Characterization of the Tupaia Rhabdovirus Genome Reveals a Long Open Reading Frame Overlapping with P and a Novel Gene Encoding a Small Hydrophobic Protein

Christoph Springfield,1 Gholamreza Darai,2 and Roberto Cattaneo1*

Molecular Medicine Program and Virology and Gene Therapy Track, Mayo College of Medicine, Rochester, Minnesota,1 and Hygiene-Institut der Universität Heidelberg, Abteilung Virologie, Heidelberg, Germany2

Received 13 December 2004/Accepted 2 February 2005

Rhabdoviruses are negative-stranded RNA viruses of the order Mononegavirales and have been isolated from vertebrates, insects, and plants. Members of the genus Lyssavirus cause the invariably fatal disease rabies, and a member of the genus Vesiculovirus, Chandipura virus, has recently been associated with acute encephalitis in children. We present here the complete genome sequence and transcription map of a rhabdovirus isolated from cultivated cells of hepatocellular carcinoma tissue from a moribund tree shrew. The negative-strand genome of tupaia rhabdovirus is composed of 11,440 nucleotides and encodes six genes that are separated by one or two intergenic nucleotides. In addition to the typical rhabdovirus genes in the order N-P-M-G-L, a gene encoding a small hydrophobic putative type I transmembrane protein of approximately 11 kDa was identified between the M and G genes, and the corresponding transcript was detected in infected cells. Similar to some Vesiculoviruses and many Paramyxovirinae, the P gene has a second overlapping reading frame that can be accessed by ribosomal choice and encodes a protein of 26 kDa, predicted to be the largest C protein of these virus families. Phylogenetic analyses of the tupaia rhabdovirus N and L genes show that the virus is distantly related to the Vesiculoviruses, Ephemeroviruses, and the recently characterized Flanders virus and Oita virus and further extends the sequence territory occupied by animal rhabdoviruses.

Tupaia rhabdovirus (TRV) was isolated from spontaneously degenerating hepatocellular carcinoma cells of a tree shrew (Tupaia belangeri) that had been imported from Thailand and was kept in captivity for about 6 years (38). The host range of the virus in vitro appears to be restricted to tupaia cells, indicating that this virus is indeed an indigenous virus of this species (39). Electron microscopy pictures clearly showed typical bullet-shaped rhabdovirus particles; however, no serological cross-reactivity to rabies virus or vesicular stomatitis virus could be demonstrated.

Tupaia, or tree shrews, are small mammals widely distributed in southeast Asia. Their taxonomical position has been controversially discussed for more than a hundred years, and after being classified as either primates or insectivores they finally were assigned a separate order, the Scandentia (reviewed in reference 42). Recent phylogenetic analyses based on genomic sequence information also showed a close relationship of the Scandentia to the primates (45), however, these results are still controversial (1). Tree shrews are used in different areas of biomedical research, especially stress research, ophthalmology, and virology (6). They have been experimentally infected with miscellaneous human viruses, including herpes simplex virus (14), hepatitis B virus (66), and hepatitis C virus (69). Several viruses have been isolated from or described for this species: herpesviruses (2), an adenovirus (56), a paramyxovirus (61), an endogenous retrovirus (23), and circoviruses (47).

This report presents the characterization of the TRV genome and transcription strategy. In addition to the typical rhabdovirus genes N-P-M-G-L, there is an additional transcription unit between the M and G genes that encodes a small hydrophobic protein of 11 kDa. Similar to vesicular stomatitis
virus and many paramyxoviruses, the P gene encodes a putative second 221-amino-acid protein from an overlapping reading frame. Phylogenetic analysis shows that TRV cannot be assigned to one of the established genera.

MATERIALS AND METHODS

Cells and viruses. The TRV sample was kindly provided by John T. Roehrig of the Arbovirus Diseases Branch, Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, U.S. Centers for Disease Control and Prevention in Fort Collins, CO, where it was deposited by G. Darai, who isolated the virus (39). The virus was grown in tupaiya baby fibroblasts (13) that were maintained in Dulbecco's modified Eagle's medium (Cellgro, Herndon, VA) supplemented with 10% fetal calf serum.

