, in nature (Niklasson, 1988) argue against this assumption. For two SF viruses that differed in their 3′-NTRs by three and a half repeats, cell type-specific differences in replication were found *in vivo* (Santagati *et al.,* 1994). These authors hypothesized that different replication rates might occur because of different secondary structures and the ability to form various circular RNA molecules during replication. However, other differences between the two SF genomes might affect replication *in vivo.* Kuhn *et al.* (1991, 1996) concluded from their studies with RR/SIN chimeras that RSE might be utilized only in specific hosts and that the entire 3′-NTR can function as an independent unit, possibly by interacting directly with host proteins. The different RSE types and lengths of the 3′-NTRs described in this paper favor the latter hypothesis.

The 19-nt conserved sequence directly adjacent to the poly(A) tract was highly conserved within all members of the genus Alphavirus. Previous studies demonstrated the limited sequence flexibility in this region (Kuhn *et al.,* 1990). The striking identity of the 19-nt conserved sequence in all alphaviruses isolated throughout the world over a period of almost 50 years supports the current understanding of this region as an essential initiation site for the viral replicase (Kuhn *et al.,* 1990; Strauss and Strauss, 1994).

Cyclization of the genomic RNA has been shown to be an intermediate stage during RNA replication (Frey *et al.,* 1979). Ou *et al.* (1982, 1983) first reported detailed sequence studies on the 3′-NTR and the 5′-NTR of several alphaviruses. They showed that parts of the 3′-NTRs can hybridize with sequences at the 5′-end of the genome and concluded that a panhandle structure of the genomic RNA might be formed during RNA replication. Sequence analyses of the 5′-end of the various VEE strains investigated in this study revealed a high level of sequence identity of the 5′-NTRs with their lengths ranging from 42 nt for Mucambo and Everglades viruses to 48 nt for Cabassou and VEE-78V viruses (Pfeffer and Kinney, unpublished data). Thus, the VEE 5′-NTRs did not reflect the variability of the VEE 3′-NTRs described here. In addition, the amino acid sequences of all 15 VEE strains (including published sequence data of 7 VEE strains) were identical up to the 51-nt conserved domain investigated. We were unable to find perfectly matched sequences longer than 6 nt for the VEE strains, and mismatching sequences are difficult to find and interpret. Although Frey *et al.* (1979) and Ou *et al.* (1983) found strong evidence for a direct 3′/5′ interaction, more recent studies using RR/SIN chimeras with their very distantly related 3′- and 5′-ends clearly demonstrated that the 3′- and 5′-NTRs do not function in the same manner (Kuhn *et al.,* 1991). Any 3′/5′ interaction therefore could only have taken place between parts of the 3′-NTR and a coding region somewhere at the 5′-terminus of the nsP1 gene. Our sequence data for the VEE strains do not support such an event. Further experimental studies and computer analyses are needed to examine the putative 3′/5′ interaction during the replication of the alphavirus genome.

ACKNOWLEDGMENTS

This work was in part supported by the "Fraunhofer Gesellschaft," Munich, Germany (No. 2087-V-4390). We thank Kiyotaka R. Tsuchiya for expert technical assistance.

REFERENCES

Bredenbeek, P. J., and Rice, C. M. (1992). Animal RNA virus expression systems. *Semin. Virol.* 3, 297–310.

Calisher, C. H., Monath, T. P., Mitchell, C. J., Sabattini, M. S., Cropp, C. B., Kerschner, J., Hunt, A. R., and Lazuick, J. S. (1985). Arbovirus investigations 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 Malayos, Resistencia, Barranqueras, and Antequera). *Am. J. Trop. Med. Hyg.* 34, 956–965.

Calisher, C. H., Karabatsos, N., Lazuick, J. S., Monath, T. P., and Wolff, K. L. (1988). Reevaluation of the western equine encephalitis antigenic complex of alphaviruses (family Togaviridae) as determined by neutralization tests. *Am. J. Trop. Med. Hyg.* 38, 447–452.

Chang, G.J. J., and Trent, D. W. (1987). Nucleotide sequence of the genome region encoding the 26S RNA of eastern equine encephalomyelitis virus and the deduced amino acid sequence of the viral structural proteins. *J. Gen. Virol.* 68, 2129–2142.

Cilnis, M. J., Kang, W., and Weaver, S. C. (1996). Genetic conservation of Highlands J viruses. *Virology* 218, 343–351.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12, 387–395.

Dower, W. T., Miller, J. F., and Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* 16, 6127–6145.

Faragher, S. G., and Dalgarno, L. (1986). Regions of conservation and divergence in the 3′ untranslated sequences of genomic RNA from Ross River virus isolates. *J. Mol. Biol.* 190, 141–148.

Faragher, S. G., Meek, A. D. J., Rice, C. M., and Dalgarno, L. (1988). Genome sequences of a mouse-avirulent and a mouse-virulent strain of Ross River virus. *Virology* 163, 509–526.

Frey, T. K., Gard, D. L., and Strauss, J. H. (1979). Biophysical studies on circle formation by Sindbis virus 49S RNA. *J. Mol. Biol.* 132, 1–18.

Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H., and Delius, H. (1980). Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. *Nature (London)* 288, 236–241.

Hahn, C. S., Lustig, S., Strauss, E. G., and Strauss, J. H. (1988). Western equine encephalitis virus is a recombinant virus. *Proc. Natl. Acad. Sci. USA* 85, 5997–6001.

