Catarina Virus, an Arenaviral Species Principally Associated with *Neotoma micropus* (Southern Plains Woodrat) in Texas

Maria N. B. Cajimat, Mary Louise Milazzo, Robert D. Bradley, and Charles F. Fulhorst*

Microbiology and Immunology Graduate Program, Graduate School of Biomedical Sciences, and Department of Pathology, University of Texas Medical Branch, Galveston, Texas; Department of Biological Sciences and Museum, Texas Tech University, Lubbock, Texas

Abstract. The purpose of this study was to define the taxonomic relationship of an arenavirus principally associated with the southern plains woodrat (*Neotoma micropus*) in southern Texas to other New World arenaviruses. The results of independent analyses of glycoprotein precursor amino acid sequences and nucleocapsid protein amino acid sequences indicated that the arenavirus in southern Texas is novel (proposed species name Catarina virus) and phylogenetically most closely related to Whitewater Arroyo virus, which is principally associated with the white-throated woodrat (*Neotoma albigula*) in northwestern New Mexico. Together, the close phylogenetic relationship between Catarina virus and Whitewater Arroyo virus and the association of these viral species with congenic rodent species support the notion that the principal host relationships of some New World arenaviruses are a product of a long-term shared evolutionary relationship between the virus family *Arenaviridae* and the rodent family Cricetidae.

INTRODUCTION

The virus family *Arenaviridae*, genus *Arenavirus* comprises two serocomplexes. The lymphocytic choriomeningitis-Lassa (Old World) complex includes Lassa virus (LASV), lymphocytic choriomeningitis virus (LCMV), Ippy virus (IPPV), Mobala virus (MOBV), and Mopeia virus (MOPV). The Tacaribe (New World) complex includes Bear Canyon virus (BCNV), Tamiami virus (TAMV), and Whitewater Arroyo virus (WWAV) in North America, Tacaribe virus (TCRV) on Trinidad in the Caribbean Sea, and Allpahuayo virus (ALLV), Amapari virus (AMAV), Cupixi virus (CPXV), Flexal virus (FLEV), Guanarito virus (GTOV), Junin virus (JUNV), Latino virus (LATV), Machupo virus (MACV), Oliveros virus (OLVV), Paraña virus (PARV), Pichindé virus (PICV), Pirital virus (PIRV), and Sabiá virus (SABV) in South America.

Arenaviruses possess genomes that consist of two single-stranded RNA segments, designated large (L) and small (S). The L segment (approximately 7.2 kb) consists of a 5′ non-coding region (NCR), the Z gene, an intergenic region that separates the Z gene from the RNA-dependent RNA polymerase (RdRp) gene, the RdRp gene, and a 3′ NCR. Similarly, the S segment (approximately 3.5 kb) consists of a 5′ NCR, the glycoprotein precursor (GPC) gene, an intergenic region, the nucleocapsid (N) protein gene, and a 3′ NCR. Presently, our most comprehensive knowledge of the phylogenetic history of the New World arenaviruses is based on the results of independent analyses of full-length glycoprotein precursor (GPC) amino acid sequences and full-length nucleocapsid (N) protein amino acid sequences.

Specific members of the subfamilies Neotominae and Sigmodontinae in the rodent family Cricetidae are the principal hosts of the New World arenaviruses for which natural host relationships have been well characterized. For example, the white-throated woodrat (*Neotoma albigula*) in northwestern New Mexico is the principal host of WWAV, the hispid cotton rat (*Sigmodon hispidus*) in southern Florida is the principal host of TAMV, Alston’s cotton rat (*Sigmodon alstoni*) in western Venezuela is the principal host of JUNV, and the drylands vesper mouse (*Calomys musculinus*) in northern Argentina is the principal host of JUNV, and the large vespertilian mouse (*Calomys callosus*) in northeastern Bolivia is the principal host of MACV. The white-throated woodrat is a neotomine rodent whereas the hispid cotton rat, Alston’s cotton rat, drylands vesper mouse, and large vespertilian mouse are sigmodontine rodents.

