UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The genome sequence and aspects of epidemiology of rabies-related Duvenhage virus

by

Charmaine van Eeden

Submitted in partial fulfilment of the requirements for the degree

of

Magister Scientiae Microbiology (M.Sc)

in the

Department of Microbiology and Plant Pathology,

Faculty of Natural and Agricultural Sciences,

University of Pretoria,

Pretoria, South Africa

Supervisor: Prof. L.H. Nel

Co-supervisor: Dr. W. Markotter

Date: 21 November 2008

© University of Pretoria

I certify that the thesis hereby submitted to the University of Pretoria for the degree M.Sc (Microbiology) has not been previously submitted by me in respect of a degree at any other University.

_____

Charmaine van Eeden

ACKNOWLEDGEMENTS

To the following people a very special word of thanks,

Prof. Louis H. Nel (University of Pretoria, South Africa)

Dr. Wanda Markotter (University of Pretoria, South Africa)

Dr. Ivan Kuzmin (The Centers for Disease Control and Prevention, USA)

Friends and fellow students (University of Pretoria, South Africa)

My deepest gratitude also goes to the following people,

Dr. Claude T. Sabeta (Rabies division, Agricultural Research Council – Onderstepoort Veterinary Institute (ARC-OVI), South Africa)

Dr. Janusz Paweska (Special Pathogens Unit, National Institute for Communicable Diseases, National Health Laboratory Services, South Africa)

Dr. PPAM van Thiel (Division of Infectious Diseases, Tropical Medicine and Aids, Academic Medical Centre, University of Amsterdam, the Netherlands)

For funding of this project,

National Research Foundation, South Africa

For their continued love and support,

Werner, Mom, Dad and Michelle; Thank you!

SUMMARY

The genome sequence and aspects of epidemiology of rabies-related Duvenhage virus

by

Charmaine van Eeden

Supervisor: Prof. L.H. Nel

Co-supervisor: Dr. W. Markotter

Department of Microbiology and Plant Pathology

University of Pretoria

For the degree M.Sc (Microbiology)

Duvenhage virus (DUVV) belongs to genotype (gt) 4 of the lyssavirus genus, in the family *Rhabdoviridae*, order *Mononegavirales*. This virus causes fatal rabies encephalitis and has only been reported from the African continent. To date there have been only five isolations of DUVV, three of which were from human fatalities and all of which were associated with insectivorous bat species. Genotype 4 lyssaviruses have not been well studied and as such little is known about them. The aim of this study was to determine the full genome sequence and investigate the epidemiology of this uniquely African lyssavirus. Standard methods of PCR and sequencing were used to determine the coding and non coding regions of various DUVV isolates. In order to determine the full genome sequence, an RNA circularization technique was used to obtain the genomic terminal sequences. Using various molecular techniques we then analyzed the sequence data, at both phylogenetic and evolutionary levels.

Our analysis showed the evolutionary forces acting against DUVV, to be similar to that which has been found for its closest relative, European bat lyssavirus type 1 (EBLV1) (gt 5). Both these viruses have strong constraints against amino acid change, with no evidence of

positive selection. Phylogenetic studies showed that not all *Lyssavirus* genes are equal for phylogenetic or lyssavirus classification analysis. High intergenotypic values at the nucleoprotein amino acid level emphasize that there is a need to reinvestigate the criteria for lyssavirus genotype classification. The strong support observed in our full genome studies suggests that full genomes may in fact be best for *Lyssavirus* analysis, so as to avoid the potential bias of individual gene analyses.

Analysis of DUVV indicates that it is an older virus within the lyssavirus genus and as shown by the discovery of the most recent isolate, the genetic diversity and incidence of this virus is greatly underestimated. Poor surveillance of rabies-related lyssaviruses as well as the poor diagnostic capabilities through most of Africa are large contributors to our lack of information. Improved surveillance of the African rabies-related lyssaviruses will extend our knowledge on the geographic distribution, host species associations and epidemiology of these viruses.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

µg	microgram
µl	microlitre
µM	micromoles
A, C, G, T	adenine, cytocine, guanine, thymine
ABLV	Australian bat lyssavirus
AMV	Avian Myeloblastosis virus
ARC-OVI	Agricultural Research Council – Onderstepoort Veterinary Institute
CDC	Centre for Disease Control and Prevention
cDNA	complementary DNA
dATP	deoxy adenosine triphosphate
DBLV	Dakar bat lyssavirus
DEPC	diethylpyrocarbonate
DNA	deoxyribonucleic acid
dNTP	deoxy nucleotide triphosphate
DTT	dithiothreitol
DUVV	Duvenhage virus
EBLV1	European bat lyssavirus type 1
EBLV2	European bat lyssavirus type 2
EDTA	ethylene diamine tetraacetic acid
ERA	Evelyn Rokitniki Abelseth
EtBr	ethidium bromide
EtOH	ethanol
FAT	fluorescent antibody test
G	glycoprotein
GHP	glycine, histidine, proline
Gts	genotypes
HCL	hydrochloric acid
HRIG	human rabies immunoglobin
i.c.	intracerebral
i.m.	intramuscular
ICTV	International Committee for the Taxonomy of Viruses
IPTG	isopropyl β-D-thiogalactosidase
KCl	potassium chloride

L	polymerase protein
LB	Luria-Bertoni
LBV	Lagos bat virus
M	matrix protein
Mabs	monoclonal antibodies
$MgCl_2$	magnesium chloride
mM	millimolar
MOKV	Mokola virus
MP	Maximum parsimony
N	nucleoprotein
NaOAC	sodium acetate
NaOH	sodium hydroxide
nt	nucleotide
ºC	degrees Celcius
P	phosphoprotein
PCR	polymerase chain reaction
PEG	poly ethylene glycol
PEP	post exposure prophylaxis
PM	Pittman Moore
PV	Pasteur virus
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNase	ribonuclease
RV	rabies virus
SAD	Street Alabama Dufferin
SDS	sodium dodecyl sulphate
SOB	super optimal broth
Tris	tris-hydroxymethyl-aminomethane
U	units
UV	ultraviolet
V	volt
VNAb	virus neutralizing antibodies
W.H.O	World Health Organization
X-gal	5-bromo-4-chloro-3-indolyl β-D-galactopyranoside

CHAPTER 1
Literature review

1.1 Introduction

Duvenhage virus (DUVV) belongs to the *Lyssavirus* genus, a group of bullet shaped viruses that have a nearly worldwide distribution. The term *Lyssavirus* is derived from the Greek word *lyssa,* meaning madness. Infection with these viruses leads to the development of rabies, which is derived from the Latin word *rabere,* meaning to rage (Wilkinson, 1988). These viruses are responsible for causing fatal encephalitis, which results in the deaths of thousands of people each year (Rupprecht *et al.,* 2002; Swanepoel, 2004; W.H.O., 2005). The *Lyssavirus* genus is one of six genera in the family *Rhabdoviridae* and constitutes a group of single stranded, negative sense RNA viruses. Rhabdoviruses infect a broad host range including plants, fish, insects and mammals. Uniquely the lyssaviruses are not associated with transmission or replication in insects as are the other Rhabdoviruses, but are adapted to replicate in the mammalian central nervous system. The lyssaviruses currently consist of seven genotypes (gts); rabies virus (RABV) (gt 1) and the rabies-related lyssaviruses; Lagos bat virus (LBV) (gt 2), Mokola virus (MOKV) (gt 3), Duvenhage virus (DUVV) (gt 4), European bat lyssavirus type 1 (EBLV1) (gt 5), European bat lyssavirus type 2 (EBLV2) (gt 6) and Australian bat lyssavirus (ABLV) (gt 7). Of these viruses only RABV, LBV, MOKV and DUVV have been identified on the African continent (Tordo *et al.,* 2005), with LBV, MOKV and DUVV being exclusive to Africa.

DUVV, for which there is to date only 5 isolates, is associated with insectivorous bats of the species; *Miniopterus* (only implicated) and *Nycteris.* Three human cases have been reported, all of which were linked to chiropteran contact. The infrequency of DUVV isolations may be the result of a number of circumstances, including host species, geographic location and poor surveillance. *Miniopterus spp* which have been implicated in the majority of cases, are nocturnal chiroptera that tend to roost in caves, rock clefts, culverts and caverns as do the *Nycteris spp,* this results in infrequent contact with humans (Van der Merwe, 1982; Gray *et al.,* 1999; Nowak, 1999). Improved surveillance of this African rabies-related lyssavirus will extend our knowledge on its geographic distribution, host species associations and epidemiology.

This project was specifically focused on gaining an increased understanding of DUVV virus through determination of a full genome sequence and investigation into the relationship between all available DUVV isolates. The close association between DUVV and EBLV1 was also explored.

1.2 Current classification of the *Lyssavirus* genus

Currently, there are seven genotypes (gts) (species) recognized in the *Lyssavirus* genus (Figure 1.1) by the International Committee for the Taxonomy of Viruses (Tordo *et al.,* 2005), these may however be expanded upon with the addition of new isolates from Eurasia (Kuzmin *et al.,* 2005).

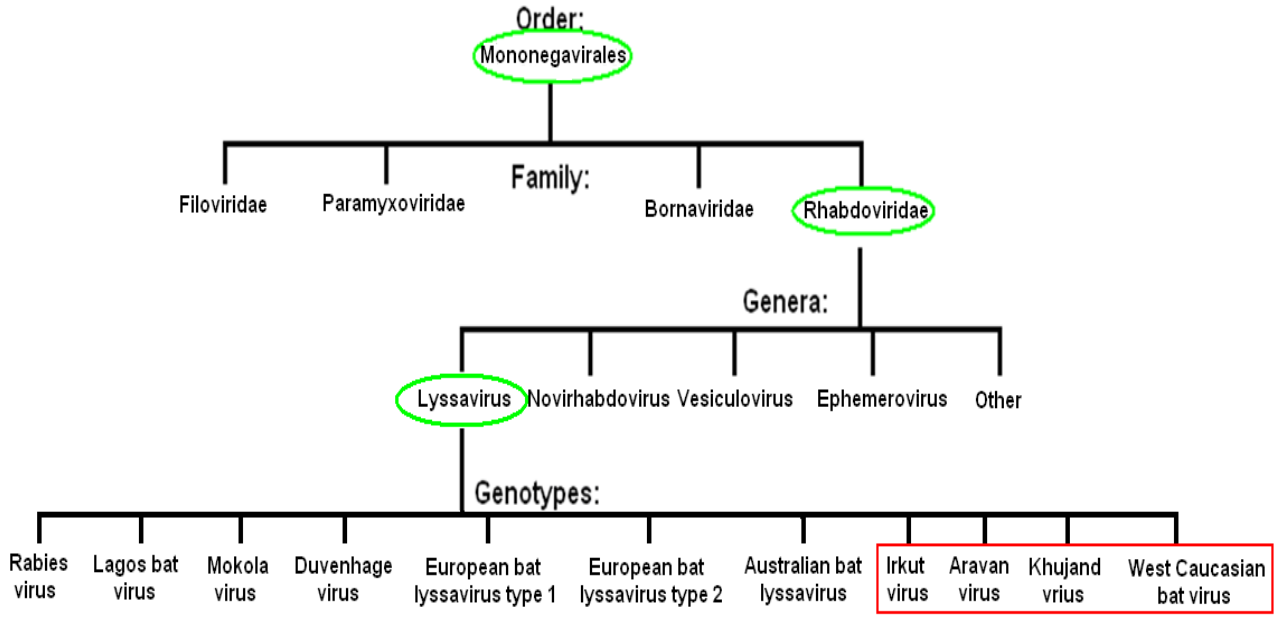


Figure 1.1 Current lyssavirus genotype classification. Putative genotypes are indicated in red block.

1.2.1 Rabies virus (RABV) – Genotype 1

RABV is the prototype lyssavirus and is one of the oldest infectious diseases known to man, with a history that can be traced back for thousands of years. The first reference to rabies appears in the *Eshunna* code in the 23rd century B.C., where it is indicated to what degree the owner of a dog was financially liable, when either a free man or a slave was bitten (Wilkinson, 1988; Steele and Fernandez, 1991; Swanepoel, 2004). The Greek philosopher Aristotle was also well aware, in the fourth century B.C, of the fatal nature of the disease and its association with the bite of an infected dog. Initial attempts at prophylaxis were described by the Roman doctor Celsus as early as the first century A.D (Wilkinson, 1988; Steele and Fernandez, 1991; Wilkinson, 2002; Swanepoel, 2004). In Europe, domestic dog (*Canis familiaris*) and red fox (*Vulpes vulpes*) rabies is still persistent, with raccoon dogs (*Nyctereutes procyanoides*) also becoming important vectors (Finnegan *et al*., 2002; Hoolmata and Kauhala, 2006).

While in Canada RABV is propagated by both red and arctic (*Alopex lagopus*) foxes (Rosatte *et al*., 2007). In the USA, striped skunks (*Mephitis mephitis*), raccoons (*Procyon lotor*), grey foxes (*Urocyon cinereoargenteus*) and coyotes (*Canis latranis*) all maintain variants of RABV (Baer, 1994; Krebs *et al*., 1999; Finnegan *et al*., 2002), which is also present in several species of insectivorous bats (Smith, 1996). In South America, RABV is principally maintained in vampire bats (*Desmodus rotundus*), although it is still prevalent in domestic dogs (Smith *et al*., 1995; Martinez-Burnes *et al*., 1997). Domestic dogs remain the principal host of RABV in Africa, although spillovers have affected a large variety of wildlife, including wild dogs (*Lycaon pictus*) (Hofmeyer *et al*., 2004; Haydon *et al*., 2006), black-backed jackals (*Canis mesomelas*) (Bingham *et al*., 2005) and kudu antelope (*Tragelaphus strepsiceros*) (Hubschle, 1998; Mansfield *et al*., 2006). In southern Africa a unique RABV variant is also well adapted to herpestid or mongoose species (Nel and Rupprecht, 2007). In Asia there remain significant cycles of canine rabies while spillover to wildlife species such as jackals and foxes also regularly reported (Bizri *et al*., 2000; Johnson *et al.,* 2003; Nadin-Davis *et al.,* 2003; Yakobson *et al*., 2004).

1.2.2 Lagos bat virus (LBV) - Genotype 2

LBV was first isolated from a fruit bat (*Eidolon helvum*) in 1956 on Lagos Island in Nigeria (Boulger and Porterfield, 1958). Most isolates to date have been from frugivorous bats (Van der Merwe, 1982; King *et al.,* 1994; Swanepoel, 2004; Markotter *et al.,* 2006a; Markotter *et al*., 2006b), although two cases in domestic cats (Crick *et al.,* 1982; King and Crick„ 1988), two in dogs (Foggin, 1988; Mebatsion *et al.,* 1992; Swanepoel et al., 2004) as well as one in an insectivorous bat (Institute Pasteur, 1985) have been identified and described. LBV is the only lyssavirus that has not been associated with human cases. It was isolated from *Epomophorus wahlbergi* bats in the Kwa-Zulu Natal province of South Africa during the 1980s (Shope, 1982; Van der Merwe, 1982) and since then further isolations have been made from *E. walhbergi* bats as well as a water mongoose (*Atilax paludinosis*) (King *et al.,* 1994; Swanepoel, 2004; Markotter *et al.,* 2006a; Markotter *et al*., 2006b). Interestingly, LBV which was thought to be exclusive to sub-Saharan Africa was also isolated from a frugivorous bat that had been imported from North Africa into France (Picard-Meyer *et al.,* 2004). A recent study by Markotter *et al.,* (2008a) found that two isolates considered to belong to genotype 2; LBVSEN1985 from Dakar, Senegal and LBVAFR1999 from either Togo or Egypt should be considered as a new lyssavirus genotype, Dakar bat lyssavirus (DBLV).

1.2.3 Mokola virus (MOKV) - Genotype 3

MOKV first isolated in 1968 from *Crocidura sp*. shrews close to the Mokola forest, in Nigeria 1968 (Kemp *et al.,* 1971), is the most divergent of the confirmed lyssavirus genotypes. This virus has also been isolated from humans, domestic cats, dogs and a single case of a rodent (Kemp *et al.,* 1971; Familusi *et al.,* 1972; Foggin, 1983; King and Crick, 1988). The first isolation of MOKV in South Africa was made from a cat near Umhlanga Rocks in Durban, 1971 (Swanepoel, 2004) with the most recent isolates coming from a cat in East London and a 6 month old puppy in Nkomazi, Mpumalanga in 2005 and 2006 respectively (Sabeta *et al.,* 2007a). The reservoir of this virus however is still unknown although it has been reported that shrews may represent potential reservoir hosts (Swanepoel, 2004). This is the only lyssavirus which has not yet been isolated from bats and is exclusive to the African continent.

1.2.4 Duvenhage virus (DUVV) – Genotype 4

Duvenhage virus was first isolated in February 1970, after a 31 year old male (Mr. Duvenhage) died in hospital after a 5 day illness diagnosed as clinical rabies. The source of exposure was reported to be a bat bite to the lip, which the victim had sustained at his home on the farm Tooyskraal, 100km north east of Pretoria, South Africa, 5 weeks prior (Meredith *et al.,* 1971). Brain tissue analysis at that point in time led to the conclusion that the victim had died from an unknown strain of rabies virus. Despite biological, morphogenetic and physicochemical comparisons that indicated DUVV to be very similar to RABV, the precise distinction of the virus came only later from serological testing (Tignor *et al.,* 1977). The isolate was named Duvenhage virus. Unfortunately the bat responsible for the bite was not collected for identification but circumstantial evidence suggests it may have been Schreiber's long-fingered bat, *Miniopterus schreibersii*, the main migration route of which falls within this area, with a maternity cave being situated only 38km away from Mr. Duvenhage's farm (Van der Merwe, 1982).

In 1981 an unidentified bat which was caught by a cat in the Louis Trichardt area of South Africa was found to be positive for DUVV (Van der Merwe, 1982). Then in 1986, DUVV was isolated from an Egyptian slit faced bat, *Nycteris thebaica*, in Bulawayo, Zimbabwe during a survey for lyssaviruses (Foggin, 1988). The fourth isolation was from a 77 year old male who died from a rabies-like illness, after being scratched on the face by what appeared to be an insectivorous bat in February 2006 in the North West province of South Africa

≈80km from the location of the first DUVV infection. The bat had flown into the victim's room at night and had landed on his spectacles, in an attempt to brush off the bat, it scratched his face and then flew away. The victim did not seek medical attention and became ill after 27 days, dying two weeks later (Paweska *et al.,* 2006). These four isolates were all obtained from a much defined geographical area in southern Africa. Most recently an isolate was obtained from Kenya in eastern Africa, December 2007 (van Thiel *et al.,* 2008). A 34 year old woman, who had been scratched on the face by an unidentified bat, whilst camping between Nairobi and Mombasa, became ill and was admitted to an Amsterdam hospital in the Netherlands. After diagnosis the 'Wisconsin rabies treatment protocol' was initiated (Willoughby *et al.,* 2005); the patient however succumbed to the virus several days' later (van Thiel *et al.,* 2008). This was the first report of DUVV outside southern Africa.

1.2.5 The European bat lyssaviruses (EBLV1) - Genotype 5 and (EBLV2) - Genotype 6

Initial isolates of the European bat lyssaviruses were classified as Duvenhage related viruses based on monoclonal antibody (Mabs) characterization (Schneider, 1982) but were subsequently recognized as independent isolates (Dietzschold *et al.,* 1988). By 1990 they were characterized as two distinct biotypes; EBLV1 and EBLV2 (Montano Hirose *et al.,* 1990; King *et al.,* 1990). Bourhy *et al.,* (1992) proposed EBLV1 and EBLV2 to be two distinct genotypes, with Amengual et al., (1997) showing both to separate into two phylogenetically distinguishable lineages (a and b). Both EBLV 1 and EBLV 2 are restricted to Europe. EBLV 1 is host adapted to *Eptesicus serotinus* (Fooks *et al.,* 2003) but EBLV1 neutralizing antibodies have been detected in *Myotis myotis, Miniopterus schreibersii* and *Rhinolophus ferrumequinum* (Serra-Cobo *et al.,* 2002). EBLV1 has also been associated with human fatalities (Bourhy *et al.,* 1992), captive (zoo) fruit bats (Ronsholdt *et al.,* 1998), sheep (Ronsholt, 2002) as well as a stone marten (*Martes foina*) (Müller *et al.,* 2001). EBLV2 is most frequently associated with insectivorous bats, mainly *Myotis daubentonii* and *M. dasycneme* (Fooks *et al*., 2003), but has also been associated with atleast two human fatalities, one in Finland (Lumio *et al*., 1986) and another in Scotland (Fooks *et al*., 2003).

1.2.6 Australian bat lyssavirus (ABLV) – Genotype 7

The last genotype to be classified within the *Lyssavirus* genus was Australian bat lyssavirus (ABLV, gt7). First isolated in 1996 from a black flying fox (*Pteropus alecto*), during a survey for equine morbillivirus (Fraser *et al.,* 1996), this virus has since been found in a large variety of bats including members of the Microchoptera and Megachoptera (Hooper *et al.,*

1997; McColl *et al.,* 2000). ABLV has also been associated with human fatalities (Allworth *et al.,* 1996; Hannah *et al.,* 2000) and is most closely related to the prototype lyssavirus, RABV. ABLV is restricted to Australia.

1.2.7 Putative genotypes (Irkut, Aravan, Khujand and WCBV)

The putative lyssavirus genotypes have all been isolated from bats in Eurasia; Aravan from *Myotis blythi* in the Osh region of Kyrgyzstan in 1991 (Arai *et al.,* 2003), Khujand from *Myotis mystacinus* in northern Tajikistan in 2001 (Kuzmin *et al.,* 2003), WCBV from *Miniopterus schreibersii* and Irkut from *Murina leucogaster* in 2002 (Botvinkin *et al.,* 2003). There is to date only a single isolate for each virus. Phylogenetic investigation into both Khujand and Aravan viruses (Kuzmin *et al.,* 2003) suggested that Khujand was certainly related to EBLV2, whereas Aravan, most closely related to Khujand, also demonstrated moderate similarity to DUVV, EBLV1 and EBLV2 (Kuzmin *et al.,* 2003). For Irkut and WCBV both antigenic typing and phylogenetic analysis linked Irkut to DUVV and EBLV1, whilst WCBV clustered with MOKV and LBV (Botvinkin *et al.,* 2003). The low bootstrap value supporting this cluster illustrated WCBV to be the most divergent lyssavirus (Botvinkin *et al.,* 2003). Based on results of phylogenetic analyses, it was proposed that these viruses should be regarded as new lyssavirus genotypes (Kuzmin *et al.,* 2005).

1.3 Phylogroup designation for the *Lyssavirus* genus

Based on phylogeny, pathogenicity and serologic cross reactivity, the lyssavirus genotypes have been proposed to divide into two phylogroups (Figure 1.2) (Badrane *et al.,* 2001). Genotypes 1, 4, 5, 6 and 7 make up phylogroup I and are considered highly pathogenic to mice both through intracerebral (i.c.) and intramuscular (i.m.) routes. Genotypes 2 and 3 make up phylogroup II and where thought to be highly pathogenic only through i.c. introduction. A recent study by Markotter *et al*., (2008b), however, showed some isolates of phylogroup II to be pathogenic via i.m. routes also. It has also been suggested that WCBV may belong to an independent phylogroup III, due to its genetic distance and absence of serologic cross-reactivity with both phylogroup I and II (Kuzmin *et al.,* 2005).

