Molecular Characterization of the Polymerase Gene and Genomic Termini of Nipah Virus


*Respiratory and Enteric Viruses Branch and †Special Pathogens Branch, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333

Received April 2, 2001; returned to author for revision May 7, 2001; accepted May 24, 2001; published online July 23, 2001

In 1998, Nipah virus (NV) emerged in peninsular Malaysia, causing fatal encephalitis in humans and a respiratory disease in swine. NV is most closely related to Hendra virus (HV), a paramyxovirus that was identified in Australia in 1994, and it has been proposed that HV and NV represent a new genus within the family Paramyxoviridae. This report describes the analysis of the sequences of the polymerase gene (L) and genomic termini of NV as well as a comparison of the full-length, genomic sequences of HV and NV. The L gene of NV is predicted to be 2244 amino acids in size and contains the six domains found within the L proteins of all nonsegmented, negative-stranded (NNS) RNA viruses. However, the GDNQ motif found in most NNS RNA viruses was replaced by GDNE in both NV and HV. The 3′ and 5′ termini of the NV genome are nearly identical to the genomic termini of HV and share sequence homology with the genomic termini of other members of the subfamily Paramyxovirinae. At 18,246 nucleotides, the genome of NV is 12 nucleotides longer than the genome of HV and they have the largest genomes within the family Paramyxoviridae. The comparison of the structures of the genomes of HV and NV is now complete and this information will help to establish the taxonomic position of these novel viruses within the family Paramyxoviridae.

INTRODUCTION

Nipah virus (NV) is an enveloped, nonsegmented, negative-stranded (NNS) RNA virus that is an unclassified member of the family Paramyxoviridae. NV was isolated in March 1999 and subsequently identified as the etiologic agent responsible for an outbreak of fatal viral encephalitis in humans and a respiratory illness and a central nervous system disease in swine in Malaysia and Singapore. Initial studies indicated that NV is antigenically and genetically closely related to Hendra virus (HV), another unclassified member of the family Paramyxoviridae (Anonymous, 1999a,b; Chua et al., 1999, 2000; Daniels et al., 2001; Paton et al., 1999). HV, formerly known as equine morbillivirus, was first isolated during an outbreak of a severe respiratory illness that killed 16 horses and two humans in Queensland, Australia in 1994 (Murray et al., 1995; Rogers et al., 1996; Selvey and Sheridan, 1994).

The ultrastructural features and genetic organization of HV and NV indicate that they are members of the subfamily Paramyxovirinae within the family Paramyxoviridae (Chua et al., 2000; Goldsmith et al., 2000; Harcourt et al., 2000; Hyatt and Selleck, 1996; Hyatt et al., 2001; Murray et al., 1995). This subfamily is divided into three genera, the Respiroviruses, the Morbilliviruses, and the Rubulaviruses, but phylogenetic analysis has shown that neither HV nor NV group within these established genera (Gould, 1996; Harcourt et al., 2000; Murray et al., 1995; Wang et al., 2001, 1998, 2000 Yu et al., 1998a,b). In addition, HV and NV are cross-reactive in immunofluorescence assays (Anonymous, 1999a) and are cross-neutralizing (Harcourt et al., 2000) but are not serologically cross-reactive with other members of the family Paramyxoviridae (Chua et al., 2000; Murray et al., 1995). For these reasons, HV and NV have been proposed to represent a fourth genus within this subfamily (Harcourt et al., 2000; Wang et al., 1998, 2000 Yu et al., 1998a). In addition to HV and NV, a number of unclassified paramyxoviruses have recently been described, including Tupaia paramyxovirus (TPMV) (Tidona et al., 1999), Salem virus (Renshaw et al., 2000), and several reptilian paramyxoviruses (Ahne et al., 1999; Manvell et al., 2000). The discovery of these novel paramyxoviruses highlights the tremendous amount of diversity present within the family Paramyxoviridae and emphasizes challenges faced by viral taxonomists.

