Molecular epidemiology of rabies virus in South Africa

L.H. NEL[1], G.R. THOMSON[2] and B.F. VON TEICHMAN[1]

ABSTRACT

NEL, L.H., THOMSON, G.R. & VON TEICHMAN, B.F. 1993. Molecular epidemiology of rabies virus in South Africa. *Onderstepoort Journal of Veterinary Research*, 60:301–306

Nucleic acid sequence analysis was used to determine the phylogenetic relationships amongst rabies viruses isolated from typical canid hosts such as bat-eared fox, jackal and dog in South Africa (SA). Geographical factors were taken into account in the selection of isolates and three different regions within the genomes of the isolates were compared for their use as phylogenetic indicators. The three genome regions, being the cytoplasmic domain of the G-gene, the G–L intergenic pseudogene and the antigenic domain II of the N–gene were found to differ in terms of the degree of nucleic acid conservation, but produced similar results when analyzed phylogenetically. The SA canid isolates were found to be closely related and could clearly be distinguished from all other rabies virus groups for which sequence data is available. In addition four SA mongoose rabies isolates were studied which were shown to be distant from the SA canid rabies virus group as well as from any other rabies viruses (or group) for which sequence data is available. Our results also indicate that spillover between the distinct canid and viverrid host reservoirs may occur.

INTRODUCTION

A wide diversity of animal species is involved in the epidemiology of rabies in South Africa. These include jackals (*Canis mesomelas*), bat-eared foxes (*Otocyon megalotis*), yellow mongooses (*Cynictis penicillata*), genets (Genetta spp.) and wild cats (*Felis lybica*). It is also known that remarkably clear associations between different host species and geographic location exist (King, Meredith & Thomson 1993). Canine rabies has, since 1976, become an ever worsening problem in Natal and KwaZulu, whereas jackals and bat-eared foxes are important hosts in the north and north-western Transvaal and the western and north-western Cape respectively. Wildlife rabies diagnosis is most common in the yellow mongoose and notably so in the Orange Free State, southern and eastern Transvaal highveld and eastern Karoo, although this species occurs over a much larger geographical area (King *et al*. 1993).

[1] Department of Microbiology and Plant Pathology, Agricultural Sciences, University of Pretoria, 0002 South Africa

[2] Foot-and-Mouth Disease Laboratory, Private Bag X6, Onderstepoort, 0110 South Africa

Historically, mongoose rabies in South Africa is considered likely to be caused by a virus different from the virus causing rabies in typical canids. This view has been supported by King *et al*. (1993) who used a panel of 80 nucleoprotein-specific monoclonal antibodies (Mab's) to differentiate between different rabies isolates. However, nothing is known about the nature and extent of genetic variation amongst these isolates. Such information on the genetic composition of rabies isolates would assist in determining the interrelationship, origin and epidemiology of rabies viruses from the different and diverse host species. Naturally, from a broad epidemiological perspective, it is of importance to establish how South African rabies isolates compare with viruses from elsewhere in the world on a molecular level.

For molecular genetic comparisons of rabies viruses, different regions of their negative stranded RNA genomes can be sequenced. For example, the nucleoprotein encoding gene which is well conserved is likely to be suited to studies of long-term evolution (Bourhy, Kissi & Tordo 1992; Smith, Lillian, Yager, Seidel & Warner 1992). A less conserved genomic region which might yield epidemiological information

over shorter time spans is the gene of the surface G glycoprotein, the major antigen involved in eliciting protective immunity (Tordo 1991). Probably the least conserved sequence in the rabies genome is the G-L intergenic region, the 423 nucleotide (nt) remnant pseudogene (Tordo, Poch, Ermine, Keith & Rougeon 1986; Sacramento, Badrane, Bourhy & Tordo 1992). This pseudogene may be an ideal indicator for "neutral" evolutionary distance, there being no gene product and thus no immunological pressure or liability to retain any specific functions.