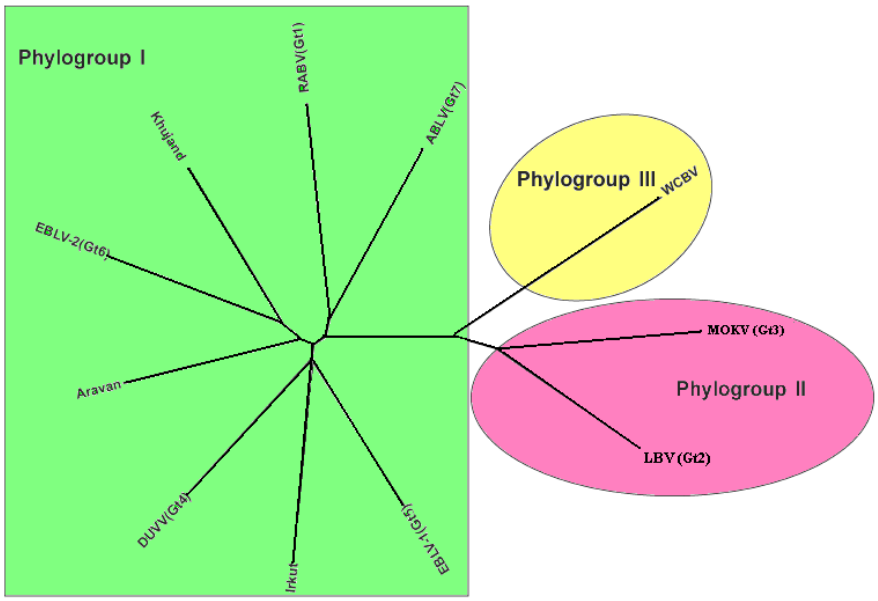


Figure 1.2 Phylogenetic tree based on comparative alignment and neighbour-joining of the 1353 nucleotides of the N gene. The separation of the genus into hypothetical phylogroups is also shown (Nel and Markotter, 2007).

1.4 Genotype classification of the *Lyssavirus* genus

Originally the lyssaviruses were subdivided into four serotypes on the basis of sero-neutralization and monoclonal antibody studies (Schneider *et al.,* 1973): RABV (serotype 1), LBV (serotype 2), MOKV (serotype 3) and DUVV (serotype 4). The EBLV's were initially proposed to constitute serotype 4, but were then subdivided into biotypes 1 and 2 (EBLV1 and EBLV2), which were finally distinguished as two distinct genotypes (Bourhy *et al.,* 1992.). In 1993, Bourhy *et al.,* undertook a study in which they sequenced the N gene of the four serotypes and the two biotypes of the lyssaviruses to determine the genetic diversity and to reinvestigate the relationships throughout the *Lyssavirus* genus. Phylogenetic analysis showed six distinct branches corresponding to the six distinct genotypes (Bourhy *et al.,* 1993). In 1998 ABLV (gt 7) was characterised using gene sequence analyses, electron microscopy and a panel of monoclonal antibodies (Gould *et al*., 1998).

It was also ascertained that the threshold below which a new genotype should be defined, was the interval given by the lowest percentage of amino acid similarity found within one genotype (97.1%) and the highest percentage similarity found between two genotypes (93.3%) (Bourhy *et al.,* 1993). The genetic diversity of the rabies virus N gene was again explored in 1995, by Kissi *et al.,* with the aim to compare the intrinsic and extrinsic genetic

diversity of the lyssavirus genotypes. Their results suggested that isolates belonging to different genotypes have less than 79.8 and 93.3% identity at the nucleotide and amino acids levels respectively. While the percentage values linking isolates within a genotype are 83.3 and 92.2% (at the nucleotide and amino acids levels respectively). It was however concluded, that no precise percentage value could be given for the classification of a definite genotype, although it was assumed that isolates sharing less than 80% nucleotide and 92% amino acid similarity would belong to different genotypes (Kissi *et al.,* 1995).

A study by Kuzmin *et al.,* (2003), found that the amino acid identity between Aravan and Khujand was 92.7%; RABV and ABLV, 92.5% and DUVV and EBLV1 93.3% and with amino acid identity between distinct RABV representatives being as low as 93.7%. It was thus concluded that this criterion for genotype differentiation i.e. Kissi *et al.,* 1995, is questionable due to the likely possibility of overlap (Kuzmin *et al.,* 2003). It was thus suggested that when a taxonomic group definition is given, that some qualitative criteria should be applied in addition to identity calculation and bootstrap support of phylogenetic tree topology, host origin and geographical distribution seeming the most logical criteria (Kuzmin *et al.,* 2003). It was however stated that due to the limited information regarding the African rabies-related viruses as well as the putative genotypes, this criteria of classification cannot yet be introduced but should be considered in the future when more data is collected (Kuzmin *et al.,* 2003). In a study by Markotter *et al.,* (2008a) the idea that nucleotide and amino acid identities should not be less between isolates of the same genotype (intragenotypic identity) than between isolates considered to belong to separate genotypes (intergenotypic identity) was used to determine which genes were best suited to genotype classification. The ratio (minimum intragenotypic identity/maximum intergenotypic identity > 1) was used to great effect, leading to the observation that two isolates of gt 2 are to be considered as a separate genotype.

1.5 Studies on the host origin of lyssaviruses

Chiroptera are reservoirs for six of the seven lyssavirus gts as well as the four putative gts. LBV circulates primarily in frugivorous bats (Markotter *et al.,* 2006b), DUVV, EBLV1 and EBLV2 in insectivorous bats (Bourhy *et al.,* 1992; King *et al.,* 1994; Paweska *et al.,* 2006) and ABLV in both insectivorous and frugivorous bats (Hooper *et al.,* 1997; Warrilow, 2005). Though there are many reservoirs and vectors for RABV worldwide, both insectivorous and

vampire bat species have been linked to its propagation (Smith *et al*., 1995; Smith, 1996; Martinez-Burnes *et al*., 1997). The reservoir species of MOKV has not yet been identified. There has been speculation that lyssaviruses evolved from insect rhabdoviruses which were then transmitted to insectivorous bats in the distant past (Shope, 1982; Badrane and Tordo, 2001). This premise is supported by the fact that members from most of the genera in the *Rhabdoviridae* family have been isolated from insects (King and Crick, 1988). Additionally, MOKV, which was first isolated from an insectivorous shrew, has been shown to replicate in *Aedes aegypti* mosquito cells (Shope, 1982; King and Crick, 1988).

In 2001, Tordo and Badrane dated the presumed common insect virus ancestor of the lyssavirus genotypes as being present approximately 7080 to 11631 years ago, by means of phylogenetic and molecular clock analysis using G gene sequence data from carnivoran and chiropteran rabies as well as the rabies-related lyssaviruses. For subsequent evolutionary events, there is little doubt RABV may have evolved as a virus of bats (Badrane and Tordo, 2001), as chiropteran lyssaviruses existed long before carnivore rabies and phylogeny of the lyssaviruses imply that at least two ancient spillover events have occurred, both within the gt1. These results provide evidence that what is today known as carnivoran rabies may indeed have resulted from various successful episodes of host switching from bats (Badrane and Tordo, 2001).

1.6 Molecular biology of lyssaviruses

1.6.1 Structure and genome organization

Rabies virus has a bullet-shaped morphology with particles ranging in size from 130-200nm in length with a diameter of 60-110nm (Tordo and Poch, 1988). The virions consist of a nucleocapsid core surrounded by a host derived-lipid envelope (Swanepoel, 2004). Lyssaviruses have a single continuous negative stranded RNA genome of about 12 kilobases. This genome encodes five viral proteins (3' to 5'): nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G), and polymerase protein (L) all of which are present in the virion (Tordo and Poch, 1988). The N, L and P proteins make up the nucleocapsid core (Kawai, 1977), while the M and G proteins are located on the viral envelope (Figure 1.3). Surface projections, consisting of glycoproteins, extend from the envelope and are anchored in the membrane by a 22 aa hydrophobic transmembrane domain.

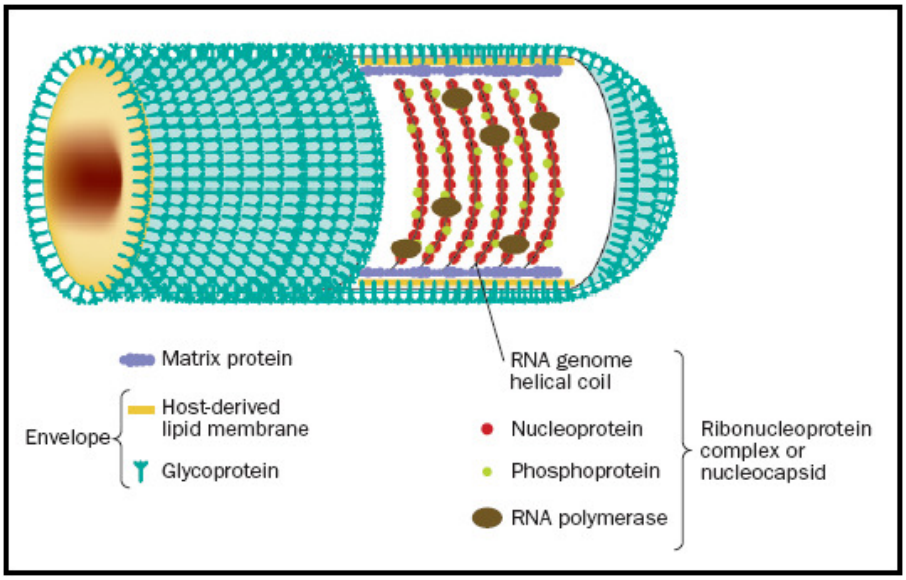


Figure 1.3 Structure of rabies virus (Warrell and Warrell, 2004).

Short non-coding regions flank the coding regions and are bordered by start and stop transcriptional signals nine nucleotides in length, that mark where initiation and termination of transcription occurs (Tordo and Koukenetzoff, 1993; Marston *et al*., 2007). Non-coding intergenic regions separate the genes from each other. In RABV there are 2 nt separating the nucleoprotein gene from the phosphoprotein gene, 5 nt each separating the phosphoprotein gene from the matrix protein gene and the glycoprotein gene, 423 nt separate the glycoprotein gene from the polymerase gene (Figure 1.4) (Wunner, 1991). The 3' terminus of RABV is 56 nt in length (Kurilla *et al*., 1984) and the encapsidation initiation sequence has been identified as the A-rich sequence spanning nucleotides 20 to 30 in the leader RNA (5' AAGAAAAAACA 3'). A similar encapsidation initiation sequence (5' AAAAATGAGA 3') is present at nucleotides 20 to 30 of the 5' end (Yang *et al.* 1998).

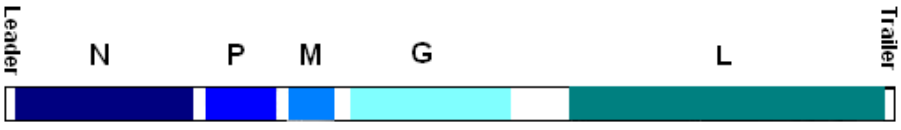


Figure 1.4 A schematic presentation illustrating the order of genes (3'-5') on the rabies virus genome

1.6.2 Lyssavirus proteins

Nucleoprotein (N)

The nucleoprotein is responsible for encapsidation of the RNA genome, ensuring protection against nucleases (Keene *et al*., 1981) and regulates viral transcription and replication through its ability to promote readthrough of viral termination signals (Tordo and Poch, 1988). The N protein is also thought to play a crucial role in transition of the RNA synthetic mode from transcription to replication (Patten *et al*., 1984). Furthermore it has been found that the N protein is a major target antigen for T-helper cells that cross-react among RABV and rabies-related viruses and has also shown to be able to act as a super antigen in humans (Dietzschold *et al*., 1987; Lafon *et al*., 1992; Kawai and Morimoto, 1994; Fu *et al*., 1994; Wunner, 2002).

Phosphoprotein (P)

The phosphoprotein is an essential component of the RNA polymerase complex in which it acts as a cofactor during transcription and replication (Wunner, 2002). It also functions as a chaperone to deliver N protein for encapsidation of viral RNA (De *et al*., 1997).

Matrix protein (M)

The matrix protein plays a central role in viral assembly and release of the viral progeny from infected cells by budding. It also functions as a regulatory protein adjusting the balance of RNP replication and mRNA synthesis (Mebatsion *et al*., 1999). It has also been found that this protein plays an important role in the induction of cellular apoptosis (Kassis *et al*., 2004).

Glycoprotein (G)

The glycoprotein is composed of four distinct domains: the signal peptide that allows the translocation of the polypeptide through the endoplasmic reticulum (Tordo and Kouknetzoff, 1993), the ectodomain which includes glycosylation, palmytolation and antigenic sites (Coulon *et al*., 1993), the transmembrane peptide which anchors the protein within the viral envelope and the cytoplasmic domain located in the inner part of the virion (Tordo and Kouknetzoff, 1993). The G protein plays an important role in virus-host interaction by mediating the attachment of virus to the host cells, whilst the transmembrane glycoprotein stimulates both humoral and cell mediated immunity against viral infection, and contains the major antigenic sites responsible for eliciting virus neutralizing antibodies (VNAb) (Tordo and Poch, 1988).

Polymerase protein (L)

The polymerase protein carries out all the enzymatic steps required for transcription, including initiation and elongation of transcripts (Patton *et al*., 1984) as well as co-transcriptional modifications of RNAs such as capping, methylation and polyadenylation (Wunner, 2002).

1.7 Pathogenicity of the lyssaviruses

Rabies is acute, incurable encephalitis which can be caused by all members of the *Lyssavirus* genus. All warm blooded mammals tested so far are susceptible to infection, which in most cases is transmitted by the exposure of wounds or cuts on the skin, to virus laden saliva, most commonly inoculated through a bite from an infected animal (Hemachudha *et al.,* 2002; Warrell and Warrell, 2004).

1.7.1 Humans

Mortality in humans after exposure to RABV varies and depends on the severity and location of the wound, as well as the presumed concentration of virus in the saliva (Hemachudha *et al.,* 2002). After inoculation, the virus can either replicate locally in muscle cells before gaining access to the central nervous system (CNS) or alternatively can attach directly to sensory nerve endings (Swanepoel, 2004; Warrell and Warrell, 2004). Having gained entry to peripheral nerves, it travels via retrograde axoplasmic transport to the brain, where massive replication of viral particles ensues within the neurons (Warrell and Warrell, 2004). Following infection of the CNS the virus spreads along peripheral nerves to sites throughout the body, including the salivary glands (Krebs *et al.,* 1995; Warrell and Warrell, 2004). The clinical features of rabies infection consist of: the incubation period; the prodrome – where the virus reaches the CNS; the acute neurological phase – where neurological and behavioral symptoms dominate, either as encephalitic (furious) rabies or paralytic (dumb) rabies; coma and death (Hermachuda *et al.,* 2002; Bishop *et al.,* 2003).

Although the symptoms of rabies in humans may be non-specific; in encephalitic rabies the victim may suffer from hyperactivity and convulsive seizures, aggravated by thirst and fear of light, noise and other stimuli (Hermachuda *et al.,* 2002; Bishop *et al.,* 2003). Within 24 hours three major cardinal signs appear, including fluctuating consciousness, phobic or inspiratory spasm as well as autonomic stimulation signs. Mental status varies between normal periods

in which the patient is lucid, to periods characterized by severe agitation, depression and aggression. Irritability is gradually followed by the deterioration of consciousness and death (Hermachuda *et al.,* 2002). Victims suffering paralytic rabies experience muscular weakness which usually starts at the bitten limb and soon spreads to encompass all limbs. Death usually results from asphyxiation, due to paralysis of the bulbar and respiratory muscles (Hemachuda *et al.,* 2002; Swanepoel, 2004).

1.7.2 Animals

Initial signs of rabies in animals are non-specific and therefore resemble a number of infectious diseases (Niezgoda *et al*., 2002). Dramatic behavioral alterations, such as wild animals losing their fear of humans, may be an indication of a lyssavirus infection (Hassel, 1982; Mansfield *et al.,* 2006; Nel and Markotter, 2007)). The clinical signs associated with EBLV1 infection in bats was recently documented by Franka *et al.,* (2008) where tremors, irritability, aggressiveness and paralysis were observed and in some cases sudden death without any apparent signs of disease.

1.8 Diagnosis

Prior to the 1950's, rabies diagnosis was based on the observation of confined animals with suspected rabies, the mouse inoculation test, as well as histological examination of brain tissue for supportive evidence of inflammation and inclusion bodies (Steele and Fernandez, 1991; Rupprecht *et al.,* 2002). This has largely been superseded by the development of the fluorescent antibody test (FAT), first described by Goldwasser and Kissling in 1958, which is based on the detection of lyssavirus antigen on touch impression brain smears using fluorescently labeled antibody conjugate (Campbell and Barton, 1958; Steele and Fernandez, 1991; Rupprecht *et al.,* 2002; Swanepoel, 2004). The advent of the reverse transcription polymerase reaction (RT-PCR) assay, provided a further useful addition to lyssavirus virus diagnosis (Rupprecht *et al.,* 2002), but remains limited in general usage due to the need for universal primers, which usually translates to several assays and where products need to be verified through sequencing. The continual development of real-time PCR procedures is however likely to expand routine diagnosis (Hughes *et al*., 2004).

1.9 Vaccines

The first rabies vaccine was developed by Louis Pasteur in 1881 and following extensive testing was successfully used in humans in 1885 (Wilkinson, 1988; Steele and Fernandez, 1991; Wilkinson, 2002; Swanepoel, 2004). This neural tissue vaccine found widespread application and established the principle of post exposure prophylaxis (PEP) against rabies. Today a variety of safe, stable and highly immunogenic inactivated cell culture and veterinary vaccines are available. These can be used to vaccinate humans and animals either before or after exposure to lyssaviruses (Swanepoel, 2004). Post-exposure prophylaxis with modern cell culture vaccines can virtually guarantee complete protection following an exposure to most lyssavirus genotypes, provided it is applied in a correct and timely manner and is combined with the correct wound treatment and recommended regimens of human rabies immunoglobin (HRIG) (W.H.O., 2005). Pre-exposure vaccination is recommended for individuals at occupational risk or those travelling in rabies endemic areas (W.H.O., 2005).

All experimental evidence to date suggests that commercial vaccines protect against only the phylogroup I lyssaviruses (RABV, DUVV, EBLV1, EBLV2 and ABLV) and not members of phylogroup II (LBV and MOKV) (Fekadu *et al.,* 1988; Bahloul *et al.,* 1998; Badrane *et al.,* 2001). Cross protection between genotypes is possible, due to shared glycoprotein antigenic sites towards which neutralizing antibodies are directed (Benmansour *et al.,* 1991). Commercial vaccine strains all belong to RABV (gt 1) and there is no evidence of their lack of efficacy against gt 1 viruses but they are less effective against the rabies-related lyssaviruses. For EBLV1 (gt 5) and EBLV2 (gt 6) varying results were obtained depending on the vaccine strain tested, a level of protection was however demonstrated, although efficiency was less than for gt 1 (Lafon *et al.,* 1986; Fekadu *et al.,* 1988; Lafon *et al*., 1988). Varied results were obtained when testing the efficacy of ERA and PM vaccine strains against DUVV (gt 4), both vaccines produced an anamnestic response to DUVV in rabbits, but only ERA produced a response in mice (Fekadu *et al.,* 1988). In the case of ABLV (gt 7) mice were shown to be protected after vaccination with various RABV vaccines (Brookes *et al*., 2001), however for LBV (gt 2) and MOKV (gt 3) no protection has been shown (Tignor and Shope, 1972; Dietzschold *et al*., 1987; Mebastion *et al.,* 1992).

Investigation into the efficacy of rabies vaccines against the unclassified lyssaviruses Aravan, Khujand, Irkut and WCBV conformed to previous findings that suggest protection is inversely proportional to the genetic distance between the viruses considered and RABV (Hanlon *et al.,* 2005). Aravan, Khujand, Irkut, are protected against, whereas WCBV is not, indicating that cross neutralization exists within but not between lyssavirus phylogroups.

1.10 Lyssavirus genome analysis

Phylogenetic analysis has become an increasingly important tool in the investigation of the epidemiology of rabies throughout the world (Smith *et al.,* 1992; Kissi *et al.,* 1995; Bingham *et al.,* 1999; Peaz *et al.,* 2003; Cohen *et al.,* 2007) and southern Africa (Coetzee *et al.,* 2007; Coetzee and Nel, 2007; Sabeta *et al.,* 2007b). The ability of phylogenetic analysis to establish speciation, geographical links and identification of common ancestry (Fitch, 1995) has made this an ideal tool to study the *Lyssavirus* genus. Evolutionary studies of lyssaviruses have focused on the nucleoprotein (N) and glycoprotein (G) genes. Infected cells have a high abundance of N mRNA and it is thus an ideal target for DNA sequencing. The N gene is also well conserved and is therefore well-suited to comparing isolates across a relatively long term of evolution (Bourhy *et al.,* 1992). The host cell receptor recognition and membrane fusion domains, which are the major targets of the host neutralizing-antibody response, make the G gene an equally apt target (Badrane *et al.,* 2001). Additionally the carboxyl terminal domain of the glycoprotein and the G-L intergenic region have been used in various studies as this region constitutes the most variable portion of the RABV genome (Tordo and Kouknetzoff, 1993). This target is considered appropriate for distinguishing closely related viral variants from each other, as it has been shown in various studies to be well suited for the investigation of the molecular epidemiology of rabies in defined geographical domains (Tordo *et al*., 1986; Sacramento *et al*., 1991; Nel *et al*., 1993; Nadin-Davis, 2000; Paez *et al*., 2003).

All five proteins are however both structurally and functionally related and there is common agreement that interacting proteins undergo co-evolution (Pazos *et al.,* 1997). Since no recombination events have been reported in lyssaviruses, Wu *et al.,* (2007) proposed that regardless of the gene chosen for phylogeny, if the same method is applied for analysis, individual genes may generate similar tree topologies.

The full genomes of all the lyssavirus genotypes have been sequenced (Tordo *et al.,* 1988; Conzelmann *et al.,* 1990; Le Mercier *et al.,* 1997; Warrilow *et al.,* 2002; Marston *et al.,* 2007; Delmas *et al.,* 2008), as have those of the four putative genotypes (Kuzmin *et al.,* 2008b). Table 1.1 lists all currently available sequences. The most widely used method for lyssavirus genomic terminal sequence determination is RACE (Rapid amplification of cDNA ends) (Warrilow *et al.,* 2002; Marston *et al.,* 2007; Delmas *et al.,* 2008), where purified cDNA is tailed with dNTPs allowing for the attachment of a RACE specific reverse primer. This method is however expensive and there are many potential problems linked to tailing of DNA fragments, thus the method of Kuzmin *et al.,* (2008b) was to be used in this study. Here total RNA is circularized and nested PCR carried out with primers specific to the virus. This method is not only more cost efficient, but both genomic termini can be determined simultaneously.

