FULL PAPER *Virology*

Genetic and Phylogenetic Analysis of Glycoprotein of Rabies Virus Isolated from Several Species in Brazil

Go SATO[1)], Takuya ITOU[1)], Youko SHOJI[1)], Yasuo MIURA[1)], Takeshi MIKAMI[2)], Mikako ITO[3)], Ichiro KURANE[3)], Samir I. SAMARA[4)], Adolorata A. B. CARVALHO[4)], Darci P. NOCITI[4)], Fumio H. ITO[5)] and Takeo SAKAI[1)]*

[1)]Departments of Preventive Veterinary Medicine and Animal Health, [2)]Veterinary Public Health, Nihon University School of Veterinary Medicine, 1866 Kameino, Fujisawa, Kanagawa 252–8510, [3)]Department of Virology I, National Institute of Infectious Disease, 1–23–1 Toyama, Shinjuku-ku, Tokyo 162–8640, Japan, [4)]Faculty of Agrarian and Veterinary Science, UNESP, Jaboticabal-SP, Via de Acesso Prof. Paulo Donato Castellane, s/n. Jaboticabal/SP 14884–900, [5)]Department of Preventive Veterinary Medicine and Animal Health, Faculty of Veterinary Medicine and Zootechny, University of São Paulo, Av. Prof. Dr. Orlando Marques de Paiva, 87, Cidade Universitária, São Paulo 05508–000, Brazil

(Received 12 June 2003/Accepted 5 February 2004)

ABSTRACT. Genetic and phylogenetic analyses of the region containing the glycoprotein (G) gene, which is related to pathogenicity and antigenicity, and the G-L intergenic region were carried out in 14 Brazilian rabies virus isolates. The isolates were classified as dog-related rabies virus (DRRV) or vampire bat-related rabies virus (VRRV), by nucleoprotein (N) analysis. The nucleotide and amino acid (AA) homologies of the area containing the G protein gene and G-L intergenic region were generally lower than those of the ectodomain. In both regions, nucleotide and deduced AA homologies were lower among VRRVs than among DRRVs. There were AA differences between DRRV and VRRV at 3 antigenic sites and epitopes (IIa, WB+ and III), suggesting that DRRV and VRRV can be distinguished by differences of antigenicity. In a comparison of phylogenetic trees between the ectodomain and the area containing the G prot ein gene and G-L intergenic region, the branching patterns of the chiropteran and carnivoran rabies virus groups differed, whereas there were clear similarities in patterns within the DRRV and VRRV groups. Additionally, the VRRV isolates were more closely related to chiropteran strains isolated from Latin America than to Brazilian DRRV. These results indicate that Brazilian rabies virus isolates can be classified as DRRV or VRRV by analysis of the G gene and the G-L intergenic region, as well as by N gene analysis.
KEY WORDS: Brazil, glycoprotein, molecular epidemiology, rabies virus, vampire bat.

J. Vet. Med. Sci. 66(7): 747–753, 2004

Rabies is one of the most serious and widespread zoonoses in the world. The rabies virus infects a wide range of mammals. Rabid animals show symptoms such as madness, fear of water and encephalitis, and fatality is almost 100% [16]. The rabies virus is bullet-shaped, with a length of 180 nm and a diameter of 75 nm. It belongs to the *Rhabdoviridae* family, and its genome consists of a non-segmented single negative strand of RNA with about 12K bases. This genome encodes 5 proteins in the following order from the 3' end: nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and RNA polymerase (L) [15]. The N protein forms an active cytoplasmic ribonucleoprotein (RNP) complex that is essential for viral propagation [20]; there have been many molecular epidemiological studies of N protein [8–11]. The G protein forms the envelope and capsid (i.e., outer shell) of the virus, which are responsible for viral attachment to the host cell surface and membrane fusion [18]. Because G protein induces production of the virus-neutralizing antibody [19], it is used to immunize animals against rabies [4], and it influences viral pathogenicity and neurotropism [17]. Therefore, analysis of the G protein is important for the investigation of pathogenicity, and is useful in molecular epidemiology because of its host adaptation potential [1]. The G-L intergenic region (pseudogene; ψ) is a noncoding region; evidence suggests that it was once functional but has become vestigial [15]. The G-L region is not subject to immunological selective pressure [13], and has therefore been used in studies of molecular epidemiology of the rabies virus [12–14].

