Sabia virus, one of five arenaviruses from South America known to cause hemorrhagic fever in humans, emerged in 1990 when it was isolated from a fatal case in Sao Paulo, Brazil. Subsequently, it has caused two laboratory-acquired infections. Its natural distribution and host are still unknown. Using viral RNA and multiple polymerase chain reaction products as templates, the nucleotide sequence of the small (S) RNA segment of Sabia virus, which codes for the nucleocapsid (N) and glycoprotein precursor, was determined. This virus shares an ambisense genome in common with other arenaviruses, although it has a unique predicted three stem-loop structure in the S RNA intergenic region. Phylogenetic analysis of a portion of the N gene sequence confirmed that Sabia virus is distinct from all other members of the Arenaviridae and shares a progenitor with Junin, Machupo, Tacaribe, and Guanarito viruses.

INTRODUCTION

The members of the family Arenaviridae have a single-stranded RNA genome composed of two segments, L (large) and S (small), with an average length of 7100 and 3400 nucleotides, respectively. Two other RNA molecules have been detected in virions and are presumed to be ribosomal RNAs since they comigrate with 28 S and 18 S ribosomal RNA (Carter et al., 1973; Southern and Bishop, 1987; Iapalucci et al., 1989). Arenaviruses utilize an ambisense coding strategy to express their genes. The genomic S RNA segment encodes the nucleocapsid (N) gene at the 3' end and the glycoprotein precursor (GPC) gene at the 5' end (Southern and Bishop, 1987). The N protein is expressed through transcription of a subgenomic mRNA from the genomic S RNA template. The GPC is expressed via transcription of a subgenomic mRNA from the antigenomic S RNA template, a replicative intermediate. The GPC undergoes posttranslational cleavage to generate the envelope glycoproteins G1 and G2 (Southern and Bishop, 1987). The L RNA segment encodes two proteins. The viral polymerase gene is located at the 3’ end of the segment and a zinc-binding protein gene at the 5’ end (Salvato, 1993); these are also expressed by an ambisense coding strategy (Salvato, 1993). In both segments, the intergenic region separating the two genes has been predicted to form hairpin structures which may play an important role in transcription termination during mRNA synthesis (Franze-Fernandez et al., 1993).

Arenaviruses have caused sporadic, hemorrhagic fever outbreaks with high mortality in South America and Africa since the 1950s and 1960s, respectively (Parodi et al., 1961; Frame et al., 1970). Sabia virus is a new member of the South American arenaviruses known to cause disease in humans (Coimbra et al., 1994); others are Junin, Machupo, Flexal, and Guanarito (Parodi et al., 1961; Johnson, 1965; Pinheiro, 1986; Salas et al., 1991). Sabia virus emerged in 1990 when it was isolated from a fatal case of hemorrhagic fever in Sao Paulo, Brazil (Coimbra et al., 1994); subsequently, two nonfatal laboratory infections have occurred (Vasconcelos et al., 1993; Barry et al., 1995). Further studies on the epidemiology, ecology, and diagnosis of this hemorrhagic fever have been severely restricted due to the biohazards associated with handling this agent and the lack of specific diagnostic reagents. Thus, the origin, geographic distribution, natural maintenance cycle, and epidemiology of this virus are unknown.

Initial characterization of Sabia virus by complement-fixation, immunofluorescence, and neutralization assays indicated that Sabia is a new member of the Tacaribe complex (Coimbra et al., 1994). A limited sequence analysis using 250 nt of the S RNA indicated that is genetically distinct from Junin, Machupo, Pichinde, Tacaribe, and Guanarito viruses (Coimbra et al., 1994). To further characterize Sabia virus, the S RNA was sequenced and compared to other arenaviruses, both genetically and phylogenetically.

MATERIALS AND METHODS

The Sabia prototype strain (SPH114202) was isolated from serum of a human fatal case by infecting newborn mice by intracerebral inoculation (Coimbra et
This virus was then passaged twice in Vero E6 cells to produce working stocks of virus. Virus RNA was purified as previously described (Gonzalez et al., 1995). The 3′ end of the S RNA of the Sabia S virus was sequenced initially by dideoxynucleotide chain termination sequencing of virus RNA (Rico-Hesse et al., 1987) using the oligonucleotide ARE/3′ END (5′-CGCACTAGTG-3′) to prime reverse transcription. This primer is complementary to the 3′ end of the S RNA of the Sabia virus. The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987), treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.

