Epidemiol. Infect. (1999), **123**, 325–336. Printed in the United Kingdom

Phylogeographic patterns exhibited by Ontario rabies virus variants

S. A. NADIN-DAVIS[1]*, M. I. SAMPATH[1], G. A. CASEY[1], R. R. TINLINE[2] AND A. I. WANDELER[1]

[1] *Rabies Centre of Expertise, Animal Diseases Research Institute, Canadian Food Inspection Agency, Nepean, Ontario, Canada*
[2] *Department of Geography, Queen's University, Kingston, Ontario, Canada*

(*Accepted 1 May 1999*)

SUMMARY

A previous study on N gene variation of rabies viruses circulating in Ontario red foxes identified four viral variants. This study confirms the geographical localization of these variants and extends the analysis to the less conserved G gene of these viruses. A greater number of regionally localized variants was revealed and their phylogenetic relationships have been examined. Ongoing surveillance on recent disease outbreaks revealed that variants do not always persist in specific areas. The distribution of these variants did however appear to be influenced by topographical features of the study area likely to affect host animal movements and contacts. The majority of G gene base changes were synonymous and limited glycoprotein sequence variation predominantly to the C-terminal transmembrane and endo-domains. These data are most readily explained by random appearance of genetic viral variants followed by their spread throughout sub-populations of the fox host according to the easiest routes of transmission.

INTRODUCTION

Rabies virus is geographically widely distributed and causes lethal neural disease in a wide range of mammalian species, although in developed countries epidemiological cycles of rabies persist in just a few specific wildlife species (e.g. red fox, skunk, raccoon). As the prototype of the *Lyssavirus* genus, genetic characterization of selected laboratory strains of rabies virus has revealed the organization of its 12 kb single-stranded negative sense RNA genome. Five genes denoted N, P (formerly NS), M, G and L which encode the nucleoprotein, phosphoprotein, matrix protein, glycoprotein and replicase activity respectively have been identified [1–4]. Until recently differentiation of rabies virus strains relied on serological methods. The examination of antigenic variation using extensive monoclonal antibody panels successfully demonstrated the existence of distinct forms of this virus according to both host species and region [5, 6]. The development of molecular genetic methodologies, and in particular the polymerase chain reaction [7], provides ready access to nucleotide sequence data which are useful not only in facilitating highly sensitive strain discrimination but permit the construction of phylogenies [8–11]. Several groups have applied this methodology to global analysis of rabies viruses [12, 13].

In the Canadian province of Ontario, rabies cases were usually dog-related and sporadic until the late 1950s when a wave of rabies moved down from Arctic regions [14]. This epizootic became established in the

* Author for correspondence.

red fox population of Southern Ontario and the resultant enzootic still persists today primarily in red foxes (*Vulpes vulpes*) although significant spillover occurs, especially to the striped skunk (*Mephitis mephitis*) [15]. Antigenic typing methods have been unable to discriminate between the virus still persisting in the arctic from that circulating in Ontario [16], a finding clearly reflecting the historical relationship between these two populations, despite their physical separation for almost 40 years. Application of genetic typing methods to the N gene locus of several Ontario specimens identified four viral variants apparently circulating in discrete regions of the enzootic area [17]. Furthermore, genetic typing clearly discriminated between the Ontario and arctic forms of this rabies strain [18].

Epidemiological discussions of fox rabies, both in Europe and southern Ontario, highlight the complex and still poorly understood interactions between host, virus and environment which must operate to determine disease persistence and movement patterns in a specific region. Most previous discussions on this subject [see 19, 20] have focused on biological aspects of the host (i.e. population density and reproductive rates, dispersion of young, animal contact rates, etc) though recently some consideration has been given to possible roles of viral variation [21]. To gain a better appreciation of the evolutionary mechanisms acting on rabies viruses on a population level in southern Ontario, a more detailed molecular epidemiological study of rabies was undertaken by targeting the G gene. Overall the G gene is one of the more variable coding regions of the virus and together with the poorly conserved contiguous G–L intergenic region, constitutes a suitable target for detailed epidemiological studies [8, 22]. The glycoprotein of rabies viruses and the more divergent non-rabies lyssaviruses consists of a well conserved ectodomain and more variable transmembrane and endo-domains [23].

This report describes the genetic characterization of a significant number of Ontario rabies virus specimens, at both the N and G loci, through restriction fragment length polymorphisms (RFLPs) and nucleotide sequencing. The regional distribution of variants thus identified is detailed, their phylogenetic relationships examined using G nucleotide sequence data and variation in the encoded glycoprotein documented. Temporal effects on variant distribution were examined as was correspondence of this localization with patterns of disease incidence and physical landscape features.

METHODS

Source of specimens

All of the virus specimens studied in this report were from animal brains diagnosed rabies positive using the fluorescent antibody test [24]. The primary study included 70 isolates, submitted in the years 1990–1, which were described previously [17] and 28 specimens submitted in 1993 (see Table 1). Twenty additional specimens from recent outbreaks (1995–6) over limited parts of the study area were also examined. Submission locations are usually documented using a locator code (UTMC) thereby allowing for their accurate mapping [25]. RNA was extracted from brain tissue using a modified acid-phenol/guanidinium procedure [26].

Reverse transcription and PCR

The primers employed for both PCR and sequencing were synthesized on an Applied Biosystems 391 DNA synthesizer and purified using Oligonucleotide Purification Cartridges as described by the manufacturer (Applied Biosystems). Reverse transcription of negative sense genomic RNA, primed using a sequence specific positive sense oligonucleotide, and the subsequent PCR amplification employing both positive and negative sense primers were performed essentially as described previously [17]. Due to the length of the G region under study, it was convenient to amplify this target sequence as two separate fragments of 1·2 kb (primers rabG1·2a/b) and 1·5 kb (primers rabG1·5a/b or rabG1·5c/b); these products which share a 0·2 kb sequence overlap, contain the entire G gene sequence, the 3′ non-coding sequence and flanking sequences of neighbouring genes. For simplicity the results from both PCRs have been combined in the Results section and the entire target sequence is referred to as the G region. In a few instances generation of the 1·5 kb product was unsuccessful using the rabG1·5c/b primer pair but successful with the rabG1·5c/G1 or rabG1·5c/G4 primer combinations. Detailed descriptions of all primers are given in Table 2. The PCR was performed in a Perkin–Elmer Cetus thermal cycler using 30 cycles of the profile: 94 °C for 1 min, 45 °C for 1 min, 72 °C for 2 min (+5 s autoextension). PCR products were purified using Geneclean (Bio 101) prior to analysis. N gene PCR was performed as detailed elsewhere [17].

