

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 Springfeld,¹ Gholamreza Darai,² and Roberto Cattaneo^{1*}

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.

Rhabdoviruses have been isolated from many different species of insects, vertebrates, and plants (15, 17). Of the more than 175 rhabdoviruses, only a minority have been characterized at the genomic and transcription level. The *Rhabdoviridae* belong to the order *Mononegavirales*, which also includes the virus families *Paramyxoviridae*, *Filoviridae*, and *Bornaviridae*. The rhabdoviruses infecting mammals are currently classified in the three genera *Vesiculovirus*, *Lyssavirus*, and *Ephemerovirus* (63), but only members of the genus *Lyssavirus*, including *Rabies virus*, were believed to cause serious human disease. The prototype of the *Vesiculovirus* genus, *Vesicular stomatitis virus*, can infect humans but occasionally causes only a mild, flu-like disease (57). However, a recent investigation of an encephalitis outbreak in children in the Indian state of Andhra Pradesh, where 183 of 329 affected children died, implicated *Chandipura virus*, a related member of the genus *Vesiculovirus*, as the most likely causative agent (54), demonstrating that rhabdoviruses other than the lyssaviruses can cause a deadly disease in humans.

Chandipura virus has been isolated in Africa and India and is probably transmitted to humans by sandflies (4). The *Chandipura virus* outbreak, together with the severe acute respiratory syndrome coronavirus epidemic (50) and the Nipah (9) and Hendra (3) virus outbreaks in Malaysia and Australia, all of which involved hitherto unknown animal viruses, demonstrate the value of characterizing animal pathogens, even if their pathogenic potential is currently unknown.

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

* Corresponding author. Mailing address: Mayo Clinic Rochester, Molecular Medicine Program, Guggenheim 1838, 200 First Street SW, Rochester, MN 55902. Phone: (507) 284-0171. Fax: (507) 266-2122. E-mail: Cattaneo.Roberto@mayo.edu.

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 tupaia 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 tupaia 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 second strand was synthesized by T4 polymerase. The cDNA was digested with RsaI, and two different specific adapters were ligated to two aliquots of cDNA from the infected cells (tester cDNA). The tester 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 hybridizes 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 tupaia baby fibroblast cells by ultracentrifugation (28,000 rpm, TH-641 rotor [Sorvall, 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 tupaia baby fibroblast cells was isolated with the RNeasy 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'- and 3'-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 phylograms were visualized with TREEVIEW (49).

Nucleotide sequence accession number. The complete TRV genome sequence has been deposited in the GenBank database under accession number AY840978.

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 rhabdovirus 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 tupaia 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 rhabdovirus 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 rhabdovirus 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-