The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987), treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.

The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987), treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.

The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987), treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.

The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987), treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.

The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987), treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.

The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987), treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.

The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987), treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.

The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987),-treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.

The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987), treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.

The majority of the remaining S RNA sequence was obtained by sequencing products obtained by reverse transcription–polymerase chain reaction (RT–PCR). For this purpose, RNA was extracted from infected cell culture supernatant by acid guanidium thiocyanate–phenol–chloroform extraction (Chomczynsky and Sacchi, 1987), treated with 10 mM methyl mercury hydroxide, and then reverse transcribed using primer ARE/3′ END. The Sabia S virus was amplified by RT–PCR using primers designed from the partial sequence determined by direct sequencing of the 3′ end of the segment which was inferred from the primer sequence.
The putative GPC gene, which begins at nt 59, is 1464 nucleotides long and encodes a 488-aa gene product. The putative GPC protein precursor is 46.1–64.7% divergent from other arenavirus GPCs (Fig. 1). The putative GPC cleavage site is located at an R–R motif, at residues 251–252, as determined by analogy with the cleavage site for the GPC of the lymphocytic choriomeningitis virus (LCM) (Buchmeier et al., 1987). The amino acid sequence of the G2 region, located at the C terminus of the GPC, is generally more conserved than the N-terminal G1 region. The most hydrophilic domain is located in the G2 region, located at the C terminus of the GPC, and four in the G2 region. An antigenic site conserved among arenavirus GPCs, KFWYL (Weber and Buchmeier, 1988), was present in the Sabiná GPC sequence at aa 251–252, as determined by analogy with the cleavage site for the GPC of the LCM virus. The predicted GPC cleavage site is located at an R–R motif, at residues 251–252, as determined by analogy with the cleavage site for the GPC of the LCM virus. The prediction of the variability in number of potential secondary structures in different arenaviruses is unknown.

The putative N protein gene is located at nt 1620 to 3308. Its 562-aa gene product is 33.6 to 41.5% divergent from the N proteins of the New World arenaviruses and it has been pointed out, CTL epitopes present several advantages in eliciting protective immunity: CTL responses may be encoded by structural or nonstructural viral proteins, may be more important in virus clearance than humoral responses (Oldstone and Dixon, 1970), and are commonly cross-reactive against serologically distinct viral strains (Whitton et al., 1989). It remains to be seen whether these regions serve as CTL epitopes in Sabiná virus.

The noncoding region which separates the stop codons of the GPC and N genes spans nt 1526 to 1619. The predicted secondary structure of this region consists of two large stem-loop structures and a third smaller stem-loop (Fig. 2). All three structures are predicted to form regardless of whether the intergenic region is analyzed in the genomic or antigenomic sense. This feature distinguishes the Sabiná S RNA from the S RNAs of other arenaviruses whose intergenic regions are predicted to form either one (LCM, Pichinde, Lassa) or two (Mopeia, Junin, Tacaribe) stem–loop structures (Auperin et al., 1984, 1986; Romanowsky et al., 1985; Franzé-Fernandez et al., 1987; Ghiringhelli et al., 1991; Wilson and Clegg, 1991). The stem–loop structures of the intergenic region are thought to play an important role in termination of transcription during mRNA synthesis (Auperin et al., 1984; Franzé-Fernandez et al., 1993). Alternatively, these structures may be involved in translational initiation of a second gene product (Auperin et al., 1984). The significance of the variability in number of potential secondary structures in different arenaviruses is unknown.

The putative N protein gene is located at nt 1620 to 3308. Its 562-aa gene product is 33.6 to 41.5% divergent from the N proteins of the New World arenaviruses and it has been pointed out, CTL epitopes present several advantages in eliciting protective immunity: CTL responses may be encoded by structural or nonstructural viral proteins, may be more important in virus clearance than humoral responses (Oldstone and Dixon, 1970), and are commonly cross-reactive against serologically distinct viral strains (Whitton et al., 1989). It remains to be seen whether these regions serve as CTL epitopes in Sabiná virus.

