JOURNAL OF VIROLOGY, Oct. 1990, p. 4873–4883
0022-538X/90/104873-11$02.00/0

Vol. 64, No. 10

Polymerase Errors Accumulating during Natural Evolution of the Glycoprotein Gene of Vesicular Stomatitis Virus Indiana Serotype Isolates

PAMUK A. BILSEL AND STUART T. NICHOL*

Cell and Molecular Biology Program, School of Veterinary Medicine, and Department of Microbiology, School of Medicine, University of Nevada, Reno, Nevada 89557

Received 22 March 1990/Accepted 22 June 1990

We report the entire glycoprotein (G) gene nucleotide sequences of 26 vesicular stomatitis virus Indiana serotype (VSV IND) type 1 isolates from North and Central America. These sequences are also compared with partial G gene sequences of VSV IND type 2 (Cocal) and type 3 (Alagoas) viruses and the complete G gene sequences of the more distantly related VSV New Jersey (NJ) and Chandipura viruses. Phylogenetic analysis of the G gene sequences by maximum parsimony revealed four major lineages or subtypes within the classical VSV IND (type 1) viruses, each with a distinct geographic distribution. A high degree of VSV genetic diversity was found in Central America, with several virus subtypes of both VSV IND and NJ serotypes existing in this mainly enzootic disease region. Nineteen percent sequence variation but no deletions or insertions were evident within the 5′ noncoding and the coding regions of the VSV IND type 1 G genes. In addition to numerous base substitutions, the 3′ noncoding regions of these viruses also contained numerous base insertions and deletions. This resulted in striking variation in G gene sizes, with gene lengths ranging from 1,652 to 1,868 nucleotides. As the VSV IND type 1 subtypes have diverged from the common ancestor with the NJ subtypes, their G mRNAs have accumulated more 3′ noncoding sequence inserts, ranging up to 303 nucleotides in length. These primarily consist of an imprecise reiteration of the sequence UUUUUAA, apparently generated by a unique polymerase stuttering error. Analysis of the deduced amino acid sequence differences among VSV IND type 1 viruses revealed numerous substitutions within defined antigenic epitopes, suggesting that immune selection may play a role in the evolution of these viruses.

Vesicular stomatitis virus (VSV) remains an economically important cause of vesicular disease in cattle, horses, and swine throughout much of the Americas (9, 10, 19). Two major VSV serotypes, Indiana (IND) and New Jersey (NJ), have been defined and can contribute to epizootic and enzootic disease activity in various geographic regions (5). The negative-strand RNA genome of VSV IND has been shown to rapidly and randomly accumulate mutations during specific in vitro passage conditions. These include serial high-multiplicity passage of virus and persistent infections of tissue culture cells (11, 12, 28, 32, 33, 40–42). However, the virus RNA genome can be maintained in a stable manner under conditions of serial low-multiplicity passage of the virus in tissue culture (4, 33). To date, there has been no thorough investigation as to how these observations relate to the genetic stability or diversity of the VSV IND RNA genome in nature. An RNase T_1 fingerprint study of the RNA genomes of four VSV IND isolates from the United States indicated that all of these viruses belonged to the same T_1 fingerprint group (i.e., had a high degree of genetic similarity), although numerous T_1 spot differences were apparent (6).

Previously, we have examined the natural genetic diversity of a large number of VSV NJ isolates. T_1 fingerprint analysis of the RNA genomes of more than 100 natural isolates together with direct nucleotide sequence analysis of numerous virus isolate glycoprotein (G), nucleocapsid (N), and phosphoprotein (P) genes indicated that extensive genetic variation existed among these viruses (2, 22–24, 27). This was most striking among isolates from Central America, where enzootic disease predominates. Three major lineages or distinct VSV NJ subtypes were defined and were found to be maintained over extended time periods in distinct geographical areas. The direct nucleotide sequence analysis of the glycoprotein (G) gene of 34 VSV NJ isolates (representing the extent of genetic diversity within the serotype) revealed up to 20% sequence variation among isolates (27). However, no base insertions or deletions were detected. Clusters of deduced amino acid substitutions were identified in the G protein hydrophobic regions, such as the signal peptide and transmembrane domain, and within the cytoplasmic domain. In addition, substitutions were identified within defined antigenic epitopes (17), suggesting that immune selection may play a role in promoting VSV NJ evolution. This is supported by recent data by Wagner and colleagues directly demonstrating antigenic variation among several VSV NJ natural isolates (18).

Although they are closely related, there are a number of significant differences between VSV of the NJ and IND serotypes. For instance, the VSV IND serotype is broader than VSV NJ in that it has previously been divided into four different types based upon antigenic cross-reactivity (35, 37, 38). These include classical Indiana (type 1), Cocal (type 2), Alagoas (type 3), and Maraba (type 4) viruses. Indiana type 1 viruses are isolated mainly from cattle but also from pigs and insects. Cocal virus was originally isolated from mites in Trinidad and subsequently from cattle, horses, and mosquitos (14, 15). Alagoas virus was originally isolated from a mule in Brazil, but it appears that cattle, horses, humans, and sandflies can also be infected (35). Although bite and transovarial transmission by *Phlebotomus* sandflies has been

* Corresponding author.

demonstrated for viruses of both serotypes (with the exception of Cocal), virus transovarial transmission rates for VSV NJ are considerably lower than for VSV IND (6a, 35, 36, 38). This may indicate a difference in maintenance and transmission mechanisms between viruses of the two serotypes. In recent years, VSV NJ outbreaks have occurred more frequently than VSV IND outbreaks in most geographic regions. Another significant difference between VSV NJ and IND type 1 viruses is the antigenic structure of the G protein. Monoclonal antibody (MAb) analysis of expressed deletion mutant and chimera G proteins has indicated striking differences in the location of antigenic epitopes between these two viruses (16). Seven of the nine defined VSV NJ G epitopes, including all four neutralization epitopes, mapped between amino acids 193 and 289 in the central portion of the protein. In contrast, 4 of the 11 defined VSV IND epitopes, including two neutralization epitopes, mapped to the amino-terminal amino acids 80 to 123. The remaining seven epitopes, including two neutralization epitopes, mapped to the carboxy-terminal domain amino acids 286 to 428. The basis of these antigenic and biological differences is unclear, as is whether these differences would be reflected in the extent and distribution of genetic variation among natural isolates of the two serotypes.

Here we report the results of phylogenetic analysis of the G genes of 26 VSV IND type 1 virus isolates from the United States, Mexico, and Central America. In addition, these sequences are compared with partial G gene sequences of VSV IND type 2 (Cocal) and type 3 (Alagoas) viruses and the complete G gene sequences of the more distantly related VSV NJ and Chandipura viruses.

MATERIALS AND METHODS

Virus isolates. The origins of the 26 natural isolates of the VSV IND serotype (type 1) viruses that were analyzed are listed in Table 1. Viruses were from the United States, Mexico, and Central America. The San Juan isolate (56-NM-B), VP-98F (69-PN-L), BT-78 (59-PN-L), and strain 903816 (84-PN-H) were provided from the VSV collection of the Yale Arbovirus Research Unit by Bob Tesh, Yale University School of Public Health, New Haven, Conn. (29). The Mudd-Summers (MS) strain of VSV IND was obtained from John Holland at the University of California, San Diego, who had received it directly from John Mudd and Don Summers, University of Utah, Salt Lake City, Utah. They originally received the virus in the late 1960s from Philip Marcus, University of Connecticut, Storrs, Conn. The history of the virus prior to this is unknown, although we suspect it was derived from the original 1925 Indiana isolate (9). Mexican VSV IND type 1 isolates were provided by John Mason and Farouk Hamdy at the Comision Mexico-Americana para la Prevencion de la Fiebre Aftosa, Mexico City, Mexico. Recent Central American VSV IND type 1 isolates were provided from the Yale Arbovirus Research Unit collection by Bob Tesh, having been originally obtained from the LADIVES facility in Panama, with the kind assistance of Luis Roquebert. Cocal virus (VSV IND type 2) was obtained from the American Type Culture Collection (14). Alagoas virus (35) (VSV IND type 3) was strain CoAr 171044, isolated from sandflies (*Lutzomyia* spp.) collected in Colombia in 1986 by Dr. Tesh.