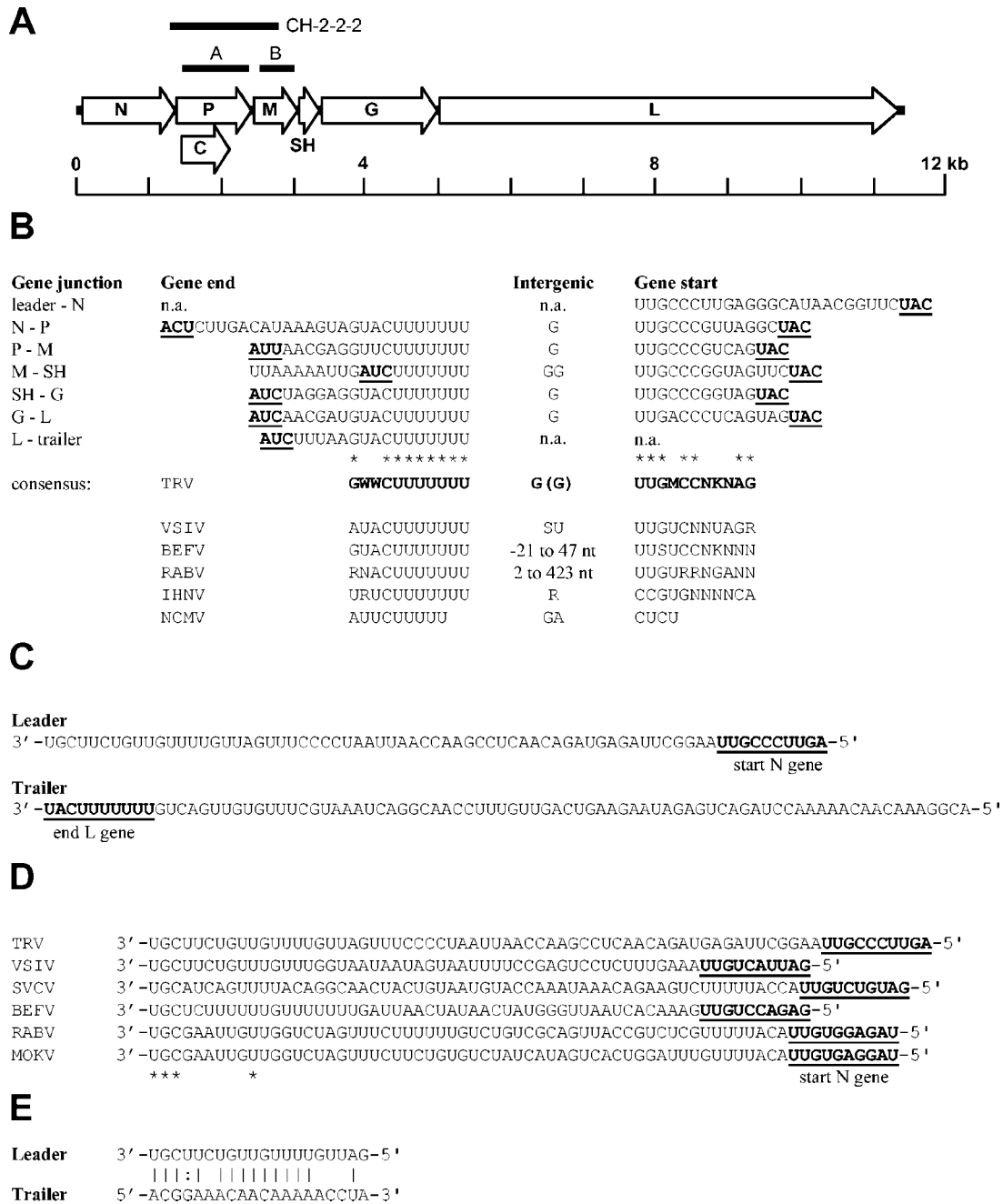


FIG. 1. TRV genome organization and transcription/replication control sequences. (A) Map of the TRV genome. The black bar represents the 11,440-nucleotides single-stranded TRV antigenome, the seven open arrows indicate the position of the open reading frames of the N, P, C, M, SH, G, and L proteins. 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.

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.

TABLE 1. Sequence identity and similarity of the tupaia rhabdovirus proteins with those of other rhabdoviruses

Virus ^a	Genus	Identity ^b (%)									
		N		P		M		G		L	
		I	S	I	S	I	S	I	S	I	S
VSIV	<i>Vesiculovirus</i>	29	17	9	9	12	10	18	14	40	18
FLAV	Unclassified	31	17	6	11	16	16	15	15	38	17
OITAV	Unclassified	30	15	n.a. ^c	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
BEFV	<i>Ephemerovirus</i>	31	21	11	11	13	14	13	13	36	16
SIGMAV	Unclassified	24	16	8	9	10	15	19	13	n.a.	n.a.
RABV	<i>Lyssavirus</i>	20	19	10	13	10	19	16	15	31	17
IHNV	<i>Novirhabdovirus</i>	8	15	9	10	8	15	14	13	18	15
NCMV	<i>Cytorhabdovirus</i>	10	14	13	9	11	13	9	12	16	15
SYNV	<i>Nucleorhabdovirus</i>	9	13	11	12	8	10	10	10	16	16

^a For abbreviations, see the legend to Fig. 5.

^b Sequence identity (I) and similarity (S) (Blosum similarity matrix) of the proteins were calculated with the Mac Vector program (Accelrys, San Diego, CA) after pairwise ClustalW alignments using the default settings. The highest values are shown in bold type.

^c n.a., sequence not available.

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 GWWCU₇ (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 protein, 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 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 VSNJV (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

A

```

C 1           M T S Q R F Q E L L K N V K T
P 1           M T D P K K I S R V A L K G Y D L T K V S R A L E E C E D
1   AACGGGCAAUCCGAUGACUGACCCUAAGAAUAUUCAAGAUUGCCUCUAAGGGGUAUGACCUCACAAAGGUUUAAGAGGCTUCUUGAAGAAUGUGAAGAC

C 16  N L G S V W V E F R S S V E E R L K H P H I W K I I D R E Q L A T
P 30  Q P G V R V G G V Q I V S G G E A E A P A H L E D N R Q G A I G D N
101  CAACUGGGGUCCGUGUGGUGGAGUUCAGAUUCGUGGAGGAGGCGAAGCACCCGCAUUUGGAAGAUAAUGACAGGGGCAAUUGCGGACA

C 49  I L M F F L G I P K M F R L S Q N M K I R Q M R N M R K P G V M E M
P 64  F D V F S R D S K N V Q I V T E H E D S S D E E Y E E A R S Y G D
201  AUUUUGAUGUUUUUCCGGGAUCCAAAAUGUUUCAGAUUGUCAAGAAUAUGUUCAGAUUGAGAAUUGAGGAAGCCAGGAGUUAUGGAGA

C 83  T P D T N Q T E Q P C P S Q M T N T E T S K E P E S V S L N S S T
P 97  D S G H Q P N R A A V P F A D D E Y G D F K R A R E R L P Q L F H
301  UGACUCCGGACACCAACCAACCAGCAGCCGUGCCUUCGACAGUACGAAUAACGGAGACTUCAAAAGACCAGAGAGGUCUCCUCAACUCUCCAC

C 116 K D G D L D P N K Q M K N L R R D S S S S Q E S A D L E E D H R M
P 130 E G W G F G S Q Q A N E E S S E G L Q Q Q S G E C G S G G G S S N A
401  GAAGGAUGGGGAUUGGAUCCCAACAAAGCAAUAUGAAGAAUUCGAGGAGUCUCCAGCAGCAGUACAGAGAGUGCGGAUCUGGAGGAGUUAUGGAAU

C 149 L G G R M R M I F K R R Q G M R K K S Q S L S H L R S Q A V Y Q D T
P 164 G G K D E D D L Q E K T G H E K E K S V I V P P K E S G S V S G Y
501  CUGGGGGAAGGAUGAUGAUUCUUAAGAGAAGACAGGGCAUGAGAAAGAAAAGUCAGUCAUUGUCCCAACCUAAGAGUCAGGCAGUGUAUCAGGAUA

C 183 T E A A I S P S T W K S F W P P Q L R L R D D N S W S S V N L S Q
P 197 H G S C D I P N R A A V P F L A S P A T T Q R R Q L L E L C Q L I A
601  CCACGGAAGCUGCAUAUCCCUUACAAUGGAAGAUUUCUGGCCUCCCCAGCUACGACUACAGAGACGACAAACUCCUGGAGUCUGUCAACUUAUGCA

C 216 T D P T K N
P 230 D R S N E E L I F P P W G F N L V K R K V H R V E P Q Q P V A V S H
701  GACAGUCCAAAGAAUUGAUUCUUCUCCUUGGGGAUUCACUUGGUUAAGAGAAAGUUCACAGAGUUGAACCUACAGCAGCCUGUAGCUGUAUCUC

P 264 R F T W E D F Q L K L K A G F T L I H K K T K A P V V L N S S S Y
801  ACAGGUUACCCUGGGAAGAUUUUACGUGAUAUAAGCAGGUUUUACCCUGAUUAAGAAAGAAUUAAGCAGUUGUGUGGAAUJCCAGCAGCUA

P 297 N L G S V P E G G I S L S P D D T E L S V L I R C L R Y L G L Y K
901  CAACUGGGGAUCUGUCCUUGAAGCGGAUUCUCUAUUCUCCAGAUAGACCCGAGUUGCGGUUCUGAUUAAUGCUUGAGGUUCUUAUGGCCUUAUAAA

P 330 F L A T Q I E F
1001 UJCCUAGCCACCAAUUGAAUUCUAAUUGCUCCAAGAAAAA
    
```