The noncoding region which separates the stop codons of the GPC and N genes spans nt 1526 to 1619. The predicted secondary structure of this region consists of two large stem-loop structures and a third smaller stem-loop (Fig. 2). All three structures are predicted to form regardless of whether the intergenic region is analyzed in the genomic or antigenomic sense. This feature distinguishes the Sabiná S RNA from the S RNAs of other arenaviruses whose intergenic regions are predicted to form either one (LCM, Pichinde, Lassa) or two (Mopeia, Junin, Tacaribe) stem–loop structures (Auperin et al., 1984, 1986; Romanowsky et al., 1985; Franzé-Fernandez et al., 1987; Ghiringhelli et al., 1991; Wilson and Clegg, 1991). The stem–loop structures of the intergenic region are thought to play an important role in termination of transcription during mRNA synthesis (Auperin et al., 1984; Franzé-Fernandez et al., 1993). Alternatively, these structures may be involved in translational initiation of a second gene product (Auperin et al., 1984). The significance of the variability in number of potential secondary structures in different arenaviruses is unknown.

The putative N protein gene is located at nt 1620 to 3308. Its 562-aa gene product is 33.6 to 41.5% divergent from the N proteins of the New World arenaviruses and it has been pointed out, CTL epitopes present several advantages in eliciting protective immunity: CTL responses may be encoded by structural or nonstructural viral proteins, may be more important in virus clearance than humoral responses (Oldstone and Dixon, 1970), and are commonly cross-reactive against serologically distinct viral strains (Whitton et al., 1989). It remains to be seen whether these regions serve as CTL epitopes in Sabiná virus.

The noncoding region which separates the stop codons of the GPC and N genes spans nt 1526 to 1619. The predicted secondary structure of this region consists of two large stem-loop structures and a third smaller stem-loop (Fig. 2). All three structures are predicted to form regardless of whether the intergenic region is analyzed in the genomic or antigenomic sense. This feature distinguishes the Sabiná S RNA from the S RNAs of other arenaviruses whose intergenic regions are predicted to form either one (LCM, Pichinde, Lassa) or two (Mopeia, Junin, Tacaribe) stem–loop structures (Auperin et al., 1984, 1986; Romanowsky et al., 1985; Franzé-Fernandez et al., 1987; Ghiringhelli et al., 1991; Wilson and Clegg, 1991). The stem–loop structures of the intergenic region are thought to play an important role in termination of transcription during mRNA synthesis (Auperin et al., 1984; Franzé-Fernandez et al., 1993). Alternatively, these structures may be involved in translational initiation of a second gene product (Auperin et al., 1984). The significance of the variability in number of potential secondary structures in different arenaviruses is unknown.
FIG. 3. Comparison of the amino acid sequence of the N protein encoded by the S RNA of Sabia virus with those of other arenaviruses (LCM arm, lymphocytic choriomeningitis Armstrong strain; LCM we, lymphocytic choriomeningitis WE strain; LAS jos, Lassa Josiah strain from Sierra Leone; LAS nig, Lassa strain from Nigeria; MOP, Mopeia; JUN, Junin; MAC, Machupo; TAC, Tacaribe; SAB, Sabia; GUA, Guanarito; PIC, Pichinde). CTL epitope and probable antigenic site are in boldface. Conserved residues among the New World arenaviruses are shown in the top consensus line; conserved residues among New World and Old World arenaviruses are shown in the bottom consensus line. Dash indicates absence of conserved residue.

46.5 to 52.1% divergent from the N proteins of Old World arenaviruses (Fig. 3). A probable antigenic site that was previously described around a conserved pair of amino acid residues (K – R) (Gonzalez et al., 1995) was identified by the first 5 amino acids of the sequence GVYMGNL was found at residues (K – R) (Gonzalez et al., 1995) was identified by the first 5 amino acids of the sequence GVYMGNL. AID VY 8006 / 6a1a $$42 06-16-96 19:10:30 viral AP: Virology
123 to 129 (GVYLNGL) on Sabia virus, where an M to L change occurs between the anchor residues of the CTL epitope (Falk et al., 1991). It remains to be seen if the potential CTL site on Sabia virus is immunogenic.

