Complete nucleotide sequence of the M RNA segment of Andes virus and analysis of the variability of the termini of the virus S, M and L RNA segments

P. J. Padula,1 A. J. Sanchez,2 A. Edelstein1 and S. T. Nichol2

1 Departamento de Virología, Instituto Nacional de Enfermedades Infecciosas, ANLIS ‘Dr Carlos G. Malbrán’, Av. Velez Sarsfield 563, 1281 Buenos Aires, Argentina
2 Special Pathogens Branch, Division of Viral and Rickettsial Disease, Center for Disease Control and Prevention, 1600 Clifton Road, Atlanta, Georgia 30333, USA

Hantavirus pulmonary syndrome (HPS) has been recognized increasingly as a significant public health problem in South America since Andes virus was first discovered in Argentina. Here, the isolation of Andes virus is reported from an infected rodent captured in Argentina in close vicinity to the place of the first HPS case, AH1. The complete nucleotide sequences of the virus M segment, partial L segment and the termini of the S, M and L segment genome RNAs were determined. The Andes virus M RNA segment is 3671 nt in length and is predicted to encode a glycoprotein precursor 1138 aa in length; it generally resembles the other HPS-associated hantaviruses in its organization. Relative to the G1 glycoprotein of other HPS-associated hantaviruses, an additional potential glycosylation site was found but this is located in the predicted cytoplasmic domain and is therefore unlikely to be glycosylated. In phylogenetic analyses, Andes virus, together with the more related hantaviruses, represented a monophyletic lineage. The S-terminal nucleotides were conserved relative to other New World hantaviruses. The M and L segment RNA termini had short deletions in the region believed to contain the sequence and structural features necessary for initiation of virus RNA replication and transcription. Clinical manifestations of Andes virus infections range from fulminant respiratory disease with high lethality to mild course without sequelae. Andes virus has also been associated with person-to-person transmission. Accumulation of Andes virus genetic data will be essential for understanding the factors that regulate virus replication and transmission and to determine the pathogenesis of HPS.

Introduction

Hantaviruses (family Bunyaviridae) constitute a genus of rodent-borne zoonotic pathogens, which includes the aetiologic agents causing haemorrhagic fever with renal syndrome and hantavirus pulmonary syndrome (HPS), a severe and often fatal respiratory disease. Andes virus was first partially characterized based on specimens from a fatal HPS case (AH1) in Argentina in 1995 (López et al., 1996). Argentina has the highest annual number of reported HPS cases in South America. These are located in three disparate regions: subtropical Salta and Jujuy Provinces in the north, Patagonia region in the south and Buenos Aires Province in the central region. Specific species of rodents serve as the natural reservoir of the various hantavirus types and are considered to be the infectious source for humans. However, person-to-person transmission of virus during an outbreak of HPS in Argentina in 1996 caused by Andes virus has been described (Enria et al., 1996; Wells et al., 1997; Padula et al., 1998). This outbreak involved 16 persons. In addition, person-to-person transmission could not be excluded in two Andes virus-related HPS family clusters in Chile (Toro et al., 1998).

In common with other members of the genus, Andes virus possesses a genome consisting of three single-stranded, negative-sense RNA segments, designated large (L), medium...
(M) and small (S), which encode the virus polymerase, glycoproteins (G1 and G2) and the nucleocapsid (N) protein, respectively. The 5′ and 3′ termini of hantavirus genome RNA segments are highly conserved and complementary to each other (Schmaljohn, 1996). To date, the termini of RNA genomes of viruses of the family *Bunyaviridae* are exactly complementary for 8 or 9 nt, whereas the complementarity extends to at least 17 nt among the hantaviruses. The 5′ and 3′ termini of virus RNAs are thought to form a double-stranded promoter regulating RNA transcription and replication, similar to that seen for other segmented, negative-stranded RNA viruses.