We previously reported the characterization of the nucleocapsid (N), phosphoprotein (P), matrix (M), fusion (F), and attachment glycoprotein (G) genes of NV (Harcourt et al., 2000). In this report we describe the sequences of the L gene and genomic termini that allowed us to analyze the complete genome of NV.
RESULTS

Genetic characterization of the L gene of NV

Inspection of an alignment of the nucleotide sequences of the L genes of HV and several other paramyxoviruses identified the regions with the greatest homology. Within these homologous regions, RT-PCR primers based on HV sequences were synthesized and used to amplify sections of the L gene of NV. Once several NV-specific amplimers were obtained, they were used as starting points to "walk up and down" the L gene, using additional RT-PCR primers derived from either HV or NV sequences. The completed sequence of the L gene of NV was derived from the sequences obtained from various amplimers as well as from cloned cDNA fragments. The complete nucleotide sequence of NV has been deposited in GenBank (Accession No. AF212302).

The L gene of NV is 6955 nucleotides in length, which is identical to the size of the L gene of HV (Wang et al., 2000). On the basis of nucleotide homology to HV and other viruses of the subfamily Paramyxovirinae, the AUG codon at nucleotide 154 is predicted to initiate translation of a 2244 amino acid protein (Fig. 1). The sizes of the 3′ (153 nucleotides) and 5′ (67 nucleotides) nontranslated regions (NTRs) of NV are identical to those of HV. The open reading frame (ORF) of the NV L gene has a homology of 73% to HV at the nucleotide level and 87% identity at the amino acid level, while the 3′ and 5′ NTRs of NV have homologies of 64 and 58%, respectively, to the same regions of HV. The transcriptional start and stop signals of the L gene of NV are identical to those found in the G gene of NV (Harcourt et al., 2000).

The L protein is predicted to have a molecular mass of 257,188 Da and a net charge of −1. The six linear domains that are found within the L proteins of all NNS RNA viruses can also be identified in NV (Poch et al., 1990) (Fig. 1). It is not surprising that the areas of greatest homology between NV, HV, and other members of the subfamily Paramyxovirinae can be found within these six domains (Fig. 1). Within these domains, HV and NV share 93% identity. The 108 amino acids between domains II and III constitute the area of least conservation, with only 45% identity between HV and NV. This hinge region is particularly variable among all NNS RNA viruses. An apparent lack of homology between NV and the published sequence of HV (GenBank Accession No. AF017149) within a 27 amino acid area of domain V (residues 1299 to 1325) was resolved after resequencing part of the L gene of HV (Fig. 1).

The domain with the highest degree of conservation among the L proteins of viruses in the order Mononegavirales is domain III (Poch et al., 1990). This domain contains four motifs (A–D) conserved between RNA-dependent RNA polymerases of negative-stranded, segmented and NNS RNA viruses, positive-stranded RNA viruses, retrotransposons, and retroviruses (Poch et al., 1990) (Fig. 2). While these four motifs are conserved in NV, motif C contains an interesting substitution. The sequence, GDNQ, is highly conserved within motif C of viruses in the order Mononegavirales; however, this sequence is GDNE in NV, HV (Wang et al., 2000), and TPMV (GenBank Accession No. AF079780).

The L protein of NV contains a putative ATP-binding site motif G-E-G-S-G and a K residue 21 amino acids proximal to this G-rich sequence. This site is identical to the putative ATP binding sites found in HV, TPMV, and the Morbilliviruses, measles virus (MV), canine distemper virus (CDV), rinderpest virus (RPV), and phocine distemper virus (PDV) (Fig. 3). This motif can also be found within domain VI of other members of the order Mononegavirales (Fig. 3), and the spacing between the K residue and the G-rich sequence is identical within the family Paramyxoviridae and varies only slightly in the Filoviridae and Rhabdoviridae.

Phylogenetic analysis based on the amino acid sequences of the L proteins of various members of the order Mononegavirales demonstrated that HV and NV form a unique cluster within the subfamily Paramyxovirinae that was more closely related to the Morbilliviruses and TPMV than to any of the other genera (Fig. 4). Bootstrap confidence levels were 100% for the unique cluster containing HV and NV. The phylogenetic analysis using the L protein mirrors the relationships described for the other genes of NV (Harcourt et al., 2000) and suggests that HV and NV should be considered as a separate genus.