B

```

TRV      MTSQRPELLKNVKTNLGWSVVEFRSSVEERLKHPIWK-IIDREQLATILMFFLGIPKMFRLSQNMKIRQMNRMRKPGVEMEMPTDNTQTEQPCP 94
VSNJV    -----MIIWILPCRMPLNLRKDERINISKTSSSKIKEINQLRHIIRKKNRQIQILIIMLNLRCCHRMKE 65
VSIV     -----MRSKHNELKSPIMSCSKRMWKSILSPLIFRQQMILTQNLNQLKTIKVCMBRITQRLSKLKYRGL 67
PIRY     -----MKWK-SRQKNHSLLSRPLYLPQTGRVLSIFSGAC 32
CHAPV    -----MTQSQNHPSVVLQREFFAITWLNKNWNRKIRKKMMTIYLLKYRILLTLCWKRLLWMLRMLTRYTLRPNRLGNR 80

TRV      SQMTNTETSKEPESVSLNSSTKDGDLDPNKQMKNLRRDSSSSQESADLEEDHRMLGGRMRMIFRRRQGMRRKKSQSLSHLSRQAVYQDTTEAAISP 189

TRV      STWKSFWPPQLRLRDDNSWSSVNLSTQDPTKN 221
    
```

FIG. 2. Predicted sequence of the proteins encoded in the TRV P gene mRNA (A) and comparison of the TRV C protein to other rhabdovirus C proteins (B). In (A) the start and stop codons of the P and C proteins are underlined and in bold type, and the deduced amino acid sequences (one-letter code) are presented above the nucleotide sequence. The alignment in B was created with the ClustalX program (59) using the default settings except for a gap penalty of 25 instead of 10. The basic amino acids arginine, lysine, and histidine are shown in bold type. TRV, tupaia rhabdovirus; VSNJV, vesicular stomatitis virus New Jersey; VSIV, vesicular stomatitis Indiana; PIRY, Piry virus; CHAPV, Chandipura virus. The accession numbers of the sequences are given in Fig. 5.

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 open 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 apopto-

sis (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') 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 certain 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 NXS/T (N283, N330, and N462).

Walker and Kongsuwan 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 polyA 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

