VIROLOGY 209, 526–537 (1995)

Genetic Polymorphism in the Rabies Virus Nucleoprotein Gene

BACHIR KISSI,* NOËL TORDO,† and HERVÉ BOURHY*,[1]

*Unité de la Rage and †Laboratoire des Lyssavirus, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

Received January 10, 1995; accepted March 27, 1995

In an attempt to compare intrinsic and extrinsic genetic diversity of the lyssavirus genotypes, 69 rabies virus isolates from various part of the world were partially sequenced and compared to 13 representative isolates of the 6 lyssavirus genotypes. The analysis of their phylogenetic relationships, performed on the complete nucleoprotein (N) coding gene (1350 bases), established that the rabies virus isolates all belonged to genotype 1 and that at least 11 phylogenetic lineages could be identified in accordance with their geographical localization and species of origin. These lineages diverged mostly by the accumulation of synonymous mutations. Stabilizing selection, possibly related to host specificity, limits amino acid sequence and antigenic drift. Analyses were also performed either on the highly variable 400-base region coding for the amino terminus of the N protein or on the 93-base noncoding region corresponding to the 3′ end of the N mRNA, the intergenic N-phosphoprotein (M1) region, and the 5′ end of the M1 mRNA. These shorter nucleotide sequences were shown to provide phylogenetic data suitable for the completion of large epidemiological studies, but with less robustness. This latter noncoding sequence, despite a 3.1 times higher mutation rate than its adjacent coding N gene, followed a parallel evolutionary pattern.

INTRODUCTION

Rabies is an enzootic disease widespread throughout the world and is a serious health problem in developing countries. Previous antigenic studies performed on both the nucleoprotein (N) and the glycoprotein have demonstrated the antigenic diversity of lyssaviruses (Wiktor *et al.*, 1980; Rupprecht *et al.*, 1987; Smith, 1989). The molecular diversity of the lyssaviruses has been reinvestigated recently by sequencing the N gene of representative isolates and six genotypes were distinguished (Bourhy *et al.*, 1992, 1993). However, little is known of the intrinsic molecular diversity of isolates of genotype 1 (rabies virus) compared with the extrinsic molecular diversity between the different genotypes. The recent studies were either geographically restricted (Nadin-Davis *et al.*, 1993, 1994; Sacramento *et al.*, 1992) or arbitrarily limited to a very small region of the genome (Rupprecht and Smith, 1994; Smith *et al.*, 1992). To date, this lack of data hampers our understanding of the evolution of lyssaviruses and in particular that of rabies virus.

The N protein of lyssaviruses is a phosphorylated protein 450 or 451 amino acids long (Tordo *et al.*, 1986a; Bourhy *et al.*, 1993), which is synthesized in large amounts during cell infection. During morphogenesis, the N protein binds tightly to genomic RNA protecting it from ribonuclease action (Sokol *et al.*, 1969; Wiktor and Clark, 1973). In the mature virion, the N protein constitutes the major component of the internal helical nucleocapsid. In rhabdoviruses, it is involved in the regulation of transcription and replication (Patton *et al.*, 1984). It also plays an important role in the T helper response, in particular against challenge with lyssaviruses antigenically distant from the vaccinal strains (Dietzschold *et al.*, 1989; Perrin *et al.*, 1991; Ertl *et al.*, 1991). Vaccine composed exclusively of N protein can protect animals from peripheral challenge (Fang Fu *et al.*, 1991; Fekadu *et al.*, 1992; Lodmell *et al.*, 1991; Sumner *et al.*, 1991). Furthermore, the recent characterization of the super antigen properties of the N protein has reinforced interest in its immunological role (Lafon *et al.*, 1992).

The N gene was chosen for several reasons. First, the amplification of the N gene using PCR as a simple diagnostic technique can be expanded to allow a precise typing method based on nucleotide sequences (Sacramento *et al.*, 1991). Second, the N sequence of representative isolates of all six genotypes of lyssavirus is available and was shown to provide sufficient information for assignment of isolates to one of the distinct genotypes (Bourhy *et al.*, 1993). Third, from a comparative point of view, the important contribution of capsid proteins to varied structural and immunological functions and the purported role of Influenza nucleoprotein as a determinant of host range (Scholtissek *et al.*, 1985) make the evolutionary analysis and genetic variability of rabies N protein interesting. The nucleoprotein coding sequences and a noncoding sequence including the 3′ end of the N mRNA, the N-phosphoprotein (M1) intergenic region, and the 5′ end of M1 mRNA, of 69 rabies viruses collected from all over the world were obtained. Extensive molecular and phylogenetic analyses were conducted to

[1] To whom correspondence and reprint requests should be addressed. Fax: (33) 1 40 61 30 20.

0042-6822/95 $6.00

further evaluate the intrinsic genetic variability and to determine the evolutionary pattern of these virus genes and the encoded N protein within genotype 1 of lyssavirus.

MATERIALS AND METHODS

Viruses

Original infected brain or suckling mouse brains infected with the original virus at the lowest number of passage available were used. Sixty-nine isolates originating from 46 different countries (Table 1) were studied and compared to the sequences of 7 isolates representing the 6 different genotypes of lyssavirus circulating in the world: 8918FRA and 8615POL, 9007FIN, 9018HOL, 8619NGA, MOK (Bourhy *et al.,* 1993), of two isolates from Ontario (Nadin-Davis *et al.,* 1993), and of four laboratory strains PV (Tordo *et al.,* 1986a), SADB19 (Conzelmann *et al.,* 1990), AVO1 (Poch *et al.,* 1988), and CVS (Mannen *et al.,* 1991).

PCR and direct sequencing

Oligodeoxynucleotide primers used for cDNA synthesis, polymerase chain reaction (PCR), and sequencing were selected by a comparison between rabies virus (PV) and Mokola virus sequence (Bourhy *et al.,* 1992) or were deduced from preliminary alignments of strains of genotype 1 (Table 2). Briefly, brains of infected animals were subjected to RNA extraction, cDNA synthesis, and PCR amplification as previously described (Sacramento *et al.,* 1991). Amplified products were purified by electrophoresis on 0.7% NuSieve GTG agarose gel (FMC) (Sacramento *et al.,* 1991) and were sequenced using appropriate primers (Table 2) on both strands and the T7 sequencing kit (Pharmacia).

Sequence alignments and phylogenetic analysis

The nucleotide and the deduced amino acid sequences were aligned initially using the multiple alignment program CLUSTALV (Higgins and Sharp, 1989). The pairwise similarity values were used to determine the mean distance, the mean transition, the mean transversion, and the mean ratio transition/transversion (Tekaia, unpublished programs) for each of the alignments. Synonymous and nonsynonymous nucleotide substitutions were computed by the method of Nei and Gojobori (1986) using the DISTANCE ESTIMATION program of the MEGA package Version 1.01 (Kumar *et al.,* 1993). Sequences were analyzed using the neighbor-joining method (Saitou and Nei, 1987) and implemented in CLUSTALV (Higgins and Sharp, 1989). The BOOTSTRAP option of CLUSTALV (1000 replicate data sets) was used to assess the robustness of the method. A phylogenetic analysis was also performed by the maximum parsimony method, implemented in the PHYLIP package Version 3.52c (Felsenstein, 1993). First, SEQBOOT was run to produce 100 bootstrapped data sets; then the phylogeny estimate for each of these sets was found using DNAPARS or PROTPARS for nucleotide or amino acid sequences, respectively. CONSENSE was used to generate the majority rule consensus tree and the bootstrap confidence limits. Finally, the TREETOOL program Version 1.0 was used to produce graphic outputs of phylogenetic trees (Maciukenas, 1992). The running average of similarity among the nucleotide and amino acid sequences in the multiple sequence alignment was plotted by PLOTSIMILARITY of the GCG package (Genetic Computer Group Inc., 1991).