Virus purification and RNA extraction. Viruses were grown at 37°C in BHK-21 cells in Eagle minimal essential medium (MEM). Virus was harvested and purified by ultracentrifugation, and RNA was extracted as described earlier (26).

TABLE 1. VSV IND isolates analyzed in this study

Virus strain	Location of isolation	Host
56-NM-B	New Mexico	Cow
MS	Unknown	Cow
86-DF-P	Mexico City, Mexico	Pig
87-OA-B	Oaxaca, Mexico	Cow
87-VC-B	Vera Cruz, Mexico	Cow
84-GM-B	Guatemala	Cow
85-GM-B	Guatemala	Cow
86-GM-B	Guatemala	Cow
85-ES-B1	El Salvador	Cow
85-ES-B2	El Salvador	Cow
86-ES-B1	El Salvador	Cow
86-ES-B2	El Salvador	Cow
87-ES-B	El Salvador	Cow
83-HD-B2	Honduras	Cow
83-HD-B1	Honduras	Cow
82-HD-B	Honduras	Cow
84-PN-H	Panama	Human
84-PN-B2	Panama	Cow
69-PN-L	Panama	Sandfly (*Lutzomyia* sp.)
84-PN-B1	Panama	Cow
84-CR-B	Costa Rica	Cow
85-CR-B1	Costa Rica	Cow
85-CR-B2	Costa Rica	Cow
87-CR-B1	Costa Rica	Cow
87-CR-B2	Costa Rica	Cow
59-PN-L	Panama	Sandfly (*Lutzomyia* sp.)

RNA sequencing. Virus RNA was sequenced by a dideoxy chain termination method described earlier (27). Oligonucleotides used as primers in the sequencing reactions were synthesized on an Applied Biosystems model 380A or NEN-DuPont Coder 300 automated synthesizer. The sequences of the primers are available from the authors upon request.

Reverse sequencing reaction. Bases 1 to 420 for virus 59-PN-L were determined by random priming synthesis of a first-strand cDNA copy, followed by specific priming of a second-strand cDNA dideoxynucleotide sequencing reaction as described previously (27) with the following modifications. First-strand reaction mixtures (30 μl) contained 50 mM Tris hydrochloride (pH 8.3), 10 mM $MgCl_2$, 70 mM KCl, 20 mM 2-mercaptoethanol, 5 to 10 μg of virus RNA, 200 ng each of 6-, 9-, and 12-nucleotide-long random primers (provided by Zymogenetics, Inc., Seattle, Wash.), 30 U of reverse transcriptase (Life Sciences), and 2 mM each dATP, dCTP, dGTP, and dTTP. Reaction mixes were incubated at 41°C for 1 h, and reactions were terminated with 2 μl of 0.5 M EDTA, phenol-chloroform extracted, and ethanol precipitated with 0.1 volume of 3 M sodium acetate (pH 5.2) at −80°C. The cDNA-RNA pellet was suspended in 20 μl of STE (10 mM Tris hydrochloride [pH 8.0], 1 mM EDTA, 100 mM NaCl) and passed through a Linkers 5 quick-spin column (Boehringer Mannheim) to remove the random primers. The RNA template was hydrolyzed in 0.2 N NaOH–0.2 mM EDTA for 5 min at room temperature. The reaction was neutralized with 0.1 volume of 3 M sodium acetate (pH 5.2) and ethanol precipitated at −80°C. The cDNA was then used as a template in sequencing reactions as described above for viral RNA.

Sequence analysis. VSV IND G gene sequences were analyzed with the Microgenie Sequence Analysis software package (Beckman Instruments, Inc., Palo Alto, Calif.) run on an IBM PC-XT microcomputer. Alignment of VSV IND, NJ, and Chandipura sequences was performed with the GENALIGN program (Hugo Martinez, University of Cali-

fornia, San Francisco) run on a SUN3 minicomputer (SUN Microsystems, Mountain View, Calif.). Maximum-parsimony analysis of G gene nucleotide sequences was performed with the PAUP program (David Swofford, Illinois Natural History Survey) run on a SUN3 minicomputer.

In vitro virus passages. In order to analyze the stability of virus G gene 3′ noncoding sequence inserts (see Results section), three virus isolates, 82-HD-B, 87-ES-B, and 85-CR-B1, were selected as representing the three major sequence insert forms described. These viruses were serially passaged 80 times in BHK-21 cells under dilute and undilute conditions. Both sets of passages for the three isolates were started with uncloned virus from original stock material. BHK-21 cells were infected with 100 μl of a 10^{-5} dilution of virus-infected culture fluid for the dilute passage series and 100 μl of undiluted infected-culture fluid for the undilute passage series. Virus was adsorbed for 45 min and grown at 37°C in Eagle MEM with 2% calf serum added. Cytopathic effect was observed 18 to 48 h postinfection (p.i.). The dilute passage series was continued with 100 μl of a 10^{-6} dilution of the culture fluid, and the undilute series was continued with 1 ml of culture fluid adsorbed directly onto confluent BHK monolayers.

RNA was extracted from virus at select passages (5, 10, 25, 50, and 80) and sequenced as described above. Virus RNA from earlier or later passages was sequenced when sequence differences from the starting material were detected at the 3′ noncoding end of the G gene. Individual plaques of 82-HD-B were isolated at passages 40, 45, 50, 55, and 60. Confluent BHK cells grown in tissue culture dishes were infected with serial 10-fold dilutions of virus. Unadsorbed virus was removed at 30 min p.i., and agar overlay (Eagle MEM with 2% calf serum and 0.4% agar) was added to the infected monolayer. Individual plaques were identified and isolated at 18 to 24 h p.i. RNA from propagated plaque virus was extracted and sequenced as described above.

Nucleotide sequence accession numbers. The nucleotide sequences presented here have been submitted to GenBank under accession no. M35207 through M35232.

RESULTS

Evolutionary relationships of VSV IND isolates. Twenty-six VSV IND type 1 field isolates (Table 1) were selected as representative of the genetic diversity within the serotype based on preliminary RNase T_1 fingerprinting studies (data not shown). The entire G gene nucleotide sequences were derived for all 26 VSV IND isolates. Four nucleotide substitutions were observed between our reported sequence for the San Juan isolate (56-NM-B) and that published previously (31).

Numerous base substitutions but no deletions or insertions were evident within the 5′ noncoding and the coding regions of the G genes. Up to 19% sequence variation was present in this region. The 3′ noncoding region was highly variable, with numerous base insertions and deletions present in addition to base substitutions (Fig. 1). This resulted in striking variation in G gene sizes, with gene lengths ranging from 1,652 to 1,868 nucleotides. Nucleotide sequences were aligned by using the Needleman-Wunsch option of the GENALIGN program (see Materials and Methods). In addition, these sequences were aligned relative to three VSV NJ serotype isolates by the same approach in conjunction with previous alignments (8). The three VSV NJ isolates selected, ../60-PN-B, ../52-GA-P, and ../49-UT-B1, represent the oldest isolates of each of the three previously determined VSV NJ subtypes (27). Phylogenetic analysis of the aligned G gene data by the maximum-parsimony method allowed construction of evolutionary trees. This was performed by using the PAUP program (see Materials and Methods) with the settings hold = 10, maxtree = 100, swap = global, and mulpars. The three VSV NJ sequences were used to outgroup root the tree.

Only one most-parsimonious tree, 3,056 steps in length, was obtained (Fig. 2). Several distinct major lineages or subtypes were observed. The VSV IND type 1 isolates could be divided into four subtypes by the criteria used previously to define the major subtypes of VSV NJ (27). Subtype I contained all isolates from the geographical area spanning from the United States through Mexico as far south as Honduras. This subgroup could be further separated into two sublineages, one containing the United States and Mexico isolates, and the other containing the Central American isolates. The U.S.-Mexico sublineage contained the older 56-NM-B VSV IND isolate. It is noteworthy that this isolate and 86-DF-P differ by only 89 nucleotide steps despite a 30-year time difference in isolation. Subtype II contained only two representative viruses from Panama. Subtype III contained several isolates from Costa Rica and Panama (located next to each other in southern Central America). Finally, subtype IV contained a single historic isolate obtained from sandflies in Panama in 1959. A striking relationship was evident between the geographical location of virus isolation and virus phylogeny.