Hajjou, M., Hill, K. R., Subramaniam, S. V., Hu, J. Y., and Raju, R. (1996). Nonhomologous RNA-RNA recombination events at the 3′ nontranslated region of the Sindbis virus genome: Hot spots and utilization of nonviral sequences. *J. Virol.* 70, 5153–5164.

Karabatsos, N. (1985). "International Catalogue of Arboviruses, Including Certain Other Viruses of Vertebrates," Third ed. The Subcommittee on the Information Exchange of the American Committee on the Arthropod-Borne Viruses and the American Society of Tropical Medicine and Hygiene, San Antonio, TX.

Kinney, R. M., Johnson, B. J. B., Welch, J. B., Tsuchiya, K. R., and 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., Trent, D. W., and 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., and Trent, D. W. (1992). Genetic evidence that epizootic Venezuelan equine encephalitis (VEE) virus may have evolved from enzootic VEE subtype I-D virus. *Virology* 191, 569–580.

Kuhn, R. J., Griffin, D. E., Owen, K. E., Niesters, H. G. M., and Strauss, J. H. (1996). Chimeric Sindbis-Ross River viruses to study interactions between alphavirus nonstructural and structural regions. *J. Virol.* 70, 7900–7909.

Kuhn, R. J., Hong, Z., and Strauss, J. H. (1990). Mutagenesis of the 3′ nontranslated region of Sindbis virus RNA. *J. Virol.* 64, 1465–1476.

Kuhn, R. J., Niesters, H. G. M., Hong, Z., and Strauss, J. H. (1991). Infectious RNA transcripts from Ross River virus cDNA clones and the construction and characterization of defined chimeras with Sindbis virus. *Virology* 182, 430–441.

Leathers, V., Tanguay, R., Kobayashi, M., and Gallie, D. R. (1993). A phylogenetically conserved sequence within viral 3′ untranslated RNA pseudoknots regulates translation. *Mol. Cell. Biol.* 13, 5331–5347.

Levinson, R. S., Strauss, J. H., and 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.

Meyers, G., Rümenapf, T., and Thiel, H.-J. (1989). Ubiquitin in a togavirus. *Nature (London)* 341, 491.

Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A., and Summers, M. D. (1995). Virus taxonomy: Classification and nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. *Arch. Virol.* 8(Suppl.), 428–433.

Nagy, P. D., and Bujarski, J. J. (1996). Homologous RNA recombination in brome mosaic virus: AU-rich sequences decrease the accuracy of crossovers. *J. Virol.* 70, 415–426.

Niklasson, B. (1988). Sindbis and Sindbis-like viruses. *In* "The Arboviruses: Epidemiology and Ecology" (T. P. Monath, Ed.), Volume 4, pp. 167–176. CRC Press, Boca Raton, FL.

Ou, J.-H., Strauss, E. G., and Strauss, J. H. (1983). The 5′-terminal sequences of the genomic RNAs of several alphaviruses. *J. Mol. Biol.* 168, 1–15.

Ou, J.-H., Trent, D. W., and Strauss, J. H. (1982). The 3′-non-coding regions of alphavirus RNAs contain repeating sequences. *J. Mol. Biol.* 156, 719–730.

Pfeffer, M., Proebster, B., Kinney, R. M., and Kaaden, O.-R. (1997). Genus-specific detection of alphaviruses by a semi-nested reverse transcription-polymerase chain reaction. *Am. J. Trop. Med. Hyg.* 57, in press.

Rümenapf, T., Strauss, E. G., and Strauss, J. H. (1995). Aura virus is a New World representative of Sindbis-like viruses. *Virology* 208, 621–633.

Sambrook, J., Maniatis, T., and Fritsch, E. F. (1989). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Santagati, M. G., Itaranta, P. V., Koskimies, P. R., Maatta, J. A., Salmi, A. A., and Hinkkanen, A. E. (1994). Multiple repeating motifs are found in the 3′-terminal nontranslated region of Semliki Forest virus A7 variant genome. *J. Gen. Virol.* 75, 1499–1504.

Shirako, Y., Niklasson, B, Dalrymple, J. M., Strauss, E. G., and Strauss, J. H. (1991). Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses. *Virology* 182, 753–764.

Sneider, J. M., Kinney, R. M., Tsuchiya, K. R., and Trent, D. W. (1993). Molecular evidence that epizootic Venezuelan equine encephalitis (VEE) I-AB viruses are not evolutionary derivatives of enzootic VEE subtypes I-E or II viruses. *J. Gen. Virol.* 74, 519–523.

Strauss, E. G., Rice, C. M., and Strauss, J. H. (1984). Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* 133, 92–110.

Strauss, E. G., and Strauss, J. H. (1986). Structure and replication of the alphavirus genome. *In* "The Togaviridae and Flaviviridae" (S. Schlesinger and M. J. Schlesinger, Eds.), pp. 35–90. Plenum, New York.

Strauss, J. H., and Strauss, E. G. (1994). Alphaviruses. *Microbiol. Rev.* 58, 491–562.

Takkinen, K. (1986). Complete nucleotide sequence of the nonstructural genes of Semliki Forest virus. *Nucleic Acids Res.* 14, 5667–5682.

Wallner, G., Mandl, C. W., Kunz, C., and Heinz, F. X. (1995). The Flavivirus 3′-noncoding region: Extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. *Virology* 213, 169–178.

White, K. A., and Morris, T. J. (1995). RNA determinants of junction site selection in RNA recombinants and defective interfering RNAs. *RNA* 1, 1029–1040.