Representational difference analysis. Polyadenylated RNA from TRV-infected and uninfected tupaiya baby fibroblast cells was isolated with the FastTrack mRNA isolation kit (Invitrogen, Carlsbad, CA). This RNA was used in the Clontech PCR-Select cDNA subtraction kit (BD Biosciences, Palo Alto, CA) according to the manufacturer's instructions. Briefly, 4 μg RNA was reverse transcribed into cDNA by avian myeloblastosis virus reverse transcriptase and the double-stranded cDNA was synthesized by T4 polymerase. The cDNA was digested with Rsal, and two different specific adapters were ligated to two aliquots of cDNA from the infected cells (tester cDNA). The test cDNA was then hybridized with an excess of cDNA from uninfected cells (driver cDNA) that was not ligated to adapters. After hybridization, the cDNA was used in a nested PCR with oligonucleotide primers binding to the adapters. Since the cDNA present in both samples hybridized with the driver cDNA that does not contain adapters, only cDNA specific for the infected cells should be amplified. The PCR products were cloned using the TOPO TA cloning procedure (Invitrogen, Carlsbad, CA).

cDNA library construction and colony hybridization. TRV virions were isolated from supernatant of infected tupaiya baby fibroblast cells by ultracentrifugation (28,000 rpm, TH-641 rotor [Sorval, Asheville, NC], 4°C, 2 h) using a 60%/20% sucrose step gradient. The virions in the interphase were then pelleted by centrifugation at 28,000 rpm. RNA was isolated from virions that were directly resuspended in RLT lysis buffer of the RNeasy kit (QIAGEN, Hilden, Germany). A cDNA library was constructed using the SuperScript choice system for cDNA synthesis (Invitrogen, Carlsbad, CA). The library was cloned into the EcoRI site of the vector pcDNA3.1 (Invitrogen, Carlsbad, CA). A digoxigenin-labeled probe was generated by amplifying TRV cDNA with the PCR DIG probe synthesis kit. Colonies of the TRV library were blotted on nylon membranes for colony and plaque hybridization and hybridized with the probe according to the manufacturer's instruction. The bound probe was visualized with alkaline phosphatase-labeled antidigoxigenin antibodies and the CDP-Star system (all from Roche Diagnostics, Mannheim, Germany).

Reverse transcription PCR/rapid amplification of cDNA ends. For reverse transcription-PCR (RT-PCR) and 5' and 3' rapid amplification of cDNA ends (RACE), RNA from infected tupaiya baby fibroblast cells was isolated with the RNAeasy kit (QIAGEN, Hilden, Germany). The RNA was transcribed with SuperScript reverse transcriptase (Invitrogen, Carlsbad, CA) using random hexamer primers, or, in the case of 5'-RACE, with a gene-specific or an oligo(dT) primer, respectively. PCRs were performed with the High-Fidelity polymerase (Roche Diagnostics, Mannheim, Germany). The 5' and 3' ends of the transcripts as well as the genomic termini were characterized with the 5'- and 3'-RACE kit (Boehringer Mannheim, Mannheim, Germany) following the manufacturer's instructions. For the N and L genes, the initial RACE procedure yielded a DNA smear. However, when the RACE product was purified (QIAquick PCR Purification Kit, QIAGEN, Hilden, Germany) and sequenced with the same primer that was used in the RACE procedure, enough sequence information could be gained to design new primers and repeat the procedure that then yielded a single band.

Northern blot. Polyadenylated RNA was isolated from infected and uninfected cells as described above, separated by electrophoresis on a 2% formaldehyde agarose gel, blotted to a nylon membrane and analyzed with DNA probes corresponding to the complete M or SH open reading frame (ORF). The probes were generated and visualized with the same system that was used for colony hybridization (Roche Diagnostics, Mannheim, Germany).