We have synthesized, amplified and sequenced cDNA specific for different portions of the genomes of a number of South African rabies isolates. In this paper we report a comparative sequence analysis for these viruses which were isolated from different canid and viverrid hosts. We have also compared the suitability of different genomic regions for phylogenetic analysis of these viruses.

MATERIALS AND METHODS

Rabies virus isolates

Rabies virus isolates from dogs, jackals, bat-eared foxes and yellow mongooses were selected from areas which conform to the typical geographical location for the given host species (Fig. 1). These virus isolates were stored in the form of lyophilized 20 % mouse brain material.

Total RNA extraction

Lyophilized mouse brain passaged material (500 μg) was dissolved in 500 μℓ of an extraction buffer (1 % SDS, 1 % NP_{40}, 1 mM EDTA pH = 8,0) and extracted three to four times with buffer saturated phenol. Total RNA was precipitated with two volumes 100 % ethanol, washed with 70 % ethanol and made up to the required concentration with ultrapure water.

Primers

Those regions of the virus genome which were analyzed and the positions of all respective oligonucleotide primers are shown in Fig. 2 a and b. For amplification of the portion of the N-gene, primers N1 (Sacramento, Bourhy & Tordo 1991) and 106 (Smith et al. 1992) were used and sequencing was carried out with the use of primer 106 (Fig. 2a). DNA encompassing the pseudogene and the cytoplasmic domain of the glycoprotein was amplified with the G and L primers (Sacramento et al. 1991). The G and L primers as well as several internal primers (Fig. 2b) were used as required for sequencing the entire pseudogene.

cDNA synthesis and amplification

Total brain RNA (0,5–3,0 μg) and either G primer (100 ng) or N1 primer (100 ng) was heated to 70 °C
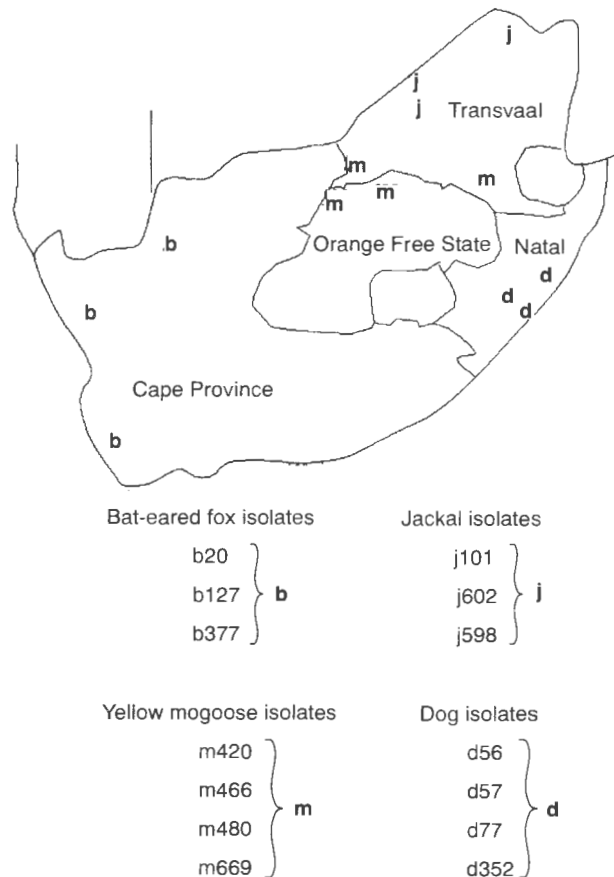

FIG. 1   Map of South Africa indicating the host species and approximate geographic origin of rabies isolates relevant to this investigation

after which a reverse transcription reaction mix was made on ice and the reaction carried out at 42 °C for 90 min in a total volume of 10 mℓ (Sacramento et al. 1992). Amplification using either the G/L or the N1/106 primer sets was carried out on one tenth of the cDNA in a Hybaid heating block using conditions similar to those described by Sacramento et al. (1991).