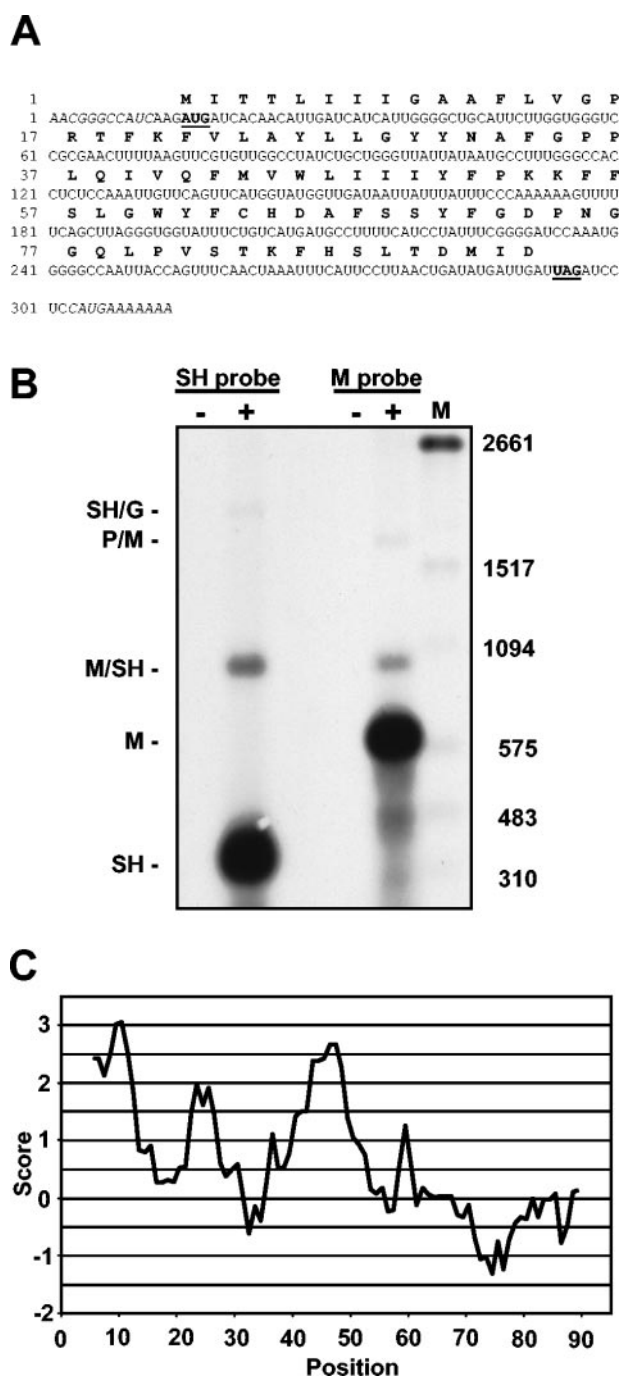


FIG. 3. Nucleotide sequence of the TRV SH gene mRNA (A), Northern blot analysis of TRV infected cells (B) and hydropathicity plot of the SH protein (C). In (A) the SH protein start and stop codons are underlined and in bold type, and the deduced amino acid sequences (one-letter code) is presented above the nucleotide sequence. In the Northern blot in B two probes against the TRV SH and M genes have been used as indicated above the lanes. Lanes containing mRNA from infected cells or uninfected cells are labeled with a plus or minus symbol, respectively, and the lane containing the molecular weight marker is marked with an "M". The sizes of the markers are given on the right, and the positions of the different TRV mRNAs are indicated on the left. The hydropathicity plot (C) was created with the program ProtScale on the proteomics server of the Swiss Institute of Bioinformatics according to the method of Kyte and Doolittle (40) with a window size of nine.

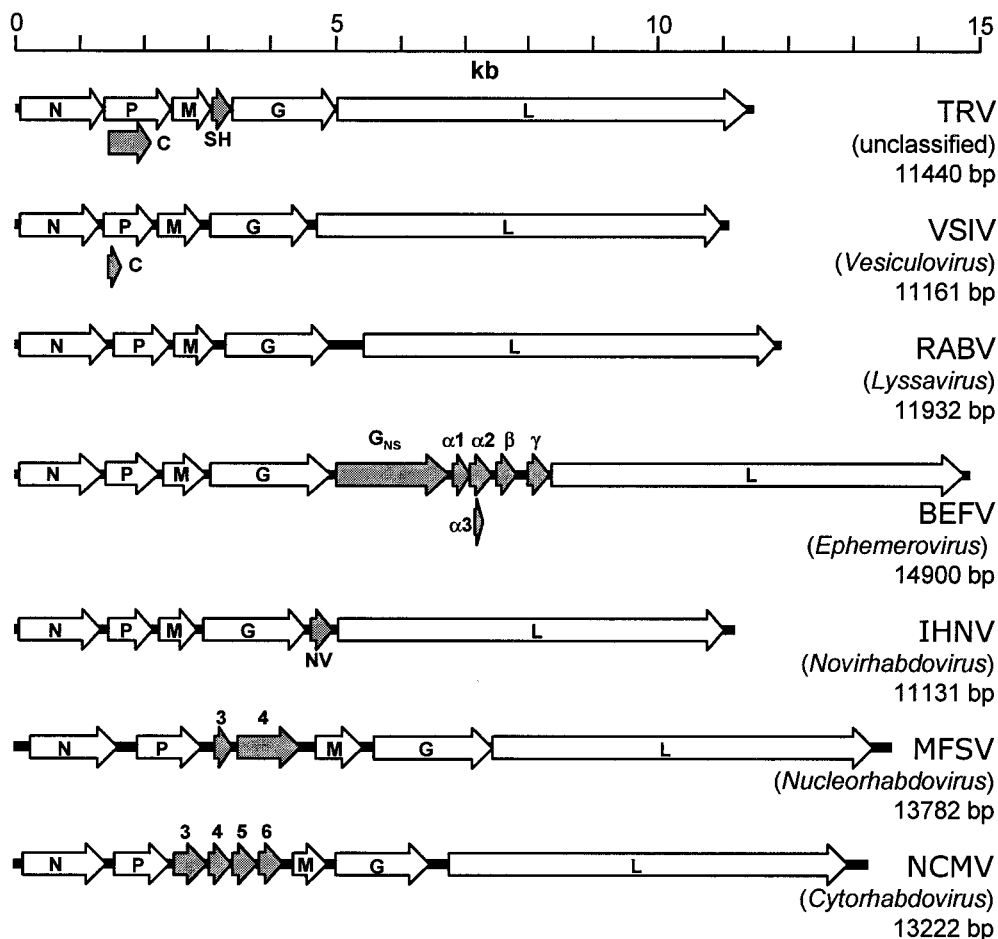


FIG. 4. Comparison of the TRV genome structure with the genome structure of representatives of the different rhabdovirus genera. The reading frames for the conserved rhabdovirus genes N, P, M, G, and L are depicted as open arrows, additional genes are shown in grey. The size of the genomes and the rhabdovirus genera are indicated. The abbreviations of virus names are the same as in Fig. 5.