Transmission of rabies by the vampire bat in South America (including Trinidad) was first reported at the beginning of the 20th century [7]. Rabies infection among dogs in urban areas of South America constitutes a well-known public health problem, and rabies transmission by vampire bats has been documented [5], but the extent of rabies infection among livestock and wild animals in South America is not well documented. In Brazil, effective control programs have reduced the number of dog rabies cases, and most urban areas are free from rabies [6]. Nevertheless, the number of cases of livestock rabies transmitted by vampire bats is increasing. There has been very little analysis of the G protein gene of vampire bat rabies in Brazil. Analysis of the G gene of Brazilian rabies virus isolates can improve understanding of the epidemiology and prevention of rabies in South America.

In the present study, we analyzed rabies strains used for studies of the N protein gene [8–10], performing detailed phylogenetic analysis of the G protein gene, which is highly susceptible to substitution. In addition, we investigated the usefulness of the G-L intergenic region of the rabies virus in

* CORRESPONDENCE TO: Prof. SAKAI, T., Department of Preventive Veterinary Medicine and Animal Health, Nihon University School of Veterinary Medicine, 1866 Kameino, Fujisawa, Kanagawa 252–8510, Japan.

phylogenetic studies.

MATERIALS AND METHODS

Samples: The 14 RNA samples used in this study were obtained from 5 cattle, 2 dogs, 2 horses, 1 sheep, 1 pig, 1 vampire bat, 1 cat and 1 human (Table 1). These samples were extracted and used to produce cDNA or PCR products in Brazil. They were previously used for the analysis of N protein by Ito *et al.* [8], and were representative of DRRV and VRRV groups on the phylogenetic tree: DRRV, 6 samples; VRRV, 8 samples. Reference samples included in the phylogenetic trees (Figs. 2 and 3) are from a report by Badrane *et al.* [1].

RT-PCR: The G protein gene was amplified by one-step RT-PCR. The area containing the G protein gene and G-L intergenic region was divided into front (nucleotides [nt] 3221–4135) and back (DRRV, nt 3995–5455; VRRV, nt 3995–5382) regions. In the front region, common primer pairs (Ga3222–40 and Gb4119–39) were used for both DRRV and VRRV (Table 2). A common sense primer (GS3994) and specific antisense primers (DRRV, rabG4; VRRV, G-antiBR2072) were used for the back region. RT-PCR was performed with 1 μl of RNA (0.2–0.4 μg) and the SuperScript One-Step RT-PCR System (Invitrogen Corp., Carlsbad, CA, U.S.A.) according to the manufacturer's instructions. Each reaction mixture was incubated at 50°C for 60 min, followed by 40 cycles of PCR at 94°C for 30 sec., 54 or 55°C for 20 sec., and 68°C for 2 min, in a PTC-0200 DNA Engine (MJ Research Inc., Waltham, MA, U.S.A.). The amplified products were confirmed by 1.5% agarose gel electrophoresis with ethidium-bromide staining, and cDNA was purified with a QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany).

Direct Sequencing: Sequences were determined with an automated sequencer (ABI Prism 3100, Applied Biosystems, Foster city, CA, U.S.A.), the PRISM Dyedeoxy Terminator Cycle Sequencing Ver. 2.0 Ready Reaction kit (Applied Biosystems) and the Big Dye Terminator Cycle Sequencing kit (Applied Biosystems). In the front region, common primers (Ga3222–40, Ga3222–40–1, Gb4119–39 and Gb4119–39–1) were used for both DRRV and VRRV. In the back region, 4 primers (GS3994, GS3994–1, rabG4 and rabG4–1) were used for DRRV, and 4 primers (GS3994, GS3994–1, G-antiBR2072 and G-antiBR2072–1) were used for VRRV (Table 2). The reaction products were purified on a Centri-Sep Spin Column (Applied Biosystems) or by ethanol precipitation.