Nucleotide sequencing

Several purification methods for the recovery of amplified DNA products were used. Besides the Geneclean (Bio 101 Inc.) and Magic PCR Prep DNA Purification System (Promega), the simplest (and most economical) way was the use of a "glass wool" column. The PCR product (20 μℓ) was electrophoresed on a 0,7 % agarose gel (Seakem), excised and spun through a column packed with a small amount of glass wool. Although loss of up to 20 % of the DNA occurred during this process, the yield was sufficient for one or more sequencing reactions.

The purified PCR product was sequenced with the use of the fmol DNA sequencing kit (Promega)

(a)

Primers

N1(+) N2(-) 106(-)
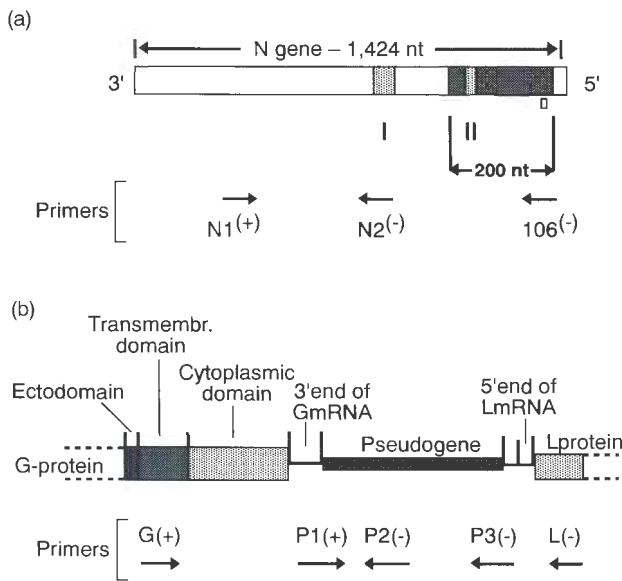
(b)

Primers

G(+) P1(+) P2(-) P3(-) L(-)

FIG. 2 Oligonucleotide primers used for amplification and sequencing of (a) a portion of the nucleoprotein gene and (b) the cytoplasmic domain of the glycoprotein gene and the remnant pseudogene. Antigenic domains on the nucleoprotein are indicated using the convention of Dietzschold, Rupprecht, Tolus, Lafon, Mattei, Wiktor & Koprowski (1987). Refer to Materials and Methods for details and other references

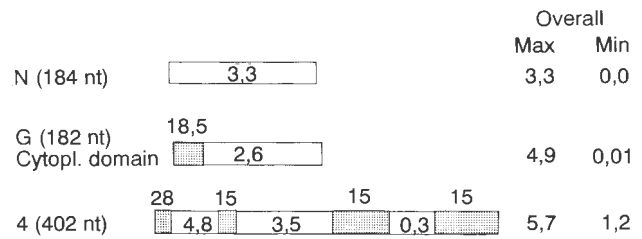|				Overall	
				Max	Min
N (184 nt)		3,3		3,3	0,0
G (182 nt) Cytopl. domain	18,5	2,6		4,9	0,01
4 (402 nt)	28 15 4,8 3,5	15 0,3	15	5,7	1,2

FIG. 3 Analysis of the nucleotide sequence divergence (maximum % between any two) within portions of the genomes (Fig. 2 a & b) of a group of 8 canid rabies virus isolates from South Africa. The overall sequence divergence for each genomic region is shown on the right, the two values indicating the maximum overall divergence, and the minimum overall divergence between any two isolates

according to the manufacturers instructions. The reaction mixture contained a minimum concentration of 50 fmoles of DNA, 30 pmoles of primer and 10 µCi ^{35}S dATP. A Hybaid thermocycler programmed for 30 cycles as specified for the fmol sequencing system was used and the reactions analyzed by standard denaturing electrophoresis and autoradiography.