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 proteins 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

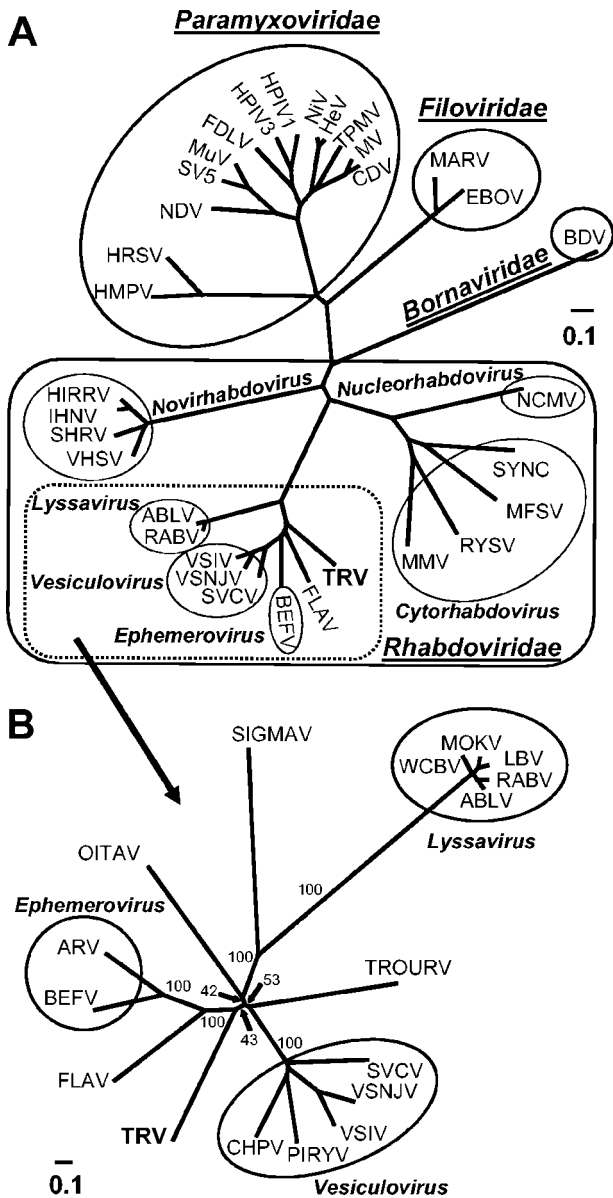


FIG. 5. Phylogenetic trees for the L proteins (conserved regions A to D) of selected members of the *Mononegavirales* (A) and N proteins of selected rhabdoviruses (B). The details of the construction of the phylogenetic trees are described in the Materials and Methods section. Family names are bolded, underlined, and italicized. Genera names within the *Rhabdoviridae* are bolded and italicized. The bar indicates 0.1 expected substitutions per site. In B, bootstrap values for the major branches from 1000 resampled data sets are indicated in percent. ABLV, Australian bat lyssavirus (accession number NC_003243); ARV, Adelaide river virus (AAC54627); BEFV, bovine ephemeral fever virus (NC_002526); BDV, Borna disease virus (NC_001607); CDV, canine distemper virus (NC_001921); CHPV, Chandipura virus (P11211); EBOV, Ebola virus (NC_002549); FDLV, fer-de-lance virus (NC_005084); FLAV, Flanders virus (AH012179); HeV, Hendra virus (NC_001906); HIPV1, human parainfluenza virus 1 (NC_003461); HIPV3, human parainfluenza virus 3 (NC_001796); HIRRV, Hirame rhabdovirus (NC_005093); HMPV, human metapneumovirus (NC_004148); HRSV, human respiratory syncytial virus (NC_001781); IHNV, infectious hematopoietic necrosis virus (NC_001652); LBV, Lagos bat virus (AAR03475); MARV, Marburg virus (NC_001608); MFSV, maize fine streak virus (NC_005974); MMV, maize mosaic virus (NC_005975); MOKV, Mokola virus (AAR03476); MuV, Mumps virus, (NC_002200); MV, measles virus (NC_001498); NCMV,

bottom). This protein is dispensable for replication in cell culture (33) but essential for pathogenicity (60).

The *Ephemerovirus* bovine ephemeral fever virus (Fig. 4, fourth from top) has a gene encoding a second, nonstructural glycoprotein (G_{NS}) downstream of the G gene (64) as well as a complex G_{NS} -L intergenic region predicted to encodes five putative small polypeptides on two mRNAs (43). The plant viruses of the *Nucleorhabdovirus* and *Cytorhabdovirus* genera have one to four additional genes between the P and M genes (Fig. 4, bottom), and the recently characterized Flanders virus also has additional genes at that position (not shown, direct submission to GenBank, accession number AF523199).

Expanding rhabdovirus family. Rhabdoviruses infect insects, mammals, fish, and plants, and this broad host range is reflected in their great genetic diversity. Although nearly all rhabdoviruses share the typical bullet-shaped morphology there is only little sequence similarity between the proteins of the more distantly related viruses, e.g., the vesiculoviruses and the plant viruses. Table 1 shows the identity and similarity values of the different TRV proteins to homologous proteins from representative rhabdoviruses of other genera. L and N are the most conserved proteins in the rhabdovirus family. For the TRV N protein, approximately 30% amino acid identities are found with members of the *Vesiculovirus* and *Ephemerovirus* genera but also the with the recently characterized Flanders virus and Oita virus. TRV L has the highest identities (40%) with the VSIV protein with slightly lower values for Flanders virus and bovine ephemeral fever virus. The low identity values for the P, M, and G proteins are not above significance levels, and no sequence similarity was detected for the TRV C and SH proteins to other proteins in the databases.