Genetic characterization of the genomic termini of NV

The genomic termini of NV are nearly identical to the genomic termini of HV (Wang et al., 2000) and are closely related to the termini of other viruses in the subfamily Paramyxovirinae, particularly the Morbilliviruses and Respiroviruses (Fig. 5). For example, the first four nucleotides of the 3′ terminus, UGGU, are shared between all viruses in the subfamily Paramyxovirinae, with the exception of HV, which has a C at position 4. However, both NV and HV are unique in having a G at the fourth position from the 5′ terminus of the genome. As in the other paramyxoviruses, the termini of NV show a high degree of complementarity.

The distance from the 3′ terminus of the genome to the start of the first gene is remarkably well conserved throughout the order Mononegavirales. The 3′ terminus of the genome of NV is predicted to contain a leader sequence of 52 nucleotides followed by a trinucleotide intergenic sequence between the end of the leader sequence and the start of the nucleocapsid gene. The length of the 3′ terminus of NV is the same size as the 3′ terminus found in other viruses in the subfamily Paramyxovirinae. The intergenic sequence immediately following the gene stop signal of the L gene and the final
FIG. 1. Comparison of the L proteins of Nipah virus and Hendra virus. The alignment shows the predicted amino acid sequence of the L protein of NV and indicates positions that differ between the L proteins of NV and HV. HV sequence was obtained from GenBank (Accession No. AF017149) with the exception of the sequence for amino acids 1299–1325, 1358, and 1359, which differed from the published sequence. The published sequence has a deletion of a C at nucleotide position 15,293, insertion of a G at position 15,372, insertion of a T at position 15,470, and deletion of a C at position 15,474. Amino acids in bold type represent residues conserved when the L proteins of the following members of the subfamily Paramyxovirinae were aligned (abbreviations and accession numbers used are in parentheses): NV, HV, human parainfluenza virus 3 (HPIV-3) (M21649), Sendai virus (X58886), canine distemper virus (CDV) (AF014953), phocine distemper virus (PDV) (Y09630), measles virus (MV) (M20865), rinderpest virus (RPV) (M21649), Sendai virus (X98886), canine distemper virus (CDV) (AF014953), phocine distemper virus (PDV) (Y09630), measles virus (MV) (M20865), rinderpest virus (RPV) (M21649).
30 nucleotides of the genome of NV constitute the trailer sequence. Unlike the 3' terminus of the genome, the lengths of the 5' termini vary widely within the order Mononegavirales (Fig. 5).

Analysis of the complete genome of NV

The genome of NV is 18,246 nucleotides in length, 12 nucleotides longer than HV, making it the largest genome in the family Paramyxoviridae. The gene order of 3'-N-P-M-F-G-L-5' is also found in the Respiroviruses, Morbilliviruses, HV, and TPMV. The intergenic sequences bracketing each gene of NV are GAA, which are identical to the intergenic sequences of HV (Wang et al., 2000) and TPMV (GenBank Accession No. AF079780). These three viruses are unusual in that they are the only viruses in the family Paramyxoviridae that do not have any variability in their intergenic sequences.

With the exceptions of NV, HV, and TPMV (17,904 nucleotides), the average size of the genomes within the subfamily Paramyxovirinae is 15.5 kb. The increased size of the P ORF in addition to the unusually large sizes of the nontranslated regions, particularly the 3' NTRs, of many of the genes of NV and HV are responsible for the increased size of the NV and HV genomes relative to the other paramyxoviruses (Table 1). The genome sizes of both NV and HV are a multiple of 6 (3041 and 3039, respectively).