A previous study established that the southern plains woodrat (*N. micropus*) in southern Texas is the principal host of an arenavirus that is phylogenetically more closely related to WWAV than to TAMV, TCRV, or the South American arenaviral species. The purpose of the present study was to define better the taxonomic and phylogenetic relationship of the arenavirus associated with the southern plains woodrat in southern Texas to WWAV and the other New World arenaviruses.

MATERIALS AND METHODS

Viruses. Strains AV A0400135 and AV A0400212 originally were isolated from southern plains woodrats captured in July 1999 at a site located approximately 18.7 km east of the town of Catarina in Dimmit County, Texas. The results of a comparison of predicted amino acid sequences in a previous study indicated that the primary structure of the N protein of AV A0400135 is different from the primary structure of the N protein of AV A0400212.

Determination of nucleotide sequences of GPC genes and N protein genes of strains AV A0400135 and AV A0400212. The nucleotide sequence of a 3,288-nucleotide fragment of the S genomic segment of AV A0400135 and the nucleotide sequence of a 3,287-nucleotide fragment of the S genomic segment of AV A0400212 were determined in a manner similar to that described previously. Each sequence includes a fragment of the 5′ NCR, the complete GPC gene (1,452 nucleotides), intergenic region (71 or 72 nucleotides), the complete N protein gene (1,686 nucleotides), and a fragment of the 3′ NCR. Total RNA was isolated from monolayers of infected Vero E6 cells using TRIzol® Reagent (Invitrogen Life Technologies, Inc., Carlsbad, CA). First-strand cDNA was generated from the arenaviral RNA by using SuperScript...
II RNase H; Reverse Transcriptase (Invitrogen Life Technologies, Inc.) in conjunction with oligonucleotide 19C-cons (5'-CGCACGWWGATCTCAGGC-3'). This oligonucleotide is a derivative of oligonucleotide ARE3'-END and was expected to anneal to the 19-nucleotide fragment at the extreme 3' end of the S segment and the 19-nucleotide fragment at the extreme 3' end of the replicative intermediate synthesized from the S segment.\textsuperscript{14,35}

Amplicons (polymerase chain reaction products) were generated from three overlapping fragments of the first-strand cDNA by using the Master Tag Kit (Eppendorf North America, Inc., Westbury, NY) in conjunction with 19C-cons and AVGPC14 (5'-GGACAGGCCTCRCCRAKTATGTCCTGTA-3'), AVGPC54 (5'-ATCTCATCTCTTGAATTCCTGA-3'), and AVNP13 (5'-GTGTGTGTCWGGYTCTGCATTA-3') and 19C-cons. Oligonucleotides 19C-cons and AVPC14 flank a 1,426-nucleotide fragment of the S segment that extends from within the 5' NCR; AVGPC54 and AVNP13 flank a 598-nucleotide fragment (strain AV A0400135) or 597-nucleotide fragment (strain AV A0400212) that extends from within the GPC gene, across the intergenic region, and through the stop codon of the N protein gene; and AVNP13 and 19C-cons flank a 1,726-nucleotide fragment that extends from within the N protein gene, through the start codon of the N protein gene, and into the 3' NCR. Amplicons of the expected size were purified from agarose gel slices by using the QIAquick Gel Extraction Kit (Qiagen, Inc., Valencia, CA). Both strands of each purified amplicon were sequenced directly, using the dye termination cycle sequencing technique (Applied Biosystems, Inc., Foster City, CA). The sequence of the 3,288-nucleotide fragment of the S segment of AV A0400135 and the sequence of the 3,287-nucleotide fragment of the S segment of AV A0400212 were deposited into the GenBank nucleotide sequence database under Accession nos. DQ865244 and DQ865245, respectively.