Knowledge of the genetic characteristics of Andes virus strains detected in patient samples is limited, as only S and partial M segment sequences are available. More recently, additional sequence information has been determined from the first Andes virus (AH1) case reported (López et al., 1996), from the southern Chilean cases (López et al., 1997) and from a variety of Andes virus lineages detected in cases or rodents in distinct geographical regions in Argentina (Levis et al., 1998; Padula et al., 2000). Here, we report the complete Andes virus M segment sequence, its phylogenetic relationship with other HPS-associated viruses and a partial Andes virus L segment sequence. We present data on the variability of the RNA termini of the S, M and L segments. In addition, we report the successful isolation of Andes virus from one *Oligoryzomys longicaudatus* rodent captured in the vicinity of the AH1 HPS case.

**Methods**

**Virus RNA RT–PCR and ligation.** Total RNA extraction, RT and first-round PCR were performed as described before (Padula et al., 1998). Briefly, RT–PCR was performed on RNA extracted from samples of serum, blood clot or tissue by using an Access RT–PCR kit (Promega). Second-round reactions were prepared using 1 µl of the first-round PCR. Primer sequences and combinations used for different fragments are available upon request. To obtain the terminal sequences of the S segment, an RNA ligation reaction was performed essentially as described previously (Johnson et al., 1997). Nested RT–PCR, with specially designed primers, was performed to span the region encompassing the ligated terminal. For the S segment, the ligated RNA was amplified using a heminested RT–PCR with one positive-sense primer and two negative-sense primers. After PCR, the fragments were cloned into the pGem-T Vector System Easy (Promega). The resulting undigested plasmids were sequenced with specific primers. After PCR, the fragments were cloned into the pGem-T Vector System Easy (Promega). The resulting undigested plasmids were sequenced with specific primers. L segment RNA terminal sequences were determined by RT–PCR. Briefly, Vero E6 cell monolayers (ATCC, CRL 1586) were inoculated with approximately 50 mg of tissue suspension, cultivated in T-125 flasks and, after 1 h adsorption, the tissue suspension inoculum was removed and maintenance medium was added to the cells. Cells were maintained at 37 °C under 5% CO2. On day 14, cells were suspended and half of them were used to infect another flask. After 14 days, a small amount of cells was scraped off for RNA extraction.

**Results and Discussion**

**Sequence features of the Andes virus M and L genome segments and their encoded proteins**

Virus nucleotide sequences were determined by direct sequence analysis of PCR products amplified from lung tissues from patient AH1 (the first Andes virus strain characterized) to avoid potential sequence changes associated with isolation procedures. The Andes virus M segment was determined to be 3671 nt in length. This is 18, 19 and 25 nt longer than Lechiguanas (LEC), Hu96994 and Oran viruses, respectively, and 27 and 25 nt shorter than Laguna Negra (LN) and Sin Nombre (SN) viruses, respectively. The major ORF (nt 52–3465) has a coding capacity of 1138 aa. The WAASA motif (aa 647–651) is conserved at the presumed cleavage site between the G1 and G2 proteins, as observed in all other hantaviruses.

The Andes virus G1 glycoprotein extends from aa 1 to 651 (nt 52–204), including the putative cleavage site, and the G2 glycoprotein extends from aa 652 to 1138 (nt 205–3465). Three of the four putative potential glycosylation sites in the Andes virus G1 protein are conserved in all HPS-associated hantaviruses (positions 138–140, 350–352 and 402–404). However, the site at positions 524–526 is conserved among hantaviruses (positions 138–140, 350–352 and 402–404). Many of the and also more minor differences, such as renal insufficiency, are seen between HPS-associated and non-HPS-associated viruses, which can alter the charge on the virus surface glycoproteins, may be relevant to the pathogenic differences seen among these viruses. This could include the differences seen between HPS-associated and non-HPS-associated viruses, and also more minor differences, such as renal insufficiency, which is often a feature of Andes virus-associated HPS but which is lacking for SN virus-associated HPS.