Table 1. The 14 isolates analyzed in this study

Identification no.	Country	Species	Year of isolation	Accession no.
BRct3[a)]	Brazil	Cat	1999	AB110656
BRhm6[a)]	Brazil	Human	1999	AB110657
BRdg10[a)]	Brazil	Dog	1987	AB110658
BRdg12[a)]	Brazil	Dog	1989	AB110659
BRhr18[a)]	Brazil	Horse	1999	AB110660
BRpg28[a)]	Brazil	Pig	1998	AB110661
BRhr31[b)]	Brazil	Horse	1998	AB110662
BRbv32[b)]	Brazil	Cattle	1994	AB110663
BRvmbt33[b)]	Brazil	Vampire bat	No data	AB110664
BRsp35[b)]	Brazil	Sheep	1992	AB110665
BRbv39[b)]	Brazil	Cattle	1999	AB110666
BRbv45[b)]	Brazil	Cattle	1999	AB110667
BRbv49[b)]	Brazil	Cattle	1989	AB110668
BRbv50[b)]	Brazil	Cattle	1999	AB110669

a) DRRV: dog related rabies virus [10].
b) VRRV: Vampire bat related rabies virus [10].

Table 2. Primers used in this study

Primer	Nucleotide sequences (5'–3')	Position[a)]	Sense	Use
Ga3222–40	CGCTGCATTTTRTCARAGT	3221–3239	+	RT-PCR/Direct sequencing
Gb4119–39	GGAGGGCACCATTTGGTMTC	4116–4135	–	RT-PCR/Direct sequencing
GS3994	CGGMTTTGTGGATGAAAGRGGC	3995–4016	+	RT-PCR/Direct sequencing
rabG4	TGGGTCAATAGGGTCATCATAGAC	5432–5455	–	RT-PCR/Direct sequencing
G-antiBR2072	TGCTGATTGCRCCTACATT	5363–5382	–	RT-PCR/Direct sequencing
Ga3222–40–1	GGGATACATCTCTGCCATA	3518–3536	+	Direct sequencing
Gb4119–39–1	GGGATTTGTCGTATGGGTC	3801–3819	–	Direct sequencing
GS3994–1	GACTTGGAACGAGATCATCC	4373–4392	+	Direct sequencing
G-antiBR2072–1	CATGAAGTATGTGAAGGGC	4978–4996	–	Direct sequencing
rabG-4–1	GACCATGTCTGGTAGATAG	5033–5052	–	Direct sequencing

a) Nucleotide positions are numbered according to the PV sequence (GenBank; accession no. M13215).

Genetic and Phylogenetic Analyses: Fragments of sequence data obtained from direct sequencing were connected with DNASIS (Ver. 2.0). Alignment of the area containing the G protein gene and G-L intergenic region was performed with ClustalX (1.81) and GENETYX-WIN (Ver. 5.0). Subsequently, homologies of the nucleotide sequence and deduced AA sequence were calculated. And 1000 bootstrap replicates of these sequencing data were analyzed by the neighbor-joining (NJ) method. From these results, we deduced the lineages of rabies virus strains.

RESULTS

The deduced AA alignments of all 14 samples are shown in Fig. 1. Among the 524 residues in the deduced AA sequences of the G protein gene, there were 63 distinct AA substitutions between DRRV and VRRV. There were apparent differences between DRRV and VRRV at AA 199 (site IIa, 198–202), AA 264 (site WB+) and AA 332 (site III, 330–338), which are antigenic sites and epitopes [2]. At 6 residues involved in pathogenicity (AA 33, 132, 147, 198, 330 and 333), there was no difference between DRRV and VRRV. The main substitutions in the present deduced AA sequences were at 6 positions in the signal peptide (SP; AA -19–1), 9 positions in the transmembrane (TM; AA 440–461), and 23 positions in the endodomain (ENDO; AA 462–) (Fig. 1).

The nucleotide and AA homologies of the area containing the G protein gene and G-L intergenic region were generally lower than those of the ectodomain (data not shown). For all DRRVs, the nucleotide homology of the area containing the G protein gene and G-L intergenic region was ≥98.2%, and AA homology of the G protein gene was ≥98.9%. For all VRRVs, nucleotide homology was ≥96.1%, and AA homology was ≥97.3%.

The NJ trees, based on the ectodomain and the area containing the G protein gene and G-L intergenic region, showed differences within carnivoran and chiropteran rabies virus groups (Figs. 2 and 3). The ectodomain tree showed 2 groups of carnivoran and chiropteran rabies viruses, and the tree for the area containing the G protein gene and G-L intergenic region showed 3 groups of carnivoran and chiropteran rabies viruses, but there was similarity within DRRV and VRRV groups. The bootstrap values were high in both trees; DRRV and VRRV were located in separate groups in both trees; and VRRV was more highly branched than DRRV in both trees. In both trees, the VRRV group (containing MEX2-VP) was closer to the chiropteran strains MEX2-BT, BRA1-BV and GUY1-BV than was the DRRV group. DRRV constituted an independent group.