Evolutionary relationships of vesiculoviruses. In order to examine the broader relationships of vesiculoviruses, the phylogenetic analysis was expanded to include partial G gene nucleotide sequences derived from the IND-related viruses Cocal (IND type 2) and Alagoas (IND type 3). The sequence of nucleotides 1 to 300 and 950 to 1100 for Cocal and 60 to 389 and 660 to 1741 for Alagoas was obtained. Cocal and Alagoas had previously been grouped in the VSV IND serotype by limited serological cross-reactivity (35, 37). Chandipura virus is a vesiculovirus with little serological cross-reactivity with either the IND or NJ serotype of VSV (2). In addition, limited nucleotide sequence analysis of the termini of the RNA genomes had demonstrated that Chandipura was more distantly related to VSV NJ and IND than they are to each other (25). The recently published Chandipura G gene sequence confirmed this relationship and provided an appropriate sequence with which to outgroup root the broader VSV tree analysis.

All the G gene nucleotide sequences were aligned relative to one another by using the Needleman-Wunsch option of the GENALIGN program. The VSV NJ, IND, and Chandipura virus nucleotide alignments were essentially identical to those expected from previous amino acid sequence alignments for these viruses (21). On analysis of one G gene representative from each of the VSV IND type 1 lineages, VSV NJ major lineages, and Cocal and Alagoas viruses, only one most-parsimonious tree 4,523 steps in length was obtained with PAUP with settings of hold = 10, maxtree = 100, swap = global, mulpars, and outgroup-rooted with the Chandipura G gene sequence (Fig. 3). This confirmed the relationships of VSV NJ and classical (type 1) VSV IND lineages with respect to the ancestral VSV node. It also indicated that Alagoas and Cocal viruses were correctly grouped as VSV IND viruses, although they were closer to the ancestral root than classical (type 1) IND isolates. It also appears that the VSV NJ serotype viruses have diverged less from the hypothetical VSV ancestor than have classical VSV IND viruses.

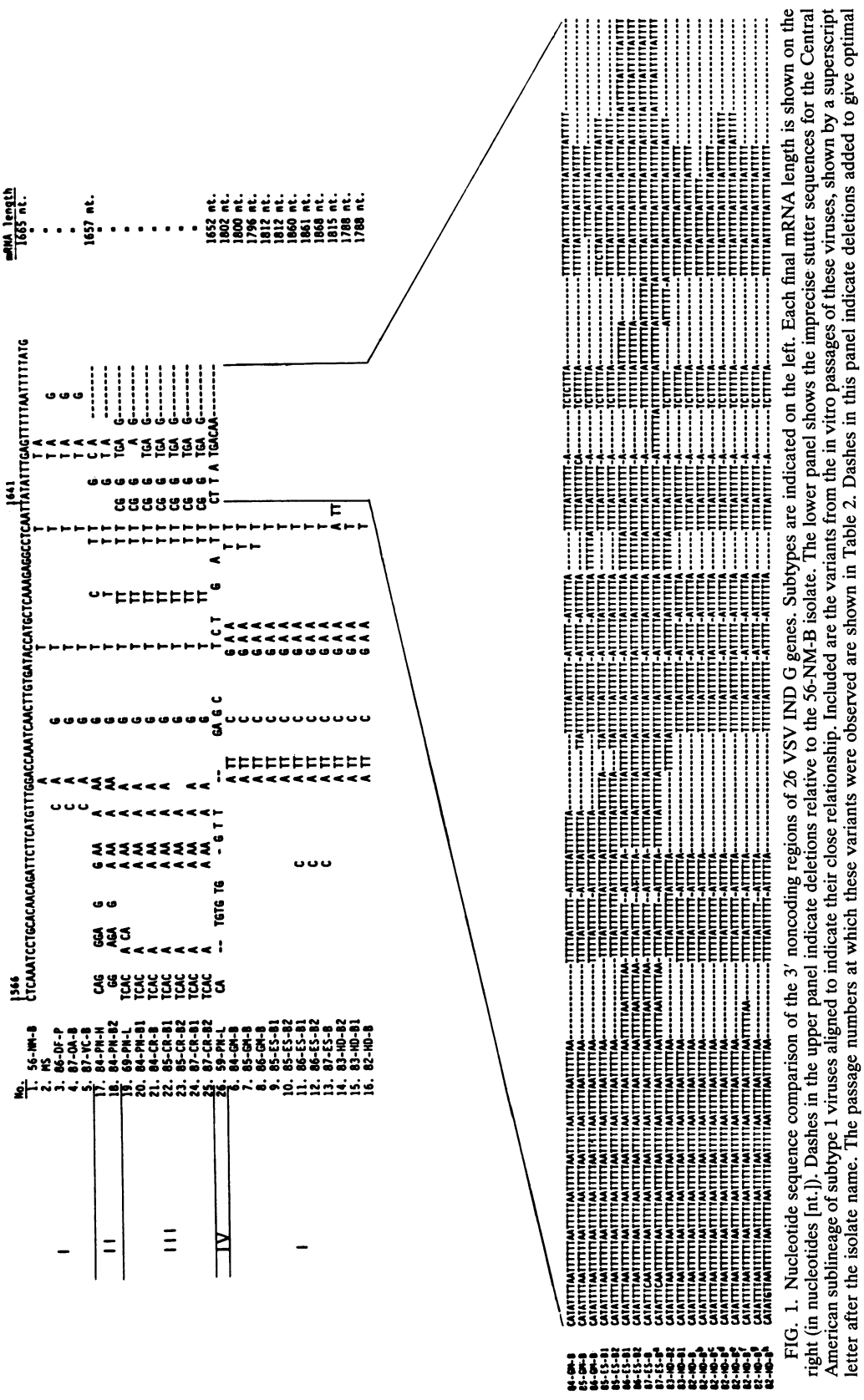

FIG. 1. Nucleotide sequence comparison of the 3′ noncoding regions of 26 VSV IND G genes. Subtypes are indicated on the left. Each final mRNA length is shown on the right (in nucleotides [nt.]). Dashes in the upper panel indicate deletions relative to the 56-NM-B isolate. The lower panel shows the imprecise stutter sequences for the Central American sublineage of subtype 1 viruses aligned to indicate their close relationship. Included are the variants from the in vitro passages of these viruses, shown by a superscript letter after the isolate name. The passage numbers at which these variants were observed are shown in Table 2. Dashes in this panel indicate deletions added to give optimal alignment of imprecise stutter sequences.