The 3′ noncoding region of the Sabia S RNA is 58 nt long and includes, at the 3′ terminus, the 19 nucleotides conserved among all arenaviruses studied to date. These 19 nt represent the only portion of the Sabia S RNA not actually sequenced in this study. Other arenaviruses have 3′ noncoding regions of varying lengths (range, 52 to 96 nt), thus demonstrating plasticity in the encoded function of this segment. Although this region is presumed to be involved in transcription (Meyer and Southern, 1994), there is no obvious structural similarity between the arenaviruses studied to date.

To determine the evolutionary relationship of Sabia to other arenaviruses, phylogeny was inferred by maximum parsimony analysis of available N gene sequences. The S RNA sequences of 10 arenaviruses [Junin (GenBank Accession No. D10072), Tacaribe (M20304), Lassa Nigeria (X52400, K03362), Lassa Josiah (J04324), Guanarito (L42001), Machupo (X62616), Mopeia (M33879), Pichinde

![Diagram of phylogenetic relationships of arenaviruses based on nucleotide sequence differences in the N gene 3′ terminus. Comparisons were done by the maximum parsimony method using the PAUP software program (Swofford, 1993) run on a Power Macintosh 8100. LCM, LAS, and MOP viruses from the Old World were used as an outgroup for rooting the tree. Horizontal branch lengths are proportional to nucleotide step differences between viruses and are indicated above the lines. Bootstrap confidence limits were calculated by 1000 replications of the analysis and limits in excess of 50% are indicated in parentheses below the branch (LCMarm, lymphocytic choriomeningitis Armstrong strain; LCMwe, lymphocytic choriomeningitis WE strain; LASjos, Lassa Josiah strain from Sierra Leone; LASnig, Lassa strain from Nigeria; MOP, Mopeia; JUN, Junin; MAC, Machupo; TAC, Tacaribe; SAB, Sabia; GUA, Guanarito; PIC, Pichinde).]
(K02734), Lymphocytic choriomeningitis Armstrong (M22138), Lymphocytic choriomeningitis WE (M22017) were obtained from the GenBank sequence database. The sequences were edited to match the shortest available sequence length, which was the homologous region of the Guanarito virus N gene open reading frame (729 nt, which encode amino acids 1 through 243, see Fig. 3). The sequences compared among arenaviruses varied in length (729 to 744 nt) because of insertions/deletions in this gene (Fig. 3). Analyses were performed with PAUP: Phylogenetic Analysis Using Parsimony, version 3.1.1 (Swofford, 1993) employing the heuristic search option. Bootstrap confidence intervals (Felsenstein, 1985) were calculated by carrying out 1000 heuristic search replicates (Fig. 4). The topology of the tree produced by this limited sequence analysis is concordant with previous estimates of arenavirus phylogeny derived using complete N or GPC gene sequences, but fewer taxa (Clegg, 1993). The position of Sabia virus relative to other taxa does not change when using the complete N gene sequences versus the homologous 732-nt fragment (data not shown). Thus, we propose classifying emerging arenaviruses by comparison of sequences derived from only a portion of the N gene.

Maximum parsimony analysis showed that arenaviruses clustered into two main geographic groups (Old World and New World complexes) as ascertained previously by antigenic characteristics (Wulff et al., 1978; Clegg et al., 1990). Sabia virus is distinct from all other arenaviruses and it shares a common ancestor with Junin, Machupo, Tacaribe, and Guanarito. Tacaribe virus is the only member of this monophyletic group that has not been associated with natural human infection. This suggests that these rodent-borne viruses have a common origin and have developed the potential to cause human outbreaks independently. Because of increased human exposure to rodent-borne virus habitats, new arenaviruses will probably infect humans and eventually cause outbreaks. Thus, it is important to develop control or vaccination strategies common to all arenaviruses. Identification of common CTL epitopes among the Tacaribe complex arenaviruses could possibly be exploited for the development of a vaccine strategy effective for all New World arenaviruses.

ACKNOWLEDGMENTS

This work was supported in part by Grants A1-10984 and A1-01124 from the National Institutes of Health, and DES-AB8004 from the Institut Français de Recherche Scientifique pour le Développement en Coopération. We thank our Brazilian colleagues for providing the Sabia prototype virus.

REFERENCES