Table 1. *Summary of N and G gene typing by restriction endonuclease analysis of Ontario rabies viruses*

		Scoring by restriction endonucleases																		
			Ms		Hp						Mb				St		No. of animals			
N gene type	G region type	BE	1	2	1	2	3	4	5	Dd	1	2	3	Ec	1	2	sk	fx	O	total
1	1	+	−	−	+	−	−	−	−	−	+	+	+	+	+	−	6	9	1	16
1	1-1	+	−	−	+	−	−	−	−	−	+	+	+	−	−	−	1	0	0	1
1	1-2	+	−	−	+	−	−	−	−	−	+	+	+	−	+	−	1	2	0	3
2	2	−	−	+	−	+	−	+	+	+	−	+	−	−	+	−	5	7	1	13
2	2-1	−	+	+	−	+	−	+	+	+	−	+	−	−	+	−	1	0	0	1
2	2-2	−	−	+	−	+	−	+	+	+	−	+	−	−	+	+	1	0	0	1
2	2-3	+	−	+	+	+	+	+	+	+	−	+	−	−	+	−	2	0	0	2
2	2-4	+	−	+	+	+	+	+	+	+	−	+	−	−	−	−	1	0	0	1
2	2-5	+	−	+	+	+	+	+	+	+	−	−	−	−	+	−	0	1	0	1
2	2-6	+	−	+	+	+	+	+	+	−	−	+	−	−	+	−	3	0	0	3
2/4	2-6	+	−	+	+	+	+	+	+	−	−	+	−	−	+	−	1	2	0	3
2	2-7	+	−	−	+	+	+	+	+	−	−	−	−	−	+	−	0	4	0	4
2	2-8	−	−	+	−	+	−	+	+	+	−	+	−	−	−	−	0	1	0	1
3	3	+	−	+	+	+	+	+	+	−	−	−	−	−	−	−	1	17	1	19
2/4	3	+	−	+	+	+	+	+	+	−	−	−	−	−	−	−	0	1	0	1
3	3-1	+	−	+	+	+	+	+	+	−	−	+	−	−	−	−	3	1	0	4
3	3-2	+	−	+	−	+	+	+	+	−	−	−	−	−	−	−	0	1	0	1
3	3-3	−	−	+	−	+	+	+	+	−	−	−	−	−	−	−	0	1	0	1
3	3-4	+	−	+	+	−	+	+	+	−	−	−	−	−	−	−	0	4	0	4
4	4	+	+	+	+	+	+	+	+	−	−	+	−	−	+	−	6	10	1	17
2/4	4	+	+	+	+	+	+	+	+	−	−	+	−	−	+	−	0	1	0	1
Totals																	32	62	4	98

The restriction endonucleases employed have been abbreviated as follows: BE, *Bst*EII; Dd, *Dde*I; Ec, *Eco*0109I; Hp, *Hph*I (five sites); Mb, *Mbo*I (three sites); Ms, *Msp*I (two sites); St, *Sty*I (two sites). All isolates except for four were from either skunks (sk) or foxes (fx). The host species of the other (O) specimens were: T1, wolf; T2, dog; T3, cat; T4, coyote. Note the inclusion of five isolates with an N gene typing of 2/4 (intermediate between variants 2 and 4), four of which originated from a relatively small area north and west of the city of London in southwestern Ontario.

PCR product analysis

N gene PCR products were analysed using the restriction endonuclease panel described previously [17]. All G gene PCR products were typed using the panel of seven restriction endonucleases developed as described in 'Results'. Two enzymes were used to map the presence or absence of 3 specific sites in the 1·2 kb product whilst the 1·5 kb fragment was analysed by 6 enzymes at 13 sites. Restriction endonuclease digestions and agarose gel electrophoresis were performed by standard procedures [27] and each result was scored as + (site present) or − (site absent). Cloning of PCR products into the vector pGEM3Zf+ (Promega) was achieved using the restriction endonuclease sites included at the ends of the PCR primers (see Table 2). Sequencing of such clones was performed with a Sequenase kit (United States Biochemical) and [^{35}S]-αthio dATP (DuPont) using single-stranded DNA template prepared from nested deletions of the pGEM clones generated according to Henikoff [28]. Direct consensus sequencing of PCR products was achieved by cycle sequencing using a fmol sequencing kit (Promega) and ^{32}P-labelled primers. DNA sequence analysis was performed using Pustell software. Representative nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence database and have been assigned the accession numbers U11734–U11758.

Sequence alignment and phylogenetic analysis

Nucleotide and protein sequences were aligned using the CLUSTALW multiple alignment package [29]. Phylogenetic analysis was performed using the PHYLIP 3.5 package of programmes [30]. One thousand bootstrap replicates of the data, generated using SEQBOOT, were analysed by neighbour joining (NJ)

Table 2. *Oligonucleotides used in these studies*

Name	Sequence	Sense/position
rabG1.2a	agccgtcgacggtaccTCTGGTGTATCAACATGAAC	+ 3000–3019
rabG1.2b	tgcggatcctgcaGACTTGGTGGTCATGATGGAC	− 4238–4258
rabG1.5a	agcgctagctgcAGACTTGCGGCTTTGTAGATGA	+ 3988–4009
rabG1.5b	acgctctagagctcGTTCAGCCTCTAACTCGATT	− 5459–5478
rabG1.5c	cctgcAG (AG) CTTGCGGATTTGTTGACG	+ 3988–4008
rabG1	gggctagctgcaGGTTCAGCCTCTAACTCGATTGGGTCA	− 5453–5479
rabG4	TGGGTCAATAGGGTCATCATAGAC	− 5436–5459
Gseq-1	GAGTTGGTCAAGAAAAGAGAGGAGTG	+ 4197–4222
Gseq-2	ATGAGATCATCCCCTCAAAGGGGT	+ 4381–4404
Gseq-3	TGTTGAGGTTCACCTTCCCGATGT	+ 4619–4642
Gseq-3a	ACAAATCTCAGGGGTTGACCTGGG	+ 4649–4672
Gseq-4	GGAAGGTCATATCTTCGTGGGAGTC	+ 4837–4861
Gseq-5	TCTAGCAGTTTCGGTGACCAACGG	+ 5088–5111
Gseq-6	CATCTACCTACTGCTCGACTAACC	+ 3869–3892
Gseq-7	TACATTTGGGGTCTCTTGGATGTG	+ 3255–3278
Gseq-8	TGTCGGTTATGTCACCACTACGTT	+ 3596–3619
Gseqrev-1	GGTTAGTCGAGCAGTAGGTAGATG	− 3869–3892

Bases shown in upper case correspond to the nucleotides present in the rabies genome of either the PV or Ontario strains in the sense indicated; the positions given refer to the corresponding bases of the PV strain [1]. Bases in lower case are not present within the rabies sequence; underlined bases contain restriction endonuclease cleavage sites.