The positions of the cysteine residues in the predicted G1 and G2 proteins are very highly conserved, as are many of the proline residues, indicating a conserved three-dimensional...
Andes virus genome M segment and RNA termini

**S segment**

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3'</td>
<td>AUCAUCAU  GAGGACU  UUCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'</td>
<td>UAGUAGUA  CUCGUUGA  AAGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**M segment**

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>---</td>
<td>CU</td>
<td>U</td>
<td>C</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3'</td>
<td>AUCAU  GAGGCGU  CUU  UU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'</td>
<td>UAG</td>
<td>UG</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>

**L segment**

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3'</td>
<td>CAUCAU  GAGGCCCU  U  UU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'</td>
<td>UA</td>
<td>UG</td>
<td>A</td>
<td>AG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Predicted secondary structure of the complementary 3' and 5' termini of Andes virus, 23 nt of the S segment (AF324902), 25 nt of the M segment (AF324901) and 24 nt of the L RNA segments (AF324900).

structure for the protein. Hydrophilicity plots of the glycoproteins of the subfamily Sigmodontinae-associated hantaviruses were very similar (data not shown). Although the virus complete M segment sequences were divergent at the nucleotide level (approximately 79% identity) and the deduced amino acid sequences were highly conserved (93% identity) when Andes virus was compared with related viruses such as LEC, Oran or Hu39694 viruses. When comparing with North American HPS hantaviruses, the identity was 70 and 76–78% [Bayou (BAY), Black Creek Canal (BCC), SN and New York (NY) viruses] at the nucleotide and amino acid levels, respectively.

The sequence of a 606 nt fragment of the Andes virus L segment was determined for the region 2832–3437 (numbered relative to SN virus). Nucleotide and amino acid sequence comparison revealed 79.5 and 90.7% identity, respectively, with LN virus (521 nt compared), the most closely related available virus, while SN virus showed 72.7% nt and 87.1% aa identity.

**Andes virus phylogenetic analysis**

Analyses of M and partial L segment nucleotide sequence differences using the maximum-parsimony method with a 2:1 weighting of transversions to transitions indicated that Andes virus represents a distinct phylogenetic lineage that is closely
Phylogenetic analysis based on the complete M (a) and partial L (b) segments of Andes virus and other hantaviruses characterized previously. The PHYLIP program was used to make 500 bootstrap replicates of the sequence data (SEQBOOT). The DNAPARS program was used to find the trees with maximum parsimony. Distance matrices were calculated using Kimura’s two parameter model (DNADIST) and analysed by the Fitch–Margoliash tree-fitting algorithm (FITCH) with a 2:1 weighting of transversions over transitions. The percentage of bootstrap supporting each node is indicated for major branches and was calculated from these trees (CONSENSE). The following published M segment sequences were included in the analysis: Blue River virus strain Indiana (AF030551); Blue River virus strain Oklahoma (AF030552); NY virus New York-1 (U36802); NY virus New York-2 (U36803); NY virus Rhode Island-1 (U36801); SN virus (L25783); SN virus (Nm R11) (L37903); Hu39694 virus strain Hu39694 (AF028023); Oran virus strain Ol22896 (AF028024); LEC virus strain Ol22819 (AF028022); LN virus related to hantaviruses of other subfamily Sigmodontinae rodents (Fig. 3). Phylogenetic analyses of both S and M segment sequences indicate that Andes virus, together with the more related hantaviruses (Oran, LEC and Hu39694 viruses) and LN viruses, represent a distinct lineage of South American HPS-associated viruses. Some support is found for the North American BAY and BCC viruses forming a clade with the other viruses associated with sigmodontine rodents (which are predominantly distributed throughout South America), suggestive of a common ancient ancestor. Our data further indicate that Andes virus sequences could have been diverged into at least four branches and these branches contain antigenically similar but genetically distinct viruses.

**Determination of the nucleotide sequences of the termini of Andes virus S, M and L RNA genomic segments**

The terminal sequences of Andes virus S and M genome RNA segments were determined for virus RNA extracted from the lung of the AH1 case patient. The S-terminal nucleotides were conserved relative to other hantaviruses associated with subfamily Sigmodontinae rodents (Schmaljohn, 1996) (Fig. 1). Prior to this analysis, the published terminal sequences of Andes virus S genome segments had been deduced from PCR primers (López et al., 1997) and two base changes were thought to exist relative to the sequence consensus for the HPS-associated hantaviruses.