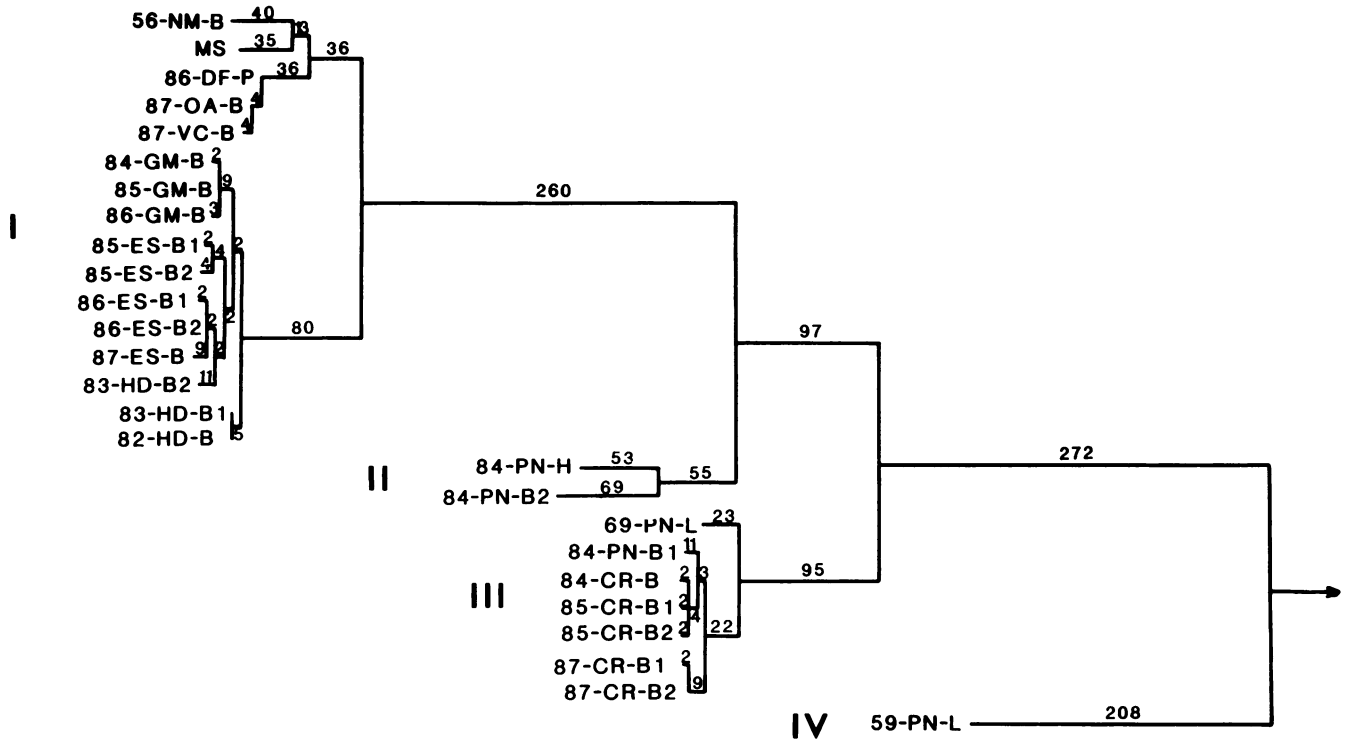

FIG. 2. Evolutionary tree for the VSV IND G gene sequences obtained by maximum-parsimony analysis. Subtype numbers are indicated next to each lineage. The lengths of the horizontal lines are proportional to the minimum number of single-nucleotide substitutions required to generate the variation observed. The numeric branch lengths are indicated. Representative VSV NJ subtypes were used to outgroup root the tree.

Deduced glycoprotein amino acid sequences. The deduced glycoprotein amino acid sequences for all VSV IND type 1 isolates are shown relative to that of the 56-NM-B isolate in Fig. 4. Four amino acid substitutions were observed between our 56-NM-B isolate G protein amino acid sequence and that published previously (31). In addition, nine amino acid changes were observed between our MS isolate G protein sequence and that published previously (39). All the isolates encoded a glycoprotein 511 amino acids in length. Amino acid sequence differences between subtypes ranged from 3 to 9%. Certain features of the IND glycoprotein were conserved among all 26 isolates. The glycosylation sites at amino acids 179 and 336 were conserved, and no new glycosylation sites were generated. The palmitate addition site at amino acid 489 (cysteine) in the hydrophilic domain was also conserved between all the isolates. Overall, the

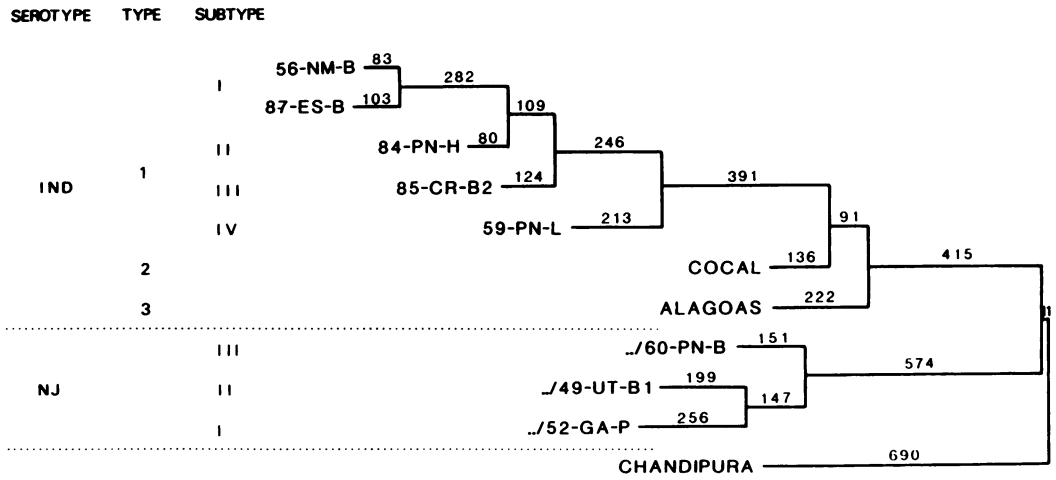

FIG. 3. Evolutionary tree for vesiculoviruses obtained by maximum-parsimony analysis of G gene sequences. Representatives of VSV IND types and subtypes and VSV NJ subtypes were analyzed by outgroup rooting with the more distant vesiculovirus Chandipura. The lengths of the horizontal lines are proportional to the minimum number of single-nucleotide substitutions required to obtain the observed relationships. The numeric branch lengths are indicated.