A phylogenetic tree derived from the conserved regions A to D of all available rhabdovirus L sequences and selected members of the *Mononegavirales* generated by the Fitch-Margoliash method (22) illustrates the large sequence territory occupied by the rhabdoviruses compared to the other virus families in this order (Fig. 5A). The *Rhabdoviridae* are considerably more diverse than the *Paramyxoviridae*. TRV is most closely related to the vertebrate and insect viruses in the *Ephemerovirus*, *Vesiculovirus*, and *Lyssavirus* genera.

Figure 5B addresses the phylogenetic origin of TRV 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 phylogram shows that TRV cannot be assigned to one of the current genera, although it was tentatively assigned to the *Vesiculovirus*

northern cereal mosaic virus (NC_002251); NDV, Newcastle disease virus (NC_002617); NiV, Nipah virus (NC_002728); OITAV, Oita virus (BAD13431); PIRYV, Piry virus (VHVNPV); RABV, rabies virus (NC_001542); RYSV, rice yellow stunt virus (NC_003746); SHRV, snakehead rhabdovirus (NC_000903); SIGMAV, Sigma virus (CAA54327); SV5, simian virus 5 (Q88434); SVCV, spring viremia of carp virus (NC_002803); SYNIV, Sonchus yellow net virus (NC_001615); TPMV, tupaia paramyxovirus (NC_002199); TROURV, trout rhabdovirus 903/87 (AAL35756); TRV, tupaia rhabdovirus; VHSV, viral hemorrhagic septicemia virus (NC_000855); VSIV, vesicular stomatitis Indiana virus (NC_001560); VSIV, Vesicular stomatitis New Jersey virus (P04881); WCBV, west Caucasian bat virus (AAR03481).

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 develop molecular detection methods for TRV and related rhabdoviruses that can be used to investigate unclear infectious diseases in humans and animals.