Table 3. *Dataset employed for cluster analysis*

Variant type	Observations in database	Missing/incorrect co-ordinates	Multiples	Sample size
N1	19	1	0	18
N2	28	2	0	26
N3	29	2	0	27
N4	17	1	3	13
Total	93	6	3	84

methods, using the DNADIST and NEIGHBOR programmes, or by DNA parsimony (MP) using DNAPARS. Consensus trees were obtained in each case using CONSENSE. Distances were reapplied to consensus trees made by distance methods using distance values of the original data and the Fitch package. Trees were drawn out with the help of TREEVIEW [31].

Cluster analysis

Of the 98 samples in the primary database (collected in 1990–3) 6 had missing or incorrect locator (UTMC) codes and had to be excluded as were the 5 N2/4 specimens. Four observations made at a wildlife compound were treated as one thereby eliminating three more observations. Table 3 illustrates the resulting data set of 84 observations used to examine the spatial distribution of the virus types for evidence of clustering. Each location within this dataset was first examined to determine the viral types of its nearest neighbours and the resulting distributions compared to the expected values assuming virus type

Table 4. *Distribution of nearest neighbour viral type for each location in the dataset*

	N1	N2	N3	N4	Total
(A) Observed					
N1	16	2	0	0	18
N2	1	18	4	3	26
N3	0	2	23	2	27
N4	0	0	2	11	13
	17	22	29	16	84
(B) Expected if viral types are randomly distributed					
N1	3·6	4·7	6·2	3·4	18
N2	5·3	6·8	9·0	5·0	26
N3	5·5	7·1	9·3	5·1	27
N4	2·6	3·4	4·5	2·5	13
	17	22	29	16	84

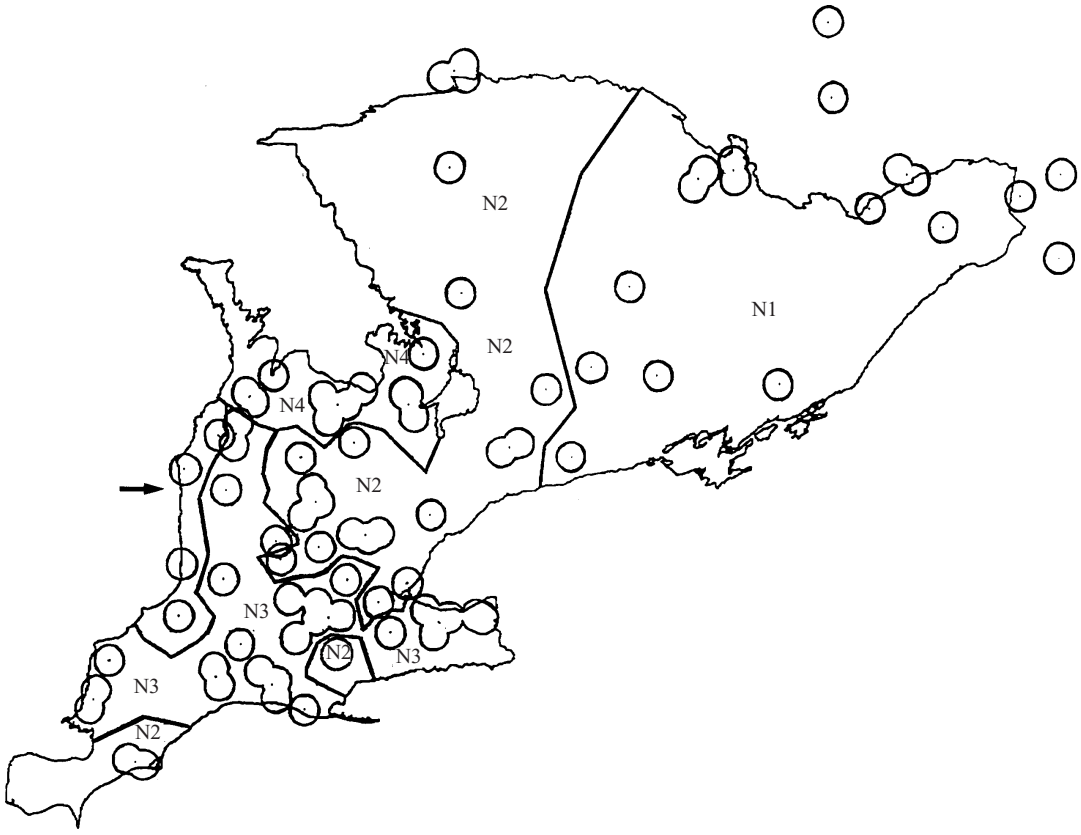

Fig. 1. Distribution of the four rabies virus N types throughout the study area as determined by application of Thiessen Polygons. The western shore line area (see arrow), which included the region where the N T2/4 variants were found, was not categorized.

had no influence on the spatial pattern. The null hypothesis of no influence was tested with a simple chi-square statistic. The second method applied Thiessen Polygons [32] to convert the point observations in the data set to an areal distribution, thereby classifying the entire space and defining regions for each of the viral types.