DISCUSSION

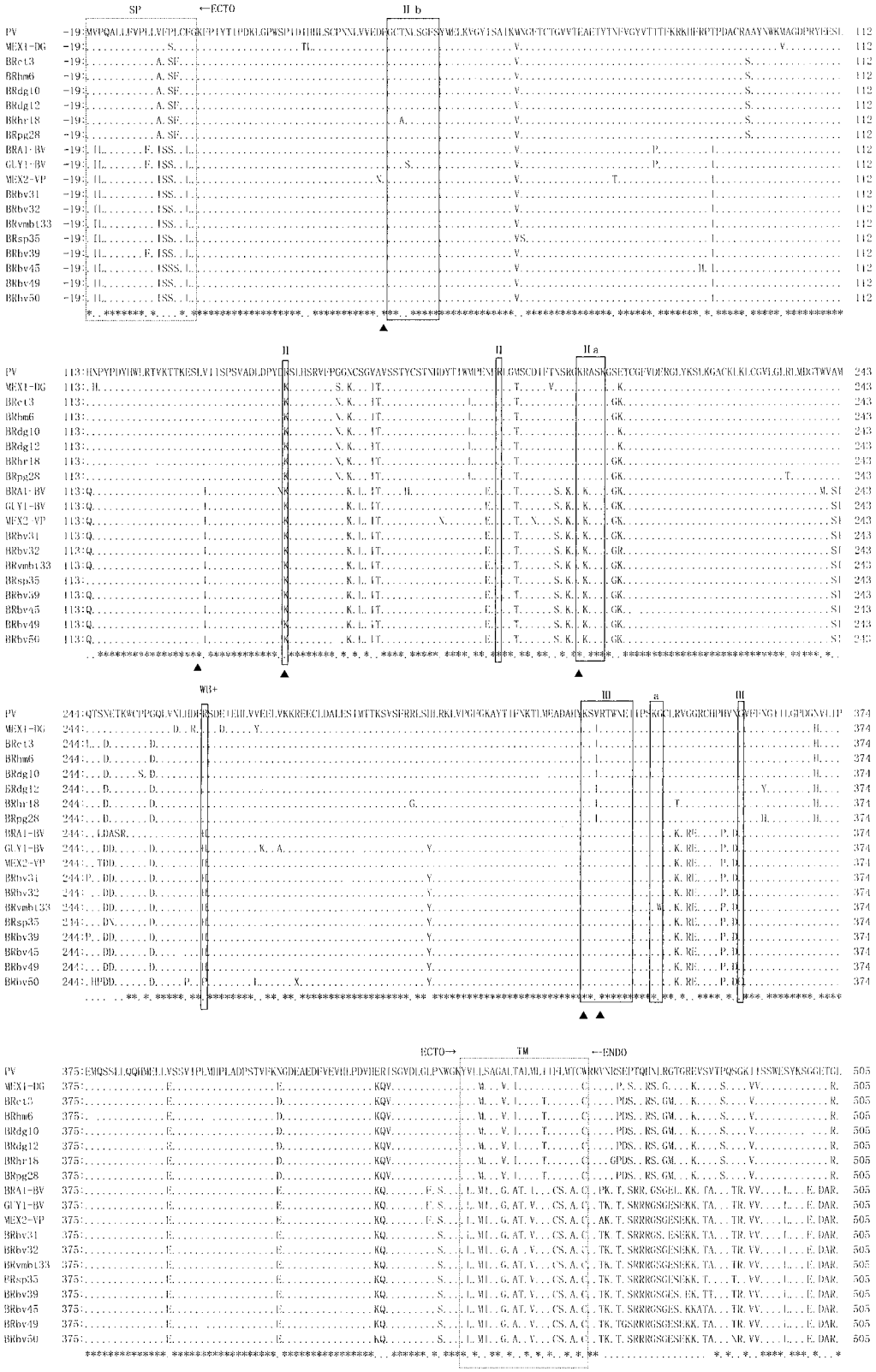
In this study, we analyzed 2066 nucleotides in the area containing the G protein gene and G-L intergenic region of 14 isolates of rabies virus from Brazil. AA alignment showed 3 differences in antigenic sites and epitopes, and many differences in SP, TM and ENDO between DRRV and VRRV. In a study of fixed rabies virus strains (ERA and CVS), Tuffereau *et al.* [17] found that AA 333 (R) is necessary for virulence of rabies. Coulon *et al.* [3] showed that AA 333 (R) and AA 330 (K) are involved in pathogenicity, using fixed rabies virus strains (CVS) and BSR cells (derived from baby hamster kidney cells). But we found no substitutions in these regions of DRRV and VRRV. The present differences at antigenic sites suggest that DRRV and VRRV have distinct antigenic properties.