Antigenic analysis

The reactivity patterns of 24 antinucleoprotein monoclonal antibodies (N MAbs) (Lafon and Wiktor, 1985; Dietzschold *et al.,* 1987) were determined by an indirect fluorescent-antibody-staining method using a fluorescein-labeled anti-mouse IgG (Biosys, Compiègne, France) directly on infected cells or suckling mouse brains (Wiktor *et al.,* 1980).

RESULTS

Genetic diversity of the nucleoprotein gene of lyssaviruses

The nucleotide sequence of the nucleoprotein coding gene of 50 isolates was determined and compared to those of 11 isolates and laboratory strains representing the 6 different genotypes of lyssaviruses. The variation in the level of similarity along the sequences was investigated (Fig. 1). Two regions show a level of similarity clearly below the average: a 199-base-long region (nucleotides 1080 to 1278) and a longer one located between nucleotides 99 and 405, which is the most variable. Phylogenetic trees were constructed using both the neighbor-joining method (Fig. 2) and the maximum parsimony method (not shown). Both methods produced trees which showed the same general pattern of branching order and which split the viruses into six major clusters, corresponding to the six genotypes previously reported (Bourhy *et al.,* 1993). All the isolates sequenced in this report as well as all the laboratory strains correspond to viruses of genotype 1. These relationships between the isolates were strongly supported by bootstrap resampling, which revealed that clusters delineating genotypes 1, 2, 4, 5, and 6 were found in all of the trees tested (Table 3). Results concerning genotype 3 were not available because Mokola virus, the most divergent lyssavirus (Bourhy *et al.,* 1993), was chosen as the outgroup. The comparison established that isolates belonging to different genotypes have less then 79.8 and 93.3% (between isolates 86132AS (Duvenhage virus) and 8918FRA (EBL1)) similarity at the nucleotide level and the amino

TABLE 2

Synthetic Oligonucleotides Used in PCR and Sequencing

Primer	Sequence	Position[a]
	Messenger sense	
N1	TTTGAGACAGCCCCTTTTG	587–605
N3	GTCTCTTTGAAGCCTGAG	113–130
N5	GAAGGCAATTGGGCTC	419–434
N7	ATGTAACACCTCTACAATG	55–73
N12	GTAACACCTCTACAATGG	57–74
N13	GTTTATACTCGAATCATGATG	1583–1603
N17	TTCTTCCACAAGAACTTTG	848–866
	Genomic sense	
N2	CCCATATAGCATCCTAC	1030–1013
N4	GCTTGATGATTGGAACTGA	1368–1350
N6	GAGCCCAATTGCCTTC	434–419
N8	AGTTTCTTCAGCCATCTC	1584–1568
N14	TTGTGAGTAGTCATTA	645–630
N23	GGTCTCTCGTCAGTTCCAT	464–446
N39	CCATCTCAAGATCGGCCA	1575–1557

[a] Primer position is indicated according to the PV strain rabies genome (Tordo *et al.*, 1986a).

acid level, respectively. Conversely, the minimum percentage similarity connecting one isolate to others within genotype 1 is 83.3% (8738THA–86118BRE) and 92.2% (8708NAM–86118BRE) at the nucleotide level and the amino acid level, respectively.

Phylogenetic clusters of geographically related isolates

The nucleoprotein genes of the 54 viruses representing members of genotype 1 were all 1350 nucleotides long. The mean distance obtained at the nucleotide level from all the comparisons of the N coding sequences was 9.6% with a standard deviation (SD) of 3.5%. The maximal distance obtained was 16.7% (86118BRE–8738THA). A remarkable feature of the phylogenetic trees is the organization of isolates of genotype 1 into at least 11 distinct clusters corresponding to genetic lineages (Fig. 3). These lineages grouped geographically or host-species-related isolates and were named according to their geographical distribution: Africa 1a, Africa 1b, Africa 2, Africa 3, Asia, Arctic, Europe/Middle East, Latin America 1, and Latin America 2. Two groups of vaccinal strains can also be distinguished. For all the branches, relatively high bootstrap values (96–100), except cluster Africa 1a, were obtained supporting the reliability of the defined lineages (Table 3). As a result of the previous evaluation of variability along the N gene, limited sequence studies of the 400 nucleotides coding for the highly variable amino terminus were performed on 19 additional isolates and 2 published sequences of isolates from Canada (Nadin-Davis *et al.*, 1993). Phylogenetic trees obtained with either the fragment or the whole N gene show correlation. Most of the lineages defined previously using the whole N sequence were found in more than 51% of the bootstrap analyses, with the exception of lineages Africa 1a, Africa 1b, and Latin America 1. These scores were very low (Table 3) due mainly to branching differences of isolates 8670NGA, 9228CAF, and 9229CAF with lineage Africa 1b and of isolates 86123BR with lineage Latin America 1 (not shown). In conclusion, these analyses allowed precise determination of the hosts and of the circulating zone of each lineage. Dog isolates from Malaya, Sri Lanka, and Thailand clustered in lineage Asia. Dog and fox isolates from Greenland, Canada (Ontario), Russia (Siberia), and Nepal grouped in the Arctic lineage, supporting a large distribution of a common ancestor of the arctic rabies virus in the Northern Hemisphere. The Europe/Middle East lineage contained all of the dog, fox, raccoon, and wolf isolates from Estonia, France, Germany, Iran, Israel, Poland, Saudi Arabia, the Sultanate of

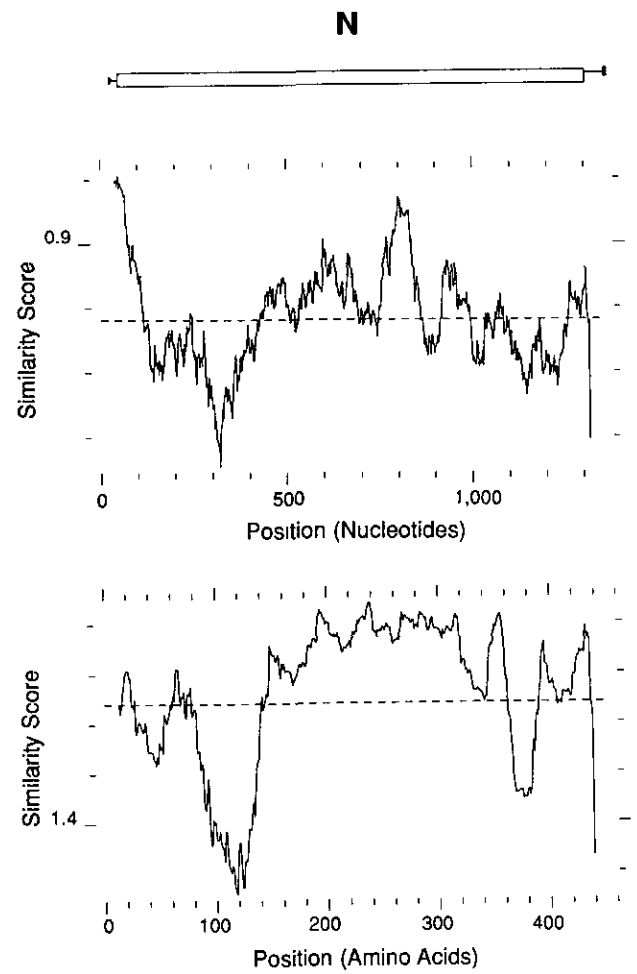

FIG. 1. Variability of the similarity score of the nucleotide (upper) and AA (lower) sequence of the N gene. The running average of the similarity among the nucleotide and amino acid sequences in the multiple sequence alignment was plotted by the PLOTSIMILARITY program (Genetic Computer Group, 1991).