**Data analysis.** The amino acid sequences of the GPCs and N proteins of AV A0400135 and AV A0400212 were compared with the homologous sequences of BCNV strain AV A0060209, TAMV strain W-10777, WWAV strain AV 9310135, ALLV strain CLHP-2472, AMAV strain BeAn 70563, CPXV strain BeAn 119303, FLEV strain BeAn 293022, GTOV strain INH-95551, JUNV strains XI13, MC2, and Romero, LATV strain MARU 10924, MACV strains Carvallo, Chicava, Mallele, and 9530537, OLVV strain 3229-1, PARV strain 12056, PICV strain An 3739, PIVR strain VAV-488, SABV strain SPH 114202, TRCV strain TRVL11573, and LCMV strain WE (GenBank Accession nos. AF512833, AF512828, AF228063, YA012687, AF512834, AF512832, AF512831, NC_005077, NC_005081, D10072, YA619641, AF485259, NC_005078, YA624355, YA619645, YA571959, U34248, AF485261, NC_006447, NC_005894, NC_006317, NC_004293, and M22138, respectively). The LCMV strain WE was included in the phylogenetic analyses to enable inference of the ancestral node within the group of New World arenaviruses. The GPC and N protein amino acid sequences were aligned independently, using the computer program CLUSTAL W1.7.\textsuperscript{14} The GPC and N protein gene nucleotide sequence alignments were constructed manually based on alignments of the GPC and N protein amino acid sequences, respectively. The analyses of the multiple amino acid sequence alignments were done by using programs in the computer software package MEGA2.\textsuperscript{17} The neighbor-joining analyses were carried out on uncorrected p-model distances generated from the multiple amino acid sequence alignments. Bootstrap support for the results of each neighbor-joining analysis was based on 1,000 repetitions of the heuristic search, with random resampling of the data.\textsuperscript{17} The analyses of the multiple nucleotide sequence alignments were done by using MRBAYES 3.1.2 and other programs in the computer software package PAUP.\textsuperscript{19,20} A general time reversible $+$ I + G model with a site-specific gamma distribution and sites partitioned by codon was used with the following options: four Markov-chains, one million generations, sample frequency = every 1,000th generation. The likelihood scores, convergence statistics, and potential scale reduction factors were reviewed, and the first 1,000 trees were then discarded. A consensus tree (50% majority rule) was constructed from the remaining trees, and clade probability values were generated to assess support for the nodes within the consensus tree.

### RESULTS

The GPC amino acid sequence alignment was 560 characters in length and the N protein amino acid sequence alignment was 576 characters in length. Sequence nonidentities (uncorrected p-model distances) between the GPCs and between the N proteins of AV A0400135 and AV A0400212 were 4.8% and 1.6%, respectively (Table 1). Sequence nonidentities between the GPCs and between the N proteins of JUNV strains XI13, MC2, and Romero ranged from 0.8% to 1.6% and from 1.2% to 3.9%, respectively. Similarly, sequence nonidentities between the GPCs and between the N proteins of MACV strains Carvallo, Chicava, Mallele, and 9530537 ranged from 1.6% to 5.6% and from 0.9% to 2.3%, respectively. Thus, AV A0400135 and AV A0400212 are conspecific, i.e., strains of the same arenaviral species.

The GPCs of AV A0400135 and AV A0400212 exhibited the lowest sequence nonidentity (33.1% and 33.5%, respectively) with the GPC of WWAV strain AV 9310135 (Table 1). Furthermore, the N proteins of AV A0400135 and AV A0400212 exhibited the lowest sequence nonidentity (13.5% and 13.3%, respectively) with the N protein of WWAV strain AV 9310135 (Table 1).

Sequence nonidentities between the GPCs of strains of different New World arenaviral species ranged from 15.8% (ALLV strain CLHP-2472 and FLEV strain BeAn 293022) to 60.3% (AMAV strain BeAn 70563 and PIVR strain VAV-488) and sequence nonidentities between the N proteins of strains of different New World arenaviral species ranged from 11.9% (JUNV strain XI13 and MACV strain 9530537) to 47.5% (AMAV strain BeAn 70563 and TAVM strain W-10777) (Table 1). Thus, the arenaviral species represented by A0400135 and AV A0400212 should be considered distinct from WWAV and other New World arenaviral species. The name Catarina virus (CTNV) is proposed to distinguish the arenaviral species associated with *N. micropus* in southern Texas from all other arenaviral species.

The neighbor-joining analysis of uncorrected p-model distances generated from the alignment of GPC sequences placed CTNV in a sister relationship to WWAV and separated the North American species (BCNV, CTNV, TAMV,
and WWAV) from TCRV and the South American species (Figure 1A). The neighbor-joining analysis of uncorrected p-model distances generated from the alignment of N protein sequences also placed CTNV in a sister relationship to WWAV and separated the North American species from TCRV and the South American species (Figure 1B). Bootstrap support for monophyly of CTNV and WWAV and for monophyly of the four North American arenaviral species in the phylograms generated from the GPC sequence data was greater than 95.0%. Bootstrap support for monophyly of CTNV and WWAV and for monophyly of the four North American arenaviral species in the phylograms generated from the N protein sequence data was greater than 97.0%.