RESULTS

All 98 rabies virus specimens included in the primary study were typed at the N gene locus by RFLP analysis and four principal viral types (N T1-4) were identified in agreement with prior findings (see Table 1). Eighty-four of these observations (see Table 3) were included in an analysis of the spatial distribution of these viral types to examine evidence of clustering as described. Table 4*a* summarizes the nearest neighbour viral type for each observation; thus for the 18 N1 observations 16 had an N1 type as nearest neighbour and 2 had an N2 type as nearest neighbour and so on. Table 4*b* shows the expected nearest neighbour distribution assuming virus type has no influence on the spatial pattern of observations. The null hypothesis of no influence was tested by a chi-square statistic with a result ($\chi^2 = 146{\cdot}3$, D.F. $= 9$) which is significant at the $P = 0{\cdot}0001$ level. Thus clearly there was a strong relationship between location and virus type since neighbours, for the most part, had the same viral type. Conversion of the point observations in this data set to an areal distribution was achieved using Thiessen Polygons which make the assumption that the attribute(s) of any unsampled location on a map are provided by the attribute(s) of the nearest data point. In effect the attribute(s) of the data points are extended towards all other data points until the expanding cells of influence meet halfway between the data points and the entire space is classified; as cells with similar attribute(s) coalesce regions with similar attribute(s) are defined. Figure 1 shows how this process delineated the regional distribution of each of these four viral types.

To more fully delineate rabies viral variants within this study area, variation at the viral G gene was also examined. In initial studies two rabies specimens, already characterized at the N locus as a T2 virus

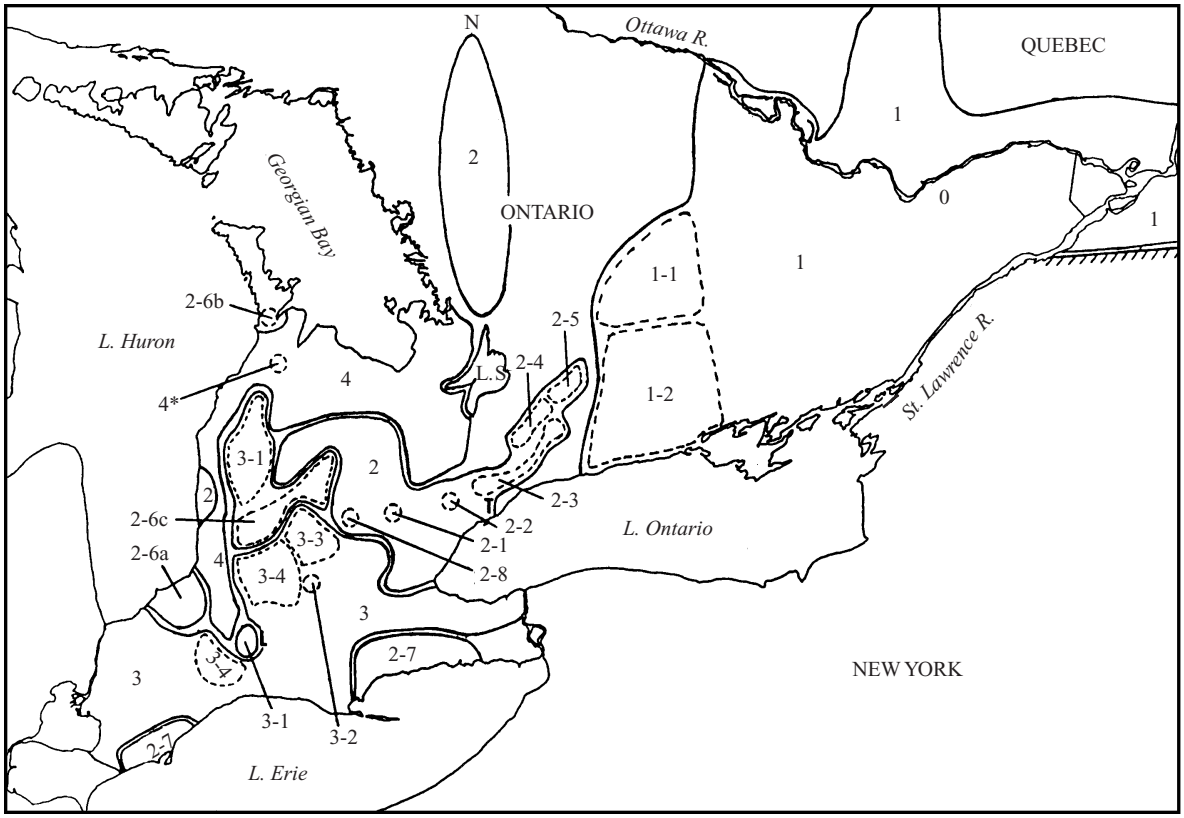

Fig. 2. Distribution of Ontario rabies virus G gene variants. Continuous lines surround areas within which all viral variants found belong to a single well defined group. Regions enclosed by broken lines indicate the areas from which subtypes (e.g. 1-1, 1-2, etc) were retrieved; their distribution has been proposed based on the same principles as applied by Thiessen Polygons but due to their limited numbers less confidence can be applied to their boundaries. Areas from which no specimens were obtained were left blank. Major cities have been labelled thus: L, London; N, North Bay; O, Ottawa; T, Toronto; L.S. indicates Lake Simcoe.

from a fox and a T3 virus from a skunk [see 17], were examined. In each case the G gene target was amplified, cloned into the pGEM3 vector and sequenced as described. These two G region sequences (GenBank accession numbers U11753, U11754) exhibited 98·4% similarity, equivalent to 39 base substitutions over the 2471 base length. These substitutions were spread throughout the sequence but were particularly concentrated in the 3′ portion of the coding region and in the contiguous noncoding and intergenic (G–L) sequences. In addition, there was a single base insertion (A) at base 2070 (non-coding region) of the skunk isolate compared to the fox isolate. These nucleotide sequences yielded distinct restriction endonuclease maps thereby allowing development of a panel of 7 restriction endonucleases for scoring a total of 15 different sites along the length of the target region. This enzyme panel was used for genetic typing of the G region for all 98 specimens (see Table 1).

Compared to N gene typing, restriction site polymorphism of the rabies virus G region identified considerably more variants; these were arbitrarily numbered according to viral N type and G sub-type. Thus the most common G gene type found in isolates with a T1 N gene was assigned G T1 and variants of this pattern were assigned numbers 1-1, etc. The data of Table 1 provide no support for the association of specific viral variants with a particular host species (i.e. fox vs. skunk). Figure 2 summarizes the regional distribution of the rabies viruses bearing these G gene variations. The localization of certain variants to discrete areas is apparent, particularly for the Type 1 group which is restricted to eastern Ontario, most type 2 variants which are located centrally and the type 3 group (3, 3-2, 3-3, and 3-4) which populate much of southeastern Ontario. The T4 viruses were restricted primarily to an area south of Georgian Bay although a small number of isolates in the southwest also yielded this typing pattern.