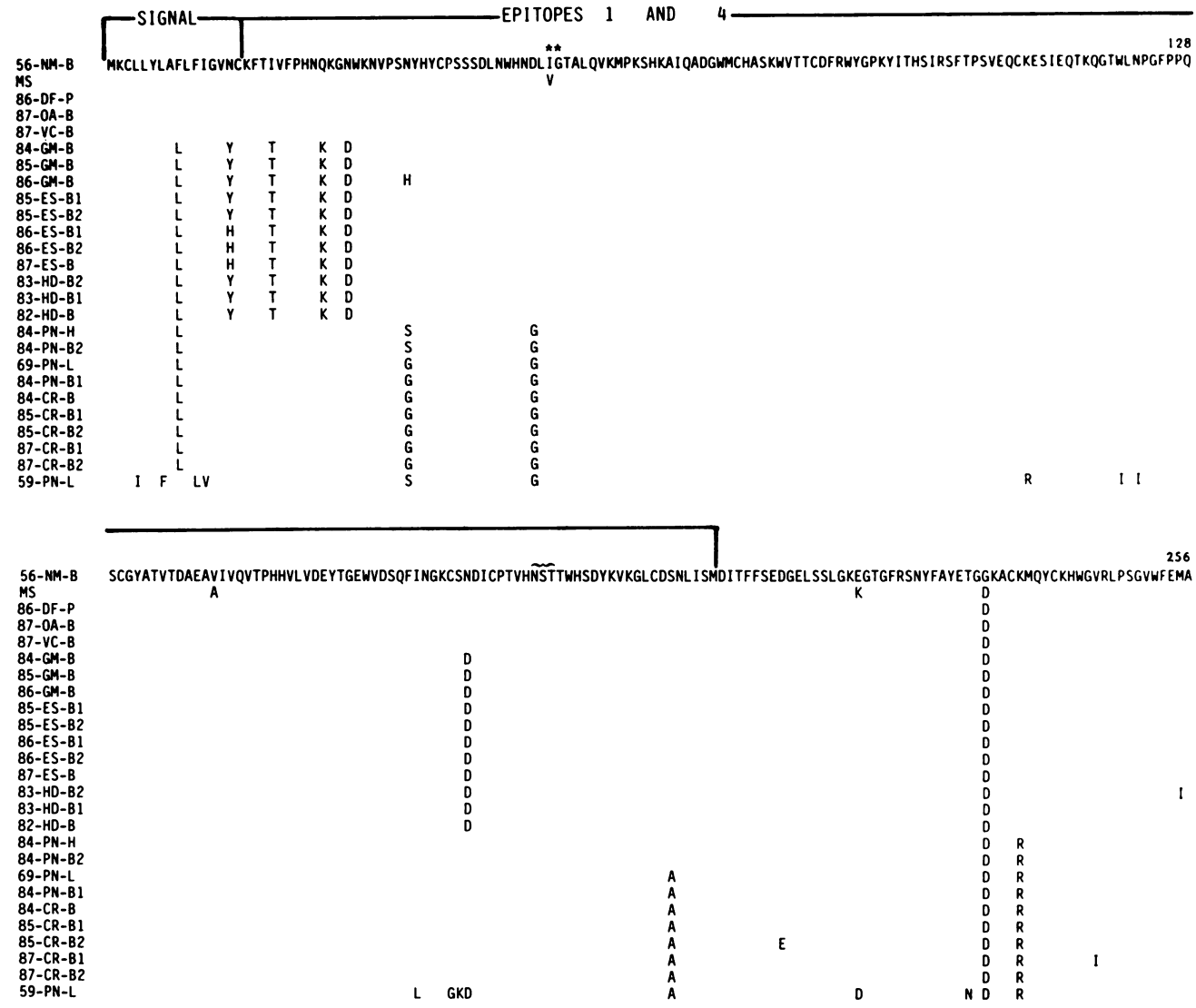

FIG. 4. Predicted amino acid sequences for G genes of 26 VSV IND isolates. Amino acid differences relative to the full sequence of San Juan (56-NM-B) are presented. Asterisks indicate neutralization-resistant mutations selected by epitope-specific MAbs. Glycosylation sites at position 179 and position 336 are indicated with brackets. The palmitate addition site at amino acid 489 is indicated by a solid circle. Only virus neutralization epitopes are shown.

G protein ectodomain was more highly conserved than other domains, including the signal, transmembrane, and cytoplasmic domains. However, amino acid differences were found in previously defined antigenic epitopes within the ectodomain.

The four major neutralization epitopes of the VSV IND glycoprotein determined previously by recombinant expressed epitopes (16) are indicated in Fig. 4. Many amino acid substitutions were present within these epitopes which may affect the antigenicity of the glycoproteins. Specific mutations induced by MAb selection are also shown. The same substitutions are found in many of the naturally occurring variants. Subtype I Central American sublineage viruses all had an Asp to Asn change at amino acid position 259, and subtype IV virus had an Ala to Gly change at position 263. Both of these changes were identified previously in MAb-resistant VSV mutants (39). Epitope two (the neutralization epitope at the carboxyl end) was determined by recombinant expression to span amino acids 339 to 428. Subtypes II, III, and IV had the same substitutions at amino acid 357 (Ser to Pro) and amino acid 362 (Met to Thr) as did MAb-selected mutants within this epitope. Epitopes at the amino end have been shown to be dependent on intact secondary structure and glycosylation at amino acid 179 (16). The 12 cysteines thought to be necessary for intact secondary structure were conserved between the isolates except for the loss of one cysteine at amino acid 169 to 59-PN-L (subtype IV). The antigenic integrity of epitopes 1, 3, and 4 has been shown to be dependent on intact secondary structure. Thus, this Cys change could potentially alter antigenicity, in addition to protein secondary structure. The Ile to Val change at amino acid 53 between 56-NM-B and MS has previously been shown not to affect binding of MAbs specific for that epitope (39). There were no other substitutions at MAb-selected sites in this region (amino acids 53 and 54).

VSV IND G mRNA 3′ noncoding sequences. The lengths of

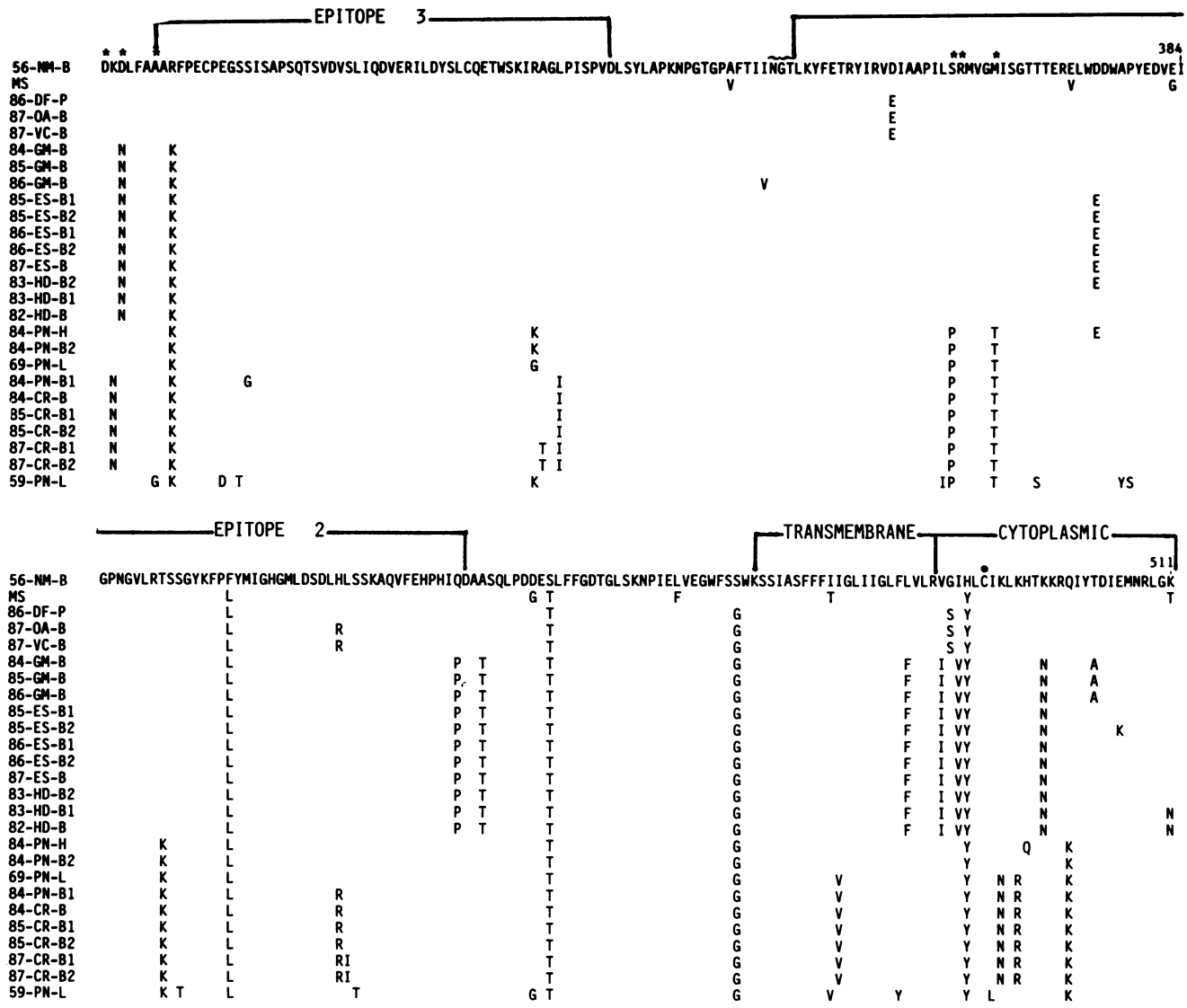

FIG. 4—*Continued.*

the 3′ noncoding ends of the VSV IND glycoprotein mRNAs varied strikingly with respect to that of the standard 56-NM-B (San Juan) strain. Subtypes II, III, and IV had deletions relative to 56-NM-B, whereas subtype I (Central American sublineage) had several insertions which accounted for the different mRNA lengths (Fig. 1). As the IND subtypes have diverged from the common ancestor with the NJ subtypes, they have accumulated more noncoding sequences at the 3′ end of the gene. This sequence length increased from subtype IV up to subtype I. The G mRNA insert in the subtype I U.S.-Mexico sublineage was an 8-nucleotide unit of sequence UUAPUUUU. The inserts in the Central American sublineage ranged from 147 to 227 nucleotides in length and consisted of the same 8-nucleotide unit in addition to an imprecise reiteration of a 7-nucleotide unit of sequence UUUUUAA (Fig. 1). Interestingly, only the subtype I Central American sublineage viruses had a 5′-GG-3′ intergenic dinucleotide between the M and G genes (data not shown). All other VSV gene junctions consisted of a 5′-AG-3′ dinucleotide with the exception of a 5′-AC-3′ that was found between the P and M genes of VSV IND type 1 viruses (7, 30).