Some studies involving lyssavirus full genomes have looked at sequence length comparisons of both the coding and non-coding regions (Table 1.2). The study by Delmas *et al.,* (2008) is the most comprehensive, including full length analysis of all the lyssavirus genomes. The average G + C content of the lyssavirus genome has been found to be 44.57%; this value is in accordance with the notion that there is G + C biasing in RNA viruses, which is directly linked to their genomic polarity. This bias is thought to be the result of host cell RNA editing, which occurs mainly on the negative strand, leading to positive strand viruses having a higher G + C content than their negatively stranded counterparts (Auewarakul, 2005; Marston *et al.,* 2007).

Table 1.1 Lyssavirus full genome sequences

VIRUS CODE	GENOTYPE	SPECIES ISOLATED FROM	YEAR OF ISOLATION	GEOGRAPHIC LOCATION	REFERENCE SOURCE	GENBANK ACCESSION NUMBER
RAVMMGN	1	Rabies virus, laboratory strain Pasteur			Tordo *et al.,* 1988	M13215
ERA	1	Rabies virus, laboratory strain Evelyn-Rokitnicki-Abelseth			Unpublished	EF206707
8743THA	1	*Homo sapiens*	1983	Thailand	Delmas *et al.,* 2008	EU293121
8764THA	1	*Homo sapiens*	1983	Thailand	Delmas *et al.,* 2008	EU293111
9147FRA	1	Fox	1991	France	Delmas *et al.,* 2008	EU293115
9001FRA	1	Dog bitten by bat	1990	France	Delmas *et al.,* 2008	EU293113
9704ARG	1	*Tadarida brasilliensis*	1997	Argentina	Delmas *et al.,* 2008	EU293116
SHBRV-18	1	*Lasionycteris noctivagans*	1983	USA	Faber *et al.,* 2004	AY705373
NNV-RAB-H	1	*Homo sapiens*	2006	India	Unpublished	EF437215
SADB19	1	Rabies virus, laboratory strain Street Alabama Dufferin B-19			Conzelmann *et al.,* 1990	M31046
8619NGA	2	*Eidolon helvum*	1956	Nigeria	Delmas *et al.,* 2008	EU293110
0406SEN	2	*Eidolon helvum*	1985	Senegal	Delmas *et al.,* 2008	EU 293108
KE131	2	*Eidolon helvum*	2007	Kenya	Kuzmin *et al.,* 2008	EU259198
MOKV	3	Cat	1981	Zimbabwe	Le Mercier *et al.,* 1997	NC_006429
86100CAM	3	Shrew	1974	Cameroon	Delmas *et al.,* 2008	EU239117
86101RCA	3	Rodent	1981	Central African Republic	Delmas *et al.,* 2008	EU293118
DUVVSA06	4	*Homo sapiens*	2006	South Africa	Paweska *et al*., 2006	EU623444
86132SA	4	*Homo sapiens*	1971	South Africa	Delmas *et al.,* 2008	EU293119
94286SA	4	*Miniopterus schreibersii*	1981	South Africa	Delmas *et al.,* 2008	EU293120
9395GER	5	*Eptesicus serotinus*	1968	Germany	Marston *et al.,* 2007	EF157976
8918FRA	5	*Eptesicus serotinus*	1989	France	Delmas *et al.,* 2008	EU293112
03002FRA	5	*Eptesicus serotinus*	2003	France	Delmas *et al.,* 2008	EU293109
9018HOL	6	*Myotis dasycneme*	1986	Holland	Delmas *et al.,* 2008	EU293114
RV1333	6	*Homo sapiens*	2002	United Kingdom	Marston *et al*. 2007	EF157977
ABLh	7	*Homo sapiens*	1998	Australia	Warrilow *et al.,* 2002	AF418014
Irkut		*Murina leucogaster*	2002	Russia	Kuzmin *et al.,* 2008b	EF614260
West Caucasian bat		*Miniopterus schreibersi*	2002	Russia	Kuzmin *et al.,* 2008b	EF614258
Khujand		*Myotis daubentonii*	2001	Tajikistan	Kuzmin *et al.,* 2008b	EF614261
Aravan		*Myotis blythi*	1991	Kyrgyzstan	Kuzmin *et al.,* 2008b	EF614259

Table 1.2 Sequence length comparisons of lyssavirus genomes in nucleotides.

	DUVV	EBLV1	EBLV2	LBV	MOKV	ABLV	RABV
3' UTR	70	70	70	70	70	70	70
N gene	1356	1356	1356	1353	1353	1353	1353
N-P	90	90-96	101	101	100-102	93-94	90-94
P gene	897	897	894	918	912	894	894
P-M	83	83	88	75	80-83	87	87-90
M gene	609	609	609	609	609	609	609
M-G	191	211	205-210	204	203-204	207-209	211-215
G gene	1602	1575	1575	1569	1569	1578-1581	1575
G-L	562-563	560	511-512	578-588	546-562	508-509	515-525
L gene	6384	6384	6384	6384	6381-6384	6384-6387	6384-6429
5' UTR	131	130-131	131	145	112-114	131	86-131
Genome (nt)	11975-11976	11966-11971	11924-11930	12006-12016	11940-11957	11 918	11923-11928

1.10.1 Antigenic domains

Antigenic sites on the nucleoprotein and glycoprotein of lyssaviruses which have been previously identified are listed in Table 1.3

Table 1.3 Antigenic domains on the lyssavirus genome

Protein	Region on genome	Reference
Nucleoprotein	aa 358-367 (site I)	Goto *et al.,* 2000; Minamoto *et al.,* 1994
	aa 313-337 (site II)	Lafon and Wiktor, 1985
	aa 374-383 (site III)	Dietzschold *et al.,* 1988
	aa 410-413 (site IV)	Ertl *et al.,* 1991
Glycoprotein	aa 14-19	Mansfield *et al.,* 2004
	aa 231 (site I)	Lafon *et al.,* 1983
	aa 34-42 & aa 198-200 (site II)	Lafon *et al.,* 1983; Prehaud *et al.,* 1988
	aa 330-338 (site III)	Lafon *et al.,* 1983
	aa 264 (site IV)	Dietzschold *et al.,* 1990
	aa 342-343 (site V)	Benmansour *et al.,* 1991

1.10.2 Conserved domains

Phosphoprotein:

The LC8 dynein light chain was found to bind strongly to the P protein of RABV and MOKV, suggesting that this interaction is important for pathogenesis (Mebatsion, 2001; Poisson *et al*., 2001). The motif (K/R)XTQT at amino acids 145-149, has demonstrated interaction with LC8, a protein which contributes to the axonal transport of RABV within neurons (Lo *et al.,* 2001). All lyssaviruses with the exception of MOKV (KSIQI) and WCBV (no motif) have this motif (Marston *et al.,* 2007).

Matrix protein:

Two classical late domain binding motifs have been identified in the RABV M protein, PPXY (aa 35-38) and PX(T/S)AP (aa 21-25) (Jayakar *et al.,* 2000). The PPXY motif has been shown to be present in all lyssaviruses (DUVV not included), with the exception of Khujand (PPES) (Marston *et al.,* 2007). Absence of this motif has been observed to reduce RABV budding (Harty *et al.,* 2001).

Polymerase protein:

Comparison of L gene sequences from the members of the order Mononegavirales, demonstrated that conserved residues are clustered into six blocks of strong conservation linked by variable regions of low conservation (Poch *et al.,* 1990). The blocks of highest amino acid conservation (II to V) are located in the central region of the protein (positions 578 to 1491) (Poch *et al.,* 1990). The study by Marston *et al.,* (2007) showed that the four conserved regions in negative stranded RNA virus proteins (A-D) (see Appendix C) are completely identical in all the lyssaviruses studied (DUVV not included), including the GG(I/L)EG (694-697) and pentapeptide QGDNQ (728-732). In block I the invariant GHP residues (373-376) were conserved. Additionally it was found that the GDGSGG motif at position 1704-1708 with a lysine residue 19 bases down, is also conserved within block V of all lyssaviruses (DUVV not included).

1.10.3 Pathogenic sites

The pathogenicity of lyssaviruses depends on the presence of various antigenic determinants on the glycoprotein and nucleocapsid proteins (Dietzschold *et al.,* 1988). The development of hybridoma technology, made possible the use of the neutralizing power of anti-glycoprotein monoclonal antibodies (Mabs) to isolate antigenic mutants which resist

neutralization (Wiktor and Koprowski, 1978). The study of these mutants associated with an analysis of their reactivity patterns with Mabs, provided an indication as to the location of antigenic sites in viral proteins (Flamand *et al.,* 1980a; Flamand *et al.,* 1980b). This enabled the mapping of epitopes and characterization of pathogenicity (Dietzschold *et al.,* 1983; Seif *et al.,* 1985). Mabs has facilitated the classification of virus strains based on their reactivity pattern and has lead to important questions regarding epidemiological studies and rabies control especially with regard to vaccination (Dietzschold *et al.,* 1988).

Nucleoprotein:

Three topographically discrete sites on the nucleoprotein; aa 358-367 (site I), aa 313-337 (site II), aa 374-383 (site III) were identified by Lafon and Wiktor, (1985) (These are listed in Table 1.3). An additional site IV (aa 410-413) was later identified by Ertl *et al.,* (1991). The putative casein-type phosphorylation site (SER^{389}) on the N protein has also been shown to be crucial for viral RNA transcription and replication by encapsidation of genomic RNA (Yang *et al.,* 1999).

Glycoprotein:

Antigenic site III of the glycoprotein has proven to have a great effect on the virulence of various rabies and rabies-related strains (Dietzschold *et al.,* 1983; Dietzschold *et al.,* 1988). Investigation revealed that the change in pathogenicity corresponded to an amino acid substitution at amino acid position 333 of the ectodomain. When the arginine or lysine at position 333 was substituted with glutamine, isoleucine, glycine, methionine or serine; pathogenicity in adult mice was lost (Dietzschold *et al.,* 1983; Seif *et al.,* 1985; Tuffereau *et al.,* 1989). Such substitutions have also shown to affect viral invasiveness into the central nervous system (CNS) (Kucera *et al.,* 1985) and the rate of cell-to-cell spread in cell culture (Dietzschold *et al.,* 1985). From the results it was concluded that this position needs to be filled by a positively charged amino acid. A new study by Sato *et al*., (2008) however indicated that substitutions at position 333, which are not positively charged may still result in a pathogenic virus. They indicated that any positive amino acid in the region aa 319-340 may be sufficient to retain pathogenicity.

In 2001, Badrane *et al.,* reported that the phylogroup I lyssaviruses, which are pathogenic to mice both intracerebrally and intramuscularly, have an arginine residue at position 333 whereas phylogroup II lyssaviruses, thought to be pathogenic only through intracerebral inoculation, have an aspartic acid at this position. It was also recently shown that a single

point mutation to the arginine at position 333, could revert an apathogenic HEP-Flurry strain back to virulence (Takayama-Ito *et al.,* 2006a), confirming that this amino acid is responsible for a change of in vitro neurotropism and the ability of the virus to spread by retrograde axonal transport to propagate in the CNS. However this is not always the case (Yan *et al.,* 2002), suggesting that the involvement of this amino acid in viral spread to the CNS is dependent on the viral strain (Takayama-Ito *et al.,* 2006a). Also included in Table 1.4 are additional amino acids of the glycoprotein which have also been suggested to have pathogenic effects, these were identified by Takayama-Ito *et al.,* (2006b).

Table 1.4 Sites implicated in pathogenesis

Protein	Amino acid	Reference
Nucleoprotein	SER^{389}	Yang *et al.,* 1999
Glycoprotein	ARG^{333} LYS^{330}	Dietzschold *et al.,* 1983; Seif *et al.,* 1985; Tuffereau *et al.,* 1989
	aa 242, 255, 268	Takayama-Ito *et al.,* 2006b

1.10.4 Genomic termini

The 3' and 5' extremities of Rhabdovirus genomes play important roles in virus replication by providing the initiation site of RNA synthesis (Emerson, 1982) and the nucleation site for the initiation of nucleocapsid assembly (Blumberg *et al.,* 1983). The distance from the 3' terminus of the genome to the start of the first gene is remarkably well conserved where as the lengths of the 5' termini vary widely (Harcourt *et al.,* 2001). Despite this variation in length the termini of viruses in the order *Mononegavirales* show high levels of complementarity (Keene *et al.,* 1979; Shioda *et al.,* 1986; Nichol and Holland, 1987; Crowley *et al.,* 1988; Tordo *et al.,* 1988; Morzunov *et al.,* 1995; Harcourt *et al.,* 2001). This complementarity however cannot result in pan handle structure *in vivo* since the RNA is encapsidated as replication proceeds (Chanda and Banerjee, 1979).

Analysis of non-variant genomic extremities of the lyssaviruses available, lead Bourhy *et al.,* 1990, to believe that the complementary 11 nucleotide long signal sequence 3' UGCGAAUUGUU 5' is genus specific. This however does not hold true for ABLV (Warrilow *et al.,* 2002) and EBLV2 (Marston *et al.,* 2007) which depart from this exact

match at position 10 (A→G), indicating that there is a greater degree of flexibility in the terminal sequences than previously thought for this group of viruses. Lyssavirus genomic termini also feature an overproduction of U and A residues. It can be speculated that these conserved nucleotides may have an important role in replication such as signaling encapsidation or replication transcription initiation (Keene *et al.,* 1981; Isaac and Keene, 1982). The fact that the conserved U residues at the 3' end do not overlap with the conserved complementary A residues at the 5' end indicates a requirement for U rich sequences rather than position for functional signals at the termini (Warrilow *et al.,* 2002).

1.11 Investigation of the European bat lyssaviruses

Various studies have shown that DUVV is most closely related to EBLV1. A study by Bourhy *et al.,* (1992) showed that EBLV1 shares more epitopes with Duvenhage virus than with EBLV2. Also notable was the fact that EBLV1 and EBLV2 did not form a monophyletic group (Davis *et al.,* 2005), as EBLV1 and DUVV were more closely related to each other than EBLV1 was to EBLV2.

1.11.1 Percentage identities

Marston *et al.,* (2007) undertook the most comprehensive analysis of EBLV sequence identities, both within genotype (intragenotypic) and between genotypes (intergenotypic). The intragenotypic similarity values for both the nucleotide and amino acid sequence of EBLV1 are given in Table 1.5. At the intergenotypic level, EBLV1 proved to have a higher identity to DUVV (92.7%) than to EBLV2 (87.8%), at the nucleoprotein (aa) (Marston *et al.,* 2007). Analysis of these similarity scores for amino acid in comparison to nucleotide alignments, suggest that the majority of changes at the nucleotide level are silent (synonymous), resulting in no amino acid changes (Marston *et al.,* 2007).

Table 1.5 Percentage identity values for EBLV1

Gene	Amino acid	Nucleotide
N	97.8-100%	95.0-99.9%
P	98.0-99.0%	98.4-99.2%
M	99.0%	99.3%
G	97.1-100%	94.8-99.6%
L	98.6-100%	94.7-99.5%

1.11.2 Analysis of selection pressure

When nucleotide substitutions among different lineages were compared, values for transitions (p) and transversions (q) showed a predominance of transitions. The ratios p/q which range between 1 and 3 for EBLV1 (Amengual *et al.,* 1997) are comparable with values determined for gt1 (Kissi *et al.,* 1995). In a study to determine the forces shaping EBLV evolution, Davis *et al*., (2005) explored whether the selection pressures and rates of evolutionary change observed in these viruses might reflect the peculiarities of their epidemiology in bats. To determine the selection pressures acting on the EBLVs, Davis *et al.,* (2005) determined the number of nonsynonymous (*d*N) and synonymous (*d*S) substitutions per site for individual codons and lineages. For both the N and G genes no evidence was found for positive selection (*d*N/*d*S > 1). The mean d_N/d_S values were low in all cases, revealing strong selective constraints. In particular, the non synonymous (*d*N)/synonymous (*dS*) values estimated for EBLV are relatively low for RNA viruses (Woelk & Holmes. 2002) even in comparison to the G gene of genotype 1 lyssaviruses in canines (Holmes *et al.,* 2002). The mean *d*N/*d*S value for the G gene was 0.088, which was nearly twice that of which was observed for the N gene 0.049 (Davis *et al.,* 2005). The stronger purifying selection against the N gene was expected, as envelope glycoproteins interact with host cell receptors and are the main targets of the immune response.

1.11.3 Rates of nucleotide substitution

The mean rates of substitution for EBLV1s N and G genes are very similar, with 6.11×10^{-5} and 5.10×10^{-5} substitutions per site per year, respectively (Davis *et al.,* 2005). These are among the lowest measurable nucleotide substitutions rates reported for RNA viruses (Jenkins *et al.,* 2002 and Hanada *et al.,* 2004). Davis *et al.,* (2005) proposed two possible explanations for the low evolutionary rate of EBLV1; a) slow replication in the bat host may allow for fewer mutational errors per unit time and b) peculiarities of the bat immune system may have altered the selection pressures faced by EBLV1. It was however concluded that the virus may have reached an adaptive peak, so that most amino acid changes reduce fitness and are therefore removed by purifying selection (Davis *et al.,* 2005).

1.12 Lyssaviruses infection in bats

The large roost sizes and high densities of many bat species make them well suited to the sustained transmission and exchange of RNA viruses (Mackenzie *et al.,* 2003), most likely through the transfer of infectious saliva during licking and biting (Ghatak *et al.,* 2000). A study by Hughes *et al.,* (2005) found that after the introduction of RABV into North American bats, there was rapid adaptation to new host species. It seemed that the biology of colonial bats (higher densities) ensured a greater number of replication cycles with time, resulting in speedier adaptation than in solitary bat species. Amengual *et al.,* (2007) showed EBLV1 infection in *M. myotis* to be characterized by high bat immunity after circularization of the virus. The high percentage of seropositive bats indicated efficient virus transmission between individuals and rapid circularization of the virus within the colony. This is not surprising as *M. myotis* are social bats, with high contact rates between individuals.

A high prevalence of virus neutralizing antibodies (VNA) against EBLV1 has been reported from colonies of insectivorous bats from Spain (Serra-Cobo *et al.,* 2002; Amengual *et al.,* 2007) and based on field observations, O'Shea *et al*., (2003) suggested that bats might acquire immunity through exposure to low doses of virus that do not result in a productive infection. In 2008, Franka *et al*. undertook a study to determine the susceptibility of insectivorous bats to infection with EBLV1, to assess the dynamics of host immune responses and to evaluate the opportunity for horizontal viral transmission within colonies. Their observations suggested that exposure to varying doses of EBLV-1 from rabid conspecifics within a colony via natural routes (frequent biting and scratching resulting from colonial behaviour and grooming) could lead to an abortive infection and serve as a natural mode of immunization resulting in the presence of VNA in free ranging bats (Franka *et al*., 2008). Stress, malnutrition, an immature immune system and immunosuppression could however enable productive infection and circulation of virus within a population with certain levels of herd immunity (Franka *et al*., 2008). This study represented the first experimental proof of the natural immunization hypothesis.

1.13 Host interaction

In 1997, Amengual *et al.,* undertook an investigation to determine the evolution of EBLV, the study included numerous EBLV isolates as well as 2 DUVV isolates. It was found that within EBLV1 and EBLV2, two lineages (a and b) could be differentiated by their nucleotide and amino acid sequences. EBLV1a and EBLV1b are the most frequently reported and are widely distributed. EBLV1a exhibited a west-east distribution whereas that of EBLV1b is north-south, indicating that EBLV1 isolates have evolved into at least two genetically distinguishable groups, following geographical drifting. From the data it was speculated that these two groups were introduced into northern Europe from two different geographic directions, no hypothetical introduction point could be found for EBLV1a, EBLV1b's results however suggested it to have come from northern Africa via the south of Spain (Amengual *et al.,* 1997; Davis *et al.,* 2005). Due to the infrequency of EBLV2 identification no conclusions regarding its geographical range could be made (Amengual *et al.,* 1997).

In EBLV1a, the phylogenetic homogeneity of isolates across geographic regions suggests that there is an established viral traffic among bat populations in northern Europe (Davis *et al.,* 2005). This process however, cannot be explained by the behavior of *Eptesicus serotinus* bats, which are not migratory (Corbet, 1991) and suffer mortality from EBLV infection. It has thus been suggested that the long distance transmission is facilitated by migratory species that roost with *E. serotinus* (Davis *et al.,* 2005). For EBLV1b data indicates that there has been less contact between bat populations from diverse regions in Europe (Davis *et al.,* 2005). This may in part be due to a decline in bat populations, for reasons that include human intervention, which leads to a reduction in contact and the enhancement of viral population subdivision (Davis *et al.,* 2005). Considering the distribution of *Tadarida teniotis* and *Miniopterus schreibersii* in southern Europe and northern Africa, Sierra-Cobo *et al.,* (2002) proposed that they may have contributed to the dispersion of EBLV1 into southern Europe. This is in agreement with the possible African origin suggested by Amengual *et al.,* (1997).

When investigating the host species involved with DUVV, EBLV1 and Irkut infection, significant overlaps can be seen with regard to geographic distribution and co-colonization in roosts (Table 1.6). *M. myotis* and *M. schreibersii* for example have been found to move between colonies and are also known to have direct contact with each other in these mixed colonies (Serra-Cobo *et al.,* 2002).

Table 1.6 Host species of DUVV, EBLV1 and Irkut

Virus	Host species	Distribution	Migration	Roosts with	References
DUVV	*Miniopterus schreibersii* (only implicated)	Southern Europe to Japan, Africa North and eastern Australia	Seasonal	*Myotis myotis* *Myotis blythii* *Rhinolophus ferrumequinum*	Nowak. 1999 Van der Merwe. 1982 Karatas *et al.,* 2003 Presetnik, 2004 Serra-Cobo *et al.,* 2002
	Nycteris thebaica	Throughout Africa, Southern Europe	Seasonal	*Miniopterus schreibersii* *Myotis tricolor* *Rhinolophus blassi* *Rhinolophus simulator*	Nowak. 1999 Gray *et al.,* 1999
EBLV 1	*Eptesicus serotinus*	Central, western and southern Europe	No	*Tadarida teniotis* *Myotis blythii* *Myotis daubentonii*	Serra-Cobo *et al.,* 2002
	Myotis myotis	Southern, central and northern Europe, North Africa, Asia	Seasonal	*Miniopterus schreibersii* *Murina leucogaster* *Rhinolophus ferrumequinum*	Karatas *et al.,* 2003 Presetnik. 2004 Ma *et al.,* 2003 Serra-Cobo *et al.,* 2002
	Miniopterus schreibersii (unconfirmed)	As described above			
	Tadarida teniotis (unconfirmed)	Southern Europe, Afghanistan, North east India and Thailand, Korea, Japan, North Africa	Seasonal	*Eptesicus serotinus* *Myotis blythii* *Myotis daubentonii*	Corbet, 1992 Serra-Cobo *et al.,* 2002
	Rhinolophus ferrumquinum (unconfirmed)	Northern India, South to north-western Africa, Eurasia	Seasonal	*Miniopterus schreibersii* *Myotis myotis*	Koopman. 1994 Karatas *et al.,* 2003 Presetnik. 2004 Ma *et al.,* 2003 Serra-Cobo *et al.,* 2002
Irkut	*Murina leucogaster*	India, Mongolia, China, Korea, Japan	Seasonal	*Myotis myotis* *Rhinolophus ferrumequinum*	Ma *et al.,* 2003

1.14 Aims of the study

With the exception of the recent study by Delmas *et al.,* 2008, which gives a brief full genome overview, the analysis of Duvenhage virus has previously focused on the study of the N, P and G genes. These analyses have also primarily focused on only two isolates; DUVVSA71 and DUVVSA81. With such limited knowledge of DUVV it is of great importance to investigate as far as is possible, the relationship between all existing DUVV isolates, so as to broaden our understanding of this African lyssavirus and its position within the *Lyssavirus* genus.