Nucleotide and deduced AA homologies were greater in the ectodomain than in the area containing the G protein gene and the G-L intergenic region. In both regions, nucleotide and deduced AA homologies were generally lower among VRRVs than among DRRVs. This suggests that overall genetic diversity is greater among VRRVs than among DRRVs.

The present NJ tree based on the ectodomain was similar to an NJ tree constructed by Badrane *et al.* [1], in which chiropteran and carnivoran rabies viruses formed distinct separate groups. The present ectodomain tree supports that carnivoran rabies viruses are derived from chiropteran rabies viruses, as described by Badrane *et al.* [1]. In contrast, in the present NJ tree based on the area containing the G protein gene and G-L intergenic region, there were 2 carnivoran and 1 chiropteran rabies virus groups; i.e., there was no clear distinction between carnivoran and chiropteran rabies viruses. This tree implies that the chiropteran rabies virus group is derived from the carnivoran rabies virus group. This difference between the 2 trees is consistent with a report in which it was concluded that it is to be expected that the G-L intergenic region would develop distinct characteristics due to accumulation of mutations [12], although it is difficult to determine which of the above-described trees accurately depicts the evolution of the rabies virus.

On the other hand, the 2 present trees showed similarities in the branching pattern within the DRRV and VRRV groups; in particular, DRRV had high bootstrap values in both trees. In both trees, the VRRV group was closer to chiropteran strains isolated from Latin America than to Brazilian DRRV; the DRRV group contained only the present 6 DRRV isolates. This suggests that DRRV maintained independent circulation in Brazil. The branches of DRRV in the NJ tree based on the area containing the G protein gene and G-L intergenic region were more highly branched than those of the NJ tree based on the N protein gene [8], suggesting that such multiple branches have contributed to the diversity of the area containing the G protein gene and G-L intergenic region. The present finding that DRRV and VRRV diverge greatly is consistent with a report that VRRV circulated independently in South America for long periods [8].

In the present study, we determined the sequences of 2066 nucleotides in the area containing the G protein gene and G-L intergenic region of 14 isolates of the rabies virus obtained in Brazil. The present nucleotide and AA homologies suggest that the VRRV group is highly diverse. Significantly, AA alignment revealed differences between DRRV


SP
←ECTO
II b
II
II
II a
WB+
III
a
III
ECTO→
TM
←ENDO
PV
MEX1-DG
BRct3
BRhm6
BRdg10
BRdg12
BRhr18
BRpg28
BRA1-BV
GLY1-BV
MEX2-VP
BRbv31
BRbv32
BRvmbt33
BRsp35
BRbv39
BRbv45
BRbv49
BRbv50