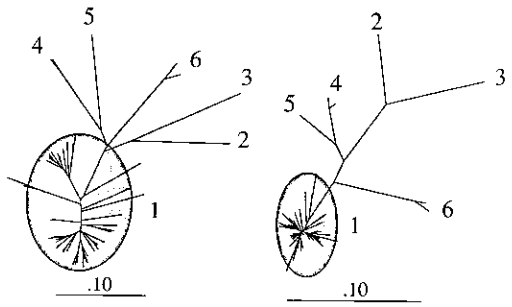

FIG. 2. A radial phylogenetic tree showing the relationships between the different genotypes of lyssavirus on the basis of the nucleotide sequence (left) or of the amino acid sequence (right). Sequences were analyzed using the neighbor-joining method (Saitou and Nei, 1987) implemented in the CLUSTALV package (Higgins and Sharp, 1989). The shaded ellipses surrounding the branches correspond to the viruses of genotype 1. The length of the branches indicates the phylogenetic distance between the different viruses.

Oman, and the former Yugoslavia. Clusters Latin America 1 and 2 included dog isolates from Brasil and Mexico and vampire bat isolates from Guyana and Brasil, respectively. Two groups of vaccinal strains (AVO1, CVS and PV, SADB19) were also distinguished. A substantial divergence can be seen for isolates from Africa (4 separate lineages). Lineage Africa 1a included all the dog and wildlife isolates from Algeria, Ethiopia, Gabon, Madagascar, Morocco, and Tunisia, while all the dog and wildlife isolates from the Central African Republic, Kenya, Mozambico, Namibia, Nigeria, Tanzania, Zaire, Zambia, Zimbabwe, and some (8721AFS, 2900AFS, and 9362AFS) of Republic of South Africa were in lineage Africa 1b. Lineage Africa 2 contained all of the dog isolates from Benin, Burkina Fasso, Cameroun, Guinea, Ivory Coast, Mauritania, Niger, Senegal, Somaliland, and Chad. Lineage Africa 3 (mongoose isolate 1500AFS) represented the rabies transmission cycle in mongoose in South Africa (King *et al.*, 1994). Finally, one human (bitten by a dog) isolate from Egypt (8692EGY) did not group with any of the previously defined lineages of Africa or of the Middle East. Further analysis will be necessary to confirm if it belongs to a putative lineage Africa 4.

Relationships between the amino acid sequence of the nucleoprotein

The deduced amino acid sequences for the 61 nucleoprotein genes of lyssaviruses were aligned and examined by a plot similarity analysis (Fig. 1). It became apparent that variability existed in the same regions in the N

TABLE 3

Bootstrap Frequencies of the Major Phylogenetic Clusters Obtained by Analysis of the N Coding Sequence (1350 b), the Noncoding Sequence Located between the Stop Codon of N Gene and Start Codon of M1 Gene (93 b), the 400-base Coding for the Amino Terminus End of N Protein, and the Amino Acid Sequence of N Protein (450 Amino Acids)

	N coding sequence		N–M1 intergenic sequence		400 b of N		N protein	
Lineages	PARS	NJ	PARS	NJ	PARS	NJ	PARS	NJ
Genotype 2	100	ND	NI	NI	NI	NI	ND	ND
Genotype 3	100	(Root)	NI	NI	NI	NI	100	(Root)
Genotype 4	ND	ND	NI	NI	NI	NI	ND	ND
Genotype 5	100	100	NI	NI	NI	NI	100	100
Genotype 6	100	100	NI	NI	NI	NI	100	100
Genotype 1	100	100	NI	NI	NI	NI	100	100
Europe/Middle East	98	100	18	32	59	76	19	32
Africa 1a	47	84	—	—	—	—	32	44
Africa 1b	98	96	74	61	8	—	11	17
Africa 2	100	100	97	98	89	98	14	45
Africa 3	ND	ND	ND	ND	ND	ND	ND	ND
Latin America 1	98	100	89	95	—	—	21	50
Latin America 2	100	100	100	100	100	100	97	93
Arctic	100	100	93	100	89	99	99	99
Asia	ND	ND	100	(Root)	51	81	ND	ND
Vaccin 1	100	100	91	97	100	100	92	92
Vaccin 2	100	100	100	100	100	100	100	99

Note. NJ values were obtained by the BOOTSTRAP option of CLUSTALV (1000 replicate data sets) (Saitou and Nei, 1987). PARS values were obtained by the maximum parsimony method (Fitch, 1971), implemented using the PHYLIP package Version 3.52c (Felsenstein, 1993) applied to 100 bootstrapped data sets. Abbreviations used: ND, not determined (when a lineage contains only one isolate); NI, not included in the analysis; —, values below 5; (root), results concerning genotype 3 (Mokola virus) and lineage Asia (8738THA) are not available because they were chosen as outgroups; genotype 2, 8619NGA; genotype 3, MOK; genotype 4, 86132AFS; genotype 5, 8615POL and 8918FRA; genotype 6, 9018HOL and 9007FIN. Bootstrap values were obtained on 61 sequences for N coding sequence and N protein sequence, on 53 sequences for N–M1 intergenic sequence and on 69 sequences for 400 bases on N gene.