3′ noncoding insert sequence stability. The high degree of variation in the insert sequences within subtype I Central American sublineage isolates suggests an extremely plastic nature for the 3′ noncoding region of these viruses. To investigate the stability and potential origin of these insert sequences, three representative virus isolates were chosen for serial passage in tissue culture under conditions of high and low virus multiplicity (see Materials and Methods). The three isolates were 82-HD-B, with the shortest insert; 87-ES-B, with the longest insert; and 85-CR-B1, with a deletion relative to the San Juan (56-NM-B) isolate (Fig. 1). RNA virus populations have previously been shown to undergo rapid and random genome evolution under high-multiplicity conditions, whereas the equilibrium population is maintained in a stable manner under low-multiplicity conditions (33, 34). The evolution of the original quasi-species present in each of these isolates was observed by serial passaging of the isolates, starting with uncloned virus. Uncloned virus

TABLE 2. Stutter sequence variants appearing during in vitro passage of viruses 82-HD-8, 87-ES-B, and 85-CR-B1[a]

Passage no.	82-HD-B		87-ES-B		85-CR-B1	
	High multiplicity	Low multiplicity	High multiplicity	Low multiplicity	High multiplicity	Low multiplicity
5	NC	NC	NC	NC	NC	NC
9				NC		
10	NC	NC	NC	NC + a	NC	NC
11				a		
25	NC	NC	NC	a	NC	NC
30	NC	NC				
35	NC + g	NC				
37		NC				
39		Mixed				
40	g	Mixed (3/6 NC, 1/6 b, 1/6 c, 1/6 d)				
45		Mixed (2/2 NC)				
50	g	Mixed (2/6 NC, 2/6 b, 2/6 f)	NC	a	NC	NC
55		Mixed (1/3 NC, 2/3 f)				
60		Mixed (2/6 b, 3/6 d, 1/6 e)				
70		Mixed				
80	h	f	NC	Mixed	NC	NC

[a] Viruses were determined to have no change (NC) or one of eight variants (a to h) in the 3′ noncoding sequence for G mRNA (see text). For strains with a mixture of no-change and variant sequences, values are no. with indicated sequence/no. tested.

would presumably consist of a mixture of genomic RNA with variant insert sequences.

Table 2 and Fig. 1 indicate the virus RNA sequences observed at the various virus passage levels analyzed. The G mRNA 3′ noncoding sequences of virus isolate 85-CR-B1 (which represented a control virus containing no insert sequence) were stably maintained for 80 passages under both high- and low-multiplicity passage conditions. The long-insert virus, 87-ES-B, had no 3′ noncoding sequence changes under undilute conditions, whereas the dilute set had an extra set of seven T's (Fig. 1, Table 2, variant a) at passage 11 that was maintained through passage 50. However, by passage 80 no single clear discernible sequence was present, the population apparently consisting of at least variant a and original sequence, as well as others. The 3′ noncoding sequence of virus isolate 82-HD-B (the short-insert virus) remained identical to the starting sequence until passage 39, when the presence of a mixed population was evident. This heterogeneous population continued through passage 70 but stabilized by passage 80 to one predominant sequence which was different from the consensus sequence of the starting material.

Individual plaques from the mixed-virus populations (at passages 40, 45, 50, 55, and 60) were sequenced to characterize the heterogeneity. Variants of 82-HD-B observed (Fig. 1) consisted of a deletion of a cluster of T's (variant b), addition of a T to a cluster of six T's (variant c), insertion of an extra cluster of T's (variant d), or an additional T_5-A_2 repeat at the beginning of the insert (variant f). By passage 80, variant f predominated in the virus population. In the undilute 82-HD-B passages, a shift to a variant with a deleted T in one of the repeats (variant g) took place between passages 30 and 40 (Fig. 1). At passage 35, both the starting sequence and variant g were equally abundant. However, by passage 80 the original sequence with a different point mutation (variant h) had become the predominant sequence. Insert sequence variation was observed under both high- and low-multiplicity conditions during in vitro passage of 82-HD-B and 87-ES-B.

DISCUSSION

The genetic diversity and epidemiological pattern described for VSV IND type 1 are remarkably similar to those outlined previously for VSV NJ (27). A striking correlation is evident between the phylogeny and geographical location of isolation of these viruses. We have defined four distinct subtypes of VSV IND type 1 viruses, with subtype I consisting of two sublineages. It is evident that VSV IND can be relatively stably maintained in distinct infection foci with limited geographical distribution. The close grouping of all the United States and Mexico VSV IND isolates in one sublineage of subtype I indicates that they have a relatively recent common ancestor. This was also clearly demonstrated previously for VSV NJ isolates from these regions (27). The disease activity in the United States and Mexico is distinct from that occurring in Central America for both VSV IND and NJ. The other sublineage of subtype I contains closely related isolates from the broadly defined northern Central American region (El Salvador, Guatemala, and Honduras). In contrast to the relative homogeneity of isolates from these areas, greater genetic diversity is present in southern Central America. Three distinct IND type 1 subtypes (II, III, and IV) have been identified in Costa Rica and Panama (adjacent countries in southern Central America). Thus, the Central American region, where VSV is generally considered enzootic, contains multiple phylogenetic lineages for both VSV IND and NJ. The reason for this high degree of genetic diversity within this relatively small geographical area is currently unknown.

VSV genetic diversity or rapid accumulation of genome mutations in tissue culture infections is dependent on factors that disrupt the equilibrium populations or quasi-species distribution of the virus. These factors include immune selection, serial high multiplicities of infection and persistent infection conditions (4, 11, 12, 28, 32, 33, 40–42). Such factors could conceivably play a role in generating extensive genetic diversity in a region where disease is enzootic. For instance, as the disease recurs in these countries each year, most animals will have some level of immunity to VSV. This may contribute to greater host immune selection on virus in these regions relative to zones where the disease is epizootic, where virtually all animals will be immunologically naive. In addition, persistence of virus in regions where it is enzootic may facilitate genetic diversity in a manner similar to that observed in persistent infections in tissue culture (11, 12, 32, 40, 42).

As it is now widely accepted that RNA viruses have tremendous potential for genetic variation, the apparent genetic stability of some VSV IND isolates is rather striking. For instance, the United States and Mexico sublineage of subtype I contains viruses isolated during at least a 31-year period (the date of isolation of the MS isolate is unknown, although it is suspected to be derived from the original 1925 VSV IND type 1 isolate). Only 89 G gene nucleotide steps separate the 56-NM-B isolate from the 86-DF-P isolate. As the 56-NM-B isolate does not appear to have been a direct ancestor of the 1986 isolate, then the rate of accumulation of virus G gene nucleotide substitutions must have been less than 89 substitutions per 30 years, or approximately 3 substitutions per year. Corresponding examples of relative genetic stability exist for VSV NJ (27).

There is a striking relationship between geographical location of virus isolation and virus phylogeny based on G gene sequences. In both the VSV IND tree described here and the previously described VSV NJ tree (represented here by three VSV NJ isolates, but shown in detail in reference 27), there is progressive movement away from the hypothetical VSV ancestral node as one moves away from the region broadly defined as southern Central America. That is, VSV IND lineages IV, III, II, and I are progressively more distant from the root with VSV NJ viruses and occur progressively further north from the southern Central American region. Similarly, VSV NJ lineage III is closest to the root with VSV IND viruses and includes isolates from the general vicinity of southern Central America (in addition to all the Panama and Costa Rica isolates analyzed, it also contains some isolates from Nicaragua and Honduras and one isolate from Ecuador). VSV NJ lineage II, containing predominantly northern Central America isolates, is much more distant from the VSV IND root, and still slightly more distant are the VSV NJ lineage I viruses, which are exclusively from the United States and Mexico. There also appears to be a greater representation of insect isolates as one nears the predicted VSV ancestral node of the tree. One interpretation of these observations is that the VSV ancestral virus may have been an insect virus from the broadly defined equatorial America. Genetic analysis of a greater number of isolates from Central and South America will be required to test the validity of such an interpretation. However, the placement of both insect VSV IND type 2 virus Cocal (isolated from mites in Trinidad) and type 3 virus Alagoas (isolated from sandflies in Colombia) close to the root of the VSV IND lineage with the ancestral node would tend to support such an interpretation.