Computer analysis

Nucleotide sequences were analyzed using the Clustal V package of multiple alignment (Higgins & Sharp 1989). For construction of phylogenetic trees the neighbour-joining method of Saitou & Nei (1987) with a minimum of 1 000 bootstrap replicates (Felsenstein 1985) was carried out.

RESULTS

Nucleotide divergence within the genomes of South African canid rabies isolates from South Africa

Eight rabies isolates comprising four dog, two jackal and two bat-eared fox viruses were analyzed with respect to intrinsic variability of portions of the nucleoprotein and glycoprotein genes and their pseudogenes, as described in Materials and Methods and

illustrated in Fig. 2. Comparison of the degree of variation between the different genomic portions of the eight isolates is presented in Fig. 3. In the case of the 184 nt region spanning antigenic domain II of the N protein, a low overall mutation rate was observed (max = 3,3 %) and in some cases (two of the dog isolates and the two jackal isolates), no sequence differences were found. In the case of the cytoplasmic domain of the glycoprotein it was found that the first approximately 30 nucleotides displayed much variation (up to 18,5 %) whereas the remainder of the G-gene appeared to be very conserved. Overall maximum and minimum percentage nucleotide differences between any two isolates in the group were found to be 4,9 and 0,01 respectively. A comparison of the respective pseudogenes revealed clearly defined regions of high and low variation, the values and areas indicated in Fig. 3. The overall variation in this region was significantly higher compared to the N- or G-gene regions, with maximum and minimum variation of 5,7 and 1,2 % respectively.

A phylogenetic analysis in which the sequence data of the relevant portions of the nucleoprotein genes were used was carried out and the result is shown in Fig. 4. Bat-eared fox, jackal and dog isolates may, on the basis of this result, be divided into three groups. Nevertheless, these isolates are very closely related and can be distinguished from other viruses of African origin (NA, TUN and CAM; see Fig. 4 for code) for which sequence data has been published (Smith et al. 1992). Further comparison of the South African sequences with those from other regions of the world indicates that they are equidistant between early European and now laboratory strains (PV, ERA) on the one hand and middle/north African strains on the other (Fig. 4). The SA group also appears to be most distant from isolates from far eastern countries such as China and the Phillipines (sequence data from Smith et al. 1992).

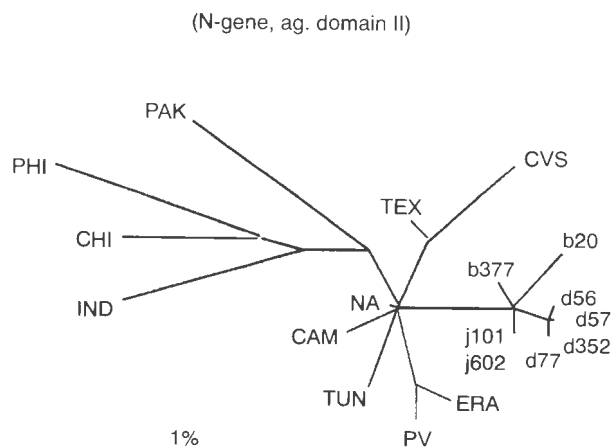Phylogenetic trees constructed on the basis of analysis of the nucleic acid sequences of the cytoplasmic

(N-gene, ag. domain II)