Specific objectives

1. To sequence and phylogenetically analyze the nucleoprotein, phosphoprotein, matrix protein and glycoprotein of the 1971, 1981, 1986 and 2006 DUVV isolates.

2. To generate the full length sequence of the 2006 South African DUVV isolate

3. To compare the newly generated sequence with all available lyssavirus full genomes to identify distinct characteristics and to determine the diversity compared to other lyssaviruses using complete genomes.

CHAPTER 2

Molecular epidemiology of Duvenhage virus

2.1 Introduction

Duvenhage virus (DUVV), a member of the *Lyssavirus* genus has a negative sense, single stranded RNA genome that codes for a nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and RNA polymerase (L) (Tordo and Poch, 1988). Comparison of the N and G genes of lyssaviruses allowed for the grouping of the *Lyssavirus* genus into seven genotypes and four putative genotypes. Gt 1 (RABV), gt 2 (LBV), gt 3 (MOKV), gt 4 (DUVV), gt 5 (EBLV1), gt 6 (EBLV2) and gt 7 (ABLV) constitute the seven lyssavirus genotypes (Tordo *et al*., 2005) and Irkut, Aravan, Khujand and West Caucasian bat virus (WCBV) the putative lyssavirus genotypes (Kuzmin *et al*., 2005). Currently the criteria suggested for classification of a new lyssavirus genotype are based on the assumption that isolates sharing less than 80% nucleotide and 92% amino acid similarity belong to different genotypes (Kissi *et al*., 1995). Based on phylogeny, pathogenicity and serological cross reactivity, the Lyssavirus genotypes have been split into three phylogroups. Phylogroup I consists of RABV, DUVV, EBLV1, EBLV2 and ABLV as well as the putative species Aravan, Khujand and Irkut; Phylogroup II, MOKV and LBV and Phylogroup III, WCBV (Badrane *et al*., 2001; Kuzmin *et al.,* 2003; Kuzmin *et al.,* 2005)..

There have been only five isolations of DUVV, all of which were from the African continent. Three of these cases led to human fatalities (Van der Merwe, 1982; Paweska *et al*., 2006; van Thiel *et al*., 2008) and all have been linked to insectivorous bats. *Nycteris thebaica* is the only confirmed host but *Miniopterus schreibersii* has also been implicated. DUVV and EBLV1 have shown great similarity to each other at both phylogenetic and antigenic levels. Although exclusive to Europe, EBLV1 is also associated with insectivorous bat species, many of which roost with *N. thebaica* and *M. schreibersii* (Serra-Cobo *et al.,* 2002; Presetnik, 2004: Karatas *et al.,* 2003). More recently the putative genotype Irkut was shown to have a close relationship with both DUVV and EBLV1 (Botvinkin *et al*., 2003). Several molecular epidemiological studies of RABV have been performed and only a few on ABLV (Guyatt *et al*., 2003), EBLV1 and EBLV2 (Amengual *et al*., 1997; Davis *et al*., 2005). Evolutionary studies also have focused on RABV, EBLV1, EBLV2 and ABLV. There has been only one molecular study to focus on the African lyssaviruses LBV and DUVV (Markotter et al., 2008b), whilst there have been a few to focus on MOKV (Nel *et al*., 2000; Sabeta *et al*., 2007a; Markotter et al., 2008b).

Thus little is known about the molecular epidemiology and evolution of DUVV and as such the objectives of this study were; 1) to determine the relationship between the DUVV isolates, as well as between DUVV, EBLV1 and Irkut viruses through comparison of full length N, P, M and G gene sequences; 2) to investigate which genes would be best suited to genotype classification with regards to DUVV, EBLV1 and Irkut and 3) to determine and compare the selective constraints acting on both DUVV and EBLV1.

(This chapter was concluded before the release of Delmas *et al.,* 2008 and those sequences were not included in this study)

2.2 Materials and methods

2.2.1 Viral isolates

The DUVVSA71, DUVVSA81 and DUVVZIM86 lyophilized mouse brain isolates were obtained from Dr C.T Sabeta of the Rabies Section, Agricultural Research Council – Onderstepoort Veterinary Institute (ARC-OVI), South Africa. The DUVVSA06 brain isolate was obtained from Dr. Janusz Paweska, Special Pathogens Unit, National Institute for Communicable Diseases, National Health Laboratory Services, South Africa. The DUVVkenya isolate was obtained from Dr. M. Schutten, Department of Virology, Erasmus Medical Centre, Rotterdam, the Netherlands (See Table 2.1).

Table 2.1 Duvenhage isolates detail

	DUVVSA71	**DUVVSA81**	**DUVVSA06**	**DUVVZIM86**	**DUVVkenya**
Year of isolation	1971	1981	2006	1986	2007
Species	Human	Unconfirmed	Human	*Nycteris thebaica*	Human
Geographic location	Bela-Bela (previously warmbaths), Limpopo province, South Africa	Louis Trichardt, Limpopo province, South Africa	Pilansberg, North West province, South Africa	Bulawayo, Zimbabwe	Kenya
Reference	Meredith *et al.,* 1971	Van der Merwe, 1982	Paweska *et al.,* 2006	Foggin, 1988	Van Thiel *et al.,* 2008

Virus isolates were amplified in suckling mice brains. Lyophilized brain material was reconstituted in sterile phosphate buffered saline (PBS) (13.7 mM NaCl, 0.27 mM KCl, 0.43 mM $Na_2HPO_4.2H_2O$, 0.14 mM KH_4PO_4, pH 7.3). Two to three day old suckling mice received 30 µl of the reconstituted material intracranially (Koprowski, 1996). Animals were monitored and collected upon death where the brain material from the dead animals was removed aseptically. The direct fluorescent antibody test was used for post-mortem diagnosis of lyssavirus infection. The standard operational procedure as indicated at (www.cdc.gov/ncidod/dvrd/Rabies/Professional/Publications/DFA_diagnosis) was followed. A polyclonal fluorescein isothiocyanate conjugated immunoglobulin (Onderstepoort Veterinary Institute, Rabies Unit, South Africa) that is capable of detecting all lyssavirus genotypes was used at a 1:20 dilution. Brain material which tested positive was pooled and used for RNA extraction.

2.2.2 RNA extraction

RNA was prepared using Trizol reagent (Invitrogen) according to the manufacturer's protocol. Briefly; 50-100 mg brain tissue was homogenized in 1 ml Trizol reagent and incubated at room temperature (about 23˚C) for 5 minutes. 0.2 ml chloroform was added, shaken vigorously for 15 seconds and incubated at room temperature for 3 minutes. The preparation was then centrifuged at 10 000 g for 15 minutes. The aqueous phase was then transferred to a fresh microcentrifuge tube and the RNA was precipitated by the addition of 0.5 ml isopropyl alcohol. The precipitate was collected after incubation at room temperature for 10 minutes by centrifugation at 10 000 g for 10 minutes. The supernatant was removed and the pellet washed twice with 1ml 75% ethanol, followed by vortexing and centrifugation at 7 500 g for 5 minutes. The supernatant was removed and the pellet left to dry. The RNA was resuspended in 50 µl diethyl-pyrocarbonate (DEPC) H_2O and stored at -20˚C.

2.2.3 Primers

Primers were designed based on DUVV and other lyssavirus sequences available in the public domain: on GenBank (www.ncbi.nlm.nih.gov), through alignment using Clustal W multiple alignment program (Thompson, *et al.,* 1994) (Table 2.2). Genome position is based on the DUVVSA06 full genome obtained in this study (EU623444).

Table 2.2 Primers used to amplify the N, P, M and G genes

Primer	Sequence	Genomic position	Gene targeted	Reference	Use
Lys001	5' ACGCTTAACGAMAAA 3'	3' non coding region (-70 to -57)	N	Markotter. *et al.,* 2006a	cDNA, PCR, sequencing
Lys304	5' TTGACAAAGATCTTGCTCAT 3'	1447-1466 rc	N	Markotter. *et al.,* 2006a	PCR, sequencing
DGF	5' CCTCAAGGAGTTCAAGCGCC 3'	3207-3226	G	This study	cDNA, PCR, sequencing
DFR	5' GGCCTCTCACTCCCTTGTTG 3'	4815-4834 rc	G	This study	PCR, sequencing
DuvN+	5' GGATCATGATGAACGGAG 3'	1223-1240	M and P	This study	cDNA, PCR, sequencing
DuvG-	5' GGCCCCAATTTGTCAGGG 3'	3304-3321 rc	M and P	This study	PCR, sequencing
DuvG1+	5'GAAGGAACCACAGGAGATGTTCG 3'	4778-4800	G-L intergenic region	This study	cDNA, PCR, sequencing
DuvL-	5' GTTGAGATTGTAGTCAGAGTTCC 3'	5485-5507 rc	G-L intergenic region	This study	PCR, sequencing

*rc = reverse complement

2.2.4 Reverse transcription

First strand cDNA synthesis was achieved by denaturing 10 µl RNA and 20 pmol of the positive sense PCR primer (For each respective PCR), at 70˚C for 5 minutes. The reaction mixtures where then cooled on ice for 2 minutes, following this 10 mM dNTP mix (10 mM), 4 µl 5x buffer (250 mM Tris-HCL, 40 mM $MgCl_2$, 150 mM KCl, 5 mM dithiothreithol) (Roche), 20 U Rnasin ribonuclease inhibitor (Promega, 20 U/µl) and 1 µl AMV (Avian Myeloblastosis Virus) (Roche Diagnostics, 20 U/µl) were added to each reaction. The reaction mix was then heated to 25˚C for 10 minutes, 42˚C for 60 minutes and 85˚C for 5 minutes.

2.2.5 Polymerase chain reaction

PCR reactions were prepared to a final volume of 50 µl. Each reaction contained 1.5 mM $MgCl_2$, 800 µM dNTPs (mixture), 5 µl 10x reaction buffer (50 mM KCl, 10 mM Tris-HCL, 0.1% Triton X-100) (Celtic Molecular Diagnostics), 20 pmol of each primer, 0.25 U Bioline Taq (Celtic Molecular Diagnostics, 5U/µl) and 5 µl template cDNA were added to each reaction. The tubes were placed in a GeneAmp thermocycler (Model 2400; PE Applied Biosystems) and the following cycling conditions were used:

1 cycle of 94˚C for 2 minutes, 30 cycles of: 94˚C for 30 seconds; 37˚C for 30 seconds; 72˚C for 90 seconds and a final elongation step of 72˚C for 7 minutes

2.2.6 Agarose gel electrophoresis

The PCR amplicons were analyzed on 0.8% (w/v) agarose gels, prepared in 1x sodium boric acid electrophoresis buffer (5 mM disodium borate decahydrate, adjusted to pH 8.5 with boric acid). The PCR amplicons were resolved against a 100 basepair molecular weight marker (Promega). 5 µl of the samples were loaded in loading dye (40% sucrose, 0.25% bromophenol blue). The gels were run at 120 V in a horizontal gel tank system using a Biorad Wide Mini Sub ™ electrophoresis cell. The gel was then stained in a 0.5 µg/ml ethidium bromide solution and the bands visualized using a UV transilluminator.

2.2.7 Purification of PCR amplicons

After gel electrophoresis, the correctly sized amplicons were cut out of the agarose gels and purified using the commercial Wizard® SV Gel and PCR Clean-Up System (Promega). The products were purified according to the manufacturer's suggestions, as follows: Membrane binding solution (4.5 M $C_2H_6N_4S$, 0.5 M CH_3COOK) was added to the gel slice at a ratio of 10 µl of solution to 10 mg of agarose gel slice and was then incubated at 65ºC for 10 minutes. The sample was applied to a SV minicolumn and centrifuged at 13 400 g for 60 seconds in a minispin® (Eppendorf), after which the flow through was discarded. The spin column was subsequently washed with 700 µl membrane wash solution (10 mM CH_3COOK, 16.7 µM EDTA, 80% ethanol) and centrifuged for 60 seconds. A second wash step was performed using 500 µl membrane wash solution and centrifugation of 5 minutes. The DNA was then eluted in 30-50 µl nuclease free H_2O. The purified DNA product was stored at -20˚C. Concentration of the purified product was determined by electrophoresis of 1 µl of the final product on an agarose gel, using a 100 bp DNA ladder (Promega) as reference.

2.2.8 Nucleotide sequencing

Sequencing of cloned insert DNA and PCR products was performed using an ABI PRISM® Big Dye® Terminator V3.1 Kit (Applied Biosystems). Reactions were prepared as follows, according the manufacturer's suggestions: Each reaction contained 3.2 pmol primer, 2 µl Terminator mix v3.1 (2.5X) (Applied Biosystems), 1 µl Sequencing buffer (5X) (Applied Biosystems) and 10 ng/100 bp template, made up to a final volume of 10 µl with nuclease free H_2O.

The reactions were cycled in an automated thermocycler as follows:

1 cycle of 94˚C for 1 minute, 25 cycles of: 94˚C for 10 seconds; 50˚C for 5 seconds and a final cycle of 60˚C for 4 minutes.

Reactions were stored at -20˚C before precipitation using the EDTA/NaOAc/EtOH method.

The EDTA/NaOAc/EtOH method according to the BigDye Terminator v3.1 cycle sequencing protocol (Applied Biosystems, 2002) is as follows: for each 10 µl reaction; 1 µl 125 mM EDTA, 1 µl 3 M sodium acetate and 25 µl of 100% ethanol were added. The tubes were then vortexed and incubated at room temperature for 15 minutes. The tubes were subsequently spun at maximum speed for 30 minutes and the supernatant removed, 100 µl 70% ethanol was added and the tubes centrifuged at maximum speed for 15 minutes and the supernatant removed. Next the DNA pellets were air dried at room temperature for 20 minutes and stored at -20˚C. The precipitated reactions were submitted to the sequencing facility of the Faculty of Natural and Agricultural Sciences, University of Pretoria and analysed on an ABI 3100 automated capillary sequencer analyzer.

2.2.9 Phylogenetic analysis

Obtained sequences were assembled using the VectorNTI 9.1.0 software package (Invitrogen); and trimmed using the Bioedit software package (Hall, 1999). Alignments were then carried out using the ClustalW subroutine (Thompson, *et al.,* 1994), which forms part of the Bioedit program. The calculation of genetic distances and construction of phylogenetic trees based on nucleotide sequence was carried out using MEGA 3.1 software (Kumar, *et al.,* 2004). Genetic distances were calculated between pairs of sequences by using the Kimura's 2-parameter method (Kimura, 1980), and based on these distances neighbour-joining (NJ) trees were constructed using the methods of Saitou and Nei, (1987). The NJ method of tree construction was chosen as it is rapid, with branch lengths being proportional to the amount of genetic change between lineages. The branching order of the trees was evaluated by using bootstrap analysis of 1 000 pseudoreplicate datasets. Results were validated by maximum parsimony as implemented in MEGA 3.1.

2.2.10 Analysis of sequences

Obtained sequences were assembled using the VectorNTI 9.1.0 software package (Invitrogen), hereafter; they were cleaved and sized using the Bioedit software package. Alignments were then carried out using the ClustalW subroutine (Thompson, *et al.,* 1994), which forms part of the Bioedit program.

Sequence similarity between isolates was determined using the distance estimation program of MEGA 3.1 (Kumar, *et al.,* 2004). Genetic distances were calculated for both the nucleotide and deduced amino acid sequences of the N, P, M and G genes using the p-distance model (Nei and Gojobori, 1986).

2.2.11 Analysis of selection pressure and nucleotide substitution patterns

The selection pressures acting on both DUVV and EBLV1 isolates were determined using the codon based Z-test (Mega 3.1), employing the Nei-Gojobori (p-distance) model (Nei and Gojobori, 1986). The variance of non-synonymous (altering) substitutions (*d*N) versus synonymous (silent) substitutions (*d*S) was computed using bootstrap resampling of 500. Three hypotheses were considered; the neutrality hypothesis (*d*N-*d*S), the positive selection hypothesis (*d*N>*d*S) and the negative (purifying) selection hypothesis (*d*N<*d*S). Hypotheses were rejected when values obtained were <0.05.

2.3 Results

2.3.1 cDNA synthesis and PCR of the N, P, M and G genes

Following propagation of the DUVV isolates in suckling mice, total RNA was extracted from the brain material and used in reverse transcriptase mediated PCR amplification of full length cDNA copies of the N, P, M and G genes. PCRs were then performed as described in section 2.2.5. Virus specific products were yielded for all isolates with the exception of DUVVZIM86.

Various attempts were made at amplifying the DUVVZIM86 isolate;

- The first brain material sample obtained from the ARC-OVI, despite all attempts was not amplified.
- Second and third samples from the ARC-OVI were then taken and successfully amplified, the products however proved to be LBV. This contamination could not be traced and thus all samples were considered redundant.

Hereafter contact was made with individuals and organizations which may have had this viral isolate in storage;

- Dr. J. Paweska and Prof. R. Swanepoel of the Special Pathogens Unit, National Institute for Communicable Diseases, National Health Laboratory Services, South Africa.
- Dr. C. Foggin, whom had originally isolated the virus in 1986. Enquiries were made at the Central Veterinary Laboratory in Zimbabwe where the original work was done.

Neither had any viral samples nor could they advise on any potential sources, consequently the DUVVZIM86 isolate was excluded from this study.

2.3.2 Purification of PCR amplicons and nucleotide sequence determination

The N gene of all three isolates was found to be 1356 nucleotides in length (451 aa) with an average GC content of 43.8%. The P gene was 897 nucleotides (298 aa) with an average GC content of 46.8%. With an average GC content of 44.8%, the M gene of all 3 isolates was found to be 609 (202 aa) and the G gene, 1602 nucleotides (533 aa) with an average GC content of 45%.

2.3.3 Sequence and phylogenetic analysis of the five Duvenhage virus isolates

Phylogenetic trees including all five DUVV isolates were constructed, using a 398 bp fragment of the nucleoprotein gene (nt 8-406), as this was the only sequence available for all isolates. The NJ method indicated low bootstrap support (67% and lower) for all major clusters representing gts 4, 5 and 6 as well as the putative genotypes (Figure 2.1). Genotype 5 split into lineages EBLV1a and EBLV1b; genotype 6 into EBLV2a and EBLV2b; and genotype 4 into lineage A (isolates from sub-Saharan Africa) and lineage B (isolate from Kenya). These groupings were supported by high bootstrap values (95% and higher). MP phylogenetic analysis also indicated the major clusters representing the different lyssavirus genotypes as well as the distinct lineages (results not shown).

In this study nucleotide identity was determined (Appendix A) using the p-distance model (Nei and Gojobori, 1986). Analysis of genetic distances between all five DUVV isolates was carried out using a well conserved 398 nt sequence from the nucleoprotein gene (nt 8-406). The intrinsic variation between DUVV isolates from southern Africa was low with a 97.7-100% nt identity, even though these isolates were isolated several years apart (1971-2006). In fact, isolate DUVVZIM86 was found to be 100% identical DUVVSA81 with respect to this part of the sequence although they were isolated 5 years apart in different countries. The east African isolate, DUVVKenya had much lower sequence identity (88.9-89.7%) to the other DUVV isolates, which supports the phylogenetic analysis that suggested this isolate to form part of a different lineage. DUVVKenya was shown to be most similar to DUVVSA71 (89.7%), the original DUVV isolate from South Africa. From this short sequence analysis, it was found that DUVV is more closely related to Irkut (77.4-78.1%) than to EBLV1 (75.1-77.9%).

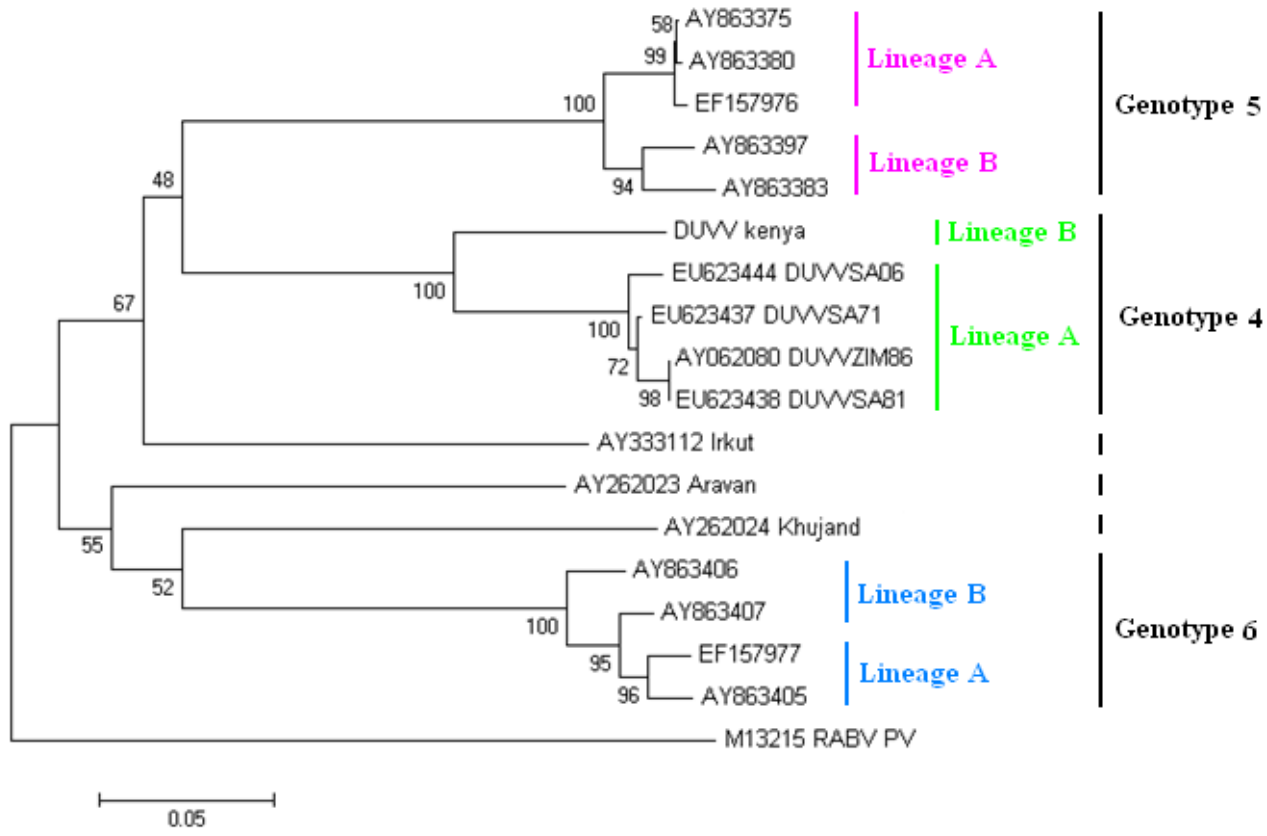


Figure 2.1 Neighbour joining tree of nt 8-406 of the nucleoprotein gene, including all DUVV isolates to date. (RABV was used as the outgroup). GenBank accession numbers are indicated for each isolate. Bootstrap values are indicated at the nodes and branch lengths are drawn to scale.