A total of 24 M clones was sequenced in order to study the heterogeneity of the M RNA segment termini (Fig. 2). The 5’ termini of M virus RNAs were highly conserved, with missing residues seen in only 5 of 24 clones. In contrast, none of the 24 clones had intact 3’ termini. Of these 24 M clones, 17 had truncated 3’ termini with the nucleotides AUC missing. Similar results were obtained for two independent preparations of Andes virus RNA. According to the cDNA sequences, the predicted secondary structure for the RNA termini matched in the same manner with 3, 6 or 9 nt missing compared to the standard sequence (Zuker, 1989).

Similar to the approach used for the S and M segment termini, multiple attempts were made to ligate the 5’ and 3’ termini of the L RNA segment either directly from the original RNA or after cloning. The terminal sequences were determined for 20 L clones. The 3’ termini of L virus RNAs were not obtained. Of the 20 L clones, 16 had truncated 3’ termini with varying degrees of truncation. The 5’ termini of L virus RNAs were highly conserved, with missing residues seen in only 6 of 20 clones. According to the cDNA sequences, the predicted secondary structure for the RNA termini matched in the same manner with 3, 6 or 9 nt missing compared to the standard sequence (Zuker, 1989).

**Fig. 3.** Phylogenetic analysis based on the complete M (a) and partial L (b) segments of Andes virus and other hantaviruses characterized previously. The PHYLIP program was used to make 500 bootstrap replicates of the sequence data (SEQBOOT). The DNAPARS program was used to find the trees with maximum parsimony. Distance matrices were calculated using Kimura’s two parameter model (DNADIST) and analysed by the Fitch–Margoliash tree-fitting algorithm (FITCH) with a 2:1 weighting of transversions over transitions. The percentage of bootstrap supporting each node is indicated for major branches and was calculated from these trees (CONSENSE). The following published M segment sequences were included in the analysis: Blue River virus strain Indiana (AF030551); Blue River virus strain Oklahoma (AF030552); NY virus New York-1 (U36802); NY virus New York-2 (U36803); NY virus Rhode Island-1 (U36801); SN virus (L25783); SN virus (Nm R11) (L37903); Hu39694 virus strain Hu39694 (AF028023); Oran virus strain Ol22896 (AF028024); LEC virus strain Ol22819 (AF028022); LN virus related to hantaviruses of other subfamily Sigmodontinae rodents (Fig. 3). Phylogenetic analyses of both S and M segment sequences indicate that Andes virus, together with the more related hantaviruses (Oran, LEC and Hu39694 viruses) and LN viruses, represent a distinct lineage of South American HPS-associated viruses. Some support is found for the North American BAY and BCC viruses forming a clade with the other viruses associated with sigmodontine rodents (which are predominantly distributed throughout South America), suggestive of a common ancient ancestor. Our data further indicate that Andes virus sequences could have been diverged into at least four branches and these branches contain antigenically similar but genetically distinct viruses.

**Determination of the nucleotide sequences of the termini of Andes virus S, M and L RNA genomic segments**

The terminal sequences of Andes virus S and M genome RNA segments were determined for virus RNA extracted from the lung of the AH1 case patient. The S-terminal nucleotides were conserved relative to other hantaviruses associated with subfamily Sigmodontinae rodents (Schmaljohn, 1996) (Fig. 1). Prior to this analysis, the published terminal sequences of Andes virus S genome segments had been deduced from PCR primers (López et al., 1997) and two base changes were thought to exist relative to the sequence consensus for the HPS-associated hantaviruses.

A total of 24 M clones was sequenced in order to study the heterogeneity of the M RNA segment termini (Fig. 2). The 5’ termini of M virus RNAs were highly conserved, with missing residues seen in only 5 of 24 clones. In contrast, none of the 24 clones had intact 3’ termini. Of these 24 M clones, 17 had truncated 3’ termini with the nucleotides AUC missing. Similar results were obtained for two independent preparations of Andes virus RNA. According to the cDNA sequences, the predicted secondary structure for the RNA termini matched in the same manner with 3, 6 or 9 nt missing compared to the standard sequence (Zuker, 1989).