**TABLE 1**

<table>
<thead>
<tr>
<th>Virus</th>
<th>Genome Size (nucleotides)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nipah</td>
<td>18,246</td>
</tr>
<tr>
<td>Hendra</td>
<td>18,090</td>
</tr>
<tr>
<td>TPMV</td>
<td>17,904</td>
</tr>
<tr>
<td>HV</td>
<td>18,034</td>
</tr>
<tr>
<td>NDV</td>
<td>17,225</td>
</tr>
</tbody>
</table>

---

(Z30698), Tupaia paramyxovirus (TPMV) (AF079780), human parainfluenza virus 2 (HPIV-2) (X57559), simian virus 5 (SV5) (M81721), mumps (D10575), La Piedad Michoacan virus (LPMV) (X98125), and Newcastle disease virus (NDV) (X05399). For the grouping of conservative amino acids, see Materials and Methods. Amino acids were considered conserved if the residue was conserved in 13 of the 14 viruses compared. Boxed domains I–VI mark the amino acid domains conserved between the NNS RNA (Poch et al., 1990). Dots above the sequences mark lengths of 10 amino acids.
respectively), suggesting that these viruses follow the rule of six (Calain and Roux, 1993). It is interesting to note that size differences within genes between NV and HV were also in multiples of 6 (Table 1). In addition, the hexameric phasing for the initiating nucleotide for each gene of NV is identical to the phasing for HV (data not shown) (Wang et al., 2000).

**DISCUSSION**

With the determination of the sequences of the L gene and genomic termini of the recently discovered paramyxovirus, NV, the sequence of the entire NV genome is now known. It has been proposed that NV and the related virus, HV, should be designated as a new genus with the subfamily Paramyxovirinae and the availability of the entire genome sequences of each virus makes it possible to complete the genetic characterization of this proposed new genus. Phylogenetic analyses of each gene of NV and HV, together with comparison of the gene start and stop signals, and the genomic terminal sequences demonstrate that NV and HV are members of the subfamily Paramyxovirinae. These viruses are more closely related to the *Morbilliviruses* and the *Respiroviruses* but have several unique genetic features that clearly distinguish them from other paramyxoviruses.

The L protein of NV has the linear domain structure found within the polymerase proteins of all of the viruses in the order Mononegavirales. The polymerases of NNS viruses are associated with a number of enzymatic functions (Poch et al., 1990). The boxed sequence is the GDN sequence that is absolutely conserved between viruses of the order Mononegavirales. Gaps (indicated by dashes) were made in the alignment to maximize the regions of homology. Abbreviations and accession numbers not given elsewhere are respiratory syncytial virus (RSV), U39661; Ebola, X67110; and vesicular stomatitis virus (VSV), X00939.

**FIG. 2.** Domain III is the most conserved domain of the L protein of the order Mononegavirales. Amino acid residues in bold type are conserved in >50% of the aligned viruses. An asterisk represents an amino acid that is conserved in all or all but one of the aligned viruses. Conservative amino acid substitutions were allowed as discussed under Materials and Methods. Underlined segments A–D represent sequences that are shared by RNA-dependent RNA polymerases of negative-stranded, segmented, and NNS RNA viruses, positive-stranded RNA viruses, retrotransposons, and retroviruses (Poch et al., 1990). The boxed sequence is the GDN sequence that is absolutely conserved between viruses of the order Mononegavirales. Gaps (indicated by dashes) were made in the alignment to maximize the regions of homology. Abbreviations and accession numbers not given elsewhere are respiratory syncytial virus (RSV), U39661; Ebola, X67110; and vesicular stomatitis virus (VSV), X00939.
functions, including RNA polymerization (De and Banerjee, 1985; Hunt and Wagner, 1974), polyadenylation (Hunt et al., 1984), capping and methylation (Hercyk et al., 1988; Horikami and Moyer, 1982), and protein kinase activities (Gao and Lenard, 1995). The functionally important regions of the polymerase proteins are thought to reside in six linear domains that are conserved between all NNS viruses (Barik et al., 1990; Blumberg et al., 1988; Poch et al., 1990). Areas of conservation within the family Paramyxoviridae have been found outside the six linear domains delineated by Poch et al. These areas are conserved in NV and HV and show more similarities to the morbilliviruses and respiroviruses than to any other viruses (McIlhatton et al., 1997; Poch et al., 1990).