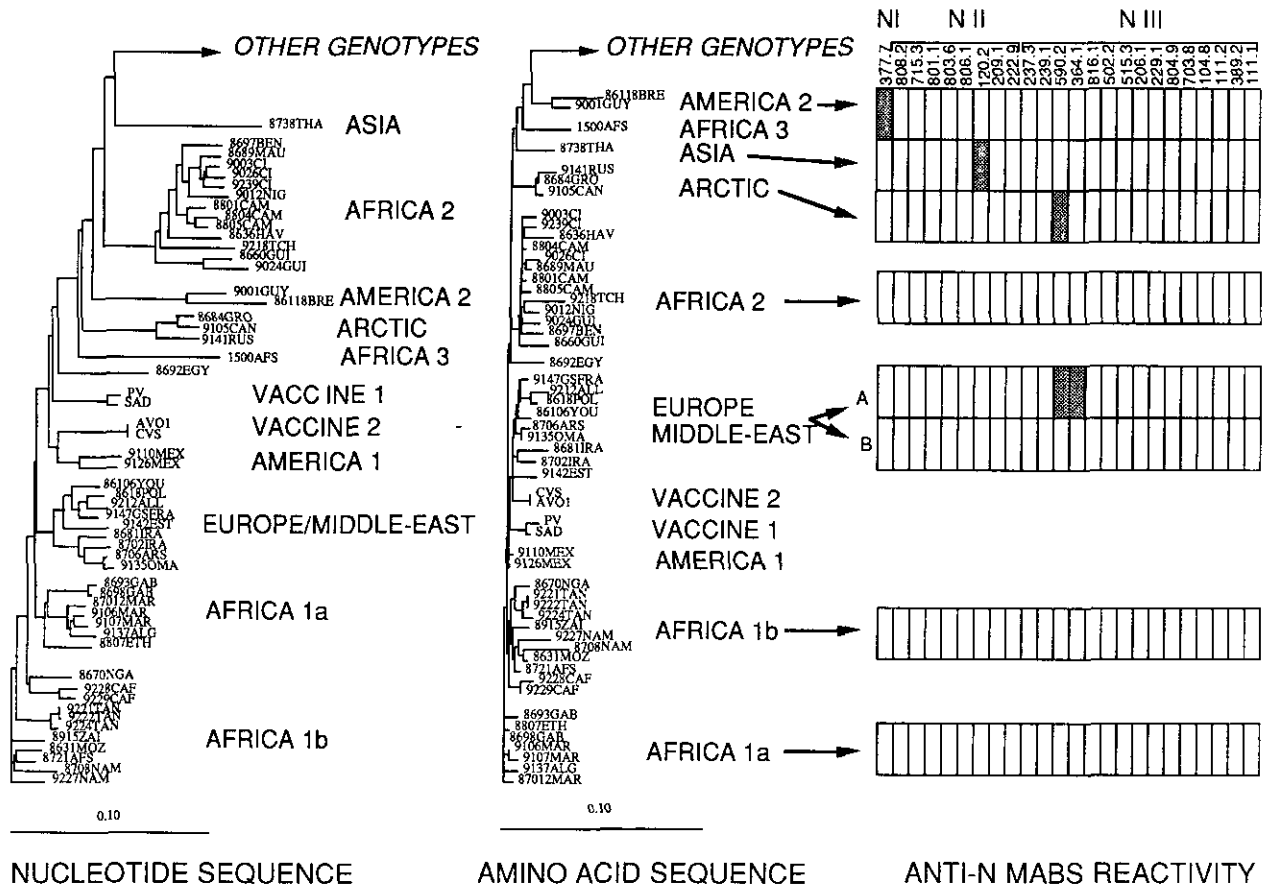

FIG. 3. A phenogram showing the phylogenetic relationships between the N coding sequences and reaction pattern of N MAbs of isolates of the different lineages of genotype 1. Sequences were analyzed using the neighbor-joining method (Saitou and Nei, 1987) implemented in the CLUSTALV package (Higgins and Sharp, 1989). Vertical branches are noninformative and set for clarity only. The length of the horizontal branches reflects phylogenetic distance relationships. The name of the 11 lineages described in the text are indicated. Twenty-four MAbs describing three antigenic sites, NI, NII, and NIII on the ERA virus nucleoprotein (Dietzschold *et al.*, 1987; Lafon and Wiktor, 1985), were used to determine the reactivity patterns of 16 representative isolates of the different lineages: 8738THA, 8636HAV, 8660GUI, 8689MAU, 8684GRO, 86118BRE, 9001GUY, 8670NGA, 8631MOZ, 8721AFS, 8693GAB, 87012MAR, 8807ETH, 9147FRA, 8706ARS, and 8702IRA. In the cluster Europe/Middle East, the reactivity pattern of fox isolates (A) is distinguished from that (B) of the other wildlife and dog isolates. Abbreviations used: open box, positive reaction; shaded boxes, negative staining.

gene at both the nucleotide and amino acid levels. We then considered the alignment of 54 nucleoprotein sequences of genotype 1. Of the 450 amino acids, 299 were conserved in all the isolates (Fig. 4). The putative phosphorylation site mapping to serine in position 389 (Dietzschold *et al.*, 1987) was one of these. A central domain of 147 AA (from AA 182 to 328) exhibited 73% similarity. Conversely, the flanking regions toward the amino and carboxy termini showed a similarity score below average at the amino acid level. Phylogenetic analyses were also performed on the deduced amino acid sequences using, in parallel, the neighbor-joining (Fig. 2) and maximum parsimony methods (not shown). This resulted in trees with the same general topology as those based on the nucleotide sequence. Clusters delineating genotypes 1–6 were found in all the trees tested (Table 3). Nevertheless, within genotype 1, the branch lengths of the different lineages of the amino acid tree were shorter than that observed in the nucleotide analysis. The branching order connecting some lineages was also not totally concordent, and the scores obtained by bootstrap analysis with lineages Europe/Middle East, Africa 1a, Africa 1b, Africa 2, and Latin America 1 were very low (Fig. 3, Table 3). The most striking difference was the close relationships seen between amino acid sequences of isolates belonging to lineage Africa 2 and those to lineages Africa 1a and 1b, Europe/Middle East, and Latin America 1, despite their relative distance at the nucleotide level. To tentatively associate certain amino acid substitutions with changes in the antigenic structure, 16 field isolates representative of the different lineages (except Africa 3 and Latin America 1) were further characterized by using 24 MAbs directed against the antigenic sites I, II, and III of the nucleoprotein. The antigenic profiles of dog isolates from Europe/Middle East and Africa 1a, 1b, and 2 were again remarkably