2.3.4 Phylogenetic analysis

In this study, phylogenetic trees were constructed using full length N, P, M and G gene nucleotide and deduced amino acid sequences. Both neighbour-joining (NJ) and maximum parsimony (MP) methods were employed.

2.3.4.1 Nucleoprotein

A set of 19 complete N gene sequences of phylogroup I lyssaviruses, consisting of representatives from genotypes 1, 4, 5, 6 and 7 and the putative genotypes Irkut, Aravan and Khujand were analysed in this study. The NJ method indicated low bootstrap support (67% and lower) for all major clusters representing gts 4, 5 and 6 as well as the putative genotypes (Figure 2.2a). Phylogenetic analysis for the deduced amino acids demonstrated the same tree topology (Figure 2.2b) for all main clusters, with the exception of Aravan and Irkut viruses.

The position of Aravan virus is unstable due to its equally moderate homology with Khujand virus and with the clade joining gts 4 and 5 (Kuzmin *et al.*, 2003). MP analysis also indicated the major clusters representing the different lyssavirus genotypes (results not shown).

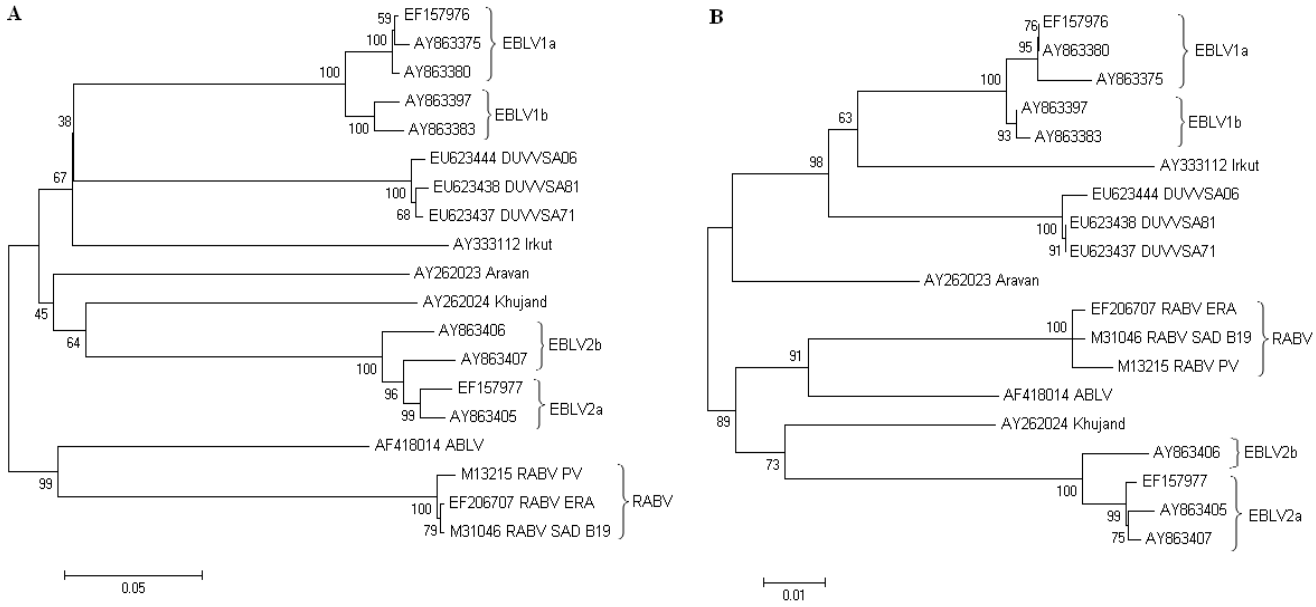


Figure 2.2 Neighbour joining tree of the full nucleoprotein A) nucleotide and B) amino acid sequence of phylogroup I lyssavirus representatives. GenBank accession numbers are indicated for each isolate. Bootstrap values are indicated at the nodes and branch lengths are drawn to scale.

2.3.4.2 Phosphoprotein

Neighbour joining phylogenetic trees based on the full length gene were constructed for both the nucleotide (Figure 2.3a) and deduced amino acid sequences (Figure 2.3b) of the phosphoprotein. A set of 15 complete P gene sequences of phylogroup I lyssaviruses was generated, consisting of representatives from genotypes 1, 4, 5, 6 and 7 and the putative genotypes Irkut, Aravan and Khujand. Tree topologies generated corresponded with the different lyssavirus genotypes as obtained in the N gene analysis. Bootstrap support for the EBLV1 and EBLV2 clusters were low in the nucleotide analysis (56% and 54%)) and high in the aa analysis (92% and 93%).

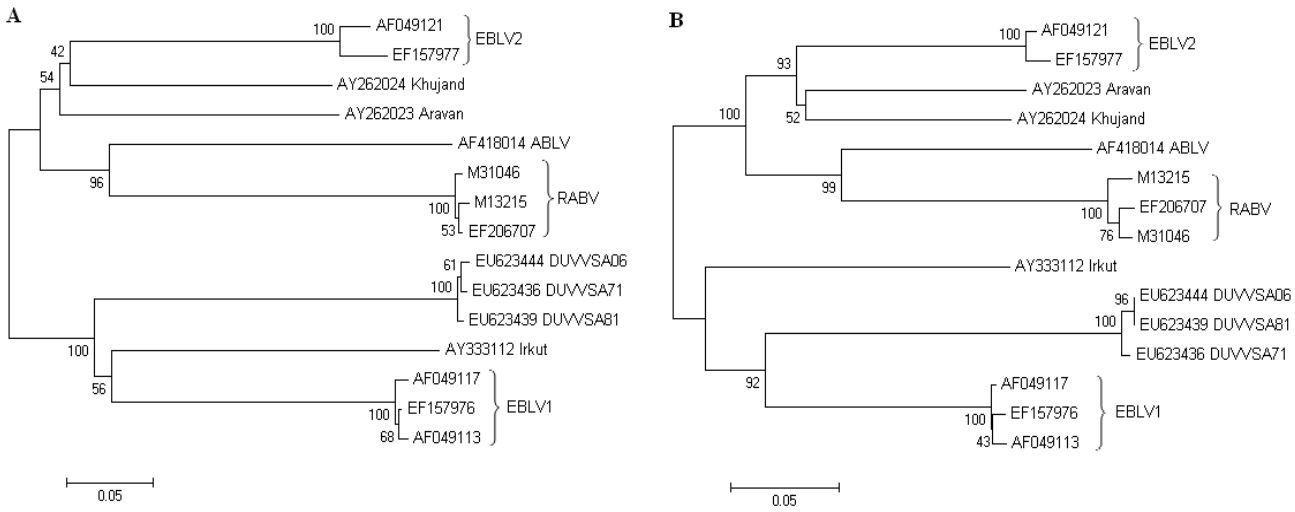


Figure 2.3 Phylogroup I representatives, NJ tree of the full length phosphoprotein A) nucleotide and B) amino acid sequences. GenBank accession numbers are indicated for each isolate. Bootstrap values are indicated at the nodes and branch lengths are drawn to scale.

2.3.4.3 Matrix protein

Neighbour joining trees deduced from the matrix protein nucleotide (Figure 2.4a) and deduced amino acid (Figure 2.4b) sequences were based on a set of 12 M genes representing the phylogroup I lyssaviruses and the putative lyssavirus genotypes. At both the nucleotide (45% and lower) and amino acid levels (45% and higher), gts 4, 5, 6, 7, Irkut, Aravan and Khujand virus clustered together with no clear distinction between them. The same tree topology was observed by MP analysis of the nt and deduced aa M gene sequences (results not shown).

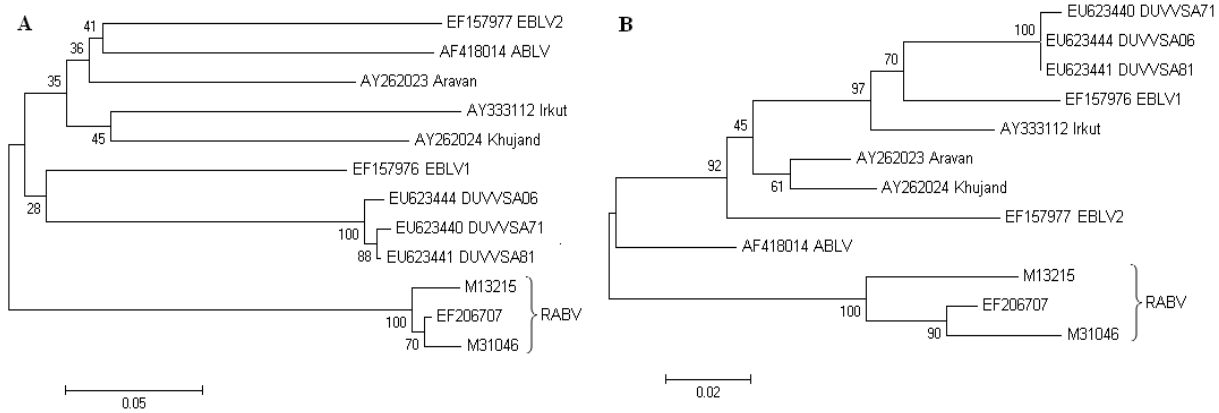


Figure 2.4 Neighbour joining tree based on full length matrix protein A) nucleotide and B) amino acid sequences for various phylogroup I representatives. GenBank accession numbers are indicated for each isolate. Bootstrap values are indicated at the nodes and branch lengths are drawn to scale.

2.3.4.4 Glycoprotein

Phylogenetic analysis of the G gene was based on a set of 20 complete G gene sequences of phylogroup I lyssaviruses, consisting of representatives from genotypes 1, 4, 5, 6 and 7 and the putative genotypes Irkut, Aravan and Khujand. Nucleotide (Figure 2.5a) and deduced amino acid (Figure 2.5b) analysis indicated the same clusters, representing the lyssavirus genotypes as described in previous studies (Kuzmin *et al.,* 2003; Kuzmin *et al.,* 2005). The nucleotide based tree however had very weak support (bootstrap 39%) for the separation of Aravan from the cluster containing EBLV1, DUVV and Irkut. Similar grouping was seen in the MP phylogenetic analysis (results not shown).

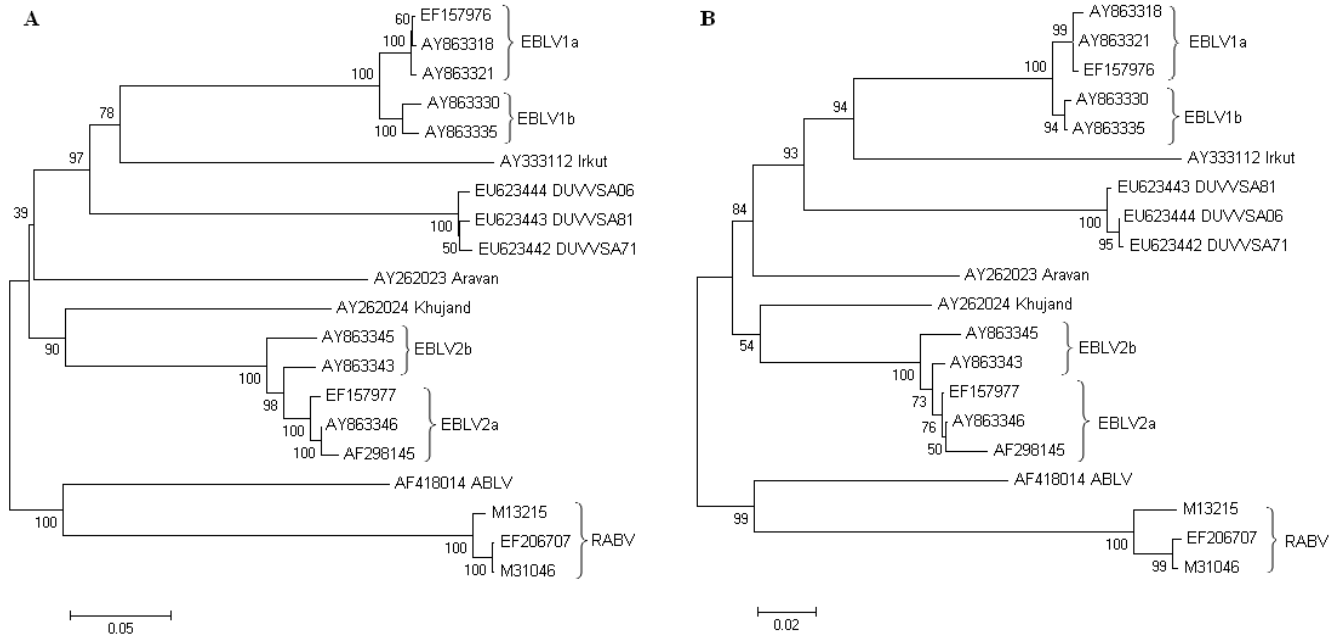


Figure 2.5 Neighbour joining tree of the full glycoprotein A) Nucleotide and B) Amino acid sequences of various phylogroup I lyssaviruses. GenBank accession numbers are indicated for each isolate. Bootstrap values are indicated at the nodes and branch lengths are drawn to scale.

2.3.5 Sequence analysis

Nucleotide and amino acid identities for the complete N, P, M and G genes of the phylogroup I and putative lyssavirus genotypes were determined (Appendix A) using the p-distance model (Nei and Gojobori, 1986) in the distance estimation program of MEGA 3.1 (Kumar *et al.,* 2004).

2.3.5.1 Intergenotypic identities

Intergenotypic identity analysis was split into two sections; A) which included the putative genotypes Aravan, Irkut and Khujand and B) which excluded the putative genotypes (Table 2.3). This was done to evaluate the influence of these viruses on percentage identity outcomes, especially with regard to the current classification criteria, where it is assumed that isolates sharing more than 80% nucleotide and 92% amino acid similarity at the nucleoprotein level would belong to the same genotype (Kissi *et al.,* 1995). In both analyses the classification criteria were only once exceeded; both times at the N gene amino acid level where DUVV and EBLV1 showed a percentage similarity of 93.3%. This was however the only gene to have yielded consistent results in both analyses. The glycoprotein was the only other gene to give the same result at both the nucleotide and amino acid levels, even though different genotypes were represented in the different analyses [(A) EBLV2 and Khujand; (B) EBLV1 and EBLV2]. Between genes results were varied and in analysis A only EBLV2 and Khujand (P aa, G aa, G nt) and Aravan and Khujand (P nt, M aa) shared highest identity in more than one gene. In analysis B, DUVV and EBLV1 (N nt, N aa, P nt, M nt) shared highest identity in three genes, making this the most frequent grouping. Results both within and between genes was equally variable, whether or not the putative genotypes were included in the analysis.

Full length analysis of the nucleoprotein amino acid sequence demonstrated high percentage identity values, many crossing the 92% intergenotypic threshold (Kissi *et al*., 1995) (Table 2.4).

2.3.5.2 Intragenotypic identity

Intragenotypic identity values for the N, P, M and G genes of DUVV are given in Table 2.5. As was described in the study by Wu *et al.,* (2007), the percentage identity order for RABV was found to be N> M> P> G.

Table 2.3 Highest intergenotypic identities. A) Includes the putative genotypes Aravan, Khujand and Irkut. B) Excludes the putative genotypes. Shaded cells indicate where the genotype classification threshold has been exceeded.

Gene		Most similar genotypes	Number of isolates	Percentage similarity	
Nucleoprotein	**nt**	DUVV & EBLV1	8	79.8%	A
	aa	DUVV & EBLV1	8	93.3%	
Phosphoprotein	**nt**	Aravan & Khujand	2	74.5%	
	aa	EBLV2 & Khujand	2	78.8%	
Matrix protein	**nt**	Aravan & ABLV	2	81.1%	
	aa	Aravan & Khujand	2	96.5%	
Glycoprotein	**nt**	EBLV2 & Khujand	5	78.9%	
	aa	EBLV2 & Khujand	5	87.4%	
Nucleoprotein	**nt**	DUVV & EBLV1		79.8%	B
	aa	DUVV & EBLV1		93.3%	
Phosphoprotein	**nt**	DUVV & EBLV1		70.6%	
	aa	ABLV & RABV		75.1%	
Matrix protein	**nt**	DUVV & EBLV1		80.3%	
	aa	ABLV & EBLV2		88.6%	
Glycoprotein	**nt**	EBLV1 & EBLV2		73.6%	
	aa	EBLV1 & EBLV2		80.8%	

Table 2.4 Intergenotypic percentage identity of the N protein at the amino acid level

Genotypes	Average identity	Identity range
DUVV – EBLV1	92.7%	91.6-93.3%
Irkut – EBLV1	92.2%	91.3-92.7%
Aravan - Khujand	92.7%	-----
ABLV - Khujand	92.2%	-----
ABLV - Aravan	92.4%	-----
ABLV- RABV	92.2%	92-92.4%

Table 2.5 Intragenotypic percentage identity values for DUVV isolates DUVVSA06, DUVVSA81 and DUVVSA71

Gene	Nucleotide	Amino acid
N	98.7-99.3%	99.6-100%
P	99-99.1%	99-100%
M	98.5-99.3%	99.5-100%
G	98.6-98.9%	99.2-99.8%

2.3.6 Genotype classification

In this study the shortcomings associated with the current proposed lyssavirus classification criteria were investigated. Nucleotide and amino acid identities for the complete N, P, M and G genes of the phylogroup I lyssavirus genotypes as well as the putative genotypes Irkut, Aravan and Khujand were determined (Appendix A). Nucleotide and amino acid identity should not be less between isolates considered as part of the same lyssavirus genotype (intragenotypic identity) than between isolates considered to belong to separate lyssavirus genotypes (intergenotypic identity) (Markotter *et al.,* 2008a). Therefore the minimum intragenotypic identity should always be higher than the maximum intergenotypic identity (Minimum intragenotypic identity/Maximum intergenotypic identity > 1). This ratio was analysed for the phylogroup I and putative lyssavirus genotypes (Table 2.6). It is important to note that the divergent DUVVKenya isolate was not included in this study and should the full gene sequences become available for this isolate, both the intragenotypic and intergenotypic identity values for DUVV may be greatly impacted. The influence of this isolate on identity values was clearly seen in the 398bp fragment analysis of the N gene (Section 2.3.3).

When Irkut virus was considered as part of either gt 5 (EBLV1) or gt 4 (DUVV), overlaps were seen between intragenotypic and intergenotypic identities (ratio<1). The same result was observed when DUVV and EBLV1 were considered as a single genotype (Table 2.6 and Figure 2.6). When considered as separate lyssavirus genotypes no overlap occurred. Thus, although the 92% aa identity threshold determined for genotype classification was often crossed by these viruses, based on N, P, M and G gene nucleotide and amino acid identities, they should all be considered as separate genotypes. Analysis of M gene amino acid identity indicated both intragenotypic and intergenotypic overlaps for gt 1 (RABV), these values may

however have been influenced by the limited number of isolates, which included mostly vaccine strains (Appendix A). Due to limited sequence availability this value is unknown for gt 5 and 6. Thus, this study has shown that the N, P and G genes could be successfully used to classify lyssavirus genotypes. The M gene however, was found to be an unsuitable candidate for lyssavirus classification due to the observed overlap, similar results were obtained by Markotter *et al*., (2008a) (Table 2.6 and Figure 2.6).

Table 2.6 Overlaps between intragenotypic and intergenotypic identity between phylogroup I and the putative lyssavirus genotypes analysed in this study. The ratio of the minimum intragenotypic identity/maximum intergenotypic identity is indicated. A ratio of < 1 indicates an overlap. Where no value is indicated only one sequence was available and intragenotypic identity could not be determined. Shaded cells indicate values < 1.

Genotype	N Gene*	N Protein§	P Gene#	P Proteinº	M Geneγ	M Protein^	G Geneβ	G Protein†
RABV	98.9/79.8 = 1.239	99.1/93.3 = 1.062	98.5/74.6 = 1.320	97.3/80.1 = 1.215	96.2/81.1 = 1.186	91/96.5 = 0.943	98.1/78.8 = 1.245	96.6/87.7 = 1.101
DUVV	98.7/79.8 = 1.237	99.6/93.3 = 1.068	99/74.6 = 1.327	99/80.1 = 1.236	98.5/81.1 = 1.215	99.5/96.5 = 1.031	98.7/78.8 = 1.252	99.2/87.7 = 1.131
EBLV1	95.8/79.8 = 1.200	98.2/93.3 = 1.053	98.5/74.6 = 1.320	98.7/80.1 = 1.232			95.9/78.8 = 1.217	98.3/87.7 = 1.121
EBLV2	95.6/79.8 = 1.198	97.8/93.3 = 1.048	95.5/74.6 = 1.280	98.3/80.1 = 1.227			94.1/78. =1.194	97.3/87.7 = 1.109
EBLV1 & Irkut	78.3/79.8 = 0.981	91.3/93.3 = 0.978	71.4/74.6 = 0.957	70.7/80.1 =0.883	79.6/81.1 = 0.994	93/96.5 =0.963	73.5/78.8 = 0.938	80.7/87.7 = 0.920
DUVV & Irkut	77.8/79.8 = 0.975	90.4/93.3 = 0.969	67.3/74.6 = 0.936	67/80.1 = 0.836	78.3/81.1 = 0.965	92/96.5 = 0.953	69.9/78.8 = 0.887	75.3/87.7 =0.858
EBLV1 & DUVV	78.9/79.8 = 0.988	91.6/93.3 = 0.982	69.8/74.6 = 0.936	71.4/78.8 = 0.891	80.1/81.1 = 0.987	92.5/96.5 = 0.959	72.5/78.8 = 0.920	79.5/87.7 = 0.906

* Maximum intergenotypic identity (79.8%) observed between Khujand (AY262024) and EBLV2 (EF157977)
§ Maximum intergenotypic identity (93.3%) observed between DUVV (EU623438) and EBLV1 (AY863397)
# Maximum intergenotypic identity (74.6%) observed between Aravan (AY262023) and Khujand (AY262024)
º Maximum intergenotypic identity (78.8%) observed between Khujand (AY262024) and EBLV2 (AF049121)
γ Maximum intergenotypic identity (81.1%) observed between ABLV (AF418014) and Aravan (AY262023)
^ Maximum intergenotypic identity (96.5%) observed between Aravan (AY262023) and Khujand (AY262024)
β Maximum intergenotypic identity (78.8%) observed between EBLV2 (AY863343) and Khujand (AY262024)
† Maximum intergenotypic identity (87.7%) observed between EBLV2 (AY863343) and Khujand (AY262024)

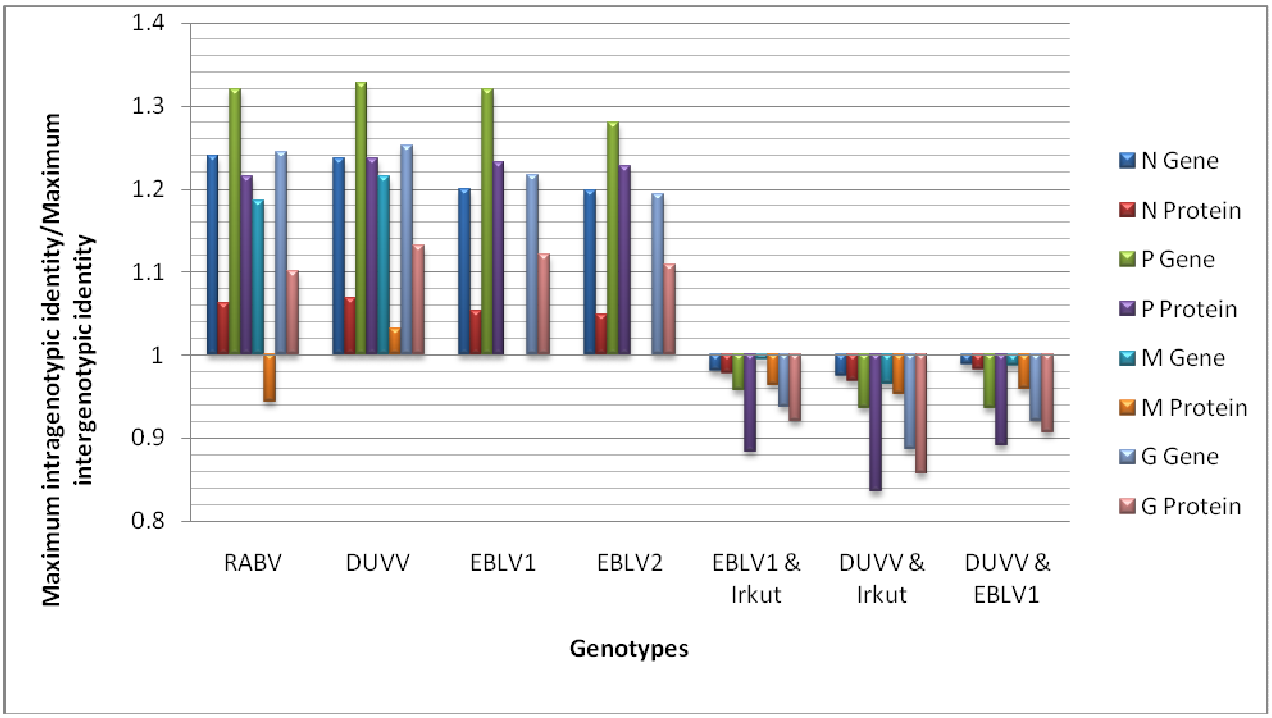


Figure 2.6 Overlaps between minimum intragenotypic and maximum intergenotypic identity observed between lyssavirus genotypes when analyzing the nucleotide and amino acid sequence identity of the N, P, M and G genes. The ratio of the minimum intragenotypic identity/maximum intergenotypic identity is indicated. A ratio of < 1 indicates an overlap. Where no value is indicated only one sequence was available and intragenotypic identity could not be determined.