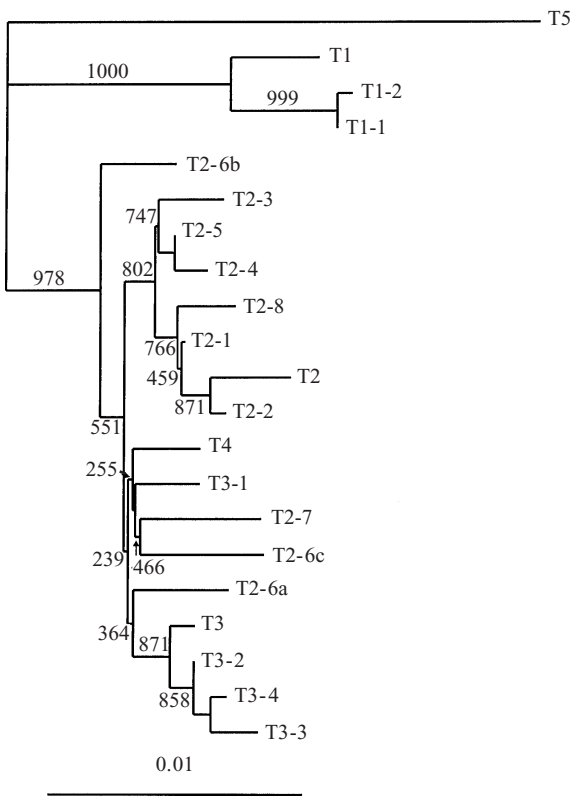

Fig. 3. Phylogram, generated by NJ analysis of G gene nucleotide sequences, illustrating the phylogenetic relationships of Ontario rabies virus variants. The number shown at each branch point represents the confidence value, expressed as number of times out of 1000 bootstrap replicates, that the grouping of specimens to the right of the fork was supported. The G gene sequence of the T5 variant of the Canadian north [18] was included as the outgroup.

To assess the degree of variation at the nucleotide level represented by these restriction mapping data, consensus sequencing was performed on the entire glycoprotein gene for at least one representative of each G gene variant. Similarity between all Ontario rabies viruses ranged from 98 to 100%, with the T1 variants representing the greatest divergence, whilst similarity to the G sequence typical of T5, the variant predominating in Northern Canada [18], was 97%. These nucleotide sequences, employing the T5 variant as outgroup, were subjected to phylogenetic analysis using both NJ and MP methods; essentially similar relationships were predicted in each case. The tree illustrated in Figure 3 was generated by NJ. The most strongly supported clade in this tree contains all T1 variants (1, 1-1, 1-2) which clearly segregate as a distinct lineage; moreover the sub-branch consisting of types 1-1 and 1-2 is well-supported. Two other groupings are supported above a 80% confidence value; one of these contains the T3 isolates and variants 3-2, 3-3 and 3-4, whilst the other association is comprised of several of the T2 variants (i.e. 2, 2-1, 2-2, 2-3, 2-4, 2-5 and 2-8). The relationships of the remaining isolates, including all the T2-6 isolates, types 2-7, 3-1 and 4, are indeterminate in this analysis due to low confidence values associated with their branching patterns. Notwithstanding, the T4 variants formed a highly homogeneous group with close to 100% similarity; their numbers and properties require their consideration as a significant group in these studies.

Figure 4*a* compares the complete glycoprotein sequence of the more commonly encountered rabies variants as predicted from nucleotide sequence data. In Figure 4*b* the sequences of the transmembrane and endo-domains of the glycoprotein of representatives of all Ontario rabies virus variants are compared. It is apparent that of the small number of coding differences occurring throughout, most are located in the transmembrane and endo-domains of this protein. In all these sequences, two glycosylation sites (at residues 37 and 319) and their neighbouring sequences were absolutely conserved as was the arginine at residue 333 known to be important for virulence [33]. In some cases amino acid substitutions parallel the proposed phylogenetic relationships. For example, variants 1-1 and 1-2 share a unique substitution at residue 128 and all T1 variants have a serine at residue 501. Note similarly the arginine amino acid at residue 488 of T2-4 and T2-5 (two isolates which segregate from the main type 2 grouping with relatively high confidence) and the distinct sequence associated with types T3, 3-2, 3-3 and 3-4. Where possible multiple isolates of each G gene variant were sequenced through this most variable region to verify the representative nature of these sequences. Of the 36 isolates characterized by nucleotide sequencing over this portion of the G gene all but one isolate predicted a protein identical to that shown in Figure 4*b* for the respective viral variant. The exception was a single T4 specimen (denoted 4*) which had a substitution for methionine at residue 452. Note also the three isolates with a T2-6 restriction profile that generated sequences differing at a small number of positions.

Since the regional pattern of viral variation, as illustrated in Figures 1 and 2, reflects the situation over a relatively short period of time (3 years) in relation to the total period (about 40 years) of rabies persistence in southern Ontario, further surveillance was undertaken to determine whether this pattern was temporally preserved. Two areas reporting significant