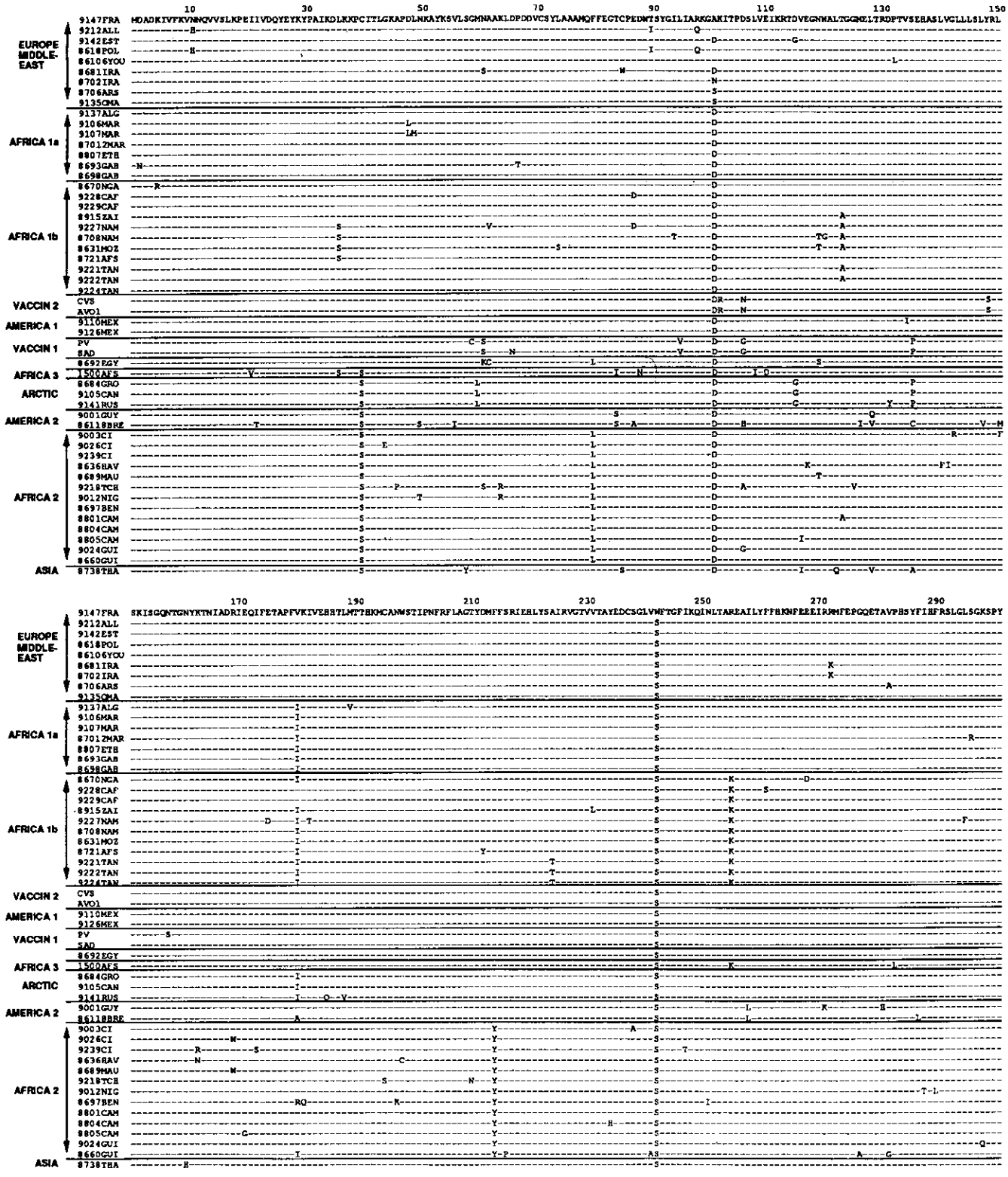

FIG. 4. Predicted amino acid sequences for the nucleoprotein of lyssaviruses of genotype 1. The sequences of 50 isolates were aligned with PV (Tordo *et al.*, 1986a), SAD B19 (Conzelmann *et al.*, 1990), AVO1 (Poch *et al.*, 1988), and CVS (Mannen *et al.*, 1991) using the multiple alignment program CLUSTALV (Higgins and Sharp, 1989). The single-letter amino acid code is used. Only amino acid changes relative to the sequence of 9147FRA are indicated. The putative phosphorylation site mapping to serine in position 389 is marked with an asterisk. Isolates belonging to the same lineages were grouped. Antigenic sites NI and NIII determined on the ERA virus nucleoprotein (Dietzschold *et al.*, 1987; Lafon and Wiktor, 1985) are boxed.

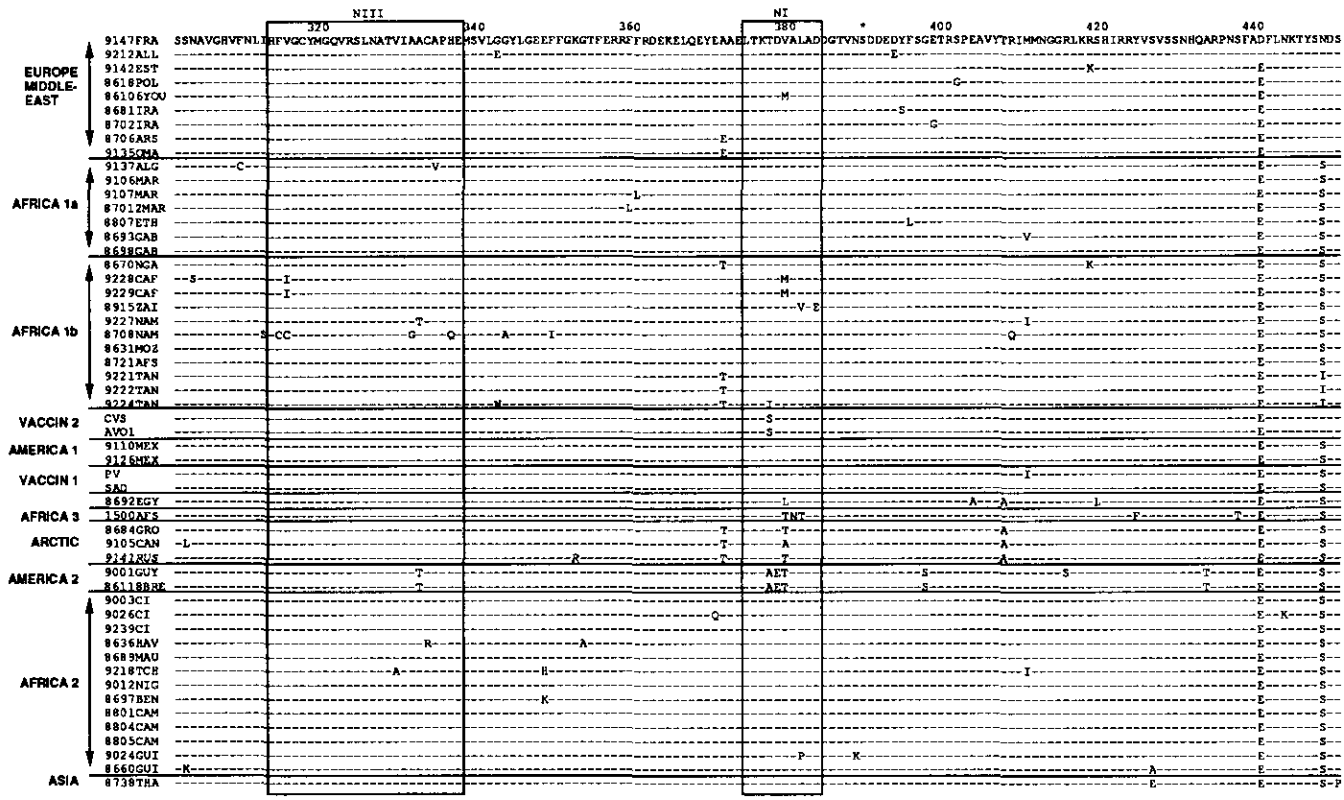

FIG. 4 — *Continued*

identical, confirming the amino acid sequence analysis (Fig. 3). They share a positive reaction with all the MAbs. This established the antigenic homogeneity of the nucleoprotein of rabies virus adapted to domestic dogs in Europe/Middle East and Africa, despite different phylogenetic origin based on the analysis of nucleotide sequences. The reactivity pattern of antigenic site III was only different in fox strains from Europe and Middle East, which exhibited a negative reaction with MAbs 590.2 and 364.1 and also in Arctic rabies virus, which showed a negative reaction with MAbs 590.2. The reactivity pattern of antigenic site II, a discontinuous epitope, was also different in the Asian isolate which had a negative reaction with MAb 120.2. In those cases, no mutation could be associated with the epitope changes. MAb 377.7 gave a negative reaction with vampire bat isolates (cluster Latin America 2). This latter could be related to the mutations TD to AE, in positions 377–378 in the antigenic site I, which are synapomorphic characteristics of the vampire bat isolates in the panel of isolates selected (Fig. 4).

TABLE 4

Proportion of Synonymous (*pS*) and Nonsynonymous (*pN*) Nucleotide Substitutions in the N Coding Sequence (×100)

Lineage	NB	*p*S (SD)	*p*N (SD)
Africa 1a	21	13.7 (6.1)	0.4 (0.2)
Africa 1b	55	14.5 (4.6)	0.9 (0.4)
Africa 2	78	17.9 (6.9)	0.9 (0.4)
America 1	2	16.4 (5.1)	0.1 (0.2)
America 2	2	19.9 (6.2)	1.6 (0.4)
Arctic	3	10.1 (3.4)	0.3 (0.2)
Eur/M E	36	15.0 (5.4)	0.6 (0.2)
Genotype 1	1431	37.1 (14.0)	1.3 (0.6)

Note. These proportions of mutations were computed by the method of Nei and Gojobori (1986) using the DISTANCE ESTIMATION program of the MEGA package Version 1.01 (Kumar *et al.*, 1993). Abbreviations used: NB, number of sequences compared; SD, standard error; Eur/M E, Europe/Middle East.