2.3.7 Analysis of selection pressures and nucleotide substitution patterns

Nucleotide substitutions were computed by the p-distance method (Nei and Gojobori, 1986) using the distance estimation program of MEGA 3.1. Values for transitions (s) and transversions (v) (Table 2.7) indicated a predominance of transitions. For EBLV1 the ratio s/v was higher for the glycoprotein than for the nucleoprotein (3.63 and 2.66). Values obtained corresponded with the findings of Amengual *et al.*, 1997. The s/v ratio for the nucleoprotein of DUVV was however significantly higher (13) than the value obtained for the glycoprotein (2.22). However when all the DUVV isolates were included (nt 8-406) the s/v ratio for the N gene decreased to 4.3. The considerably different results obtained for the DUVV N gene may be explained by s/v rate bias, where low sequence divergence leads to ratio over estimation (Yang and Yoder, 1998).

Table 2.7 Comparison of nucleotide substitutions between DUVV and EBLV1

	DUVV	EBLV1a	EBLV1b	EBLV1
Full nucleoprotein gene				
***d*S**	0.0366	0.0239	0.0849	0.117
SD	0.0106	0.0087	0.0159	0.167
***d*N**	0.0012	0.0025	0.0010	0.0042
SD	0.0009	0.0012	0.0010	0.0018
s	0.0088	0.0034	0.0140	0.0212
SD	0.0025	0.0016	0.0032	0.0036
V	0.0004	0.0039	0.0059	0.074
SD	0.0004	0.0017	0.0021	0.0023
R=s/v	13.0*	1.02	2.38	2.66
Nucleoprotein gene nt 8-406				
R=s/v	4.3*	1.5	2.5	3.3
Full glycoprotein gene				
***d*S**	0.0445	0.018	0.0717	0.097
SD	0.0107	0.0069	0.0135	0.0144
***d*N**	0.0025	0.0016	0.0017	0.0047
SD	0.0014	0.0011	0.0012	0.0018
s	0.0085	0.0034	0.0140	0.0211
SD	0.0023	0.0014	0.0030	0.0033
v	0.0038	0.0021	0.0038	0.0051
SD	0.0016	0.0011	0.0016	0.0017
R=s/v	2.22	1.64	3.67	3.63

*Possible s/v rate bias

Proportions of synonymous (*d*S) and non synonymous (*d*N) nucleotide substitutions (Table 2.5) were comparable between DUVV and EBLV1. Values obtained for *d*N and *d*S indicated both the N and G genes to be subject to purifying selection (*d*N<*d*S) and this was confirmed with the codon based Z test. The average *d*N/*d*S values were higher for the G protein (0.056 and 0.048) than for the N protein (0.033 and 0.034) for both DUVV and EBLV1 respectively. The values obtained for the N gene of EBLV1 were equivalent to those found by Davis *et al*., (2005), our values for the G gene were however lower and was most likely due to sample size. Upon investigation of the *d*N/*d*S values between these two genotypes we found that at the N gene a value of 0.053 was obtained and at the G gene 0.139. These values are both greater than those that were found within both DUVV and EBLV1.

2.4 Discussion

This chapter includes the first phylogenetic study of all the known DUVV isolates. Analysis based on partial nucleoprotein sequences showed clear separation of the DUVV isolates from those of EBLV1, although these two groups are most closely related within the lyssaviruses. As previously demonstrated by Amengual *et al.,* (1997), EBLV1 and EBLV2 each split into two lineages (a and b). The DUVV isolates also split into two lineages, the longer branch lengths suggesting that these two lineages split from each other earlier than those of the EBLV's. Intrinsic heterogeneity between the DUVV isolates also clearly differentiated between these two lineages. Lineage A isolates, which are from southern Africa, showed less than 2% nucleotide variation, even though isolates were obtained a number of years apart. Lineage B, at present consisting solely of the DUVV isolate from Kenya, showed an 11% variation to the lineage A isolates, again highlighting the distance between these two lineages. It has been shown for EBLV1a that there is phylogenetic homogeneity between isolates across geographical regions, possibly due to viral traffic among bat populations (Davis *et al.,* 2005). For EBLV1b however, geographic origin plays a significant role in phylogenetic clusters, as there is less contact between bat populations (Davis *et al.,* 2005). These observations may also hold true for DUVV, though more isolates are needed to fully understand the dynamics of this African lyssavirus.

This study also included the molecular analysis of DUVV using complete N, P, M and G gene and protein sequences. Phylogenetic results were similar for the N, P and G genes with tree topologies being in agreement with the current classification of lyssaviruses as described in previous studies (Kuzmin *et al.,* 2003; Kuzmin *et al.,* 2005). Phylogenetic analysis for the deduced amino acids demonstrated the same tree topology. The position of Aravan virus was however unstable due to its equally moderate homology with Khujand virus and with the clade joining gts 4 and 5 (Kuzmin *et al.,* 2003). The M gene however showed unusual grouping. In the majority of cases there was only low bootstrap support for the DUVV, EBLV1 and Irkut cluster, especially at the nucleotide level, which is likely due to the limited sample size that included the only isolate of Irkut virus. The high level of similarity between these viruses suggests that the full picture will only become clear once additional isolates are included. Analysis of sequence identity gave very varied results, not only between genes but also between the nucleotide and amino acid sequences of individual genes. Due to the possible influence of single isolates on results, we conducted two studies; one which included the putative genotypes (Aravan, Irkut, Khujand and WCBV) and one which did not.

Although percentage identity values were greatly decreased with the exclusion of the putative genotypes, this did not influence the inconsistency seen between the genes or between the nucleotide and amino acid sequences of individual genes. The only gene to show uniformity was the nucleoprotein, which in both studies, at both the nucleotide and amino acid level, showed highest similarity between the same two genotypes (DUVV and EBLV1).

Nucleoprotein amino acid analysis showed much overlap between genotypes using current lyssavirus classification criteria (Kissi *et al.,* 1995). These criteria became problematic with the discovery of the four putative genotypes Irkut, Aravan, Khujand and WCBV and it became apparent that with all the new information available these criteria needed to be reviewed. This study indicated that the analysis of the N, P and G gene intragenotypic and intergenotypic nucleotide identities supported the classification of phylogroup I lyssavirus genotypes (RABV, DUVV, EBLV1, EBLV2) as well as the putative genotype Irkut as separate genotypes. A high level of intragenotypic variation was observed between RABV isolates, where overlap between intragenotypic and intergenotypic identity was found when analyzing the M amino acid sequences. The ratio of minimum intragenotypic/maximum intergenotypic identity is however dependent on the number of viral isolates analyzed for each genotype and as such may vary with the addition of new isolates. Only a single matrix protein sequence was available for each of the EBLV's, so intragenotypic and intergenotypic identity values could not be obtained. The intragenotypic and intergenotypic identity values obtained for DUVV were also very low, making the M gene an unsuitable candidate for lyssavirus classification. The variation in results between the different genes implies they may not all be equal for phylogenetic analysis as was suggested by Wu *et al.,* (2007). As indicated by the study only the nucleoprotein nucleotide identity provided a clear distinction between both the lyssavirus and putative lyssavirus genotypes, where the current criteria suggesting <80% nucleotide identity constitutes a new lyssavirus genotype (Kissi *et al*., 1995) still applies.

When nucleotide substitutions among DUVV and EBLV1 were investigated the values obtained for transitions (s) and transversions (v) were comparable to those found by Amengual *et al.,* (1997), where a predominance of transitions was shown. This higher rate of transitions may suggest that these viruses are at an early stage of divergence (Jukes, 1987). The s/v values obtained for DUVV (13) however greatly differed from those that were found for EBLV1 (2.66), which may have been due to s/v rate bias, where the s/v ratio is overestimated due to low sequence divergence (Yang and Yoder, 1998). Our data supports

this as the s/v ratio decreased to 4.3 when the more divergent lineage B DUVV isolate was included; this was however based on partial nucleoprotein sequence. On investigation of the selective constraints acting on both DUVV and EBLV1, we found the *d*N/*d*S values to be very similar in both viruses. Values obtained for the glycoprotein were higher than those obtained for the nucleoprotein, indicating stronger purifying selection against the nucleoprotein gene. This observation was also made by Davis *et al*., (2005), and is to be expected as the glycoprotein serves as the main target for the immune response, where greater amino acid diversity is more likely. The *d*N/*d*S values obtained for both the N and G genes of both DUVV and EBLV1 were low, indicating strong selective constraints against amino acid change. On analysis of the *d*N/*d*S values between these two viruses, the values obtained for both the N and G genes, were higher than those obtained in the individual analysis. This suggests that the evolutionary changes between these two viruses were not due to random drift but rather, natural selection (Ridley, 2004). As the results obtained were the same for both the nucleoprotein and glycoprotein, the evolutionary changes could have been explained by positive selection (*d*N>*d*S) for these two viruses. However we know this is not the case. Thus their evolution is more likely to be explained by the nearly neutral theory, where a population bottleneck (the effect of genetic drift on a temporarily small population) occurred during speciation (Ridley, 2004). Such a 'bottleneck-like' transmission mechanism for EBLV was mentioned by Amengual *et al*., (1997).

Different insectivorous bat species have been associated with DUVV, EBLV1 and Irkut, with the exception of *M. schreibersii* which has been linked to both DUVV and EBLV1, although not confirmed. Many of these species are known to co-colonize roosts (Nowak, 1999; Ma *et al.,* 2003) where close contact allows for the spread of RNA viruses between species with relative ease (Mackenzie *et al.,* 2003). Increasing the potential role of these bat species in viral spread and evolution is the ability of species such as *M. schreibersii, N. thebaica* and *M. myotis* to migrate (Van der Merwe, 1982; Nowak, 1999; Ma *et al.,* 2003). The distribution of these species in both southern Europe and northern Africa (Gray *et al.,* 1999; Nowak, 1999; Ma *et al.,* 2003) implies that these host species may have facilitated the spread of these viruses between Africa and Europe as was suggested by both Amengual *et al.,* (1997) and Davis *et al.,* (2005).

CHAPTER 3
Characterization of the full genome of Duvenhage virus

This study was initially started with the aim to produce the first full length genome sequence of rabies-related Duvenhage virus, as at the time only N, P and G full gene sequences were available for the DUVVSA71 and DUVVSA81 isolates. Sequencing of the DUVVSA06 isolate's full genome for this study was completed in January 2008.

It was however at this time that Delmas and colleagues (Delmas *et al*., 2008) submitted their paper describing the full genome sequences of DUVVSA71 and DUVVSA81. This paper was published in April 2008. Thus our sequence was no longer the first full genome for DUVV and as such the focus of this study had to be slightly altered.

As the genomic properties of DUVV had already been described (Delmas *et al*., 2008), our focus turned to; A) the investigation of antigenic, pathogenic and conserved domains of the lyssavirus genome and B) evaluate whether full genome analysis would be better for lyssavirus classification than individual gene analysis (Chapter 2). This is the first study to include both the lyssavirus (RABV, LBV, MOKV, DUVV, EBLV1, EBLV2 and ABLV) and putative lyssavirus (Aravan, Khujand, Irkut and WCBV) genotypes in full genome phylogenetic analysis.

3.1 Introduction

Duvenhage virus (DUVV) is a member of the *Lyssavirus* genus in the *Rhabdoviridae* family. Seven lyssavirus genotypes are currently recognized by the International Committee of the Taxonomy of Viruses (ICTV), these include RABV (gt 1), LBV (gt 2), MOKV (gt 3), DUVV (gt 4), EBLV1 (gt 5), EBLV2 (gt 6) and ABLV (gt 7) (Tordo *et al.,* 2005). The genus may however be expanded upon with the addition of the putative genotypes; Irkut, Aravan, Khujand and West Caucasian bat virus (WCBV) (Kuzmin *et al.,* 2005). Some LBV isolates have also been shown to form a separate genotype and this may also lead to further expansion of the genus (Markotter *et al.,* 2008a). The full genomes of all the lyssavirus genotypes have been sequenced (Tordo *et al.,* 1988; Conzelmann *et al.,* 1990; Le Mercier *et al.,* 1997; Warrilow *et al.,* 2002; Marston *et al.,* 2007; Delmas *et al.,* 2008), as have those of the four putative genotypes (Kuzmin *et al.,* 2008b).

The genome comprises a single negative stranded RNA molecule of approximately 12kb that is transcribed into five non-overlapping mRNAs encoding five structural proteins, N (nucleoprotein), P (phosphoprotein), M (matrix protein), G (Glycoprotein) and L (RNA polymerase). With the exception of MOKV and WCBV, the intergenic sequences (IGS), which are eluded by the transcriptase between one (transcription terminal signal) TTP and the following (transcription initiation signal) TIS, are an invariant; N-P 2 nt, P-M 2 nt, M-G 5 nt and the G-L a variable 19 - 28 nt (Marston *et al.,* 2007; Kuzmin *et al.,* 2008b). The 5′ and 3′ genomic termini are highly conserved both in length and sequence.

The termini of the lyssaviruses are complementary to each other along the first 11 (RABV, LBV, EBLV1, MOKV, DUVV, Irkut and WCBV) or 9 (EBLV2, ABLV, Aravan and Khujand) nucleotides. Various sites with antigenic (Lafon *et al.,* 1983; Lafon and Wiktor, 1985; Prehaud *et al.,* 1988 ; Dietzschold *et al.,* 1990; Benmansour *et al.,* 1991 ; Ertl *et al.,* 1991) and pathogenic (Dietzschold *et al.,* 1983; Seif *et al.,* 1985; Tuffereau *et al.,* 1989; Yang *et al.,* 1999; Takayama-Ito *et al.,* 2006b) properties have been identified for the lyssaviruses. Domains of high conservancy have also been recognized for this genus with the polymerase gene demonstrating conserved residues which are clustered into six blocks of strong conservation linked by variable regions of low conservation (Poch *et al.,* 1990).

Evolutionary studies of lyssaviruses have so far mostly focused on the N and G proteins. All five proteins are however both structurally and functionally related and there is common agreement that interacting proteins undergo co-evolution (Pazos *et al.,* 1997). Since no

recombination events have been reported in lyssaviruses, Wu *et al.,* (2007), hypothesized that in the *Lyssavirus* genus each individual gene may generate similar tree topology for phylogenetic analysis. This study will be the first to include full genome analysis of all the lyssavirus representatives, including both the lyssavirus and putative lyssavirus genotypes. The objectives of this study were; 1) to determine and describe the full length sequence of the DUVVSA06 isolate; 2) to investigate the phylogenetic relationship between all the lyssavirus representatives at the full genome level; 3) to determine whether full genomes are better than individual genes for lyssavirus analysis and 4) to investigate antigenic, pathogenic and conserved domains on the DUVV genome.

3.2 Materials and methods

3.2.1 Viral isolate

The DUVVSA2006 human brain isolate was obtained from Dr. J. Paweska of the Special Pathogens Unit, National Institute for Communicable Diseases, National Health Laboratory Services, South Africa. The virus isolate was grown up as previously described in section 2.2.1.

3.2.2 RNA extraction

RNA was isolated using the Trizol method as explained in section 2.2.2.

3.2.3 Primer design

Refer to section 2.2.3. Primers were designed for both PCR and sequencing, details given in Table 3.1. The relative positions of the primers on the genome of the 2006 DUVV isolate are shown in Figure 3.1.

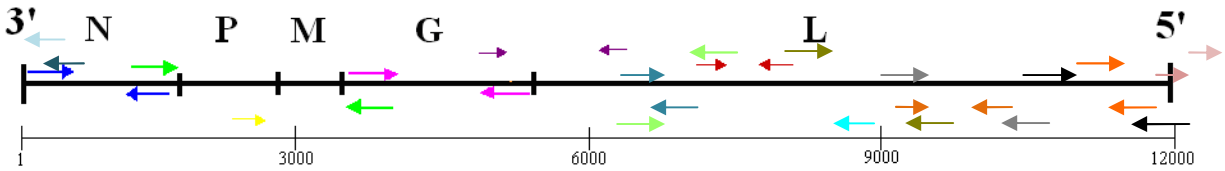


Figure 3.1 Relative position of primers on DUVV genome

Table 3.1 Primers used to amplify the DUVV genome. Genome position is based on the DUVVSA06 full genome obtained in this study (EU623444).

Primer	Sequence	Genomic position	Reference	Use
Lys001	5' ACGCTTAACGAMAAA 3'	3' non coding region (-70 to -57)	Markotter. *et al.,* 2006a	cDNA, PCR, sequencing
Lys304	5' TTGACAAAGATCTTGCTCAT 3'	1447-1466 rc	Markotter *et al.,* 2006a	PCR, sequencing
DGF	5' CCTCAAGGAGTTCAAGCGCC 3'	3207-3226	This study	cDNA, PCR, sequencing
DFR	5' GGCCTCTCACTCCCTTGTTG 3'	4815-4834 rc	This study	PCR, sequencing
DuvN+	5' GGATCATGATGAACGGAG 3'	1223-1240	This study	cDNA, PCR, sequencing
DuvG-	5' GGCCCCAATTTGTCAGGG 3'	3304-3321 rc	This study	PCR, sequencing
DuvG1+	5'GAAGGAACCACAGGAGATGTTCG 3'	4778-4800	This study	cDNA, PCR, sequencing
DuvL-	5' GTTGAGATTGTAGTCAGAGTTCC 3'	5485-5507 rc	This study	PCR, sequencing
DuvLint+	5'GTCATCACAGAGAAGCTTTTGGCC 3'	7086-7109	This study	cDNA, PCR, sequencing
DuvLint-	5' GTCCACCGTCCTGACCGTTCCAGC '3	7450-7473 rc	This study	PCR, sequencing
DuvP+	5' CCACCCAGACTGTTACTG 3'	1880-1896	This study	Sequencing
DpolF1	5' GGACAAGGGTTGTTAGAC 3'	7194-7211	This study	cDNA, PCR, sequencing
DpolR1	5' GATAAGGCCCTCTTGACCACATG 3'	8949-8970 rc	This study	PCR, sequencing
DpolF2	5' GCTCTTCCGAGAGGGCAG 3'	8667-8685	This study	cDNA, PCR, sequencing
5020B	5' GCCCTGATATCAATATCAG 3'	10351-10369 rc	This study	PCR, sequencing
DpolR4	5' CAGAGGCTCCACAGACC 3'	8443-8459 rc	This study	Sequencing
DuvL+	5' GTACCGCTCTTAAGTGATGAGG 3'	5943-5964	This study	cDNA, PCR, sequencing
LintRev	5' GCCCGAATACCTTATCTAG 3'	7302-7320 rc	This study	PCR, sequencing
DuvL3F2	5' CAAGAGGTCCGCCATGCAGC 3'	10209-10228	This study	cDNA, PCR, sequencing
DuvL3Rb	5' GCCAACGAGTCTGGTAGTCTTCAC 3'	11637-11660 rc	This study	PCR, sequencing
DuvL-si	5' GCTGGAGTCCACAGAGGTG 3'	5393-5411	This study	cDNA, PCR, sequencing
DuvL+si	5' GCTTCTGGAGGTGAGAGC 3'	6170-6187 rc	This study	PCR, sequencing
DpolF5	5' CCATCCGAGATGTTGTCC 3'	8742-8759	This study	cDNA, PCR, sequencing
DpolR5	5' GGAGATACTCTCTTGTATATG 3'	9614-9634 rc	This study	PCR, sequencing
DuvL5'1	5' CCATGAACTTTACAACAACCC 3'	11468-11488	This study	cDNA, PCR, sequencing
DuvL5'2	5' GGAAGCAGATGATAGGAGGG 3'	11506-11525	This study	cDNA, PCR, sequencing
DuvN3'1	5' GCATCCATTGTAGGGGTGTTAC 3'	3' non-coding region (-13 to -08)	This study	PCR, sequencing
DuvN3'2	5' GCTGTTACGGACCTTAAAG 3'	21-40 rc	This study	PCR, sequencing
DpolF6	5' CAAGACTTACGGGACAATGTTGG'3	10874-10896	This study	cDNA, PCR, sequencing
DpolR6	5' CAGCCGAATCCAGTGCGCGG 3'	11605-11624 rc	This study	PCR, sequencing

rc = reverse complement

3.2.4 Generation of sequence information

DNA sequences were obtained as described in sections 2.2.4-2.2.8.