DNA sequencing. All sequencing reactions were performed in the Mayo Molecular Biology Core Facility using the ABI PRISM Big Dye Terminator v1.1 Cycle Sequencing Ready Reaction Kit and an ABI PRISM 3700 DNA Analyzer (Perkin-Elmer Applied Biosystems, Foster City, CA). Assignments for most positions were clear-cut, with two exceptions. The signal for nucleotide 1648 (P gene, antigenome) was heterogeneous and nearly equally strong for cytosine and thymidine. While this difference does not affect the P protein, it results in an amino acid difference in the C protein (S63L). Sequence heterogeneity was also observed for the very last nucleotide (11,440) of the genome where the majority of the cloned RACE products had an adenine, but some clones also contained other nucleotides.

Analysis of the nucleotide and deduced amino acid sequences. The Sequencher program (Gene Codes, Ann Arbor, MI) was used to assemble DNA fragments and identify open reading frames. The mass and isoelectric point of the deduced proteins were predicted with Vector NTI 9.0 (Invitrogen, Carlsbad, CA). Proteins were further analyzed with programs on the website of the Swiss Institute of Bioinformatics (www.expasy.com) including ScanProsite (Release 18.39), NetNGlyc 1.0, PredictProtein, and Phobius (34). Protein sequences were aligned with ClustalX (59). Phylogenetic analyses were carried out using the programs PROTDIST, KITSCH, SEQBOOT, and CONSENSE from the PHYLIP package 3.62 (21), and phylogenograms were visualized with TREEVIEW (49).

Nucleotide sequence accession number. The complete TRV genome sequence has been deposited in the GenBank database under accession number AJ840978.

RESULTS AND DISCUSSION

Identification of TRV-specific sequences and completion of the genomic sequence. Attempts to amplify TRV-specific sequences by RT-PCR with degenerate primers against conserved regions in the rabdovirus N and L genes failed, and therefore representational difference analysis was used to enrich viral cDNAs from infected cells. When cDNA libraries from infected and uninfected tupaiya baby fibroblast cells were subtracted, two sets of clones that did not show significant homology to sequences in the databases were identified (Fig. 1A, clones labeled A and B). When forward and reverse primers corresponding to the ends of these clones were used to attempt cDNA amplification by RT-PCR a 1,556-nucleotides fragment covering both initial clones and a short intermediate segment was amplified. Sequence analysis revealed two partial open reading frames separated by a sequence reminiscent of a rabdovirus intergenic region (Fig. 1B, P-M gene junction). The missing parts of the two genes were amplified by 5'- and 3'-RACE and sequenced. A comparison with rabdovirus proteins revealed some similarity with P and M (see below).

To extend this sequence, since several clones generated by representational difference analysis contained only smaller fragments than clones A and B, we isolated genomic RNA from viral particles purified from the supernatant of infected cells and generated another cDNA library (Materials and Methods). The library contained mostly small inserts, but we identified a clone (CH2-2-2) that covered most of the known sequence and contained 50 additional nucleotides upstream of the P gene (Fig. 1A, top). This sequence allowed amplification of progressively larger segments of the N gene by 5'-RACE (see Materials and Methods).

In parallel, we attempted to obtain more sequence information in the direction of the L gene using a primer located directly upstream of the polyadenylation signal of the M gene. Based on 3'-RACE, the complete mRNA of an unexpected small gene (see below) was amplified, probably from bicistronic mRNA. When this strategy was repeated with a primer directly upstream of the end of this gene, the G gene was amplified and sequenced. A combination of rescreening the cDNA library and 3'-RACE was then used to complete the L gene sequence.

The leader sequence was amplified with an oligonucleo-
tide primer directly adjacent to the 5′ end of the N gene and 5′-RACE, and the trailer was amplified and sequenced by 5′-RACE using primers complementary to the antigenome. To obtain a consensus sequence, the whole genome including leader and trailer was finally covered with overlapping PCR amplicons that were directly sequenced in both directions. For the amplification of the 5′ and 3′ ends of the antigenome, primers corresponding to the first 24 and last 25 nucleotides in combination with primers in the N and L genes were used.