Chandipura, a vesiculovirus found in Asia and Africa, was used as the outgroup for the broader phylogenetic analysis of these viruses. VSV IND, NJ, and Chandipura viruses were found to be approximately equidistant from the ancestral node. A previous comparison of nucleocapsid gene sequences of the San Juan (56-NM-B) isolate of VSV IND and the Ogden (../49-UT-B) isolate of VSV NJ with Chandipura suggested that Chandipura was much more distant from both of these viruses than they are from each other (20). From this phylogenetic tree determined by maximum-parsimony analysis of G genes, it appears that type I IND isolates are the most distant from Chandipura. 56-NM-B is 2,318 nucleotide steps from Chandipura, whereas ../49-UT-B is 1,621 steps (Fig. 3). In addition, 56-NM-B and ../49-UT-B (separated by 2,537 steps) are more distant from each other than either is from Chandipura. However, VSV IND types 2 and 3 are closer to Chandipura than VSV NJ. Overall, the genetic diversity of the VSV IND serotype is much greater than that of the VSV NJ serotype.

The natural antigenic variation of VSV IND type 1 viruses also correlates with the virus phylogeny and geographic location. Nucleotide sequence analysis of MAb-resistant VSV IND type 1 variants has identified several point mutations that alter epitope recognition (39). Several of the naturally occurring variants analyzed possess point mutations in these positions (indicated with asterisks in Fig. 4), which would be expected to alter the antigenicity of these viruses. In addition, antigenic epitopes of VSV IND type 1 have been defined by deletion mapping of recombinant expressed epitopes (16). Each of the subtypes exhibit amino acid substitutions within neutralization epitopes (Fig. 4). It is likely that some of these substitutions will affect the antigenicity of these viral glycoproteins. This is strengthened by recent work by Wagner and colleagues which directly demonstrated antigenic variation among VSV NJ natural isolates with amino acid variation within similarly determined epitopes (18).

The VSV IND type 3 (Alagoas) virus appears to be antigenically more distantly related to the VSV IND type 1 viruses than they are to one another. This virus shows many amino acid substitutions within neutralization epitopes, including changes at MAb-induced sites (though not the same substitutions as observed in VSV IND type 1 viruses) (data not shown). The frequency of amino acid substitutions at MAb selection sites and within neutralizing epitopes suggests that host immunological pressure could be playing an important role in selecting VSV antigenic variants and driving the evolution of the virus in nature. Direct determination of antigenic differences among the viruses will clarify this issue.

VSV IND type 1 viruses appear to show a high degree of plasticity in the 3′ noncoding sequences of their G mRNAs as they have evolved from the common ancestral root (Fig. 1). This is particularly evident in subtype I viruses, which have acquired sequence insertions relative to the other subtype viruses, in addition to numerous nucleotide substitutions. The 3′ noncoding sequences of subtypes II, III, and IV, which lie closer to the ancestral root, are 92, 92, and 87 nucleotides in length, respectively. The G mRNAs of the subtype I U.S.-Mexico sublineage have acquired an 8-nucleotide U-rich insert similar to an immediately preceding U-rich sequence (Fig. 1). The subtype I Central American sublineage G mRNAs have acquired remarkably long insert sequences consisting of an imprecise reiteration of the 7-nucleotide sequence UUUUUAA, in addition to the 8-nucleotide U-rich insert. These sequences vary in length, but all share a number of specific features and represent minor variations of one another. It appears likely that these insert sequences are the result of a unique polymerase stuttering or slippage error. The variation evident among these sequences represents accumulation of subsequent polymerase errors following the initial stutter error. From the phylogenetic tree, it appears that 82-HD-B and 83-HD-B1 (possessing the shortest inserts) are most representative of an ancestral imprecise-stutter sequence.

The negative-sense RNA genomes of both sublineages of the subtype I viruses have a series of AAAAAUU-like sequences before the G gene transcription termination signal (Fig. 1). This is the complement of the basic repeat unit in the inserts found in the Central American sublineage G mRNAs; therefore, it seems likely that these sequences served as the template initiating the polymerase stuttering event. Whether the original polymerase stutter event took place during replication of the full-length minus- or positive-sense template is unknown. However, we favor the latter for two reasons. First, the polymerase would be stuttering or slipping on a poly(U) tract in a similar manner to polymerase slippage during polyadenylation of virus mRNAs (1). Second, only the Central American sublineage positive-strand RNA templates possessed a 3′-AUAC-5′ following the slippage sequences. As this is considered part of the normal transcription termination/polyadenylation signal (3′-AUAC UUUUUUU-5′), the presence of the 3′-AUAC-5′ in the template prior to the slippage event may have facilitated the original polymerase slippage.

The plasticity of G mRNA 3′ noncoding sequences of the subtype I Central American sublineage isolates is in contrast to the high degree of sequence similarity in the coding regions of these viruses, which varies at most by only 1%. The fact that the stutter sequences vary highly between

closely related isolates indicates that these viruses can rapidly accumulate polymerase errors in this region and/or that the virus quasi-species populations contained a high degree of sequence variation in this genome region. This is supported by the experimental data that demonstrated rapid evolution or quasi-species fluctuations of these sequences during in vitro passage of these viruses. This in vitro stutter sequence variation was analogous to that seen among natural virus isolates in that variation in the number of nucleotide stutter units could be seen in addition to single nucleotide insertions, deletions, and substitutions.

The question arises why these polymerase stutter errors only appear in the subtype I Central American sublineage viruses. Several reasons appear plausible. It may be that the fixation of such a genetic error in a virus population is extremely rare and has occurred in this group of viruses by random chance. Alternatively, accumulation of sequence alterations close to the G-L gene junction may have predisposed these viruses to a polymerase stuttering event as discussed briefly above. In addition, accumulation of amino acid changes in the polymerase-associated proteins could also have predisposed these viruses to polymerase stuttering errors. Previous studies have shown that a VSV L polymerase mutation can cause aberrant mRNA polyadenylation, a process involving polymerase slippage (13). Further work will be required to clarify these issues and to determine the biological significance of these large stutter sequence inserts.

ACKNOWLEDGMENTS

We thank Bob Tesh, Luis Roquebert, John Mason, and Farouk Hamdy for their kind assistance in provision of virus isolates.

This work was supported by U.S. Department of Agriculture Animal Molecular Biology grant 87-CRCR-1-2428.

LITERATURE CITED

1. **Banerjee, A. K.** 1987. Transcription and replication of rhabdoviruses. Microbiol. Rev. **51**:66–87.
2. **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.
3. **Bilsel, P. A., J. E. Rowe, W. M. Fitch, and S. T. Nichol.** 1990. Phosphoprotein and nucleocapsid protein evolution of vesicular stomatitis virus New Jersey. J. Virol. **64**:2498–2504.
4. **Brand, C., and P. Palese.** 1980. Sequential passage of influenza virus in embryonated eggs or tissue culture: emergence of mutants. Virology **107**:424–433.
5. **Cartwright, B., and F. Brown.** 1972. Serological relationships between different strains of vesicular stomatitis virus. J. Gen. Virol. **16**:391–398.
6. **Clewley, J. P., D. H. L. Bishop, C. Y. Kang, J. Coffin, W. M. Schnitzlein, M. E. Reichmann, and R. E. Shope.** 1977. Oligonucleotide fingerprints of RNA species obtained from rhabdoviruses belonging to the vesicular stomatitis virus subgroup. J. Virol. **23**:152–166.

6a. **Comer, J. A., R. B. Tesh, G. B. Modi, J. L. Corn, and V. F. Nettles.** 1990. Vesicular stomatitis virus, New Jersey serotype: replication and transmission by *Lutzomyia shannoni* (Diptera: psychodidae). Am. J. Trop. Med. Hyg. **42**:483–490.