3.2.5 Sequence analysis at the full genome level

Sequence similarity between isolates was determined using the distance estimation program of MEGA 3.1 (Kumar *et al.,* 2004). Genetic distances were calculated for both the nucleotide and deduced amino acid sequences of the L gene and the full genome of the DUVVSA06 isolate, using the p-distance model (Nei and Gojobori, 1986).

3.2.6 Phylogenetic analysis

Sequences were edited using the Bioedit software package (Hall, 1999) hereafter; they were assembled using the VectorNTI 9.1.0 software package (Invitrogen). Alignments were then carried out using the ClustalW subroutine (Thompson *et al.,* 1994), which forms part of the Bioedit program.

The calculation of genetic distances and construction of phylogenetic trees was carried out using the MEGA 3.1 software (Kumar *et al.,* 2004). Genetic distances were calculated between pairs of sequences by using the Kimura's 2-parameter method (Kimura, 1980), and based on these distances neighbour-joining (NJ) trees were constructed using the methods of Saitou and Nei, (1987). The bootstrap option of 1000 replicate datasets was used to assess the robustness of the method. Bootstrap values of more than 70% were regarded as providing evidence for a phylogenetic grouping. Results were validated by maximum parsimony as implemented in MEGA 3.1 (Kumar *et al.,* 2004).

3.2.7 Determination of genomic 3′ and 5′ terminal sequences

Based on the method described by Kuzmin *et al.,* (2008b), the genomic termini were determined as indicated in Figure 3.2.

- Circularization of the genome

To determine the terminal sequences, total brain RNA was extracted by the Trizol method (refer to section 2.2). The RNA was circularized using T4 RNA ligase (Promega) according to the manufacturer's instructions with the following modifications: 20U T4 RNA ligase, 40U Protector RNase Inhibitor (Roche), 20µl PEG 40% solution, 4 µl Ligase buffer (50 mM Tris, 10 mM $MgCl_2$, 5 mM DTT and 1 mM ATP) and 13 µl RNA (1-2 µg) were incubated at 37˚C for 30 minutes. The ligated genomic RNA was ethanol precipitated; 4 µl NaAOH (3 M) and 80 µl 100% ethanol were added and incubated at room temperature for an hour and

then centrifuged for 30 minutes, the pellet was washed in 70% ethanol and the dried pellet resuspended in 20 µl DEPC water. Reverse transcription was carried out using 10 pmol oligonucleotide primer DuvN3'2 (Table 3.1). The first round PCR (refer to section 2.2.5) was carried out using the primers DuvN3'2 and DuvL5'1 and a second round of nested amplification was performed using the primers DuvN3'1 and DuvL5'2. Products were then analysed by agarose gel electrophoresis and purified (see section 2.2.6-2.2.7) before cloning.

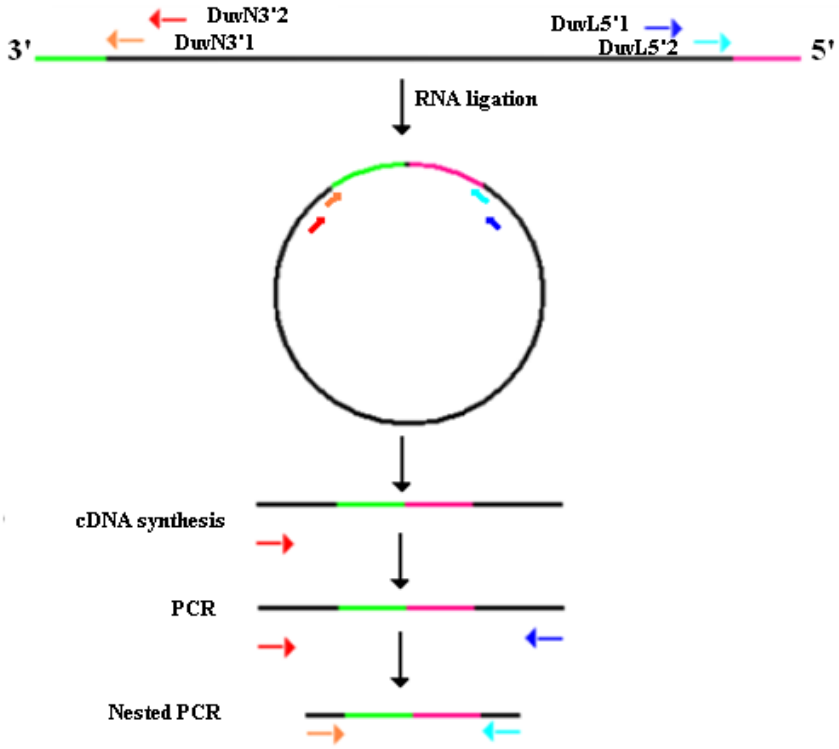


Figure 3.2 Method used to determine the genomic 3′ and 5′ terminal sequences.

3.2.8 Cloning of PCR products of genome ends

- *Ligation*

Purified amplicons were ligated with pGEM®-T Easy vector (Promega), using the manufacturer's instructions as follows: Approximately 50 ng pGEM®-T easy vector and 300 ng PCR product (refer to section 2.2.5) were ligated with 5 µl rapid ligation buffer (60 mM Tris-HCL, 20 mM DTT, 20 mM $MgCl_2$, 2 mM ATP), 3 U T4 DNA ligase and were mixed by pipetting and incubated overnight at 4˚C.

- *Preparation of competent cells*

Competent *E.coli* JM109 cells were prepared as described in Hanahan *et al.,* (1991). Briefly, *E.coli* was streaked onto M9 minimal media agar plates (20% 5X M9 minimal salt solution, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.1% glucose) and grown overnight at 37˚C. Colonies were picked and streaked on Luria Bertani agar plates and grown overnight at 37 ˚C. Several colonies were then inoculated into 1 ml SOB medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl) and vortexed. Next, 1ml of the inoculated SOB was added to 50 ml SOB and grown at 37 ˚C with shaking until an OD_{600} of 0.4-0.6 was reached. The cells were incubated on ice for 10 minutes and centrifuged at 1000 g for 15 minutes. The pellets were resuspended in 1/3 of volume of CCMB (80 mM $CaCl_2.2H_20$, 20 mM $MnCl_2.4H_20$, 10 mM $MgCl_2.6H_20$, 10 mM K-acetate, 10% glycerol) medium and incubated on ice for 20 minutes. Cells were pelleted at 1000 g for 10 min at 4˚C and resuspended in 1/12 volume of CCMB. The competent cells were aliquoted and stored at -70 ˚C for further use.

- *Transformation of competent E.coli JM109 cells*

A test transformation was performed by adding 10 ng control plasmid (pUC18) to 100 µl competent cells. A negative control with no plasmid was also prepared. 2 µl of each ligation reaction was added to a sterile tube with 100 µl of cells. The cells were incubated on ice for 30 minutes and heat shocked at 42˚C for 30 seconds, 900 µl pre-warmed LB broth was added and the cultures incubated at 37 ˚C with shaking for an hour. 100 µl of the culture broth was then plated out on LB-agar plates supplemented with 100 µg/ml ampicillin. To allow for blue-white colour selection, based on insertional inactivation of the *lac Z* gene, 40 µl of X-gal (2% stock solution) and 10 µl IPTG (100mM stock solution) were also spread over the surface of the plates. The plates were incubated overnight at 37˚C to allow for the observation of transformants with the Gal^- phenotype (white). These were selected for further characterization and grown overnight in 5 ml LB-broth, supplemented with ampicillin.

3.2.9 Plasmid purification

Purification of these presumed recombinant clones was achieved using the commercial Wizard® Plus SV Minipreps DNA Purification System (Promega). 10 ml of the overnight cultures were pelleted for 5 minutes by centrifugation at 12 000 g. The pellet was then resuspended in 250 µl cell resuspension solution (50 mM Tris-HCl, 10 mM EDTA, 100 µg/ml Rnase A), then 250 µl cell lysis solution (0.2 M NaOH, 1% SDS) was added and the tubes inverted 4 times to mix. 10 µl alkaline protease solution was added, the tubes inverted

4 times and let to incubate at room temperature for 5 minutes. Thereafter 350 µl neutralization solution (4.09 M CH_5N_3.HCl, 0.759 M CH_3COOK, 2.12 M $C_2H_4O_2$; pH 4.2) was added and the tube inverted, followed by centrifugation for 10 minutes. The reactions were placed in spin columns and centrifuged, allowing the DNA to bind to the silica. This was followed by washing with 750 µl column wash solution (60 mM CH_3COOK, 8.3 mM Tris-HCl, 0.04 mM EDTA, and 60% ethanol) followed by centrifugation. The DNA was eluted in 100 µl nuclease free H_2O. Recombinant clones were verified by automated DNA sequencing.

3.2.10 Nucleotide sequencing of genomic termini

Sequencing of cloned insert DNA was carried out as previously described in section 2.2.8, using 500ng/5kb plasmid as template. The T7 primer (5' TAATACGACTCACTATAGGG 3') specific to the pGEM® T-Easy vector was used.

3.3 Results

3.3.1 cDNA synthesis and PCR amplification of the full genome

RNA extraction, cDNA synthesis and PCR were performed as described in sections 2.2.2-2.2.5. The primer sets yielded specific products (Table 3.2) which were then purified as described in section 2.2.7.

Table 3.2 Primer sets used to obtain the full genome sequence of DUVVSA06, with the approximate product size indicated in nucleotides.

Primer set	Product size	Primer set	Product size
Lys001-Lys304	1500nt	DuvL+- LintRev	1400nt
DGF-DGR	1600nt	DuvL3F2- DuvL3Rb	1400nt
DuvN+-DuvG-	2100nt	DuvL-si- DuvL+si	800nt
DuvG1+-DuvL-	700nt	DpolF5- DpolR5	700nt
DuvLint+- DuvLint-	450nt	DuvL5'1- DuvN3'2	540nt
DpolF1- DpolR1	1850nt	DuvN3'1- DuvL5'2	470nt
DpolF2-5020B[b]	1700nt	DpolF6- DpolR6	750nt

3.3.2 Construction of recombinant pGEM®T-Easy vectors containing the circularized genome ends

Circularization of total RNA gave reproducible results. The initial round of amplification did not result in visible bands by agarose gel electrophoresis. However after nested amplification bands of approximately 470 nt were clearly visible. The ligated genomic termini were frequently truncated and as such cloning was an obligatory prerequisite to sequencing, this was previously observed by Kuzmin *et al.,* (2008b). Following plasmid DNA extractions, the plasmid DNA was analyzed for the 5'-3'end inserts, using the restriction endonuclease EcoR1. Fragments of approximately 520 nt were excised (results not shown). Recombinant clones containing the insert were then purified using the commercial Wizard® Plus SV Minipreps DNA Purification System (Promega) as discussed in section 3.2.7.

3.3.3 Sequencing of the full genome

The N protein of the 2006 Duvenhage virus isolate (DUVVSA2006) was found to be 451 aa, the P protein 298 aa, the M protein 202 aa, the G protein 533 aa and the polymerase protein 2127 aa with the full genome reading 11975 nucleotides. The total coding capacity was 90.6% with the polymerase protein accounting for 53.3%, which is similar to what has been previously described by Delmas *et al*., (2008) as well as Marston *et al*., (2007). Sequence length comparison of the coding and non coding regions of the three DUVV genomes are given in Table 3.3. Termination and initiation signals were conserved between all DUVV isolates (Table 3.4). The termination signals were an invariant TGAAAAAAA, whilst two initiation signals were identified; AACACCCT (N, G and L genes) and AACACCACT (P and M genes).

The N-P, P-M and M-G IGSs (intergenic spacer regions) were conserved and followed the 2, 5, 5 nt pattern observed by (Marston *et al.,* 2007).The G/L IGS ranged between 36 and 37 nucleotides which makes it the largest G/L IGS in the *Lyssavirus* genus with only putative genotype WCBV exceeding this number (100 nt) (Kuzmin *et al.,* 2008b).

Table 3.3 Comparison of the coding and non coding regions of the three DUVV genomes.

	DUVVSA06 (EU623444) (This study)	**86123SA (EU293119)** (Delmas *et al.*, 2008)	**9486SA (EU293120)** (Delmas *et al.*, 2008)
3' UTR	70	70	70
N gene	1356	1356	1356
N-P	90	90	90
P gene	897	897	897
P-M	83	83	83
M gene	609	609	609
M-G	191	191	191
G gene	1602	1602	1602
G-L	562	563	562
L gene	6384	6384	6384
5' UTR	131	131	131
Genome	11975	11976	11975

Table 3.4 Transcription initiation and termination signals for all DUVV isolates

Region	**Termination**	**IGS**	**Initiation**
Leader/N	_____	_____	AACACCCCT
N/P	TGAAAAAAA	CT	AACACCACT
P/M	TGAAAAAAA	CATGC	AACACCACT
M/G	TGAAAAAAA	CAGGC	AACACCCCT
G/L	TGAAAAAAA	36-37nt	AACACCCCT
L/Trailer	TGAAAAAAA	_____	_____

For determination of the genomic termini, 70 clones were sequenced, all of which contained inserts, the size of the inserts however varied greatly due to degradation of the 3' end. Figure 3.3 shows 25 clones which produced sequence past the polymerase termination signal TGAAAAAAA. Full length termini were obtained from clones MT74 and MT75. These sequences however showed an additional 5 nt at the ligation site (i.e. between the 5′ and 3′ ends) (Figure 3.4), which was most likely due to the presence of non specific RNA at the time of ligation. Due to the fact that all lyssavirus termini to date are conserved for at least 9 nucleotides at both ends and that these 9 nucleotides are complementary to each other the additional 5 nt obtained were not considered as part of either the 5′ or 3′ termini.

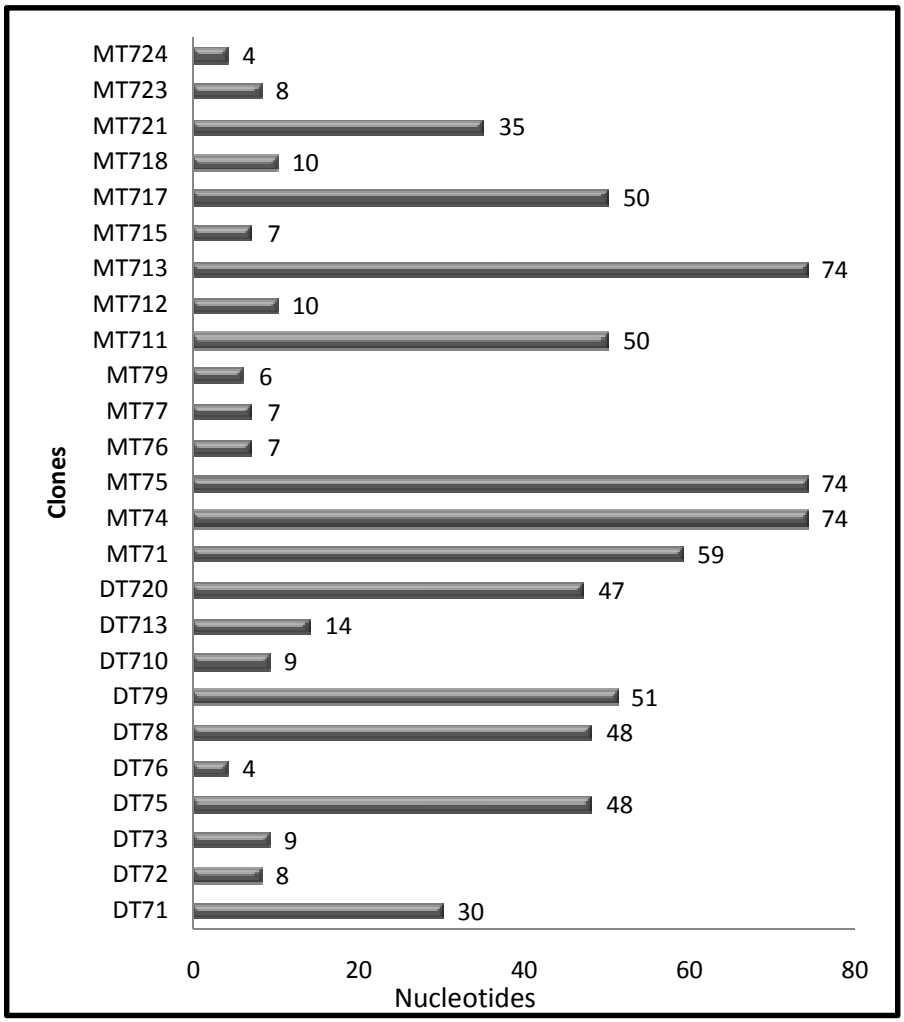


Figure 3.3 Sequence length of genomic termini 3' end from the termination signal TGAAAAAAA of polymerase gene. Clones MT74 and MT75 code the full 3' terminal sequence.

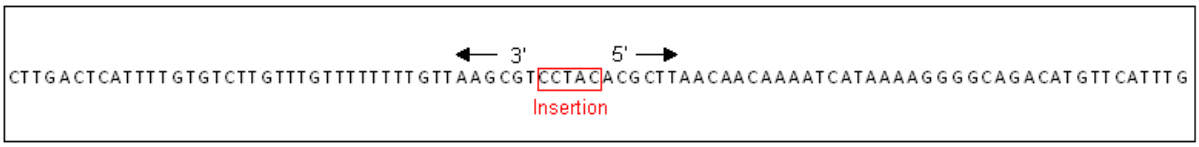


Figure 3.4 Five nucleotide insertion (red block) present in full length ligated genomic termini sequences.

The 5′ and 3′ termini were highly conserved in both length and sequence (Figure 3.5). The DUVVSA06 isolate was similar to other DUVV isolates in that the first 11 nucleotides showed complementarity to the 5′ terminus, nucleotides 14 and 16 were also complementary, which is in agreement with the study by Delmas *et al.*, 2008.

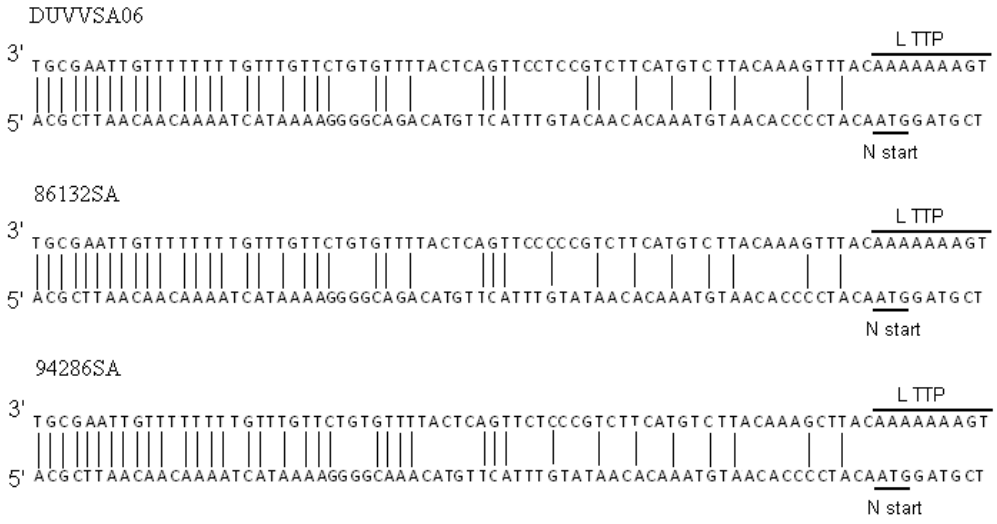


Figure 3.5 Complementarity of the 5′ and 3′ genomic termini of the Duvenhage virus genomes sequenced to date. Complementary nucleotides are indicated by a vertical line. L TTP: polymerase transcription termination signal; N start: nucleoprotein start codon.

3.3.4 P-distances

Nucleotide identity for the full genomes of both the lyssaviruses (gt 1-7) and putative lyssaviruses (Irkut, Aravan, Khujand and WCBV) was determined (Appendix A) using the p-distance model (Nei and Gojobori, 1986) in the distance estimation program of MEGA 3.1 (Kumar *et al.,* 2004). Percentage identity values obtained for the full genomes demonstrated DUVV to have the highest intragenotypic value (99%) and LBV the lowest (75.6%). The highest intergenotypic value, 78.4% was seen for EBLV2 and Khujand virus whilst the lowest intergenotypic value 63.7% was between RABV and WCBV. Analysis of polymerase gene nucleotide and deduced amino acid sequence similarity showed DUVV to have the highest intragenotypic values in both cases (99.1% and 99.6% respectively) while LBV had the lowest intragenotypic values (76.9% and 91.3% respectively). Intergenotypic analysis revealed EBLV2 and Khujand to have the highest similarity at both the nucleotide and amino acid levels (79.7% and 94%) and RABV and WCBV the lowest (67.3% and 74.7%). As stated in section 2.3.4.3, nucleotide and amino acid identity should not be less between isolates considered as part of the same lyssavirus genotype (intragenotypic identity) than between isolates considered to belong to separate lyssavirus genotypes (intergenotypic identity). Therefore the minimum intragenotypic identity should always be higher than the maximum intergenotypic identity (Minimum intragenotypic identity/Maximum intergenotypic identity > 1). This ratio was analysed for the polymerase gene of all the lyssavirus representatives (Table 3.5 and Figure 3.6).

Table 3.5 Overlaps between intragenotypic and intergenotypic identity of the polymerase gene between the lyssavirus genotypes and the putative lyssavirus genotypes analysed in this study. The ratio of the minimum intragenotypic identity/maximum intergenotypic identity is indicated. A ratio of < 1 indicates an overlap (shaded cells).

Genotype	L Gene*	L Protein§
RABV	81.7/79.7 = 1.025	93.6/94 = 0.996
DUVV	98.8/79.7 = 1.240	99.5/94 = 1.059
EBLV1	95.7/79.7 = 1.201	99.3/94 = 1.056
EBLV2	98.2/79.7 = 1.232	99.4/94 = 1.057
LBV	76.9/79.7 = 0.965	91.2/94 = 0.970
DBLV (without LBV)	98.7/79.7 = 1.238	99.7/94 = 1.060
MOKV	86.8/79.7 = 1.089	96.7/94 = 1.029

* Maximum intergenotypic identity (79.7%) observed between Khujand (EF614261) and EBLV2 (EF157977)
§ Maximum intergenotypic identity (94%) observed between Khujand (EF614261) and EBLV2 (EU293114)

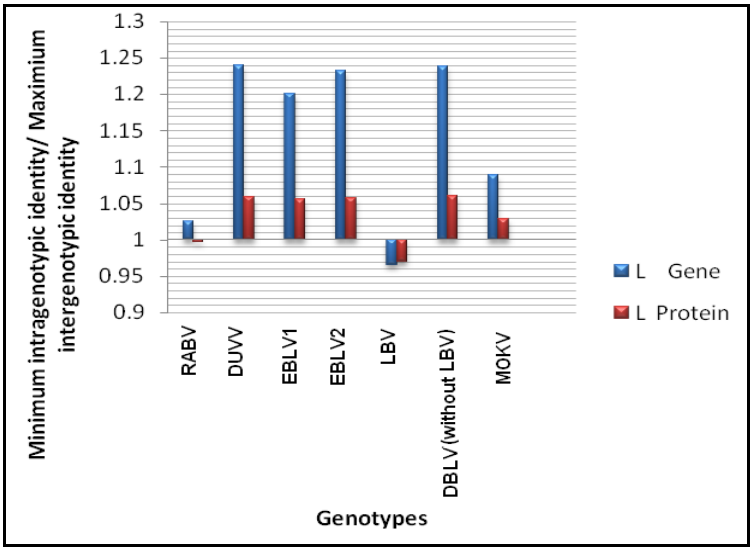


Figure 3.6 Overlaps between minimum intragenotypic and maximum intergenotypic identity observed between lyssavirus genotypes when analyzing the nucleotide and amino acid sequence identity of the L gene. The ratio of the minimum intragenotypic identity/maximum intergenotypic identity is indicated. A ratio of < 1 indicates an overlap.