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

1. Arnason, U., J. A. Adegoke, K. Bodin, E. W. Born, Y. B. Esa, A. Gullberg, M. Nilsson, R. V. Short, X. Xu, and A. Janke. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc. Natl. Acad. Sci. USA* **99**:8151–8156.
2. Bahr, U., and G. Darai. 2001. Analysis and characterization of the complete genome of tupaia (tree shrew) herpesvirus. *J. Virol.* **75**:4854–4870.
3. Barclay, A. J., and D. J. Paton. 2000. Hendra (equine morbillivirus). *Vet. J.* **160**:169–176.
4. Bhatt, P. N., and F. M. Rodrigues. 1967. Chandipura: a new Arbovirus isolated in India from patients with febrile illness. *Indian J. Med. Res.* **55**:1295–1305.
5. Bras, F., D. Teninges, and S. Dezelee. 1994. Sequences of the N and M genes of the sigma virus of *Drosophila* and evolutionary comparison. *Virology* **200**:189–199.
6. Cao, J., E. B. Yang, J. J. Su, Y. Li, and P. Chow. 2003. The tree shrews: adjuncts and alternatives to primates as models for biomedical research. *J. Med. Primatol.* **32**:123–130.
7. Cathomen, T., B. Mrkic, D. Spehner, R. Drillien, R. Naef, J. Pavlovic, A. Aguzzi, M. A. Billeter, and R. Cattaneo. 1998. A matrix-less measles virus is infectious and elicits extensive cell fusion: consequences for propagation in the brain. *EMBO J.* **17**:3899–3908.
8. Chattopadhyay, D., and A. K. Banerjee. 1987. Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro. *Cell* **49**:407–414.
9. Chua, K. B., K. J. Goh, K. T. Wong, A. Kamarulzaman, P. S. Tan, T. G. Ksiazek, S. R. Zaki, G. Paul, S. K. Lam, and C. T. Tan. 1999. Fatal encephalitis due to Nipah virus among pig-farmers in Malaysia. *Lancet* **354**:1257–1259.
10. Coll, J. M. 1995. The glycoprotein G of rhabdoviruses. *Arch. Virol.* **140**:827–851.
11. Collins, P. L., and G. W. Wertz. 1985. The 1A protein gene of human respiratory syncytial virus: nucleotide sequence of the mRNA and a related polycistronic transcript. *Virology* **141**:283–291.
12. Conzelmann, K. K. 2004. Reverse genetics of mononegavirales. *Curr. Top. Microbiol. Immunol.* **283**:1–41.
13. Darai, G., B. Matz, R. M. Flügel, A. Gräfe, H. Gelderblom, and H. Delius. 1980. An adenovirus from tupaia (tree shrew): growth of the virus, characterization of viral DNA, and transforming ability. *Virology* **104**:122–138.
14. Darai, G., A. Schwaier, D. Komitowski, and K. Munk. 1978. Experimental infection of tupaia belangeri (tree shrews) with herpes simplex virus types 1 and 2. *J. Infect. Dis.* **137**:221–226.
15. de Mattos, C. A., C. C. de Mattos, and C. E. Rupprecht. 2001. Rhabdoviruses, p. 1245–1277. In D. M. Knipe and P. M. Howley (ed.), *Fields Virology*. Lippincott Williams & Wilkins, Philadelphia, Pa.
16. Devaux, P., and R. Cattaneo. 2004. Measles virus phosphoprotein gene products: conformational flexibility of the P/V protein amino-terminal domain and C protein infectivity factor function. *J. Virol.* **78**:11632–11640.
17. Dietzgen, R. G., P. J. Walker, Y. Gaudin, A. Flamand, J.-A. C. Leong, G. Kurath, M. M. Goodin, A. O. Jackson, G. W. Wertz, and L. A. Ball. 2002. Rhabdoviridae, p. 1049–1083. In C. A. Tidona and G. Darai (ed.), *The Springer index of viruses*. Springer Verlag, Heidelberg, Germany.
18. Einer-Jensen, K., T. N. Krogh, P. Roepstorff, and N. Lorenzen. 1998. Characterization of intramolecular disulfide bonds and secondary modifications of the glycoprotein from viral hemorrhagic septicemia virus, a fish rhabdovirus. *J. Virol.* **72**:10189–10196.
19. Elango, N., J. Kovamees, T. M. Varsanyi, and E. Norrby. 1989. mRNA sequence and deduced amino acid sequence of the mumps virus small hydrophobic protein gene. *J. Virol.* **63**:1413–1415.
20. Emerson, S. U., and Y. Yu. 1975. Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus. *J. Virol.* **15**:1348–1356.
21. Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
22. Fitch, W. M., and E. Margoliash. 1967. Construction of phylogenetic trees. *Science* **155**:279–284.
23. Flügel, R. M., H. Zentgraf, K. Munk, and G. Darai. 1978. Activation of an endogenous retrovirus from tupaia (tree shrew). *Nature* **271**:543–545.
24. Giorgi, C., B. M. Blumberg, and D. Kolakofsky. 1983. Sendai virus contains overlapping genes expressed from a single mRNA. *Cell* **35**:829–836.
25. Hansen, J. L., A. M. Long, and S. C. Schultz. 1997. Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure* **5**:1109–1122.
26. Harty, R. N., J. Paragas, M. Sudol, and P. Palese. 1999. A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding. *J. Virol.* **73**:2921–2929.
27. He, B., G. P. Leser, R. G. Paterson, and R. A. Lamb. 1998. The paramyxovirus SV5 small hydrophobic (SH) protein is not essential for virus growth in tissue culture cells. *Virology* **250**:30–40.
28. He, B., G. Y. Lin, J. E. Durbin, R. K. Durbin, and R. A. Lamb. 2001. The SH integral membrane protein of the paramyxovirus simian virus 5 is required to block apoptosis in MDBK cells. *J. Virol.* **75**:4068–4079.
29. Hiebert, S. W., R. G. Paterson, and R. A. Lamb. 1985. Identification and predicted sequence of a previously unrecognized small hydrophobic protein, SH, of the paramyxovirus simian virus 5. *J. Virol.* **55**:744–751.
30. Hiebert, S. W., C. D. Richardson, and R. A. Lamb. 1988. Cell surface expression and orientation in membranes of the 44-amino-acid SH protein of simian virus 5. *J. Virol.* **62**:2347–2357.
31. Iwasaki, T., S. Inoue, K. Tanaka, Y. Sato, S. Morikawa, D. Hayasaka, M. Moriyama, T. Ono, S. Kanai, A. Yamada, and T. Kurata. 2004. Characterization of Oita virus 296/1972 of Rhabdoviridae isolated from a horseshoe bat bearing characteristics of both lyssavirus and vesiculovirus. *Arch. Virol.* **149**:1139–1154.
32. Johansson, T., S. Nylund, N. J. Olesen, and H. Björklund. 2001. Molecular characterization of the nucleocapsid protein gene, glycoprotein gene and gene junctions of rhabdovirus 903/87, a novel fish pathogenic rhabdovirus. *Virus Res.* **80**:11–22.
33. Johnson, M. C., B. E. Simon, C. H. Kim, and J. A. Leong. 2000. Production of recombinant snakehead rhabdovirus: the NV protein is not required for viral replication. *J. Virol.* **74**:2343–2350.
34. Kall, L., A. Krogh, and E. L. Sonnhammer. 2004. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**:1027–1036.
35. Kouznetzoff, A., M. Buckle, and N. Tordo. 1998. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. *J. Gen. Virol.* **79**:1005–1013.
36. Kretzschmar, E., R. Peluso, M. J. Schnell, M. A. Whitt, and J. K. Rose. 1996. Normal replication of vesicular stomatitis virus without C proteins. *Virology* **216**:309–316.
37. Kurath, G., and J. C. Leong. 1985. Characterization of infectious hemato-