Of the six conserved linear domains, domain III is the most conserved among NNS viruses. In fact, domain III contains four motifs (A–D, see Fig. 2) that are conserved among the polymerases from a broad range of RNA viruses, including NNS viruses, negative-stranded segmented viruses, positive-stranded segmented and nonsegmented viruses, double-stranded RNA viruses, and retroviruses (Kamer and Argos, 1984; Poch et al., 1990; Ribas and Wickner, 1992). Of particular note is motif C, which has a highly conserved tripeptide sequence that may be required for cation binding, template specificity, and/or phosphodiester bond formation (Poch et al., 1990; Schnell and Conzelmann, 1995). In all known NNS viruses, with the exceptions of NV, HV, TPMV, and the rhabdovirus, infectious hematopoietic necrosis virus (IHNV), this tripeptide is a GDN motif followed by a Q residue. In HV, NV, and TPMV, the Q is replaced by an E while INHV has a V residue at this position (Le Mercier et al., 1997). Mutating the GDNQ motif to GDNN in the polymerases resulted in a complete loss of polymerase activity, whereas mutating the motif to GDNE resulted in a polymerase with very low activity (Schnell and Conzelmann, 1995). A GDNN to GDQNN mutation also resulted in a severe loss of polymerase activity using a VSV in vitro transcription assay system (Sleat and Banerjee, 1993). Although NV and HV contain the QDNE motif, what effect, if any, this motif has on polymerase function remains to be seen.

The K-X18–21-G-X-G-X-G motif within domain VI of all known NNS RNA viruses is highly conserved and closely resembles the consensus ATP-binding site, G-X-G-X-X-G-X15–20-K, found in a number of kinases (Barik et al., 1990).

FIG. 4. Phylogenetic analysis of the L proteins of members of the order Mononegavirales. Phylogenetic analysis using PAUP 4.02 was performed on the amino acid sequence of L proteins from various members of the order Mononegavirales. The tree shown was based on maximum parsimony; however, analysis of the same data using maximum likelihood produced a tree with nearly identical topology (data not shown). Scale representing the number of amino acid changes is shown at bottom left. Accession number used to predict the amino acid sequences not given elsewhere is human parainfluenza virus 1 (HPIV-1), AF117818.
This motif is identical in NV, HV, TPMV, and the Morbilliviruses. Mutational analysis of the G-rich motif in VSV demonstrated that it may also function as an S-adenosylmethionine binding site (J. Perrault, personal communication); thus, this highly conserved motif may have multiple functions.

The N protein of viruses of the subfamily Paramyxovirinae encapsidates viral genomic and antigenomic RNA and is thought to cover precisely six nucleotides (Egelman et al., 1989). As a consequence, paramyxoviruses replicate most efficiently when the length of their genomes are evenly divisible by 6 (Calain and Roux, 1993) and the "rule of six" has now been demonstrated for several paramyxoviruses (Bellini et al., 1998; Kolakofsky et al., 1998). The genome sizes of both HV and NV are divisible by 6; therefore, it is highly probable that these viruses follow the rule of six; however, experimental evidence for this is lacking. Within a genus, the transcription initiation start sites are highly conserved within a hexameric phase (Kolakofsky et al., 1998). Likewise, the hexameric phasing pattern for HV and NV is unique among the paramyxoviruses. The pattern for HV and NV is identical except

### TABLE 1
Comparison of the Sizes of the 5' Nontranslated Regions, Open Reading Frames, and 3' Nontranslated Regions of Nipah Virus, Hendra Virus, Tupaia Paramyxovirus, and Other Members of the Subfamily Paramyxovirinae