FIG. 4 Unrooted phylogenetic tree showing the genetic relation-
ships among South African canid rabies isolates (Fig. 1)
and between the South African isolates (numbers) and
isolates from other continents or African countries. The
genetic relationships were inferred from sequence data
for a portion of the nucleoprotein gene (Fig. 3a). Se-
quence data other than for the South African isolates was
obtained from Smith et al. 1992, abbreviated as follows:
PAK, PKDG1345 from Pakistan; PHI, RPLZDG727 from
the Philippines; CHI, CHGXDG1141 from China; IHD,
IDJVDG-853 from Indonesia; PV, Pasteur virus; ERA,
Evelyn-Rokitnicki-Abelseth; TEX, TXUSFX 289 from Tex-
as, USA; CVS, challenge virus standard; NA, MEDG1172
from North Africa; TUN, TSDG1292 from Tunisia and
CAM, NYUSDG343 from Cameroon. The tree was con-
structed and 1 000 bootstrap replicates performed as indi-
cated in Materials and Methods. An arbitrary cut-off point
of 70 % was taken as significant and all nodes with less
than 70 % significance was collapsed onto the next sig-
nificant node. All distances are proportional to the per-
centage nucleotide divergence, shown by the scale

TABLE 1 Isolates from yellow mongooses in South Africa used
in the investigation

No.	Nucleotide sequences available		Mab reaction pattern[a]	Viverrid group[b]
	G (cytoplasmic)	Pseudogene		
m480	+		17	2
m466	+		20	3
m420	+	+	6	5
m669	+		12	5
b 127[c]	+		12	5

[a] & [b] N-specific Mab reaction patterns and grouping into different
viverrid groups on this basis from King et al. (1993)

[c] Serologically an atypical isolate from bat-eared fox. Refer
to text for details

with a panel of monoclonal antibodies (Mabs) to N-
protein (King et al. 1993), mongoose isolates fell into
different reaction pattern groups, as indicated in
Table 1. The isolate b127 was included as it is the
only bat-eared fox isolate (of 15) which does not con-
form to the Mab reaction pattern typical for canid iso-
lates (King et al. 1993).

In the cases where pseudogene, glycoprotein gene
(cytoplasmic domain) and nucleoprotein gene (anti-
genic domain II) sequences for a particular isolate
were available, similar results, in terms of phylo-
genetic grouping, were obtained in such analysis.
For that reason, and since we had the G-cytoplasmic
domain sequence available for all isolates, the phylo-
genetic analysis using this region of the genome is
presented (Fig. 6). From this analysis it is evident
that the isolates m420, m466, m480 and m669, are
phylogenetically related to each other but very differ-
ent from all other SA rabies isolates. The distance
between these mongoose isolates and the SA canid
group also appears to be significantly greater than
the distance between the European/vaccine strains
and the SA canid strains. The bat-eared fox strain
b127 also appear to be significantly different from all
other isolates included in the study (Fig. 6).

DISCUSSION

Nucleic acid sequence analyses of three different
regions of the genomes of eight canid rabies isolates
from geographically distinct areas in South Africa
were carried out. The 423 nt G-L intergenic pseudo-
gene displayed most variation, followed by the cyto-
plasmic domain of the glycoprotein gene while the
antigenic domain II region of the nucleoprotein gene
was found to be most conserved. Despite these ex-
pected differences in nucleic acid conservation, phylo-
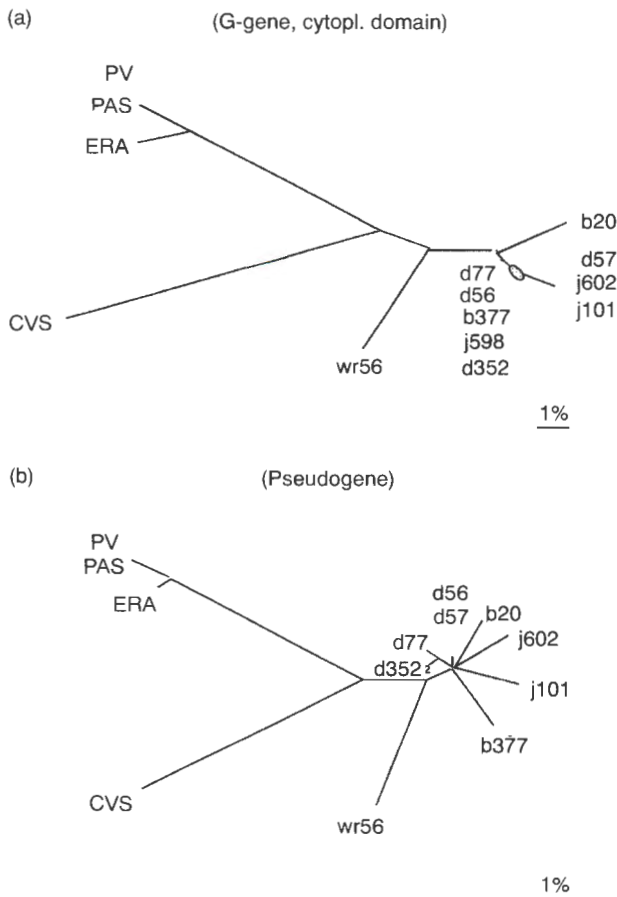genetically similar results were obtained in analyses