The two trees generated by simultaneous Bayesian analyses of the GPC gene nucleotide sequence data were identical. The two trees generated by simultaneous Bayesian analyses of the N protein gene nucleotide sequence data also were identical. In the analyses of the GPC gene sequences, CTNV was sister to WWAV, the CTNV-WWAV clade was sister to BCNV, the BCNV-CTNV-WWAV clade was sister to TAMV, and the North American (TAMV-BCNV-CTNV-WWAV) clade was sister to a clade that comprised ALLV, CPXV, GTOV, JUNV, MACV, SABV, and TCRV. The clade probability values for monophyly of CTNV and WWAV and monophyly of the North American viruses was 100; however, the clade probability value for the sister relationship between BCNV and the CTNV-WWAV clade was less than 95. In the analyses of the N protein gene sequences, CTNV was sister to WWAV, the CTNV-WWAV clade was sister to TAMV, the TAMV-CTNV-WWAV clade was sister to BCNV, and the North American (BCNV-TAMV-CTNV-WWAV) clade was sister to a clade that comprised ALLV, CPXV, FLEV, PARV, PICV, and PIRV. The clade probability values for the North American clade and for each node within the North American clade were 100.

**DISCUSSION**

The present-day principal host relationships of some South American arenaviruses appear to be a consequence of a long-term shared evolutionary relationship between the *Arenaviridae* and the subfamily Sigmodontinae. Evidence for this ancient virus-rodent host relationship includes the present-day association of phylogenetically closely related arenaviruses with phylogenetically closely related sigmodontine rodents, e.g., JUNV with the drylands vespertine mouse (*C. musculinus*) and MACV with the large vespertine mouse (*C. callosus*). The results of the analyses of the amino acid sequence data and the analyses of the nucleotide sequence data in this study indicate that CTNV is phylogenetically most closely related to WWAV. Together, the association of CTNV with the southern plains woodrat in southern Texas and WWAV with the white-throated woodrat in northwestern New Mexico is evidence that the present-day principal host relationships of some New World arenaviruses are a consequence of a long-term shared evolutionary relationship between the *Arenaviridae* and the subfamily Neotominae, which is almost exclusively North American.

Previous studies showed that some arenaviruses in association with their principal hosts are geographically widely dis-

**Table 1**

<table>
<thead>
<tr>
<th>Species</th>
<th>CTNV-135</th>
<th>CTNV-212</th>
<th>BCNV</th>
<th>TAMV</th>
<th>WWAV</th>
<th>ALLV</th>
<th>AMAV</th>
<th>CPXV</th>
<th>GTOV</th>
<th>JUNV</th>
<th>MACV</th>
<th>SABV</th>
<th>TCRV</th>
<th>TCRV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonidentities</td>
<td>4.8</td>
<td>35.7</td>
<td>35.7</td>
<td>18.1</td>
<td>18.1</td>
<td>18.5</td>
<td>18.5</td>
<td>18.5</td>
<td>18.5</td>
<td>18.5</td>
<td>18.5</td>
<td>18.5</td>
<td>18.5</td>
<td>18.5</td>
</tr>
<tr>
<td>CTNV-135</td>
<td>1.6</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>44.7</td>
<td>44.7</td>
<td>44.7</td>
<td>44.7</td>
<td>44.7</td>
<td>44.7</td>
<td>44.7</td>
<td>44.7</td>
<td>44.7</td>
</tr>
<tr>
<td>CTNV-212</td>
<td>1.6</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
</tr>
<tr>
<td>BCNV</td>
<td>1.6</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
</tr>
<tr>
<td>TAMV</td>
<td>1.6</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
</tr>
<tr>
<td>WWAV</td>
<td>1.6</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
</tr>
<tr>
<td>ALLV</td>
<td>1.6</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
</tr>
<tr>
<td>AMAV</td>
<td>1.6</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
</tr>
<tr>
<td>CPXV</td>
<td>1.6</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
</tr>
<tr>
<td>GTOV</td>
<td>1.6</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
</tr>
<tr>
<td>JUNV</td>
<td>1.6</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
</tr>
<tr>
<td>MACV</td>
<td>1.6</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
<td>28.9</td>
</tr>
<tr>
<td>SABV</td>
<td>1.6</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
<td>47.7</td>
</tr>
<tr>
<td>TCRV</td>
<td>1.6</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
</tr>
</tbody>
</table>