Evolution and pattern of nucleotide substitution

The regions corresponding to the 3′ noncoding sequence of the N mRNA, the N–M1 intergenic region, and the 5′ noncoding sequence of M1 mRNA are identical in size (62, 2, and 29 nucleotides, respectively) in all the isolates of genotype 1 (Fig. 5) The polyadenylation signal of the N mRNA (HGA7), the N–M1 intergenic region (CT), and a pentanucleotide (AACAC) located in the putative transcription start signal of the M1 gene are also identical between all of the sequences, confirming previous results (Tordo *et al.*, 1986a,b; Bourhy *et al.*, 1993). A phylogenetic analysis was also performed on these 93 nucleotides and the region appeared to be discriminative despite its very short length. The isolates were grouped in

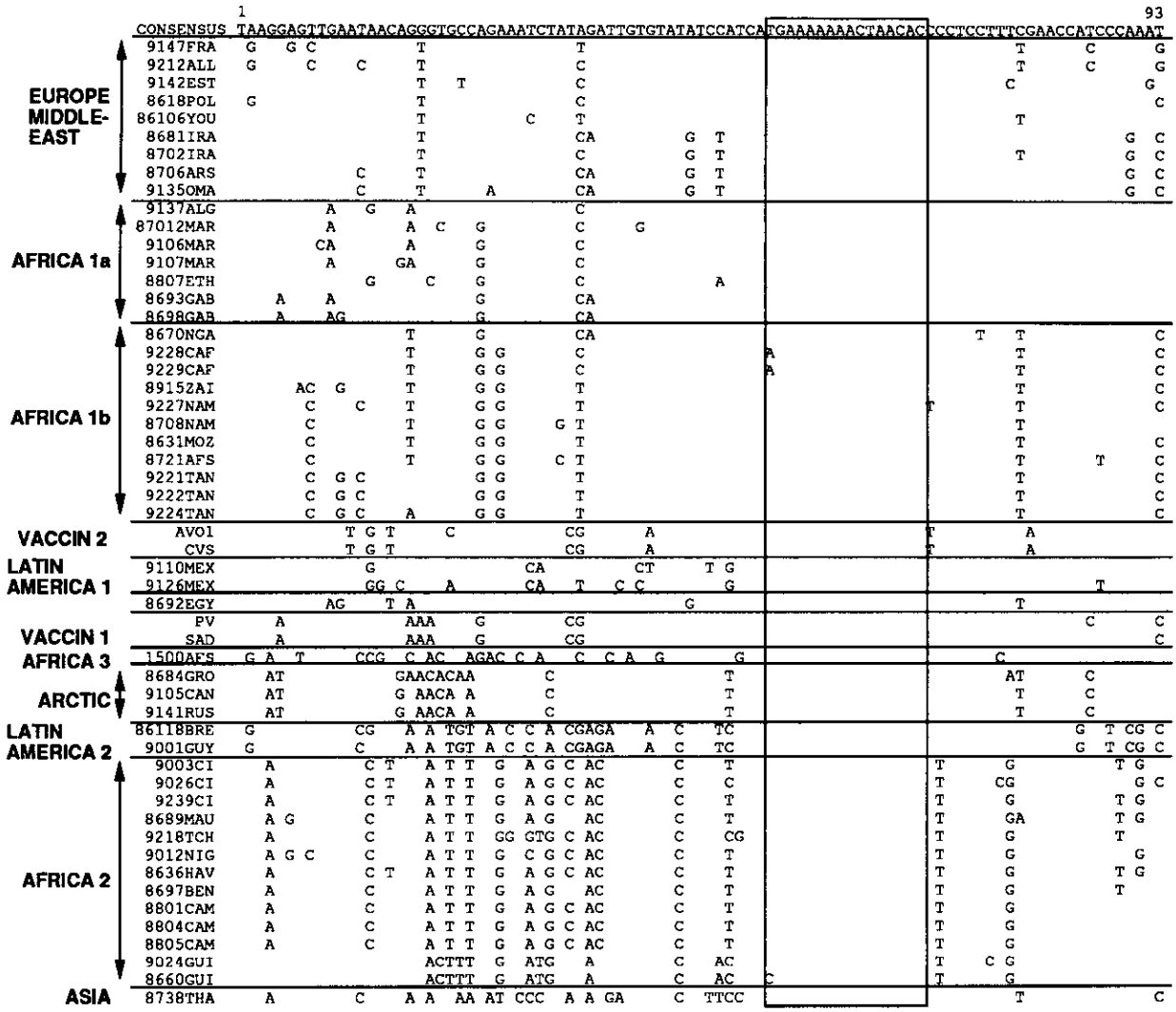

FIG. 5. Alignment of the 3′ stop sequence of the N mRNA, the N–M1 intergenic region, and the 5′ start sequence of M1 mRNA. Only nucleotide changes relative to the consensus sequence calculated by PRETTY (Genetic Computer Group, 1991) are indicated. The 16 nucleotides conserved in almost all the sequence are boxed (see details in the text).

the same clusters as obtained using the total N gene coding sequence (not shown), although the bootstrap score of lineage Europe/Middle East was low (Table 3). However, there were two exceptions: isolate 8807ETH was no longer directly associated with lineage Africa 1a but stood alone and, conversely, isolate 8692EGY branched within this lineage. The mean distance obtained at the nucleotide level from all the comparisons is 18.3% (SD, 8.1%). This was 1.9 times higher than that found for the N gene coding sequence. The ratio of transitions (p) to transversions (q) (p/q, 2.5; SD, 2.2) was 1.6-fold lower than that found with the N coding sequence ($p/9$, 4.1; SD, 1.6). This indicated a higher proportion of transitions, rather than transversions, involved in the divergence of both coding and noncoding sequence and, also, that the proportion of transversions was lower in the N coding gene. As expected, the proportion of synonymous or silent substitutions (pS) was, in all the comparisons greater than that of nonsynonymous or amino acid altering substitutions (pN). However, the particular increase of pS when all the isolates of genotype 1 were compared indicated that the participation of synonymous substitutions is increasing with the phylogenetic distance.

DISCUSSION

The lack of proofreading among RNA polymerase is a source of replication errors potentially resulting in enormous genetic variation (Holland *et al.*, 1982; Steinhauer and Holland, 1987). Genetic heterogeneity, rapid evolution, and consequent antigenic diversity, if confirmed, will offer the rabies virus significant opportunity to colonize new ecological niches, i.e., new vectors, and to evade the host immunity induced by current rabies vaccines. The genetic diversity presented in this study is almost

certainly underestimated, due to sample bias. However, it presents the first comparison between the intragenotype and the intergenotype molecular diversity of the N gene. The phylogenetic analyses of N protein gene of 82 lyssaviruses confirmed the existence of 6 different genotypes (Bourhy *et al.,* 1993) and also provided good predictive value for assignment of rabies isolates of genotype 1 to distinct genetic lineages. In contrast, no precise limit concerning the percentage of nucleotide and amino acid similarity to enable one isolate to be classified in a definite genotype could be given, as it was established that some overlap was possible. From our present state of knowledge of the N protein gene, we can only assume that isolates sharing less than 80% of nucleotide and 92% of amino acid similarity belong to different genotypes. Some of the lineages, containing only one or two isolates each, may not warrant consideration as distinct groups. However, the robustness of the bootstrap resampling and their phylogenetic distance to other clusters obtained by two independent phylogenetic methods, establishing them as distinct entities, should be emphasized. These studies also showed that two short regions, the 400 nucleotides coding for the amino terminus of the N protein and the 93 nucleotides of the N–M1 noncoding region, can be used to determine the geographical distribution of the major virus lineages previously defined by the study of the whole N coding sequence. These short and variable regions can be recommended to facilitate phylogenetic analysis when rapid information from many samples is requested. However, we have demonstrated here that these limited sequence analyses should be used with caution, as they are less reliable than an analysis of the whole N coding gene.

At least 11 cocirculating phylogenetic lineages of isolates in accordance with their geographical origin or species of isolation were identified. This analysis provided a precise delimitation of the circulating zone and data concerning the origin and the spread of rabies for some lineages, particularly those of Africa. The lineages Africa 1, 2, and 3 have probably arisen independently in this continent as a result of different progenitor viruses. Although the presently circulating viruses of lineages Africa 1a and 1b were different from the viruses enzootic in Europe and the Middle East, there was a striking genetic relatedness between them, indicating a common origin. This could result from colonization in the 19th century which may have contributed to the African spread of European canine rabies (Smith *et al.,* 1992). Geographic isolation and canid population modulation would have then favored the emergence of different variants of this ancestor. Lineages Africa 2 and 3 therefore represented older African lineages of rabies virus. Two clusters of vaccinal strains could also be distinguished, the first corresponding to the PV and its related SAD, the second to CVS and its related PM. The genetic relatedness seen between the vaccinal strains contradicted the origin and passage history of the viruses (Clark and Wiktor, 1972; Lafon *et al.,* 1988; Lépine and Gamet, 1974), but confirmed previous genetic studies (Sacramento *et al.,* 1992; Smith *et al.,* 1992).