domain of the glycoprotein and the pseudogene are
shown in Fig. 5a and b. In both cases it is noticed
that, as in the case of analysis of the nucleoprotein
gene, the SA isolates clearly form a group which is
distinguishable from laboratory strains and from a
French fox isolate (wr56; sequence from Sacramento
et al. 1992). In the case of the pseudogene analysis,
a different "internal" pattern between the SA types
is evident and it appears that, in comparison with the
G-region there is less evolutionary distance between
all the other isolates and the South African isolates.
For example, the average sequence difference be-
tween the SA types and CVS is 16,2 % in the case
of the G-region and 11,5 % for the pseudogene.

Comparative analysis of genetic material from South African mongoose rabies isolates

Four rabies isolates from yellow mongoose which
were included in the present study are shown in
Table 1. Whereas almost all South African canid ra-
bies isolates conformed to only one reaction pattern

(a)

(G-gene, cytopl. domain)

(b)

(Pseudogene)
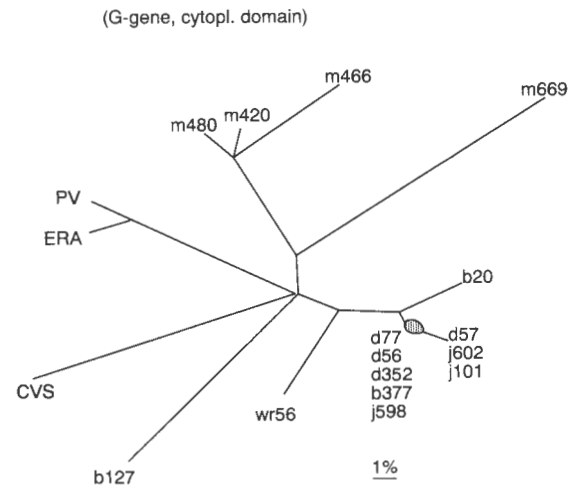

(G-gene, cytopl. domain)

FIG. 6 Phylogenetic tree indicating the genetic relationships between South African mongoose rabies isolates (prefix m, Fig. 1, Table 1) and South African canid rabies isolates (numbers) and laboratory strains as well as European fox isolates (wr 56). Calculations and construction of the tree was as described in the legend to Fig. 4, using in this case the nucleic acid sequence of the cytoplasmic domain of the glycoprotein. Distances along the solid lines indicate percentage nucleotide divergence proportional to the scale

FIG. 5 Phylogenetic trees indicating the genetic relationships among South African canid rabies isolates (capitals) and their relationship with laboratory rabies strains and a European fox isolate (wr 56) on the basis of (a) sequence data for the cytoplasmic domain of the glycoprotein and (b) the G-L intergenic pseudogene. Sequence data other than that of the South African isolates was from Sacramento et al. (1992) and standard abbreviations are used. Construction of the trees was carried out as described in the legend to Fig. 4 and all distances are proportional to percentage nucleotide divergence as shown by the scale

of the three different genomic regions. Relative evolutionary distances as indicated for the different regions of the genome varied, but with debatable significance. In all cases it was shown that these SA canid rabies isolates are closely related and that they, as a group, can be distinguished from other African, European and Asian rabies isolates for which published sequence data were available.