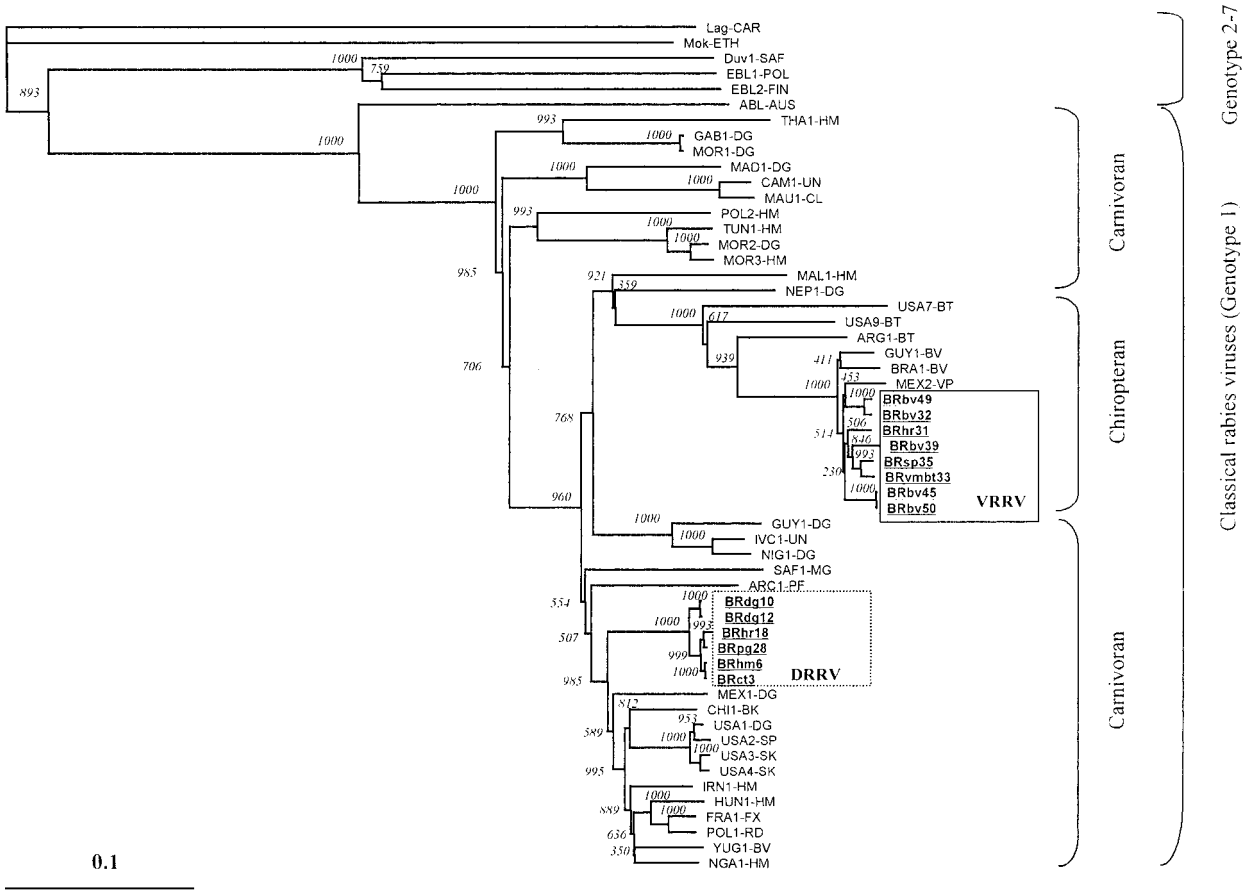


Fig. 2. NJ tree based on the rabies virus G protein and G-L intergenic nucleotides (3291–5363). Nucleotide positions are numbered according to the PV sequence (GenBank; accession no. M13215). Samples in the box of solid lines belong to the VRRV group, and those in the box of dotted lines belong to the DRRV group. The other strains were retrieved from the GenBank and the data of Badrane *et al.* [1]. Numbers show bootstrap values of 1000 replicates.

and VRRV at the positions of antigenic sites and epitopes, suggesting that DRRV and VRRV can be distinguished by differences of antigenicity. Additionally, in the NJ tree, the branching patterns of the ectodomain and the area containing the G protein gene and G-L intergenic region generally differed between the chiropteran and carnivoran rabies virus groups, whereas there was similarity in these branching patterns within the DRRV and VRRV groups. These results indicate that Brazilian rabies virus isolates can be classifies as DRRV or VRRV by analysis of the G gene and the G-L intergenic region, as well as N gene analysis.

ACKNOWLEDGEMENT. This work was financially supported in part by a Grant-in-Aid for Scientific Research (No. 14405032) from the Japan Society for the Promotion of Science.

REFERENCES

1. Badrane, H. and Tordo, N. 2001. Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders. *J. Virol.* **75**: 8096–8104.
2. Badrane, H., Bahloul, C., Perrin, P. and Tordo, N. 2001. Evidence of two Lyssavirus phylogroups with distinct pathogenicity and immunogenicity. *J. Virol.* **75**: 3268–3276.
3. Coulon, P., Ternaux, J., Flamand, A. and Tuffereau, C. 1998. An avirulent mutant of rabies virus is unable to infect motoneurons *in vivo* and *in vitro*. *J. Virol.* **72**: 273–278.
4. Cox, J.H., Dietzschold, B. and Schneider, L.G. 1977. Rabies

Fig. 1. Deduced amino acid alignment of rabies virus G protein. The amino acid position corresponds to nucleotides 3318–4892 of PV (GenBank; accession no. M13215). The notations are the same as those used by Badrane *et al.* [2]. Boxes of solid lines indicate antigenic sites and epitopes. Boxes of dotted lines indicate signal peptide (SP) and transmembrane (TM) domain. The triangles (▲) indicate residues involved in pathogenicity. Arrows indicate ectodomain (ECTO) and endodomain (ENDO).