A study of the distribution of variability of the 1443-b sequence corresponding to the whole nucleoprotein gene and the adjacent noncoding sequence revealed three regions to be highly variable. Two of them, coding for the amino and carboxy terminus regions of the nucleoprotein, surrounded a highly conserved central region, a common feature in all the rhabdoviruses so far studied (Tordo *et al.,* 1986a; Bourhy *et al.,* 1993). The third region, which was the most variable, corresponded to the 3′ noncoding sequence of the N mRNA, the N–M1 intergenic region, the 5′ noncoding sequence of M1 mRNA and included transcription signals which were consistent with the consensus sequence of those lyssaviruses defined previously (Tordo *et al.,* 1986b; Bourhy *et al.,* 1993). This noncoding sequence was characterized by a 1.9 times higher rate of mutation than the N coding sequence, due mainly to a 3.1-fold increase of transitions. A similar higher rate of mutation has already been observed in the noncoding region, separating the G and L cistrons, compared to the adjacent L protein coding region (Sacramento *et al.,* 1992). The high rate of nucleotide substitution found in noncoding sequences in the rabies genome was therefore consistent with the less stringent requirements on these regions and with the neutral mutation hypothesis (Kimura, 1968, 1991; Li *et al.,* 1981; Gojobori *et al.,* 1990). However, the RNA base substitutions in each rabies virus generation were not random and the similarity of phylogenetic analysis performed on both coding and noncoding regions did not support an independent evolution of these regions, confirming previous data (Yamamoto and Yoshikura, 1986). This suggested that identical, or parallel selective forces, related to presently unknown structural requirements may be involved.

The data on the rabies virus N protein presented here suggests that it is not subjected to overdominant selection (Hughes and Nei, 1989) but rather to stabilizing selection which acts to maintain an optimally adapted phenotype. Over time, the accumulation of most neutral variations in geographically separated virus populations has led to significant lineage divergence in nucleotide sequences of the N gene. However, there was no evidence for cumulative effects of the nonsynonymous mutations and epitope changes within each of the clusters, certainly due to severe constraints on N protein evolution. This corroborates previous findings obtained from capsid viral genes of RNA viruses: vesiculoviruses (Bilsel *et al.,* 1990), eastern equine encephalomyelitis virus (Weaver *et al.,* 1991), foot and mouth disease virus (Martinez *et al.,* 1992), and lentiviruses (Myers *et al.,* 1992). However, rabies virus does not follow the complex pattern of evolution of influenza A virus N gene (Gorman *et al.,* 1990,

1991; Shu *et al.*, 1993). Whether this evolutionary stasis reflects constraints in adaptation to different major host species or to other phenomena is presently unclear. Additional and older isolates may be necessary to generate a fully encompassing genetic database and to answer in more detail the question of whether and if how selection acts in lyssaviruses under natural conditions.

ACKNOWLEDGMENTS

We gratefully acknowledge all our correspondents who provided the isolates for this study (see Table 1). We are deeply indebted to the late T. J. Wiktor, Wistar Institute (Philadelphia, PA), for the generous gift of the MAbs used in this study. We particularly thank Fredj Tekaia from the scientific data processing service at Pasteur Institute, Paris, for helpful advice and fruitful discussions. We are grateful to Jan Whitby and Iain Old for critical reading of the manuscript. We thank H. Badrane for his help and L. Audry and N. Vachet for expert technical assistance.

REFERENCES

Bilsel, P. A., Rowe, J. A., Fitch, W. M., and Nichols, S. T. (1990). Phosphoprotein and nucleocapsid protein evolution of vesicular stomatitis virus New Jersey. *J. Virol.* **64,** 2498–2504.

Bourhy, H., Kissi, B., Lafon, M., Sacramento, D., and Tordo, N. (1992). Antigenic and molecular characterization of bat rabies virus in Europe. *J. Clin. Microbiol.* **30,** 2419–2426.

Bourhy, H., Kissi, B., and Tordo, N. (1993). Molecular diversity of the Lyssavirus Genus. *Virology* **194,** 70–81.

Clark, H. F., and Wiktor, T. J. (1972). Rabies virus. *In* "Strains of Human Viruses" (M. Majer, S. A. Plotkin, M. Majer, and S. A. Plotkin, Eds.), pp. 177–182, Karger, Basel.

Conzelmann, K. K., Cox, J. H., Schneider, L. G., and Thiel, H. J. (1990). Molecular cloning and complete nucleotide sequence of the attenuated rabies virus SAD B19. *Virology* **175,** 485–489.

Dietzschold, B., Lafon, M., Wang, H., Otvos, L. J., Celis, E., Wunner, W. H., and Koprowski, H. (1987). Localization and immunological characterization of antigenic domains of the rabies virus internal N and NS proteins. *Virus Res.* **8,** 103–125.

Dietzschold, B., Gore, M., Ertl, H., Celis, E., Otvos, L., and Koprowski, H. (1989). Analysis of protective immune mechanisms induced by rabies nucleoprotein. *In* "Genetics and Pathogenicity of Negative Strand Viruses" (D. Kolakofsky and B. W. J. Mahy, Eds.), pp. 295–305, Elsevier Science Publ BV, Amsterdam.

Ertl, H. C. J., Dietzschold, B., and Otvos, L. (1991). T Helper cell epitope of rabies virus nucleoprotein defined by tri- and tetrapeptides. *Eur. J. Immunol.* **21,** 1–10.

Fang Fu, Z., Dietzschold, B., Schumacher, C. L., Wunner, W. H., Ertl, H. C. J., and Koprowski, H. (1991). Rabies virus nucleoprotein expressed in and purified from insect cells is efficacious as a vaccine. *Proc. Natl. Acad. Sci. USA* **88,** 2001–2005.

Fekadu, M., Sumner, J. W., Shaddock, J. H., Sanderlin, D. W., and Baer, G. M. (1992). Sickness and recovery of dogs challenged with a street rabies virus after vaccination with a vaccinia virus recombinant expressing rabies virus N protein. *J. Virol.* **66,** 2601–2604.

Felsenstein, J. (1993). "PHYLIP: Phylogeny inference package," Version 3.52c, Univ. of Washington, Seattle, WA.

Genetic Computer Group Inc. (1991). Sequence analysis software package, Version 7. University Research Park, Madison, WI.

Gojobori, T., Moriyama, E. N., and Kimura, M. (1990). Molecular clock of viral evolution, and the neutral theory. *Proc. Natl. Acad. Sci. USA* **87,** 10015–10018.

Gorman, O. T., Bean, W. J., Kawaoka, Y., and Webster, R. G. (1990). Evolution of the nucleoprotein gene of influenza A virus. *J. Virol.* **64,** 1487–1497.

Gorman, O. T., Bean, W. J., Kawaoka, Y., Donatelli, I., Guo, Y., and Webster, R. G. (1991). Evolution of influenza A virus nucleoprotein genes: implications for the origin of H1N1 human and classical swine viruses. *J. Virol.* **65,** 3704–3714.

Higgins, D. G., and Sharp, P. M. (1989). Fast and sensitive multiple sequence alignments on a microcomputer. *Cabios* **5,** 151–153.