- poietic necrosis virus mRNA species reveals a nonvirion rhabdovirus protein. *J. Virol.* **53**:462–468.
38. **Kurz, H. W.** 1984. Isolation und Charakterisierung von tupaia Paramyxovirus und Rhabdovirus. Medical doctoral thesis. Heidelberg University, Heidelberg, Germany.
 39. **Kurz, W., H. Gelderblom, R. M. Flügel, and G. Darai.** 1986. Isolation and characterization of a tupaia rhabdovirus. *Intervirology.* **25**:88–96.
 40. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**:105–132.
 41. **Lin, Y., A. C. Bright, T. A. Rothermel, and B. He.** 2003. Induction of apoptosis by paramyxovirus simian virus 5 lacking a small hydrophobic gene. *J. Virol.* **77**:3371–3383.
 42. **Martin, R. D.** 1990. Primate origins and evolution: a phylogenetic reconstruction. Princeton University Press, Princeton, N.J.
 43. **McWilliam, S. M., K. Kongsuwan, J. A. Cowley, K. A. Byrne, and P. J. Walker.** 1997. Genome organization and transcription strategy in the complex GNS-L intergenic region of bovine ephemeral fever rhabdovirus. *J. Gen. Virol.* **78**:1309–1317.
 44. **Mould, J. A., R. G. Paterson, M. Takeda, Y. Ohigashi, P. Venkataraman, R. A. Lamb, and L. H. Pinto.** 2003. Influenza B virus BM2 protein has ion channel activity that conducts protons across membranes. *Dev. Cell.* **5**:175–184.
 45. **Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally, C. J. Douady, E. Teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, and M. S. Springer.** 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**:2348–2351.
 46. **Nagai, Y., and A. Kato.** 2004. Accessory genes of the paramyxoviridae, a large family of nonsegmented negative-strand RNA viruses, as a focus of active investigation by reverse genetics. *Curr. Top. Microbiol. Immunol.* **283**:197–248.
 47. **Okamoto, H., T. Nishizawa, M. Takahashi, A. Tawara, Y. Peng, J. Kishimoto, and Y. Wang.** 2001. Genomic and evolutionary characterization of TT virus (TTV) in tupaia and comparison with species-specific TTVs in humans and non-human primates. *J. Gen. Virol.* **82**:2041–2050.
 48. **O'Reilly, E. K., and C. C. Kao.** 1998. Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. *Virology* **252**:287–303.
 49. **Page, R. D.** 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**:357–358.
 50. **Peiris, J. S., K. Y. Yuen, A. D. Osterhaus, and K. Stöhr.** 2003. The severe acute respiratory syndrome. *N. Engl. J. Med.* **349**:2431–2441.
 51. **Peluso, R. W., J. C. Richardson, J. Talon, and M. Lock.** 1996. Identification of a set of proteins (C' and C) encoded by the bicistronic P gene of the Indiana serotype of vesicular stomatitis virus and analysis of their effect on transcription by the viral RNA polymerase. *Virology* **218**:335–342.
 52. **Pinto, L. H., L. J. Holsinger, and R. A. Lamb.** 1992. Influenza virus M2 protein has ion channel activity. *Cell* **69**:517–528.
 53. **Poch, O., B. M. Blumberg, L. Bougueleret, and N. Tordo.** 1990. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J. Gen. Virol.* **71**:1153–1162.
 54. **Rao, B. L., A. Basu, N. S. Wairagkar, M. M. Gore, V. A. Arankalle, J. P. Thakare, R. S. Jadhav, K. A. Rao, and A. C. Mishra.** 2004. A large outbreak of acute encephalitis with high fatality rate in children in Andhra Pradesh, India, in 2003, associated with Chandipura virus. *Lancet* **364**:869–874.
 55. **Schmitt, A. P., and R. A. Lamb.** 2004. Escaping from the cell: assembly and budding of negative-strand RNA viruses. *Curr. Top. Microbiol. Immunol.* **283**:145–196.
 56. **Schöndorf, E., U. Bahr, M. Handermann, and G. Darai.** 2003. Characterization of the complete genome of the tupaia (tree shrew) adenovirus. *J. Virol.* **77**:4345–4356.
 57. **Shope, R. E., and R. B. Tesh.** 1987. The Ecology of Rhabdoviruses That Infect Vertebrates. *In* R. R. Wagner (ed.), *The rhabdoviruses*. Plenum Press, New York, NY.
 58. **Spiropoulou, C. F., and S. T. Nichol.** 1993. A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus. *J. Virol.* **67**:3103–3110.
 59. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
 60. **Thouloze, M. I., E. Bouguyon, C. Carpentier, and M. Bremont.** 2004. Essential role of the NV protein of Novirhabdovirus for pathogenicity in rainbow trout. *J. Virol.* **78**:4098–4107.
 61. **Tidona, C. A., H. W. Kurz, H. R. Gelderblom, and G. Darai.** 1999. Isolation and molecular characterization of a novel cytopathogenic paramyxovirus from tree shrews. *Virology* **258**:425–434.
 62. **Van Ranst, M.** 2004. Chandipura virus: an emerging human pathogen? *Lancet* **364**:821–822.
 63. **van Regenmortel, M. H. V., C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner.** 2000. Virus taxonomy: the classification and nomenclature of viruses. The Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, CA.
 64. **Walker, P. J., K. A. Byrne, G. A. Riding, J. A. Cowley, Y. Wang, and S. McWilliam.** 1992. The genome of bovine ephemeral fever rhabdovirus contains two related glycoprotein genes. *Virology* **191**:49–61.
 65. **Walker, P. J., and K. Kongsuwan.** 1999. Deduced structural model for animal rhabdovirus glycoproteins. *J. Gen. Virol.* **80**:1211–1220.
 66. **Walter, E., R. Keist, B. Niederost, I. Pult, and H. E. Blum.** 1996. Hepatitis B virus infection of tupaia hepatocytes in vitro and in vivo. *Hepatology* **24**:1–5.
 67. **Whelan, S. P., J. N. Barr, and G. W. Wertz.** 2004. Transcription and replication of nonsegmented negative-strand RNA viruses. *Curr. Top. Microbiol. Immunol.* **283**:61–119.
 68. **Whitney, E.** 1964. Flanders Strain, an Arbovirus Newly Isolated from Mosquitoes and Birds of New York State. *Am. J. Trop. Med. Hyg.* **13**:123–131.
 69. **Xie, Z. C., J. I. Riezu-Boj, J. J. Lasarte, J. Guillen, J. H. Su, M. P. Civeira, and J. Prieto.** 1998. Transmission of hepatitis C virus infection to tree shrews. *Virology* **244**:513–520.