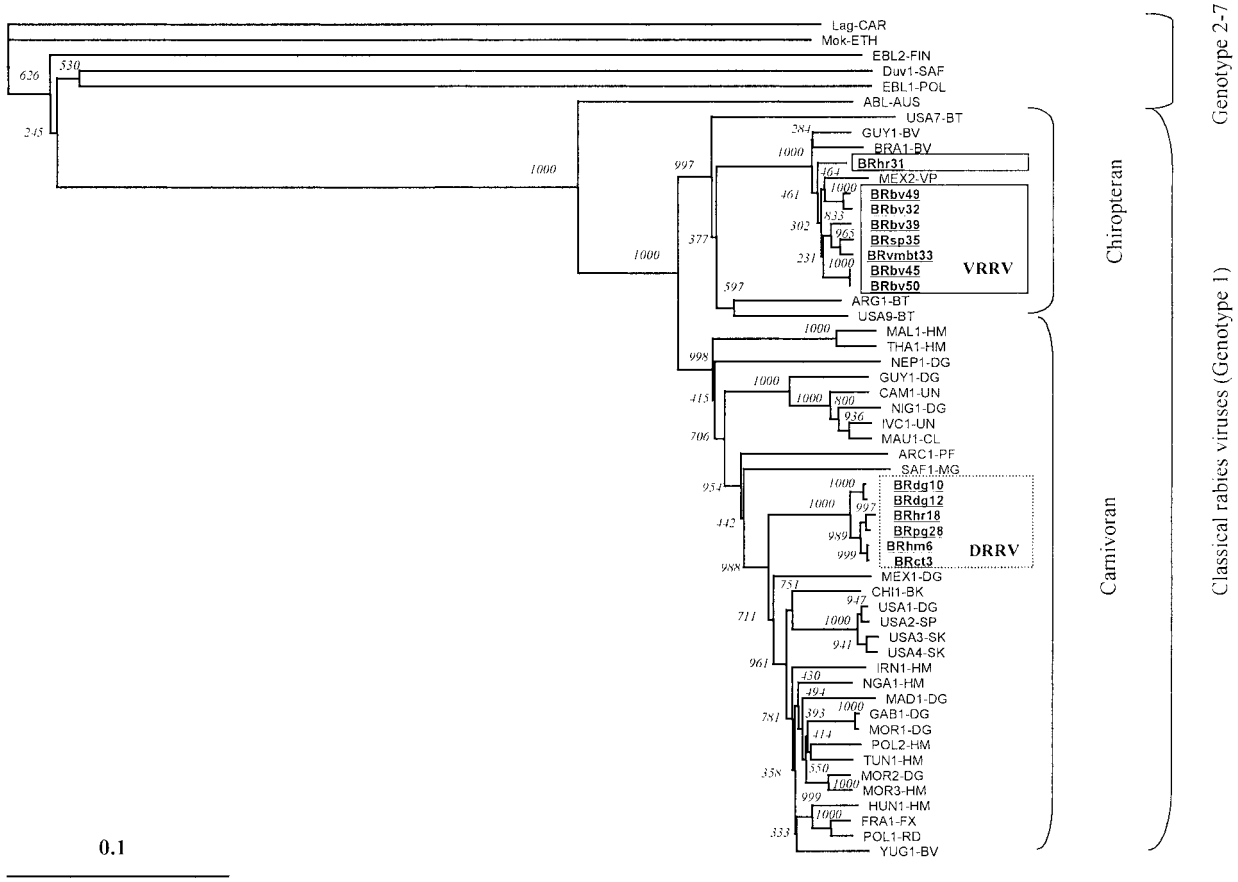


Fig. 3. NJ tree based on the rabies virus G protein ectodomain nucleotides (3375–4691). Nucleotide positions are numbered according to the PV sequence (GenBank; accession no. M13215). Samples in the box of solid lines belong to the VRRV group, and those in the box of dotted lines belong to the DRRV group. The other strains were retrieved from the GenBank and the data of Badrane *et al.* [1]. Numbers show bootstrap values of 1000 replicates.

virus glycoprotein. II. Biological and serological characterization. *Infect. Immun.* **16**: 754–759.