Similar to the approach used for the S and M segment termini, multiple attempts were made to ligate the 5’ and 3’ termini of the L RNA segment either directly from the original RNA or after cloning. The terminal sequences were determined for 20 L clones. The 3’ termini of L virus RNAs were not obtained. Of the 20 L clones, 16 had truncated 3’ termini with varying degrees of truncation. The 5’ termini of L virus RNAs were highly conserved, with missing residues seen in only 6 of 20 clones. According to the cDNA sequences, the predicted secondary structure for the RNA termini matched in the same manner with 3, 6 or 9 nt missing compared to the standard sequence (Zuker, 1989).
AH1 lung or from lung tissue of infected O. longicaudatus rodents. These all failed, forcing us to attempt to isolate Andes virus from three virus RNA-positive rodents trapped in South Argentina near the place of the reference AH1 patient. The virus was successfully isolated from the lung of one rodent and a reference stock was prepared. Virus RNA was obtained from culture supernatant and infected cells of Andes virus harvested from the fifth passage in Vero E6 cells.

The 5’ and 3’ termini of the virus L RNA segment had sequences similar to those reported for other hantaviruses. A panhandle structure at least 18 bp in length could be potentially formed by complementary regions of the 5’ and 3’ termini. When compared with SN virus, few exceptions to the panhandle structure were found. Differences included a single nucleotide at position 16 (for SN virus, U→A and for Andes virus, C→G), one each at positions 20 and 24 at the 5’ end, which result in complementary nucleotides for Andes virus, and another one at position 23 at the 3’ end. To determine if the termini were full-length or deleted, the PCR products were cloned and 12 clones were sequenced. All 12 L clones had 5’- or 3’-terminal deletions. Three clones had only two bases deleted but the exact position could not be determined because both bases could belong either to the 3’ or to the 5’ end. However, based on earlier studies (Meyer & Schmaljohn, 2000), the deletions would appear to belong to the 3’ end.

It has to be noticed that L segment terminal sequences were recovered from the virus particles obtained from infected cells together with culture supernatant, while the sources for the S and M segment terminal sequences were obtained from virus RNA extracted from lung tissue of patient AH1. Although differences in the panhandle structure between Andes virus and the rest of the hantaviruses were observed, it has not been proven that those were connected to the unique properties of Andes virus.

The finding that the Andes virus M and L RNA segment termini had short deletions in the region believed to contain the sequence and/or structural features necessary for initiation of replication and transcription is intriguing. According to a proposed prime-and-realign model for Hantaan virus initiation, virus RNAs with very short 3′ deletions could be expected potentially to replicate (Garcin et al., 1995). The Andes virus M virus RNAs with 3′ deletions that were longer than a few nucleotides are probably not replication competent, because deletions were very rare (5 of 24) at the 5′ termini. Terminally deleted RNAs were proposed previously to have a role in downregulating virus gene expression for a persistent virus in the Arenaviridae family, lymphocytic choriomeningitis virus (Meyer & Southern, 1994, 1997). Several possible explanations of why the 3′ termini of genome RNAs are more heterogeneous than the 5′ termini were also discussed previously for Tula virus (Kukkonen et al., 1998). In addition, it has also been proposed that accumulation of terminally deleted RNAs may play a role in Seoul virus persistence, as the percentage of 3′-deleted virus RNAs increases in the population, they could potentially compete with standard virus and downregulate virus replication (Meyer & Schmaljohn, 2000). It is possible that the differences in the relative abundance of G1–G2 and N proteins seen in virus-infected cells could reflect deletions found in the Andes virus M segment that are not present in the S segment termini and which could downregulate G1–G2 synthesis relative to N synthesis.

Clinical manifestations of human infections with Andes virus range from fulminant respiratory disease with high lethality to mild course without sequelae. It is generally assumed that humans usually become infected with hantaviruses by close contact with infected rodents or infectious rodent secretions or excretions. In addition, Andes virus has been associated with person-to-person transmission. Accumulation of Andes virus genetic data will be essential for understanding the factors that regulate virus gene expression and transmission and to determine HPS pathogenesis.

References


Received 14 February 2002; Accepted 18 April 2002