<table>
<thead>
<tr>
<th>Gene</th>
<th>Virus</th>
<th>5' NTR</th>
<th>ORF</th>
<th>3' NTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>NV</td>
<td>57</td>
<td>1599</td>
<td>586</td>
</tr>
<tr>
<td></td>
<td>HV</td>
<td>57</td>
<td>1599</td>
<td>568</td>
</tr>
<tr>
<td></td>
<td>TPMV</td>
<td>58</td>
<td>1659</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Morbillovirinae</td>
<td>52–96</td>
<td>1530–1650</td>
<td>43–111</td>
</tr>
<tr>
<td>P</td>
<td>NV</td>
<td>105</td>
<td>2130</td>
<td>469</td>
</tr>
<tr>
<td></td>
<td>HV</td>
<td>105</td>
<td>2124</td>
<td>469</td>
</tr>
<tr>
<td></td>
<td>TPMV</td>
<td>214</td>
<td>1584</td>
<td>346</td>
</tr>
<tr>
<td></td>
<td>Morbillovirinae</td>
<td>33–79</td>
<td>1176–1812</td>
<td>66–123</td>
</tr>
<tr>
<td>M</td>
<td>NV</td>
<td>100</td>
<td>1059</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>HV</td>
<td>100</td>
<td>1059</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>TPMV</td>
<td>30</td>
<td>1059</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>Morbillovirinae</td>
<td>32–36</td>
<td>1008–1134</td>
<td>61–426</td>
</tr>
<tr>
<td>F</td>
<td>NV</td>
<td>284</td>
<td>1641</td>
<td>412</td>
</tr>
<tr>
<td></td>
<td>HV</td>
<td>272</td>
<td>1641</td>
<td>418</td>
</tr>
<tr>
<td></td>
<td>TPMV</td>
<td>258</td>
<td>1662</td>
<td>375</td>
</tr>
<tr>
<td></td>
<td>Morbillovirinae</td>
<td>28–583</td>
<td>1590–1989</td>
<td>38–137</td>
</tr>
<tr>
<td>A</td>
<td>NV</td>
<td>233</td>
<td>1809</td>
<td>504</td>
</tr>
<tr>
<td></td>
<td>HV</td>
<td>233</td>
<td>1815</td>
<td>516</td>
</tr>
<tr>
<td></td>
<td>TPMV</td>
<td>158</td>
<td>1998</td>
<td>269</td>
</tr>
<tr>
<td></td>
<td>Morbillovirinae</td>
<td>20–78</td>
<td>1698–1854</td>
<td>66–111</td>
</tr>
<tr>
<td>L</td>
<td>NV</td>
<td>153</td>
<td>6735</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>HV</td>
<td>153</td>
<td>6735</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>TPMV</td>
<td>34</td>
<td>6813</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Morbillovirinae</td>
<td>8–47</td>
<td>6552–6786</td>
<td>34–137</td>
</tr>
</tbody>
</table>

**Notes:**
- Paramyxovirinae shows the ranges of sizes found in a comparison of the Respiroviruses, Sendai and HPIV-3, the Morbilliviruses, MV and CDV, and the Rubulaviruses, mumps and SIV.
- Abbreviations: N, nucleocapsid; P, phosphoprotein; M, matrix; F, fusion; A, G, H, or HN attachment glycoprotein; and L, polymerase. The SH gene of the Rubulaviruses is not shown.
- The wide range in this category is due to the large 3' NTRs of the Morbilliviruses.
- The wide range in this category is due to the large 5' NTR of MV.

1990; Fry et al., 1986). This motif is identical in NV, HV, TPMV, and the Morbilliviruses. Mutational analysis of the G-rich motif in VSV demonstrated that it may also function as an S-adenosylmethionine binding site (J. Perrault, personal communication); thus, this highly conserved motif may have multiple functions.
MATERIALS AND METHODS

Virus culture and RNA isolation

The NV strain used in this study was isolated directly from human brain tissue in Vero E6 cells. The virus was passaged in Vero E6 cells and harvested when the cytopathic effect was maximal. All work with live virus was performed under biosafety level 4 conditions at the Centers for Disease Control and Prevention in Atlanta, GA. RNA was extracted from the infected cells by using the guanidinium acid-phenol technique (Chomczynski and Sacchi, 1987).

Amplification of viral sequences

RT-PCR was done as previously described (Rota et al., 1994) except that the cycling conditions were 95°C for 3 min, followed by 35 cycles of 95°C for 1 min, 50°C for 2 min, and 72°C for 3 min. The reaction was then incubated at 72°C for an additional 5 min. PCR products were visualized using agarose gel electrophoresis and were purified by use of the Wizard PCR Prep DNA Purification System (Promega, Madison, WI).