FIG. 1. TRV genome organization and transcription/replication control sequences. (A) Map of the TRV genome. The black bar represents the 11,440-nucleotides single-stranded TRV antigenome, the seven open arrows indicate the position of the open reading frames of the N, P, M, SH, G, and L proteins. Relevant cDNA clones are indicated by small bars labeled A, B, and CH-2-2-2, respectively. (B) Sequences of the gene junctions. The gene start and end sequences (right and left columns, respectively) and the intergenic nucleotides (middle column) of the six TRV genes are indicated. A consensus sequence is indicated below the gene sequences (shown in bold type, nucleotide base code of the Nomenclature Committee of the International Union of Biochemistry). This sequence is compared to the sequences of other rhabdoviruses (bottom). (C) The TRV leader and trailer sequences. The leader sequence includes the first ten nucleotides of the N gene, the trailer sequence the ten last nucleotides of the L gene (underlined and in bold type). (D) Alignment of the TRV leader with the leaders of other rhabdoviruses. The first three and the tenth nucleotide are perfectly conserved. (E) Complementarity of the 3′ and 5′ end of the TRV genome.
Two unexpected reading frames in the TRV genome. The TRV genome comprises 11,440 nucleotides, and the antigenome contains seven open reading frames (ORFs) with a coding capacity of more than 90 amino acids. In analogy to other rhabdoviruses, these proteins were named nucleocapsid (N, 430 amino acids), phosphoprotein (P, 337 amino acids), matrix (M, 200 amino acids), glycoprotein (G, 531 amino acids), and polymerase (L, 2107 amino acids). The unexpected sixth ORF encodes a small protein of 93 amino acids that is hydrophobic and that we call SH protein, for small hydrophobic (Fig. 1A, center). The second surprise was a large putative C protein (221 amino acids) encoded in a reading frame overlapping with P. The gene order is N-P/C-M-SH-G-L as in other rhabdoviruses but with the additional SH gene between the M and G genes (Fig. 1A).

The six transcription units are flanked by typical rhabdovirus initiation and polyadenylation sequences (Fig. 1B, top). The transcription initiation sites of the N and P genes and all transcription termination sites were mapped by 5′- and 3′-RACE. A consensus for the TRV gene start signal was derived from the sequence of the six genes (Fig. 1B, right column, bold). The consensus is UUGMCCNKNAG (where M = C or A and K = U or G), which is similar to the corresponding Vesiculovirus and Ephemerovirus sequences and also related to the sequences of the Lyssaviruses and other rhabdovirus genera (Fig. 1B, bottom). The TRV polyadenylation signal is GWWCU, (where W = A or U, Fig. 1B, left column, bold), also similar to that of other rhabdoviruses. The 5′ nontranslated regions of the mRNAs are 11 to 24 nucleotides long, while the nontranslated 3′ regions of the mRNA (between the stop codon and the poly-U stretch) are 0 to 18 nucleotides in length (Fig. 1B). Thus, these sequences are remarkable short even for rhabdovirus standards.

The nontranscribed intergenic regions consist of a single G except for the M-SH junction with the dinucleotide GG. The leader and trailer regions comprise the first 61 nucleotides and last 74 nucleotides of the genome, respectively (Fig. 1C), a size typical of rhabdoviruses (67). The first three and the tenth genomic nucleotides are the same in all rhabdoviruses infecting mammals characterized so far, and the TRV leader shares 14 of the first 16 nucleotides with the vesicular stomatitis virus leader (Fig. 1D). Similar to other rhabdoviruses, 14 out of 15 terminal nucleotides are complementary (Fig. 1E) (reviewed in reference 67).

Replication and transcription apparatus: the N, P and L proteins. All members of the order Mononegavirales encode N, P, and L proteins that, together with the genomic RNA, form the ribonucleoprotein complex (12). The N protein is tightly associated with the genomic RNA and forms the RNase-resistant ribonucleocapsid. The TRV N protein consists of 430 amino acids and has a predicted molecular mass of 49 kDa. When the TRV protein was compared to other rhabdovirus N proteins, highest amino acid identities (31%) were found with the corresponding proteins of bovine ephemeral fever virus and Flanders virus (Table 1, I). In addition, 21% and 17% of the aligned amino acids, respectively, were of similar chemical nature (Table 1, S). The longest stretch of conserved amino acids in vertebrate rhabdoviruses, the motif SPYS (one-letter amino acid code) that may be directly involved in viral RNA binding (35), is also present in the TRV N sequence (amino acids 291 to 294).