King et al. (1993) have shown that, on the basis of the reaction of SA rabies isolates with a panel of 80 N-Mab, most canid isolates conform to one particular reaction pattern whereas viverrid rabies isolates group to into several different reaction patterns. Our results based on the nucleic acid sequences of the

cytoplasmic domain of the glycoprotein support the differences described between canid and viverrid isolates from South Africa. All four mongoose isolates investigated, formed a distinct group quite distant (> 10 % nucleic acid sequence divergence) from the group comprised by canid isloates. Furthermore, our sequence data for these rabies isolates does not indicate any close relationship with other rabies virus isolates for which sequence data is available.

One isolate from a bat-eared fox which was serologically different from other bat-eared fox isolates, was shown to have more than 13 % sequence divergence in the cytoplasmic domain (glycoprotein) when compared with typical canid isolates. Thus, apart from confirming a different and perhaps archaic origin of SA mongoose rabies viruses, our data also suggests that spillover between distinct animal species reservoirs may occur as was also found with Mab studies (King et al. 1993). To date our information is insufficient to permit reliable interpretation on the molecular epidemiology of rabies in South Africa. However, it is evident that nucleic acid sequence analyses will make an important contribution to improving our understanding of the epidemiology of rabies in South Africa and elsewhere as more data is generated.

ACKNOWLEDGEMENTS

REFERENCES

BOURHY, H., KISSI, B. & TORDO, N. 1993. Molecular diversity of the *Lyssavirus* genus. *Virology*, 194:70–81.

DIETZSCHOLD, B., RUPPRECHT, C.E., TOLUS, M., LAFON, M., MATTEI, J., WIKTOR, T.J. & KOPROWSKI, H. 1988. Antigenic diversity of the glycoprotein and nucleocapsid proteins of rabies and rabies-related viruses: Implications for epidemiology and control of rabies. *Reviews of Infectious Diseases*, 10(4):785–798.

FELSENTSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39:783–791.

HIGGINS, D.G. & SHARP, P.M. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *Cabios*, 5:151–153.

KING, A.A., MEREDITH, C.D. & THOMSON, G.R. 1993. The biology of southern African lyssavirus variants. *Current Concepts in Microbiology and Immunology*. In press.

SACRAMENTO, D., BOURHY, H. & TORDO, N. 1991. PCR technique as an alternative method for diagnosis and molecular epidemiology of rabies virus. *Molecular and Cellular Probes*, 5:229–240.

SACRAMENTO, D., BADRANE, H., BOURHY, H. & TORDO, N. 1992. Molecular epidemiology of rabies virus in France: comparison with vaccine strains. *Journal of General Virology*, 73: 1149–1158.

SAITOU, N. & NEI, M. 1987. The neighbour-joining method: a new method for reconstructing phylogentic trees. *Molecular Biology Evolution*, 4(4):406–425.

SMITH, J.S., LILLIAN, A.O., YAGER, P.A, SEIDEL, H.D. & WARNER, C.K. 1992. Epidemiologic and historical relationships among 87 rabies virus isolates as determined by limited sequence analysis. *The Journal of Infectious Diseases*, 166:296–307.

TORDO, N., POCH, O., ERMINE, A., KEITH, G. & ROUGEON, F. 1986. Walking along the rabies genome: is the G-L intergenic region a remmant gene? *Proceedings of the National Academy of Sciences*, USA, 83: 3914–3918.

TORDO, N. 1991. Contribution of molecular biology to vaccine development and molecular epidemiology of rabies disease. *Memorias. Instituto Butantan*, 53:31–51.