Cloned fragments of the L ORF were obtained using RT-PCR (Rota et al., 1994). AMV reverse transcriptase (Roche Molecular Biochemicals, Indianapolis, IN) was used, and PCR was done with the Elongase enzyme mix (Life Technologies, Gaithersburg, MD) using the manufacturer’s recommendations for long PCR. The L ORF was amplified in two overlapping fragments of 3.9 and 3.4 kb using the following primers (restriction sites are in italics): fragment 1 (3.9 kb), 5’-GATCTACCAGGATGCTCCGATQAATTATCAATATCC-3’ (NcoI) and 5’-ATTACCCGGGTTCATCAACACTCTTCTTGTG-3’ (XmaI); and fragment 2 (3.4 kb), 5’-GGATAATTCTGTACAGGTGC-3’ and 5’-GGTCCAACGGGTTCAGAATAGATGTATCTC-3’ (XmaI). Each PCR fragment was cloned into pCR2.1 (Invitrogen, San Diego, CA) and sequenced.

Rapid amplification of cDNA ends (5’ RACE) and cloning

The protocol for 5’ RACE was carried out according to the recommendations of the manufacturer (Life Technologies). To obtain the termini of the genome, first-strand synthesis of cDNA was primed using an NV-specific primer and 1 μg of infected cell RNA. The primer used for the 5’ terminus was 5’-GACAAACAGATCACCTTCCAC-3’, located ~430 nucleotides from the 5’ terminus of the genome, and the primer used for the 3’ terminus was 5’-CTGCTTTAGGATCAGATCC-3’, located ~650 nucleotides from the 3’ terminus of the genome. PCR amplification of the tailed first-strand cDNAs was done using nested NV-specific primers. The primer used for the 5’ terminus was 5’-GTGTTGTTCTATCATTAC-3’, which is ~320 nucleotides from the 5’ terminus of the genome, and the primer used for the 3’ terminus was 5’-CTGCTTT-
TAGCGATCAGTATCC-3’, located ∼500 nucleotides from the 3’ terminus of the genome. These primers were used in conjunction with the Abridged Anchor primer supplied by the manufacturer (Life Technologies). PCR products were purified (see above) and cloned into the plasmid vector, pCR2.1 (Invitrogen) and sequenced. Plasmid clones containing the NV 3’ terminus were identified by colony hybridization using a [γ-32P]ATP (ICN Biomedicals, Costa Mesa, CA) end-labeled oligonucleotide consisting of the 3’ terminal 21 nucleotides of the HV virus genome.

Sequence analysis

Plasmid DNA or purified PCR products were sequenced using a cycle sequencing reaction with fluorescent dye terminators (Perkin-Elmer, Applied Biosystems Division, Foster City, CA), and the reaction products were analyzed by use of an ABI 373 or ABI 3100 automatic sequencer (Perkin-Elmer). Sequencing primers were based either on the previously published sequence of HV (GenBank Accession No. AF017149) or on the sequence of NV. Sequence data were analyzed using version 10.0 of the Genetics Computer Group package (Devereaux et al., 1984), PHYLIP version 3.4 (Felsenstein, 1988), and PAUP (Swofford and Olsen, 1990). For the analysis of conservative amino acid changes, amino acids were grouped into families as follows: (F, Y), (M, L, I, V), (A, G), (T, S), (E, D), (K, R), and any changes involving an H, P, W, or C residue were not considered conservative.

Note added in proof. After acceptance of the manuscript, the Executive Committee of the ICTV approved Henipavirus as the name for the genus containing HV and NV.

ACKNOWLEDGMENTS

We are grateful to the staff of the DNA Chemistry Section, Biotechnology Core Facility, CDC, for expedient synthesis of numerous oligonucleotide primers and we thank Anthony Sanchez for the oligo(dT) primer. We are also grateful for technical assistance provided by Stephanie Liflick and Sandi McCoy and thank Bettina Bankamp for critical reading of the manuscript. B.H.H. and K.H. are ASM/NCID Research Associates.

REFERENCES