The rhabdovirus P protein is a polymerase cofactor necessary for replication and transcription (20). The best-characterized P proteins are those of the vesicular stomatitis virus strains Indiana (VSIV) and New Jersey (VSNJV). For these proteins, a structure with domains I, II, and III has been proposed. The amino-terminal domain I is phosphorylated by the cellular casein kinase II, whereas domain II is phosphorylated by another cellular kinase (8). The P proteins of TRV, VSIV, and VSNJ share low homology, but the TRV protein has potential casein kinase II phosphorylation sites (PROSITE motif PS00006, [ST]-x2-[DE]) in domain I (T78, S83, and S84, Fig. 2, P reading frame) that are close to those phosphorylated by this kinase in the VSIV (S60, T62, and S64) and VSNJ (S59 and S61) P proteins. Also, in domain II, the serine residues phosphorylated in VSIV (S226 and S227) and VSNJV (S236 and S242) can be aligned with serine residues in the TRV protein (S294 and S300). The TRV protein is longer than those of VSIV and VSNJV (337 amino acids as opposed to 265 and 274 amino acids, respectively).

The RNA-dependent RNA polymerases are the largest and most conserved proteins of the Mononegavirales. The four
highly conserved regions A to D (53) forming the “palm” subdomain of the polymerase structure (25) can be identified in the TRV protein: the invariable aspartates in motifs A and C (D601 and D710 in TRV L) may be involved in the coordination of the divalent cations necessary for catalysis; the specific functions of the conserved glycine in motif B (G678) and the lysine in motif D (K788) are not yet known (48).

Large C protein is encoded by an overlapping reading frame in the P gene. An additional protein encoded in an overlapping P gene reading frame accessed by ribosomal choice has been first described for the paramyxovirus Sendai (24). The paramyxovirus C proteins are between 153 and 218 amino acids in length, and recently some of their functions have been elucidated. While being dispensable for virus growth in cultured cells, they counteract the host immune response by antagonizing the effects of interferon and suppression of apoptosis (reviewed in reference 46). They also have effects on viral transcription and might serve as infectivity factors (16). Proteins encoded by an overlapping P gene reading frame have also been identified in the two vesiculoviruses, VSNJV (58) and VSIV (51), and named C in analogy to the paramyxovirus proteins. The proteins of VSIV and VSNJV share little homology but are small (65 and 67 amino acids, respectively) and highly basic with a pI of approximately 11. In both viruses, shorter isoforms of these proteins that start at a downstream AUG (C' /H11032) have been identified. C proteins have also been predicted for the other vesiculoviruses Chandipura virus (80 amino acids, pI 11.8) and Piry virus (32 amino acids, pI 11.1) (58). No C proteins have been identified in members of other rhabdovirus genera so far. The C protein of VSNJV is a strong transcriptional activator in vitro (51), but a recombinant VSIV lacking the C proteins replicates normally in cultured cells (36).
The overlapping ORF in the TRV P mRNA (nucleotides 57 to 719, Fig. 2A) has the potential to code for a 221-amino-acid protein that is therefore more than 3 times larger than the vesicular stomatitis virus proteins and slightly larger than the paramyxovirus proteins. The TRV C protein has a predicted pI of approximately 10, similar to the paramyxovirus and vesicular stomatitis virus proteins. Except for a high content in the basic amino acids arginine, lysine, and histidine no significant sequence similarity was found between the TRV and vesicular stomatitis virus C proteins which is not surprising considering that even the C proteins of the different vesiculoviruses show minimal or no similarity (Fig. 2B). Nevertheless, we consider TRV C a candidate host defense evasion protein.