Holland, J. J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and Vandepol, S. (1982). Rapid evolution of RNA genomes. *Science* **215,** 1577–1585.

Hughes, A. L., and Nei, M. (1989). Nucleotide substitutions at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA* **86,** 958–962.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* **217,** 624–626.

Kimura, M. (1991). Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proc. Natl. Acad. Sci. USA* **88,** 5969–5973.

King, A. A., Meredith, C. D., and Thomson, G. R. (1994). Canid and viverrid rabies viruses in South Africa. *Onderstepoort J. Vet. Res.* **60,** 295–299.

Kumar, S., Tamura, K., and Nei, M. (1993). "MEGA: Molecular Evolutionary Genetic Analysis," Version 1.01. The Pennsylvania State University, University Park, PA.

Lafon, M., Bourhy, H., and Sureau, P. (1988). Immunity against the European bat rabies (Duvenhage) virus induced by rabies vaccines: An experimental study in mice. *Vaccine* **6,** 362–368.

Lafon, M., Lafage, M., Martinez-Arends, A., Ramirez, R., Vuillier, F., Charron, D., Lotteau, V., and Scott-Algara, D. (1992). Evidence for a viral superantigen in humans. *Nature* **358,** 507–510.

Lafon, M., and Wiktor, T. J. (1985). Antigenic sites on the ERA rabies virus nucleoprotein and non-structural protein. *J. Gen. Virol.* **66,** 2125–2133.

Lépine, P., and Gamet, A. (1974). Sur l'évolution des virus fixes et de la souche Pasteur en particulier. *In* "Symposia Series in Immunobiological Standardization. International Symposium on Rabies" (R. H. Regamey, W. Hennessen, R. Lang, F. T. Perkins, and R. Triau, Eds.), pp. 60–66, Karger, Basel.

Li, W. H., Gojobori, T., and Nei, M. (1981). Pseudogenes as a paradigm of neutral evolution. *Nature* **292,** 237–239.

Lodmell, D. L., Sumner, J. W., Esposito, J. J., Bellini, W. J., and Ewalt, L. C. (1991). Raccoon poxivirus recombinants expressing the rabies virus nucleoprotein protect mice against lethal rabies virus infection. *J. Virol.* **65,** 3400–3405.

Maciukenas, M. (1992). "Treetool." Univ. of Illinois Board of Trustees, Urbana, IL.

Mannen, K., Hiramatsu, K., Mifune, K., and Sakamoto, S. (1991). Conserved nucleotide sequence of rabies virus cDNA encoding the nucleoprotein. *Virus Genes* **5,** 69–73.

Martinez, M. A., Dopazo, J., Hernandez, J., Mateu, M. G., Sobrino, F., Domingo, E., and Knowles, N. J. (1992). Evolution of the capsid protein gene of foot-and-mouth disease virus: Antigenic variation without accumulation of amino acid substitutions over six decades. *J. Virol.* **66,** 3557–3565.

Myers, G., MacInnes, K., and Korber, B. (1992). The emergence of simian/human immunodeficiency viruses. *AIDS Res. Hum. Retrovir.* **8,** 373–386.

Nadin-Davis, S. A., Casey, G. A., and Wandeler, A. (1993). Identification of regional variants of the rabies virus within the Canadian province of Ontario. *J. Gen. Virol.* **74,** 829–837.

Nadin-Davis, S. A., Casey, G. A., and Wandeler, A. (1994). A molecular epidemiological study of rabies virus in central Ontario and western Quebec. *J. Gen. Virol.* **75,** 2575–2583.

Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3,** 418–426.

Patton, J. T., Davis, N. L., and Wertz, G. W. (1984). N protein alone

satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus. *J. Virol.* **49,** 303–309.

Perrin, P., Joffret, M. L., Zanetti, C., Bourhy, H., Gontier, C., Fritzell, C., Leclerc, C., and Sureau, P. (1991). Rabies-specific production of interleukin-2 by peripheral blood lymphocytes from human rabies vaccines. *Vaccine* **9,** 549–558.

Poch, O., Tordo, N., and Keith, G. (1988). Sequence of the 3386 3′ nucleotides of the genome of the AVO1 strain rabies virus: structural similarities of the protein regions involved in transcription. *Biochimie* **70,** 1019–1029.

Rupprecht, C. E., Glickman, L. T., Spencer, P. A., and Wiktor, T. D. (1987). Epidemiology of rabies virus variants. *Am. J. Epidemiol.* **126,** 298–309.

Rupprecht, C. E., and Smith, J. S. (1994). Raccoon rabies: the re-emergence of an epizootic in a densely populated area. *Semin. Virol.* **5,** 155–164.

Sacramento, D., Badrane, H., Bourhy, H., and Tordo, N. (1992). Molecular epidemiology of rabies in France: Comparison with vaccinal strains. *J. Gen. Virol.* **73,** 1149–1158.

Sacramento, D., Bourhy, H., and Tordo, N. (1991). PCR technique as an alternative method for diagnosis and molecular epidemiology of rabies virus. *Mol. Cel. Probes* **6,** 229–240.

Saitou, N., and Nei, M. (1987). The neighbour joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425.

Scholtissek, C., Bürger, H., Kistner, O., and Shortridge, K. F. (1985). The nucleoprotein as a possible major factor in determining host specificity of influenza H3N2 viruses. *Virology* **147,** 287–294.

Shu, L. L., Bean, W. J., and Webster, R. G. (1993). Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990. *J. Virol.* **67,** 2723–2729.

Smith, J. S. (1989). Rabies virus epitopic variation: use in ecological studies. *Adv. Virus Res.* **36,** 215–253.

Smith, J. S., Orciari, L. A., Yager, P., Seidel, H. D., and Warner, C. K. (1992). Epidemiologic and historical relationships among 87 rabies virus isolates as determined by limited sequence analysis. *J. Infect. Dis.* **166,** 296–307.

Sokol, F., Schlumberger, H. D., Wiktor, T. J., Koprowski, H., and Hummeler, K. (1969). Biochemical and biophysical studies on the nucleocapsid and on the RNA of rabies virus. *Virology* **38,** 651–665.

Steinhauer, D. A., and Holland, J. J. (1987). Rapid evolution of RNA viruses. *Annu. Rev. Microbiol.* **41,** 409–433.

Sumner, J. W., Fekadu, M., Shaddock, J. H., Esposito, J. J., and Bellini, W. J. (1991). Protection of mice with vaccinia virus recombinants that express the rabies nucleoprotein. *Virology* **183,** 703–710.

Tordo, N., Poch, O., Ermine, A. and Keith, G. (1986a). Primary structure of leader RNA and nucleoprotein genes of the rabies genome: Segmented homology with VSV. *Nucleic Acids Res.* **14,** 2671–2683.

Tordo, N., Poch, O., Ermine, A., Keith, G., and Rougeon, F. (1986b). Walking along the rabies genome: Is the large G–L intergenic region a remnant gene? *Proc. Natl. Acad. Sci. USA* **83,** 3914–3918.

Weaver, S. C., Scott, T. W., and Rico-Hesse, R. (1991). Molecular evolution of Eastern equine encephalitis virus in North America. *Virology* **182,** 774–784.

Wiktor, T. J., and Clark, H. F. (1973). Comparison of rabies virus strains by means of the plaque reduction test. *Ann. Microbiol.* **124A,** 283–287.

Wiktor, T. J., Flamand, A., and Koprowski, H. (1980). Use of monoclonal antibodies in diagnosis of rabies virus infection and differentiation of rabies and rabies-related viruses. *J. Virol. Methods* **1,** 33–46.

Yamamoto, K., and Yoshikura, H. (1986). Relation between genomic and capsid structures in RNA viruses. *Nucleic Acids Res.* **14,** 389–396.