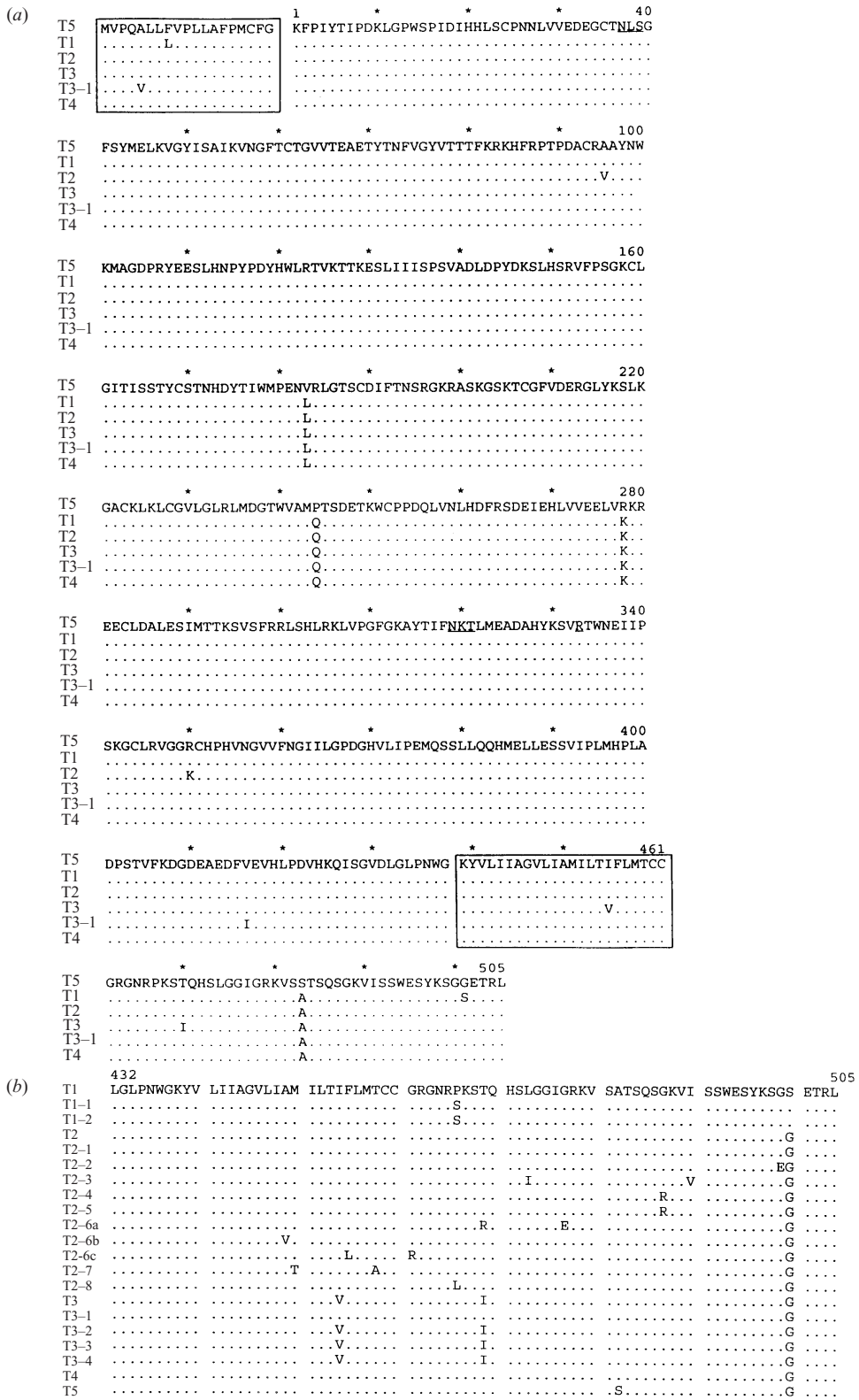

Fig. 4. (*a*) Comparison of the predicted glycoprotein sequence of the most common Ontario rabies virus variants with that of the T5 variant [18]. Residue numbering begins at the N terminus of the mature product. The 19 amino acid signal sequence and the transmembrane domain are boxed, residue 333 (arginine responsible for conferring virulence) is underlined as are conserved glycosylation sites. (*b*) Comparison of the amino acid sequences of the transmembrane and endo-domains of the glycoprotein of all identified Ontario rabies virus variants.

rabies activity in 1995–6, and which had not been sampled since 1991/2 were examined. The first was a region south of Georgian Bay which included areas harbouring variants 2 and 4 previously. Of 16 rabid samples typed at both the N and G gene loci only one T4 variant was identified; the other 15 were T2. Five of these 15 were located in an area previously populated with this variant whilst the other 10 specimens originated from regions in which T4 had predominated in earlier times. The second outbreak studied occurred in the vicinity of the town of Peterborough in eastern Ontario where viruses of N T1/G T1-2 had been identified previously. Of 4 isolates examined 3 were N T1 of which 1 was G T1-2 and the other 2 specimens typed as G T1. The fourth specimen was from a more westerly location at the border of G T1-2 and T2-5 as shown in Figure 2; it was closely related to the N T2/G T2-5 variant.

DISCUSSION

In this study we report the genetic variation observed for Ontario rabies viruses for two non-contiguous regions of the viral genome (N and G) thereby permitting a detailed molecular epidemiological description of this viral population.

Despite the high N gene similarity amongst these viruses ($\geqslant 98{\cdot}5\%$), four main viral types (N T1-4) and a small group of intermediate type (N T2/4) were described here and elsewhere [17]. Cluster analysis clearly demonstrated the regional distribution of these types and it is noteworthy that these clusters (see Fig. 1) apparently reflect the paths of the original rabies virus invasion of Ontario in the 1956–9 period [14]. In the 1950s fox rabies rapidly moved south down either side of the Ottawa River. One stream then moved south towards the Toronto area (N T2) and then subsequently invaded much of south-western Ontario (N T3); in addition a split in the rabies stream at the north end of Lake Simcoe yielded a front that moved west into the Georgian Bay area (N T4). The stream on the Quebec side of the Ottawa River entered eastern Ontario in an easternmost region and then proceeded to move westward (N T1). It is unfortunate that viral samples from this epizootic period, which constitute the ancestors of all current variants in Ontario, are unavailable for molecular analysis.

At the rather less conserved G gene locus similarity within this rabies virus population was 98% or higher, again clearly indicating the close evolutionary links of these specimens compared, for example, to the 85% similarity they exhibit with rabies of laboratory strains (eg. PV, CVS) or the Western Canada skunk strain or the 76% similarity to the US raccoon strain (34). The higher variation of the G gene permitted identification of a larger number of viral variants and by analogy with the N typing results these variants did appear to be spatially clustered in most instances (Fig. 2) although the numbers observed for each type did not allow as comprehensive an examination of this distribution. Phylogenetic analysis using the G gene sequence data supported three main branches consistent with the historical accounts of rabies invasion of the region. The most divergent branch includes all T1 viruses, together with the T1-1 and T1-2 subgroups. The two other well-supported branches are composed respectively of T2 and many T2 variants and all type 3 viruses (excluding T3-1). Isolates of some variants, notably all T2-6 isolates, T2-7, T3-1 and T4 viruses, could not be placed within the constructed phylogeny with any confidence. These findings may reflect the inherent difficulties in performing phylogenetic analysis on sequences which are so similar. Due to the considerable restraints on glycoprotein change [23, and see Fig. 4*a*], many of the mutations in the coding region are synonymous in nature, i.e. free to mutate and revert with no phenotypic consequence. The high degree to which such reversion is probably occurring likely distorts the nucleotide sequence data to the point that the algorithms used in the phylogenetic analysis do not sufficiently compensate. Furthermore, whilst nucleotide sequence determination confirmed the similarity of most isolates with identical or closely-related RFLP patterns, T2-6 variants, which were recovered from a number of geographically separated regions of the study area, yielded base sequences which did not segregate as a distinct clade. We suspect that the T2-6 variant emerged independently in these distinct regions by such mutation and reversion processes. Where such events mask the true phylogenetic relationships it becomes impossible to place much significance on infrequent variants, identified either by nucleotide sequencing or RFLP analysis.