The envelope associated M and G proteins. The M proteins of negative-stranded RNA viruses are the assembly organizers and interact with the ribonucleocapsids, lipid membranes and cytoplasmic tails of the glycoproteins (7, 55). There is very little homology between the rhabdovirus M proteins, and the TRV M protein has 8 to 16% identity and 10 to 19% similarity to the M proteins of other rhabdoviruses (Table 1). The late-domain motif of vesicular stomatitis virus and rabies virus (PPXY) necessary for efficient budding (26) is not conserved in the TRV protein. The TRV M gene contains a small overlapping open reading frame starting at the second ATG on the mRNA at nucleotide 39 (data not shown). The putative protein has a predicted size of 8.2 kDa and does not show significant similarity to any proteins in the databases.

Rhabdovirus G proteins are type I glycoproteins with an N-terminal signal peptide that is removed after translocation into the endoplasmic reticulum, a C-terminal transmembrane domain, a short cytoplasmic tail and two to six potential N-glycosylation sites (10). They form trimers and are responsible for attachment to the cellular receptors as well as for fusing viral and cellular membranes at low pH in the endosome. Although rhabdovirus G proteins share only limited sequence similarity, the overall structure and some residues, especially cysteines, are conserved. The TRV G protein is 531 amino acids, an average size. The first 21 amino acids are predicted to be the signal peptide, while amino acids 495 to 519 are hydrophobic and likely to span the membrane (34). There are three possible N-glycosylation sites with the consensus sequence NXST (N283, N330, and N462).

Walker and Kongswan have aligned 14 rhabdovirus G protein amino acid sequences, numbered the conserved cysteines, and predicted the intramolecular disulfide bridges (65). Of the highly conserved cysteines I to XII, all except VIII and X are found in the TRV protein. Interestingly, these two cysteines form disulfide bridges with cysteines XI and IX, respectively, and not with each other, in the viral hemorrhagic septicemia virus G protein (18). Thus, not all rhabdoviruses may follow the predicted disulfide bonding strategy (65).

Novel small hydrophobic (SH) transcription unit and protein. The TRV SH protein is encoded by an additional transcription unit that starts and ends with the conserved TRV initiation and polyadenylation signals (Fig. 3A, italics). The production of polyadenylated SH mRNAs in infected cells was demonstrated by 3′-RACE, and transcripts of the expected size (313 nucleotides plus poly(A) tail) were detected in infected cells by Northern blot (Fig. 3B, SH probe). The larger transcript of about 900 nucleotides detected by the SH probe was not sequenced.
also visible when a probe against the M gene was used, indicating that it corresponds to bicistronic M-SH mRNA (Fig. 3B, M probe, expected size 936 nucleotides plus polyA tail). Weaker bands corresponding to other bicistronic transcripts are also visible (Fig. 3B, SH-G detected by the SH probe and P-M detected by the M probe). Thus, SH transcription is conform to that of standard genes of nonsegmented negative-strand RNA viruses.

A Kyte-Doolittle hydrophobicity plot of the SH protein shows two hydrophobic amino acid stretches, one at the amino terminus and one in the center of the sequence (Fig. 3C). The combined transmembrane topology and signal peptide prediction program Phobius (34) identified a type I transmembrane protein with a signal peptide (amino acids 1 to 20), a small extracellular domain, a transmembrane region (amino acids 30 to 51), and a cytoplasmic tail of 42 amino acids. This protein is reminiscent of the SH proteins of the pneumoviruses (11) and rubulaviruses (19, 29), two paramyxovirus genera.

The SH proteins of the different paramyxovirus genera have no significant sequence similarity to each other or the TRV SH protein. The simian virus 5 SH protein is not required for replication in vitro (27) but blocks apoptosis (28), leading to an increased cytopathic effect in cells infected with a SH-knockout virus (41). Remarkably, the TRV protein is predicted to be a type I transmembrane protein whereas the paramyxovirus proteins are type II proteins (30). Small integral membrane protein with another function are found in the Orthomyxoviridae: the M2 protein of influenza A virus (52) and the BM2 protein of influenza B virus (44) have ion channel activities necessary for successful virus uncoating in the endosome. Whether the TRV SH protein has functions similar to the influenza proteins or the simian virus 5 SH protein remains to be determined.