7. **Gallione, C. J., J. R. Greene, L. E. Iverson, and J. K. Rose.** 1981. Nucleotide sequence of the mRNAs encoding the vesicular stomatitis virus N and NS proteins. J. Virol. **39**:529–535.
8. **Gallione, C. J., and J. K. Rose.** 1983. Nucleotide sequence of a cDNA clone encoding the entire glycoprotein from the New Jersey serotype of vesicular stomatitis virus. J. Virol. **46**:162–169.
9. **Hanson, R. P.** 1952. The natural history of vesicular stomatitis virus. Bacteriol. Rev. **15**:179–204.
10. **Hanson, R. P.** 1986. Vesicular stomatitis: introduction and overview, p. 1–11. *In* J. Mason (ed.), Proceedings of an International Conference on Vesicular Stomatitis, Mexico City. U.S. Commission for the Prevention of Foot-and-Mouth Disease, Mexico City.
11. **Holland, J. J., E. A. Grabau, C. L. Jones, and B. L. Semler.** 1979. Evolution of multiple genome mutations during long-term persistent infection by vesicular stomatitis virus. Cell **16**:495–504.
12. **Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol.** 1982. Rapid evolution of RNA genomes. Science **215**:1577–1585.
13. **Hunt, D. M., E. F. Smith, and D. W. Buckley.** 1984. Aberrant polyadenylation by a vesicular stomatitis virus mutant is due to an altered L protein. J. Virol. **52**:515–521.
14. **Jonkers, A. H., R. E. Shope, T. H. G. Aitken, and L. Spence.** 1984. Cocal virus, a new agent in Trinidad related to vesicular stomatitis virus, type Indiana. Am. J. Vet. Res. **25**:236–242.
15. **Karabatsos, N.** 1985. International catalogue of arboviruses, including certain other viruses of vertebrates, p. 337–338. American Society of Tropical Medicine and Hygiene, San Antonio, Tex.
16. **Keil, W., and R. R. Wagner.** 1989. Epitope mapping by deletion mutants and chimeras of two vesicular stomatitis virus glycoprotein genes expressed by a vaccinia virus vector. Virology **170**:392–407.
17. **Luo, L., Y. Li, R. M. Snyder, and R. R. Wagner.** 1988. Point mutations in glycoprotein gene of vesicular stomatitis virus (New Jersey serotype) selected by resistance to neutralization by epitope-specific monoclonal antibodies. Virology **163**:341–348.
18. **Luo, L., Y. Li, R. M. Snyder, and R. R. Wagner.** 1990. Spontaneous mutations leading to antigenic variations in the glycoproteins of vesicular stomatitis virus field isolates. Virology **174**:70–78.
19. **Mason, J.** 1986. Vesicular stomatitis in Mexico, p. 118–147. *In* J. Mason (ed.), Proceedings of an International Conference on Vesicular Stomatitis, Mexico City. U.S. Commission for the Prevention of Foot-and-Mouth Disease, Mexico City.
20. **Masters, P. S., and A. K. Banerjee.** 1987. Sequences of Chandipura virus N and NS genes: evidence for high mutability of the NS gene within vesiculoviruses. Virology **157**:298–306.
21. **Masters, P. S., R. S. Bhella, M. Bucher, B. Patel, H. P. Ghosh, and A. K. Banerjee.** 1989. Structure and expression of the glycoprotein gene of Chandipura virus. Virology **171**:285–290.
22. **Nichol, S. T.** 1987. Molecular epizootiology and evolution of vesicular stomatitis virus New Jersey. J. Virol. **61**:1029–1036.
23. **Nichol, S. T.** 1988. Genetic diversity of enzootic isolates of vesicular stomatitis virus New Jersey. J. Virol. **62**:572–579.
24. **Nichol, S. T.** 1988. Vesicular stomatitis virus fingerprinting. Foreign Animal Dis. Rep. **16**:8–11.
25. **Nichol, S. T., and J. J. Holland.** 1987. Genome RNA terminus conservation and diversity among vesiculoviruses. J. Virol. **61**:200–205.
26. **Nichol, S. T., P. J. O'Hara, J. J. Holland, and J. Perrault.** 1984. Structure and origin of a novel class of defective interfering particle of vesicular stomatitis virus. Nucleic Acids Res. **12**: 2775–2790.
27. **Nichol, S. T., J. E. Rowe, and W. M. Fitch.** 1989. Glycoprotein evolution of vesicular stomatitis virus New Jersey. Virology **168**:281–291.
28. **O'Hara, P. J., F. M. Horodyski, S. T. Nichol, and J. J. Holland.** 1984. Vesicular stomatitis virus mutants resistant to defective-interfering particles accumulate stable 5′-terminal and fewer 3′-terminal mutations in a stepwise manner. J. Virol. **49**:793–798.
29. **Quiroz, E., N. Moreno, P. H. Peralta, and R. B. Tesh.** 1988. A human case of encephalitis associated with vesicular stomatitis virus (Indiana serotype) infection. Am. J. Trop. Med. Hyg. **39**:312–314.
30. **Rose, J. K.** 1980. Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. Cell **19**:415–421.
31. **Rose, J. K., and C. J. Gallione.** 1981. Nucleotide sequences of the mRNAs encoding the vesicular stomatitis virus G and M

proteins determined from cDNA clones containing the complete coding regions. J. Virol. **39**:519–528.
32. **Rowlands, D., E. Grabau, K. Spindler, C. Jones, B. Semler, and J. Holland.** 1980. Virus protein changes and RNA terminal alterations evolving during persistent infection. Cell **19**:871–880.
33. **Spindler, K. R., F. M. Horodyski, and J. J. Holland.** 1982. High multiplicities of infection favor rapid and random evolution of vesicular stomatitis virus. Virology **119**:96–108.
34. **Steinhauer, D. A., J. C. de la Torre, E. Meier, and J. J. Holland.** 1989. Extreme heterogeneity in populations of vesicular stomatitis virus. J. Virol. **63**:2072–2080.
35. **Tesh, R. B., J. Boshell S., G. B. Modi, A. Morales A., D. G. Young, A. Corredor A., C. Ferro de Carrasquilla, C. de Rodriguez, L. L. Walters, and M. O. Gaitan.** 1987. Natural infection of humans, animals, and phlebotomine sand flies with the Alagoas serotype of vesicular stomatitis virus in Colombia. Am. J. Trop. Med. Hyg. **36**:653–661.
36. **Tesh, R. B., B. N. Chaniotis, and K. M. Johnson.** 1972. Vesicular stomatitis virus (Indiana serotype): transovarial transmission by phlebotomine sandflies. Science **175**:1477–1479.
37. **Tesh, R. B., A. P. A. Travassos da Rosa, and J. S. Travassos da Rosa.** 1983. Antigenic relationship among rhabdoviruses infecting terrestrial vertebrates. J. Gen. Virol. **64**:169–176.
38. **Travassos da Rosa, A. P. A., R. B. Tesh, J. F. Travassos da Rosa, J. P. Herve, and A. J. Main.** 1984. Carajas and Maraba viruses, two new vesiculoviruses isolated from phlebotomine sand flies in Brazil. Am. J. Trop. Med. Hyg. **33**:999–1006.
39. **VandePol, S. B., L. LeFrancois, and J. J. Holland.** 1986. Sequences of the major antibody binding epitopes of the Indiana serotype of vesicular stomatitis virus. Virology **148**:312–325.
40. **Wilusz, J., J. S. Youngner, and J. D. Keene.** 1985. Base mutations in the terminal noncoding regions of the genome of vesicular stomatitis virus isolated from persistent infections of L cells. Virology **140**:249–256.
41. **Youngner, J. S., E. V. Jones, M. Kelly, and D. W. Frielle.** 1981. Generation and amplification of temperature-sensitive mutants during serial undilute passages of vesicular stomatitis virus. Virology **108**:87–97.
42. **Youngner, J. S., O. T. Preble, and E. V. Jones.** 1978. Persistent infection of L cells with vesicular stomatitis virus: evolution of virus population. J. Virol. **28**:6–13.