5. Delpietro, H.A., Gury-Dhomen, F., Larghi, O.P., Mena-Segura, C. and Abramo, L. 1997. Monoclonal antibody characterization of rabies virus strains isolated in the River Plate Basin. *Zentralbl. Veterinarmed. B* **44**: 477–483.
6. Heinemann, M.B., Fernandes-Matioli, F.M.C., Cortez, A., Soares, R.M., Sakamoto, S. M., Bernardi, F., Ito, F.H., Madeira, A.M.B.N. and Richtzenhain, L.J. 2002. Genealogical analyses of rabies virus strains from Brazil based N gene alleles. *Epidemiol. Infect.* **128**: 503–511.
7. Hurst, E.W. and Pawan, J.L. 1931. An outbreak of rabies in Trinidad without history of bites and with the symptoms of acute ascending myelitis. *Lancet* **221**: 622–628.
8. Ito, M., Arai, Y.T., Itou, T., Sakai, T., Ito, F.H., Takasaki, T. and Kurane, I. 2001. Genetic characterization and geographic distribution of rabies virus isolates in Brazil: identification of two reservoirs, dogs and vampire bats. *Virology* **284**: 214–222.
9. Ito, M., Itou, T., Sakai, T., Santos, M.F., Arai, Y.T., Takasaki, T., Kurane, I. and Ito, F.H. 2001. Detection of rabies virus RNA isolated from several species of animals in Brazil by RT-PCR. *J. Vet. Med. Sci*. **63**: 1309–1313.
10. Ito, M., Itou, T., Shoji, Y., Sakai, T., Ito, F.H., Arai, Y.T., Takasaki, T. and Kurane, I. 2003. Discrimination between dog-related and vampire bat-related rabies viruses in Brazil by strain-specific reverse transcriptase-polymerase chain reaction and restriction fragment length polymorphism analysis. *J. Clin. Virol*. **26**: 317–330.
11. Nadin-Davis, S.A., Huang, W., Armstrong, J., Casey, G.A., Bahloul, C., Tordo, N. and Wandeler, A.I. 2001. Antigenic and genetic divergence of rabies viruses from bat species indigenous to Canada. *Virus Res.* **74**: 139–156.
12. Sabeta, C.T., Bingham, J. and Nel, L.H. 2003. Molecular epidemiology of canid rabies in Zimbabwe and South Africa. *Virus Res.* **91**: 203–211.
13. Sacramento, D., Badrane, H., Bourhy, H. and Tordo, N. 1992. Molecular epidemiology of rabies virus in France: comparison with vaccine strains. *J. Gen. Virol.* **73**: 1149–1158.
14. Sacramento, D., Bourhy, H. and Tordo, N. 1991. PCR technique as an alternative method for diagnosis and molecular epidemiology of rabies virus. *Mol. Cell. Probes* **5**: 229–240.
15. Tordo, N., Poch, O., Ermine, A., Keith, G. and Rougeon, F. 1986. Walking along the rabies genome: is the large G-L intergenic region a remnant gene? *Proc. Natl. Acad. Sci. U.S.A.* **83**:

3914–3918.
16. Tordo, N., Charlton, K. and Wandeler, A. 1998. Rhabdovirus; Rabies. pp. 665–692. *In*: Topley and Wilson,s Microbiology and Microbial Infections, 9th ed. (Brian, WJ. M. and Leslie, C. eds.), Univ. Oxford Press, New York.
17. Tuffereau, C., Leblois, H., Benejean, J., Coulon, P., Lafay, F. and Flamand, A. 1989. Arginine or lysine in position 333 of ERA and CVS glycoprotein is necessary for rabies virulence in adult mice. *Virology* **172**: 206–212.
18. Whitt, M.A., Buonocore, L., Prehaud, C. and Rose, J.K. 1991. Membrane fusion activity, oligomerization, and assembly of the rabies virus glycoprotein. *Virology* **185**: 681–688.
19. Wiktor, T.J., Gyorgy, E., Schlumberger, D., Sokol, F. and Koprowski, H. 1973. Antigenic properties of rabies virus components. *J. Immunol.* 110: 269–276.
20. Yang, J., Hooper, D.C., Wunner, W.H., Koprowski, H., Dietzschold, B. and Fu, Z.F. 1998. The specificity of rabies virus RNA encapsidation by nucleoprotein. *Virology* **242**: 107–117.