**Rhabdovirus genome structure.** A comparison of the genomes of TRV (Fig. 4, top line) and other rhabdoviruses illustrates that TRV is the only virus encoding an additional protein between the M and G genes. While the vesicular stomatitis virus and rabies virus genomes have the simplest genome structure in the order Mononegavirales with the gene order N-P-M-G-L (Fig. 4, second and third from top), other rhabdoviruses have additional genes at different positions in their genomes. In particular, the fish viruses of the Novirhabdovirus genus have a gene coding for a nonstructural (NV for nonvirion) protein gene between the G and L genes (37) (Fig. 4, third from top).
This page contains a phylogenetic tree depicting the relationships among different genera of rhabdoviruses. The tree is divided into several branches representing different virus families, with rhabdoviruses being the focal point. The text explains the phylogenetic relationships and highlights the diversity within the rhabdovirus family.

The tree is labeled with various virus names and abbreviations, indicating the evolutionary connections among different species. For example, "Mumps virus," "Measles virus," "Mokola virus," "Maize fine streak virus," and "Maize mosaic Lagos bat virus," among others.

The text describes the phylogenetic analysis method used to construct the tree, mentioning the Fitch-Margoliash method and the large sequence territory occupied by rhabdoviruses compared to other virus families.

The tree shows that rhabdoviruses share the typical bullet-shaped morphology and have a broad host range, infecting insects, mammals, fish, and plants. The text notes that there is only little sequence similarity between the proteins of the more distantly related viruses, such as the vesiculoviruses and the plant viruses.

The phylogenetic tree derived from the conserved regions A to D of all available rhabdovirus L sequences and selected members of the Mononegavirales is described in the Materials and Methods section. Family names are bolded, underlined, and italicized. Genera names within the Rhabdoviridae are bolded and italicized. The bar indicates 0.1 expected substitutions per site. Bootstrap values for the major branches from 1000 resampled data sets are indicated in percent.

The phylogenetic tree is used to address the phylogenetic origin of TRV (tupaia rhabdovirus) based on N protein sequences available for more rhabdovirus species (the fish rhabdoviruses of the Novirhabdovirus genus and the plant rhabdoviruses are omitted for clarity). This phylogenetic analysis shows that TRV cannot be assigned to one of the current genera, although it was tentatively assigned to the Vesiculovirus.
genus by the International Committee on Taxonomy of Viruses (63). Genetic evidence is consistent with the fact that the serological cross-reactivity observed in the Vesiculoviruses (63) was not detected for TRV and vesicular stomatitis virus (39). Four other rhabdoviruses also do not fit the current rhabdovirus classification: the Drosophila virus Sigma virus (5); Flanders virus, a virus isolated from mosquitoes and birds in the United States (68); Oita virus, a virus isolated from bats in Japan (31); and trout rhabdovirus 903/87 (32). The TRV, Oita virus, and trout rhabdovirus branches all originate close together from the root of the tree, and bootstrap support for the exact origin is low (Fig. 5B, bootstrap support in percent from 1000 resampled data sets). These results demonstrate that the current classification of rhabdoviruses with the genera Vesiculovirus, Ephemerovirus, and Lyssavirus needs some extension. However, the establishment of novel genera should probably await the characterization of more viruses.

As a summary, we have characterized the complete genomic sequence and transcription strategy of TRV and identified a novel gene encoding a small hydrophobic putative type I transmembrane protein. Furthermore, the P gene contains a large overlapping ORF encoding a C protein more than three times bigger than the known vesicular stomatitis virus proteins. The recent identification of Chandipura virus as the likely cause of fatal encephalitis in humans demonstrates that our understanding of the pathogenic potential of rhabdoviruses is incomplete (62). The TRV genomic sequence makes it possible to await the characterization of more viruses.

ACKNOWLEDGMENTS

We thank Sompong Vongpunsawad for excellent technical support, Patricia Devaux for sharing her Northern blot expertise, and Joe Felsenstein for information on his PHYLIP package. This work was supported by research grants of the Mayo and Siebens Foundation and by a research scholarship (SP 694/1-1) from the German research foundation (DFG) to C.S.

REFERENCES