The emergence of distinct viral variants is readily explained by the quasi-species nature of rabies virus [35], a property shared by all RNA viruses which have replicase activities with little proof-reading capabilities. However, additional factors must be operating to lead to regional localization of these variants rather than their random distribution throughout the study

area. The clustering of variants representing synonymous changes is the result of clonal growth and spread and the same may be true for variants with non-synonymous changes. However, one might also consider the possibility that subtle phenotypic differences give distinct variants selective advantages in different habitats.

Through an analysis of rabies incidence at the county level over a period of many years, Tinline and colleagues have identified several 'rabies units' within the enzootic region of southern Ontario [20, 36]. Each of these units corresponds to an area exhibiting a particular disease incidence pattern distinct from that of neighbouring units. These patterns differ primarily in the regularity and strength of the disease cycle. It was intriguing to explore whether the boundaries of these rabies units corresponded to the 'borders' between the viral variants described here. Indeed, for certain parts of the study area this appeared to be the case. For example, the disease in eastern Ontario (corresponding to the area harbouring T1 variants) exhibited a very regular and strongly fluctuating cycle of incidence which was quite distinctive from that found in the rest of the province. Also notable is the existence of a very distinct incidence pattern around Georgian Bay where the T4 virus was predominant. However, a correlation of the distribution of viral variants and county units of disease incidence patterns was not immediately evident in other parts of the province, possibly reflecting the rather different time periods considered by these two lines of investigation.

The genetic differences seen in the N gene of these viral variants are virtually all synonymous and therefore could not confer phenotypic variation. Even at the G gene locus the coding differences are small due to the functional constraints imposed upon the ectodomain. However, the comparatively high divergence of the glycoprotein's endo-domain is intriguing. In addition to its role as a membrane anchor, this domain may influence viral budding efficiency [37]. Subtle differences in its sequence could thus affect the rates of viral propagation within a host, transmission between hosts and hence could potentially influence disease incidence patterns. As yet however, such a mechanism remains highly speculative.

The extent to which topographical features of the landscape might influence rabies virus variant distribution was considered by examining physical features which might separate areas harbouring different rabies virus variants. The region of eastern Ontario, populated with T1 rabies variants, is bounded on three sides by the Ottawa and St Lawrence rivers and their convergence and also by Lake Ontario. Although these two rivers have not totally contained this rabies variant, as evidenced by the presence of fox rabies indistinguishable from the T1 variant (data not shown) in Quebec province and New York State, migration of rabies epizootics out of Ontario is relatively limited. On the western boundary of this region the landscape consists of wooded, hilly areas with a high density of lakes, some of substantial size. To the northwest lies the Canadian shield where rabies is rarely reported, although it is unclear to what extent this is due to low fox population density in this habitat or poor surveillance because of the sparsity of human settlements [18].

The area west of the T1 group and north of Lake Ontario forms the eastern boundary of type 2 variants; the most easterly part of this region, east of Toronto, in which only T2-4, 2-5 variants were recovered, is marshy and bound by the Kawartha Lakes and Lake Simcoe. The range of most T2 variants (2, 2-1, 2-2, 2-3) appears to be divided between two regions: an area west of Toronto and Lake Ontario and a region northeast of Georgian Bay which extends northwards from Lake Simcoe to North Bay. The wave of rabies which moved north from the Lake Simcoe area towards North Bay in the late 1980's, and which appeared to consist of T2 viruses, seemed to follow the paths of a CN rail line and a major highway – the features of this valley system which permitted building of these structures may also have supported extensive fox movement and/or contacts and hence rabies spread.

Most T4 variants occupy an area bounded by Lake Simcoe and the Severn River in the east, by Georgian Bay to the north and by Lake Huron to the west. To the south lies the Niagara escarpment and several associated valley systems which appear to define the border with T2 variant viruses. Much of the area occupied by T3 variants (3, 3-2, 3-3, 3-4) is farmland with few physical features which might be expected to impede fox movements. T3-1 viruses were found only in the neighbouring western area bordered by Greenock and Saratoga swamps. Several river systems, i.e. the Thames River in the west and the Grand River in the east, physically separate the range of T3 variants from the two small areas in which T2-7 variant viruses were found. It is apparent from the foregoing that many of the boundaries separating these rabies virus variants appear to follow waterways. Rivers, lakes, and other physiographical features are

likely to provide fox territory boundaries that are infrequently crossed.

Observations on the status of these variant patterns with time are revealing. The study in eastern Ontario did not indicate any dramatic change in variant distribution, though the number of samples examined here was not large. However, the larger sample sizes in the Georgian Bay study region did strongly suggest invasion of the area by the neighbouring rabies virus variant to the south (type 2) at the later time period. It would thus appear that the patterns of viral variant distribution described here are not stable but subject to periodic changes as new variants emerge and spread. We therefore speculate that these viral variations are essentially neutral genetic markers tracing the paths of host–host contacts and that topographical features of the landscape, acting to constrain host animal movements and hence rabies transmission, was likely significant in determining the extent of their spread. Moreover, judging from the phylogenetic relationships of these variants, these same landscape features may also have influenced the routes of rabies invasion observed in the 1950s.

ACKNOWLEDGEMENTS

These studies were supported by a contract from the Ontario Ministry of Natural Resources. We thank L. Rapaport and M. Prud'homme for expert technical assistance.