**Note:** Numbers below the diagonal line are the amino acid sequence nonidentities of the nucleocapsid proteins of 19 New World arenaviruses.
FIGURE 1. Phylogenetic relationships among 24 New World arenaviruses based on neighbor-joining analyses of p-model distances generated from A, full-length glycoprotein precursor amino acid sequences and B, full-length nucleocapsid protein amino acid sequences. The scale bar indicates a sequence divergence of 0.05. The numerical value at the node indicates the percentage of 1,000 bootstrap replicates that supported the interior branch. Bootstrap support values < 70% are not listed. The branch labels in B include (in the following order) viral species, strain, host species, and country. ALLV = Allpahauyo virus; AMAV = Amapari virus; BCNV = Bear Canyon virus; CPXV = Cupiti virus; CTNV = Cataria virus; FLEV = Flexal virus; GTOV = Guanarito virus; JUNV = Junin virus; LATV = Latino virus; LCMV = lymphocytic choriomeningitis virus; MACV = Machupo virus; OLVV = Olivers virus; PARV = Parana virus; PICV = Pichinde virus; PIRV = Pirital virus; SABV = Sabia virus; TCRV = Tacaribe virus; TAMV = Tamiami virus; WWAV = Whitewater Arroyo virus. Arg = Argentina; Bol = Bolivia; Bra = Brazil; Col = Colombia; Par = Paraguay; Per = Peru; Tri = Trinidad; USA = United States (Ca = California; Fl = Florida; NM = New Mexico; Tx = Texas); Ven = Venezuela. Asp = Artibeus species (frugivorous bats); Bobs = Bolomys obscurus (dark bolo mouse); Ccal = Calomys callosus (large vester mouse); Cmus = Calomys musculinus (drylands vester mouse); Csp = Calomys species; Haas = Homo sapiens (human); Naib = Neotoma albigula (white-throated woodrat); Ngui = Neacomys guianae (Guiana bristly mouse); Nmic = Neotoma micropus (southern plains woodrat); Obic = Oryzomys bicolor (bicolored arborescent rice rat); Obuc = Oryzomys bינוicus (Paraguayan rice rat); Ocap = Oryzomys capito (large-headed rice rat); Os = Oryzomys species; Pcal = Peromyscus californicus (California mouse); Sals = Sigmodon alstoni (Alston’s cotton rat); Shis = Sigmodon hispidus (hispid cotton rat). The LCMV strain WE is an Old World arenavirus and was included in the analyses to infer the ancestral node within the group of New World arenaviruses.

LASV. The human health significance of CTNV and other arenaviruses associated with neotomine or sigmodontine rodent species in North America is the subject of ongoing research supported by the United States Public Health Service, National Institutes of Health.

Received March 1, 2007. Accepted for publication June 25, 2007.

Acknowledgments: Maria N. B. Cajimat and Mary Louise Milazzo contributed equally to this study.

Financial support: This study was supported by National Institutes of Health grant AI-41435 (Ecology of Emerging Arenaviruses in the Southwestern United States).

Authors’ addresses: Maria N. B. Cajimat, Department of Epidemiology, University of Michigan, School of Public Health, 109 South Observatory, Ann Arbor, MI 48109, E-mail: ncajimat@umich.edu. Mary Louise Milazzo and Charles F. Fulhorst, Department of Pathology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-0609, E-mails: marinacajimat@utmb.edu and dfulhorst@utmb.edu. Robert D. Bradley, Department of Biological Sciences and Museum, Texas Tech University, Lubbock, TX 79409-3313, E-mail: robert.bradley@ttu.edu.

Reprint requests: Charles F. Fulhorst, Department of Pathology, Delivery Code L20706, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-0609.
REFERENCES