REFERENCES

1. Tordo N, Poch O, Ermine A, Keith G, Rougeon F. Walking along the rabies genome: Is the large G–L intergenic region a remnant gene? Proc Natl Acad Sci USA 1986; **83**: 3914–8.
2. Tordo N, Poch O, Ermine A, Keith G, Rougeon, F. Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative-strand RNA viruses. Virology 1988; **165**: 565–76.
3. Conzelmann K, Cox JH, Schneider LG, Thiel H. Molecular cloning and complete nucleotide sequence of the attenuated rabies virus SAD B19. Virology 1990; **175**: 485–99.
4. Tordo N. Characteristics and molecular biology of the rabies virus. In: Meslin F-X, Kaplan MM, Koprowski H, eds. Laboratory techniques in rabies, 4th ed. Geneva: WHO, 1996: 28–51.
5. Smith JS. Rabies virus epitopic variation: use in ecologic studies. Adv Virus Res 1989; **36**: 215–53.
6. Rupprecht CE, Dietzschold B, Wunner WH, Koprowski H. Antigenic relationships of lyssaviruses. In: Baer GM, ed. The natural history of rabies. Boca Raton: CRC Press, 1991: 69–100.
7. Saiki RK, Gelfand DH, Stoffel S, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 1988; **239**: 487–91.
8. Sacramento D, Badrane H, Bourhy H, Tordo N. Molecular epidemiology of rabies virus in France: comparison with vaccine strains. J Gen Virol 1992; **73**: 1149–58.
9. Bourhy H, Kissi B, Tordo N. Molecular diversity of the lyssavirus genus. Virology 1993; **194**: 70–81.
10. von Teichman BF, Thomson GR, Meredith CD, Nel LH. Molecular epidemiology of rabies virus in South Africa: evidence for two distinct virus groups. J Gen Virol 1995; **76**: 73–82.
11. Smith JS, Orciari LA, Yagar PA. Molecular epidemiology of rabies in the United States. Sem Virol 1995; **6**: 387–400.
12. Kissi B, Tordo N, Bourhy H. Genetic polymorphism in the rabies virus nucleoprotein gene. Virology 1995; **209**: 526–37.
13. Smith JS, Orciari LA, Yagar PA, Seidel HD, Warner CK. Epidemiologic and historical relationships among 87 rabies virus isolates as determined by limited sequence analysis. J Infect Dis 1992; **166**: 296–307.
14. Tabel H, Corner AH, Webster WA, Casey GA. History and epizootiology of rabies in Canada. Can Vet J 1974; **15**: 271–81.
15. Rosatte RC. Rabies in Canada: history, epidemiology and control. Can Vet J 1988; **29**: 362–5.
16. Webster WA, Casey GA, Charlton KM, Wiktor TJ. Antigenic variants of rabies virus in isolates from Eastern, Central and Northern Canada. Can J Comp Med 1985; **49**: 186–8.
17. Nadin-Davis SA, Casey GA, Wandeler A. Identification of regional variants of the rabies virus within the Canadian province of Ontario. J Gen Virol 1993; **74**: 829–37.
18. Nadin-Davis SA, Casey GA, Wandeler AI. A molecular epidemiological study of rabies in Central Ontario and Western Quebec. J Gen Virol 1994; **75**: 2575–83.
19. Steck F, Wandeler A. The epidemiology of fox rabies in Europe. Epidem Rev 1980; **2**: 71–96.
20. Tinline RR. Persistence of rabies in wildlife. In: Campbell JB, Charlton KM, eds. Rabies. Boston: Kluwer Academic Publishers, 1988: 301–22.
21. Wandeler AI, Nadin-Davis SA, Tinline RR, Rupprecht CE. Rabies epidemiology: some ecological and evolutionary perspectives. In: Rupprecht CE, Dietzschold B, Koprowski H, eds. Lyssaviruses. Berlin: Springer-Verlag, 1994: 297–324.
22. Ravkov EG, Smith JS, Nichol ST. Rabies virus glycoprotein gene contains a long 3′ noncoding region which lacks pseudogene properties. Virology 1995; **206**: 718–23.
23. Tordo N, Badrane H, Bourhy H, Sacramento D. Molecular epidemiology of Lyssaviruses: focus on the glycoprotein and pseudogenes. Onderstepoort J Vet Res 1993; **60**: 315–23.

24. Webster WA, Casey GA. Diagnosis of rabies infection. In: Campbell JB, Charlton KM, eds. Rabies. Boston: Kluwer Academic Publishers, 1988: 201–22.
25. Tinline RR, Gregory D. The Universal Transverse Mercator Code: a location code for disease reporting. Can Vet J 1988; **29**: 825–9.
26. Kamolvarin N, Tirawatnpong T, Rattanasiwamoke R, Tirawatnpong S, Panpanich T, Hemachudha T. Diagnosis of rabies by polymerase chain reaction with nested primers. J Infect Dis 1993; **167**: 207–10.
27. Sambrook J, Fritsch EF, Maniatis T, eds. Molecular cloning: a laboratory manual, 2nd edn. New York: Cold Spring Harbor Laboratory, 1989.
28. Henikoff S. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 1984; **28**: 351–9.
29. Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 1988; **73**: 237–44.
30. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 1985; **39**: 783–91.
31. Page RD. TREEVIEW: An application to display phylogenetic trees on personal computers. Comp Appl Biosci 1996; **12**: 357–8.
32. Burroughs P, McDonnell R. Principles of geographical information systems. Oxford: Oxford University Press, 1998: 114–5.
33. Tuffereau C, Leblois H, Benejean J, Coulon P, Lafay F, Flamand A. Arginine or lysine in position 333 of ERA and CVS glycoprotein is necessary for rabies virulence in adult mice. Virology 1989; **172**: 206–12.
34. Nadin-Davis SA, Huang W, Wandeler AI. The design of strain-specific polymerase chain reactions for discrimination of the raccoon rabies virus strain from indigenous rabies viruses of Ontario. J Virol Methods 1996; **57**: 1–14.
35. Benmansour A, Brahimi M, Tuffereau C, Coulon P, Lafay F, Flamand A. Rapid sequence evolution of street rabies glycoprotein is related to the highly heterogeneous nature of the viral population. Virology 1992; **187**: 33–45.
36. Voigt DR, Tinline R. Fox rabies and trapping. A study of disease harvest interaction. In: Sanderson G, ed. Proceedings of the Midwest Fish and Wildlife Conference. Kansas, 1982: 139–56.
37. Mebatsion T, Konig M, Conzelmann K-K. Budding of rabies virus particles in the absence of the spike glycoprotein. Cell 1996; **84**: 941–51.