Analysis of L gene nucleotide identity indicated an overlap between intragenotypic and intergenotypic identities (ratio<1) for LBV, whilst amino acid analysis indicated overlaps of intragenotypic and intergenotypic values for both for gt 1(RABV) and gt 2 (LBV). RABV is by far the most diverse lyssavirus, having many host species spanning worldwide. It is this

diversity which leads RABV to have an overlap at the L aa level, where the greatest intragenotypic variation was observed between the vaccine strain SADB19 (M31046) and isolate EU293113 isolated from a dog in France, 1990. In the case of LBV however, it has previously been shown that some isolates within this genotype should be considered a new, separate lyssavirus genotype (Markotter *et al.,* 2008a). One such isolate is that from *Eidolon helvum,* Senegal 1985. In this study the Senegal isolate (EU293108) proved to be most similar to isolate EU259198 from Kenya, as was previously documented by Kuzmin *et al.,* (2008a) and these two isolates were thus considered as a separate genotype. Our analysis showed that this may indeed be the case as no overlap was seen when the LBV isolates were split. These results indicate that the L gene may be a good candidate for lyssavirus classification.

3.3.5 Sites of antigenicity and pathogenicity

The amino acid sequences of the N and G genes of the lyssavirus and putative lyssavirus genotypes were aligned using ClustalW and regions previously suggested to have antigenic or pathogenic properties analyzed. This analysis was based on a total of 29 full genome sequences off representatives of the lyssavirus and putative lyssavirus genotypes.

3.3.5.1 Nucleoprotein

Antigenic sites I-IV are shown in Figure 3.7.

- Antigenic site I (aa 355-369) (Goto *et al.,* 2000; Minamoto *et al.,* 1994), was conserved within gt 1 (RABV). Conservancy was also seen between gt 4 (DUVV), gt 5 (EBLV1), Irkut and Aravan virus as well as between gt 2 (LBV) and gt 3 (MOKV).
- Antigenic site II (aa 313-337) (Lafon & Wiktor., 1985), was conserved between gt 4 (DUVV) and gt 5 (EBLV1) and between Aravan and Khujand virus but differed for other lyssavirus representatives.
- Antigenic site III (aa 374-383) (Dietzschold *et al.,* 1988), was very variable between the different lyssavirus genotypes.
- Antigenic site IV (aa 410-413) (Ertl *et al.,* 1991), was conserved between gt 1 (exception M13215, ERA), gt 4, gt 5, gt 7, Khujand and Aravan virus.

Site of pathogenicity (Yang *et al.,* 1999) is shown in Figure 3.7.

- 389^{Ser} was conserved in all the lyssavirus as well as the putative lyssavirus genotypes.

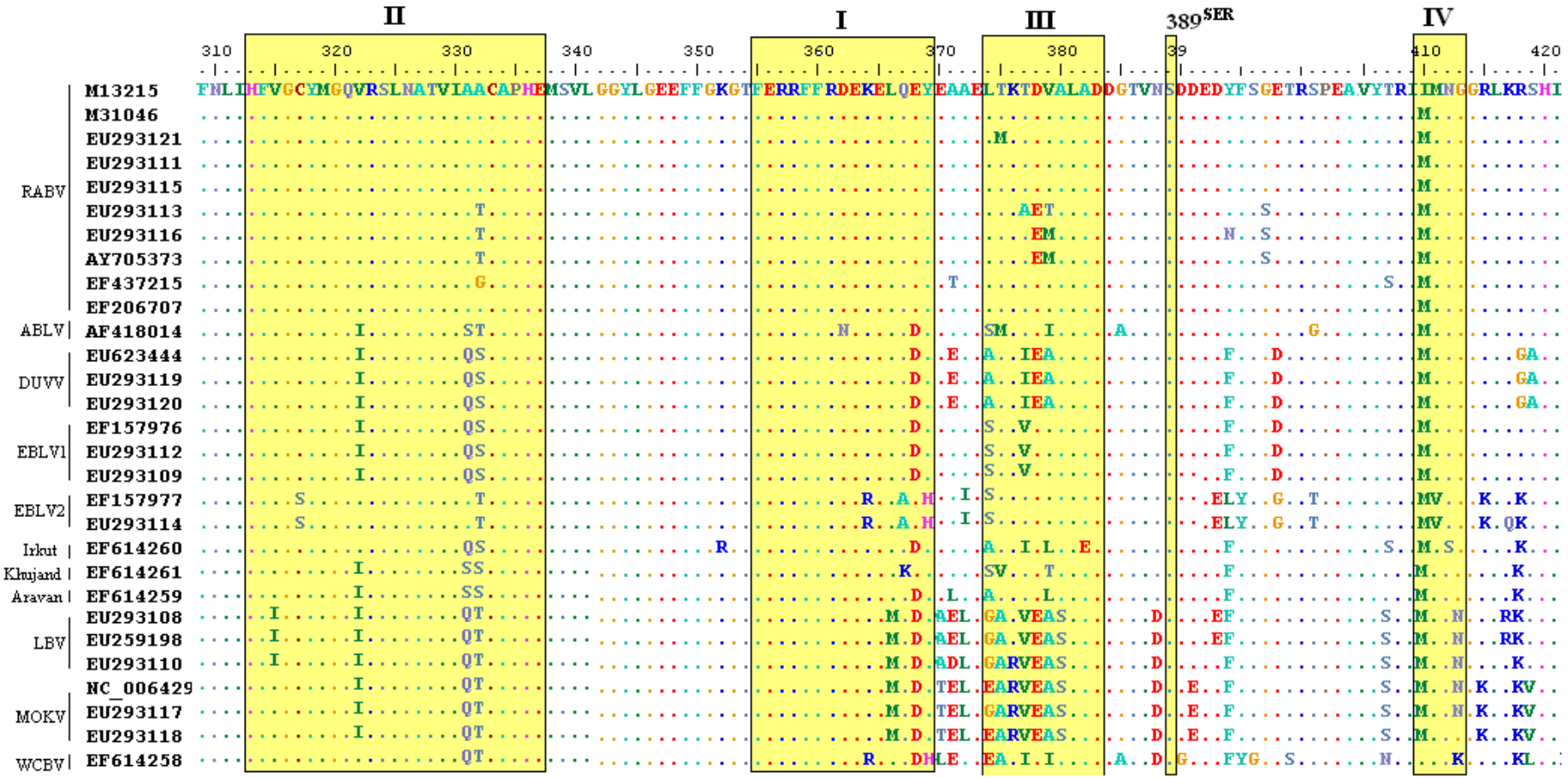


Figure 3.7 Multiple alignment of the lyssavirus genotypes and putative lyssavirus genotypes, indicating nucleoprotein antigenic and pathogenic sites. Dots represent identity to PV.

3.3.5.2 Glycoprotein

Antigenic sites I-V are shown in Figure 3.8.

- Antigenic site (aa 14-19) (Mansfield *et al.,* 2004), was conserved between all the phylogroup I lyssaviruses (gts 1, 4, 5, 6, 7) as well as the putative genotypes Aravan, Irkut and Khujand. Gt 2 , 3 and WCBV also shared conservancy at this site.
- Antigenic site I (aa 231) (Lafon *et al.,* 1983), was conserved between gts 4, 6, 7 and Khujand virus. Conservancy was also seen between some representatives of gt 1 and gts 2, 3 and 5.
- Antigenic site II (aa 34-42 and aa 198-200) (Prehaud *et al.,* 1988), was conserved between gt 4 (DUVV) and gt 5 (EBLV1). This site was not well conserved within the lyssavirus and putative lyssavirus genotypes.
- Antigenic site III (aa 330-338) (Lafon *et al.,* 1983), was conserved between gt 1 (RABV) (exception AY705373) and gt 7 (ABLV) as well as between gt 4 and Irkut virus.
- Antigenic site IV (aa 264) (Dietzschold *et al.,* 1990), was conserved between gt 4, gt 5, gt 6, gt 7, Irkut, Khujand and Aravan virus. Gt 2 and gt 3 were also conserved at this site.
- Antigenic site V (aa 342-343) (Benmansour *et al.,* 1991), was conserved between all lyssavirus representatives with the exception of gt 3 isolate EU293117 and WCBV.

Sites of pathogenicity are shown in Figure 3.8.

- 330^{Lys} (Dietzschold *et al.,* 1983; Seif *et al.,* 1985), was conserved in all lyssavirus representatives with the exception of gt 2 isolate EU293110 and WCBV.
- 333^{Arg} (Dietzschold *et al.,* 1983; Seif *et al.,* 1985), was conserved in all the phylogroup I representatives but not in gt 2, gt 3 (both phylogroup II) and WCBV (phylogroup III).
- aa 242 (Takayama-Ito *et al.,* 2006b), with the exception of gt 1 this site is conserved in all lyssavirus and putative lyssavirus genotypes.
- aa 255 (Takayama-Ito *et al.,* 2006b), was conserved between gt 1 (exception M13215, ERA), gt 4, gt 5, gt 6, gt 7 and Khujand virus.
- aa 268 (Takayama-Ito *et al.,* 2006b), was conserved between all lyssavirus representatives with the exception of gt 6, gt 2, gt 3 and WCBV.

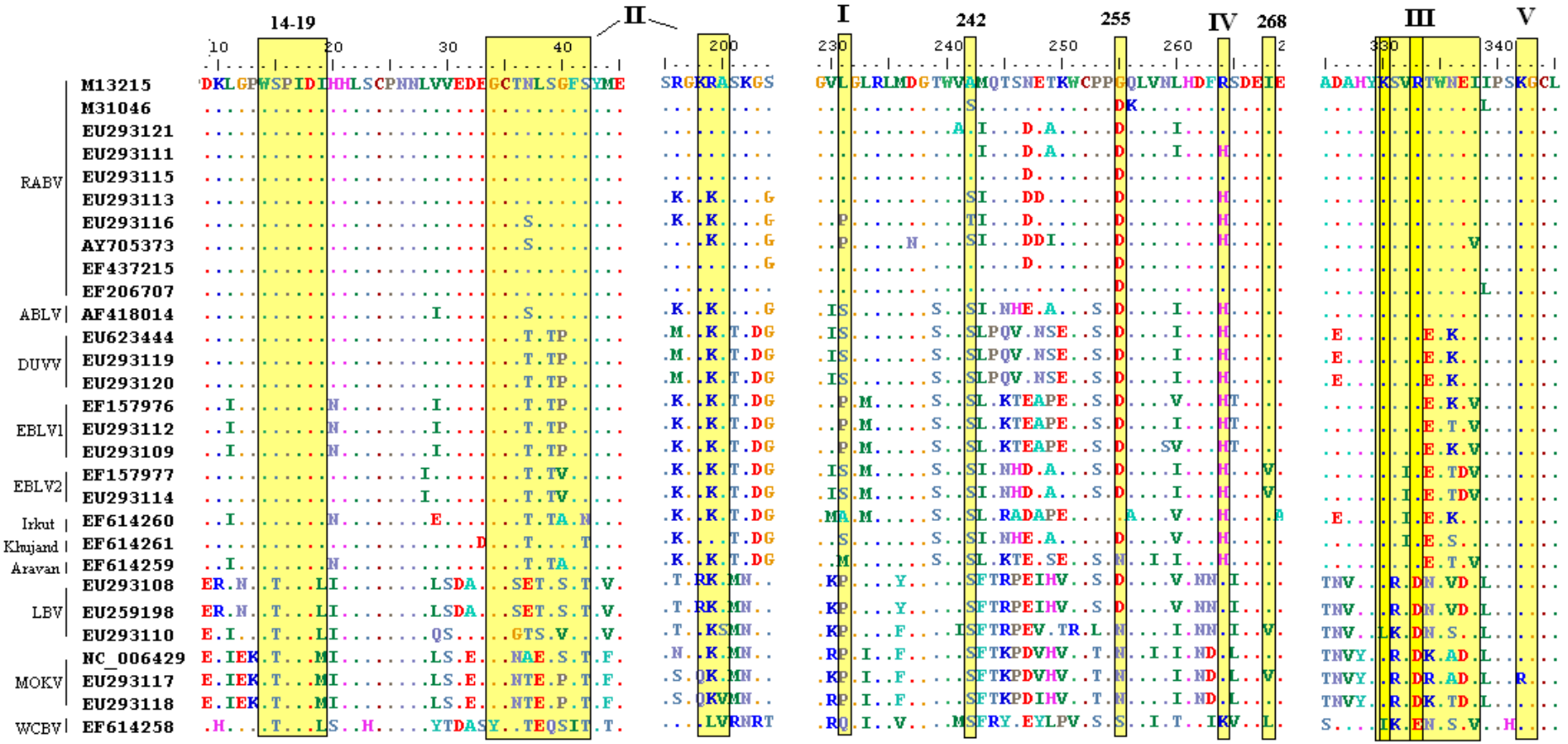


Figure 3.8 Multiple alignment of the lyssavirus genotypes and putative lyssavirus genotypes, indicating glycoprotein antigenic sites (I-V) and other sites implicated in pathogenicity. Dots represent identity to PV.

3.3.6 Conserved domains

3.3.6.1 Phosphoprotein

The P protein has shown to interact with LC8 (cytoplasmic dynein light chain) and of specific importance is the (K/R)XTGQT motif (Lo *et al.,* 2001) (Figure 3.9).

- (K/R)XTGQT (aa 145-149), was conserved between gt 4, gt 5 (exception EU157976), gt 6, gt 7, Aravan, Khujand and Irkut virus (KSTQT).

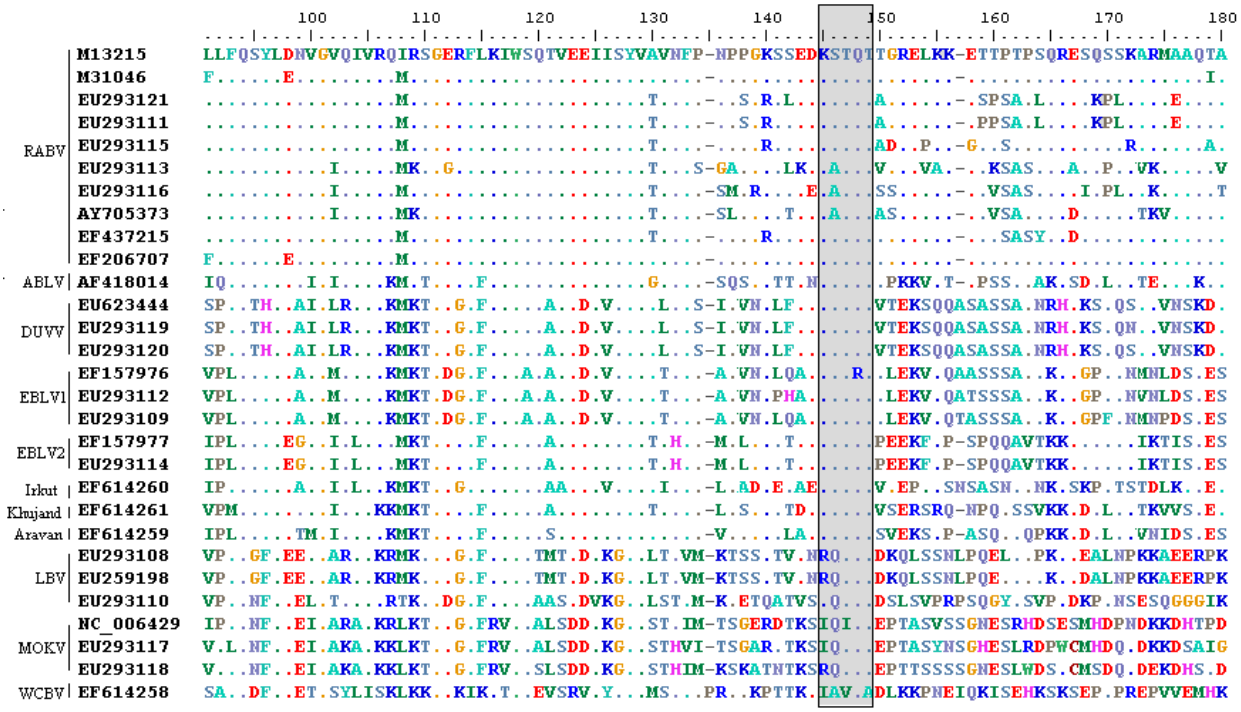


Figure 3.9 Phosphoprotein amino acid alignment. Dots represent identity to PV; hyphens are gaps for optimal alignment.

3.3.6.2 Matrix protein

Two motifs have been identified in the M protein (Jayakar *et al.,* 2000) (Figure 3.10).

- PPXY (aa 35-38), was conserved in all lyssavirus representatives with the exception of Khujand virus.

- PX(T/S)AP (aa 21-25), this site was conserved between gt 4 and gt 5 (PVSAP), gt 6, gt 7, Aravan and Khujand virus (LVSAP).

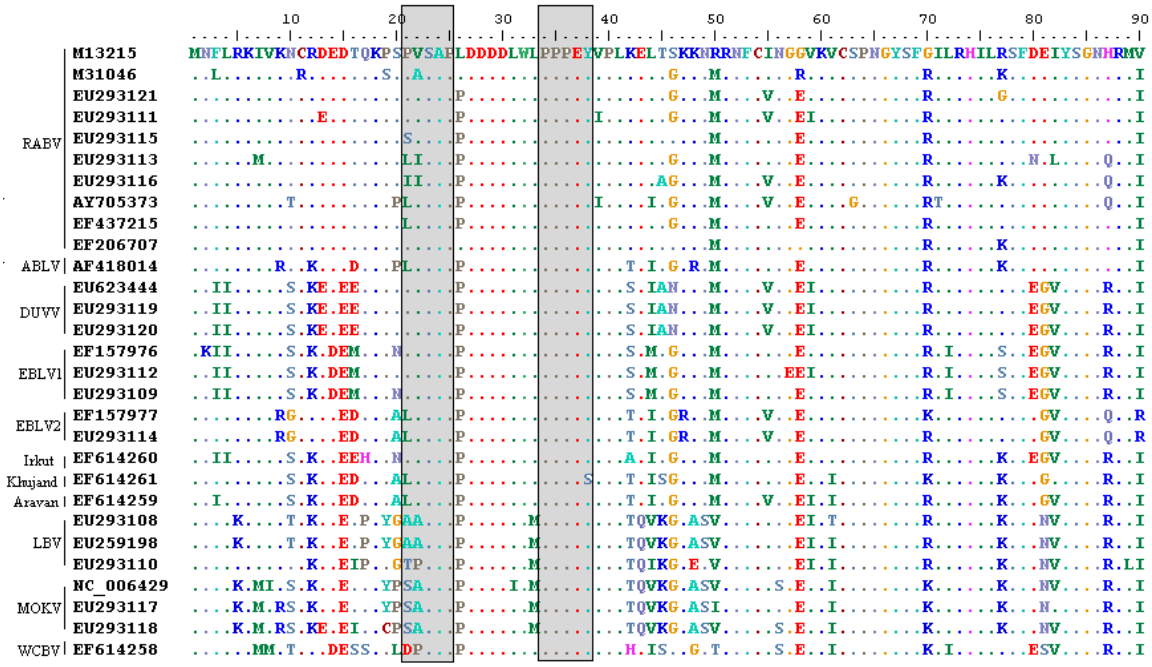


Figure 3.10 Matrix protein amino acid alignment. Dots represent identity to PV. Two classical late domain binding motifs were identified, PX(T/S)AP (aa 21-25) and PPXY (aa 35-38).

3.3.6.3 Polymerase protein

The six conserved domains (Poch *et al.,* 1990) in the polymerase gene were investigated (Appendix C) for both the lyssavirus (gt 1-7) and the putative lyssavirus genotypes (Irkut, Aravan, Khujand and WCBV).

- Block I, the invariant GHP residues (aa 373-376) were conserved within all the lyssavirus representatives.

- Block II, was found to contain the Pre-A motif which has shown to be involved in the positioning and binding of the RNA template. This motif was highly conserved within the lyssaviruses with only Irkut and MOKV having one substitution each.

- Block III, was found to contain four conserved regions in negative strand RNA viruses (A-D).

    - A, was highly conserved in all lyssavirus representatives with the exception of gt 1 isolate AY705373.

    - B, was conserved in all phylogroup I lyssaviruses (exception gt 1 EU 293111), gt 2 and gt 3 (phylogroup II) have a single amino acid substitution, whilst WCBV (phylogroup III) had two amino acid substitutions.

  - C and D, were conserved in all lyssavirus representatives.

- Block IV, was well conserved and was rich in proline as was described by Marston *et al.,* 2007.

- Block V, was relatively well conserved within the lyssaviruses and was rich in cysteine and histine.

- Block VI, was less conserved and as previously described by Marston *et al.,* 2007, all lyssavirus representatives had a conserved GDGSGG at position 1704-1708 and a lysine 19 residue downstream.

3.3.7 Phylogenetic analysis

3.3.7.1 Full genome

This was the first study to include all the current lyssavirus genotypes (gt 1-7) as well as the putative genotypes (Irkut, Aravan, Khujand and WCBV) in full genome phylogenetic analysis. Neighbour joining phylogenetic analysis (Figure 3.11) revealed three major branches, separating the lyssaviruses into the three previously defined phylogroups (Kuzmin *et al.,* 2005). The phylogroup I lyssaviruses cluster into their respective genotypes (gt 1, gt 4, gt 5, gt 6 and gt 7); putative genotypes Irkut, Aravan and Khujand also formed part of this group. Gt 2 and gt 3 cluster in phylogroup II whilst WCBV formed the third group representative of the possible phylogroup III. The isolates of gt 2 however, formed two very distinct lineages which when compared to the distance analysis (Section 3.3.4), further emphasized that these may be two separate genotypes (as previously suggested by Markotter *et al.,* 2008a). All branches were supported by high bootstrap values (92 - 100%) with the exception of the DUVV/EBLV1 branch which had a lower 71% bootstrap value. As was indicated by Delmas *et al.,* (2008), even though EBLV1 and EBLV2 both circulate in European bats, EBLV1 is most closely related to DUVV which circulates in African bats. The putative genotype Irkut was shown to cluster with these most similar lyssavirus genotypes (gt 4 and gt 5). The putative genotypes Aravan and Khujand clustered with EBLV2. These tree topologies are in agreement with the current classification of lyssaviruses as described in previous studies (Kuzmin *et al.,* 2003; Kuzmin *et al.,* 2005).

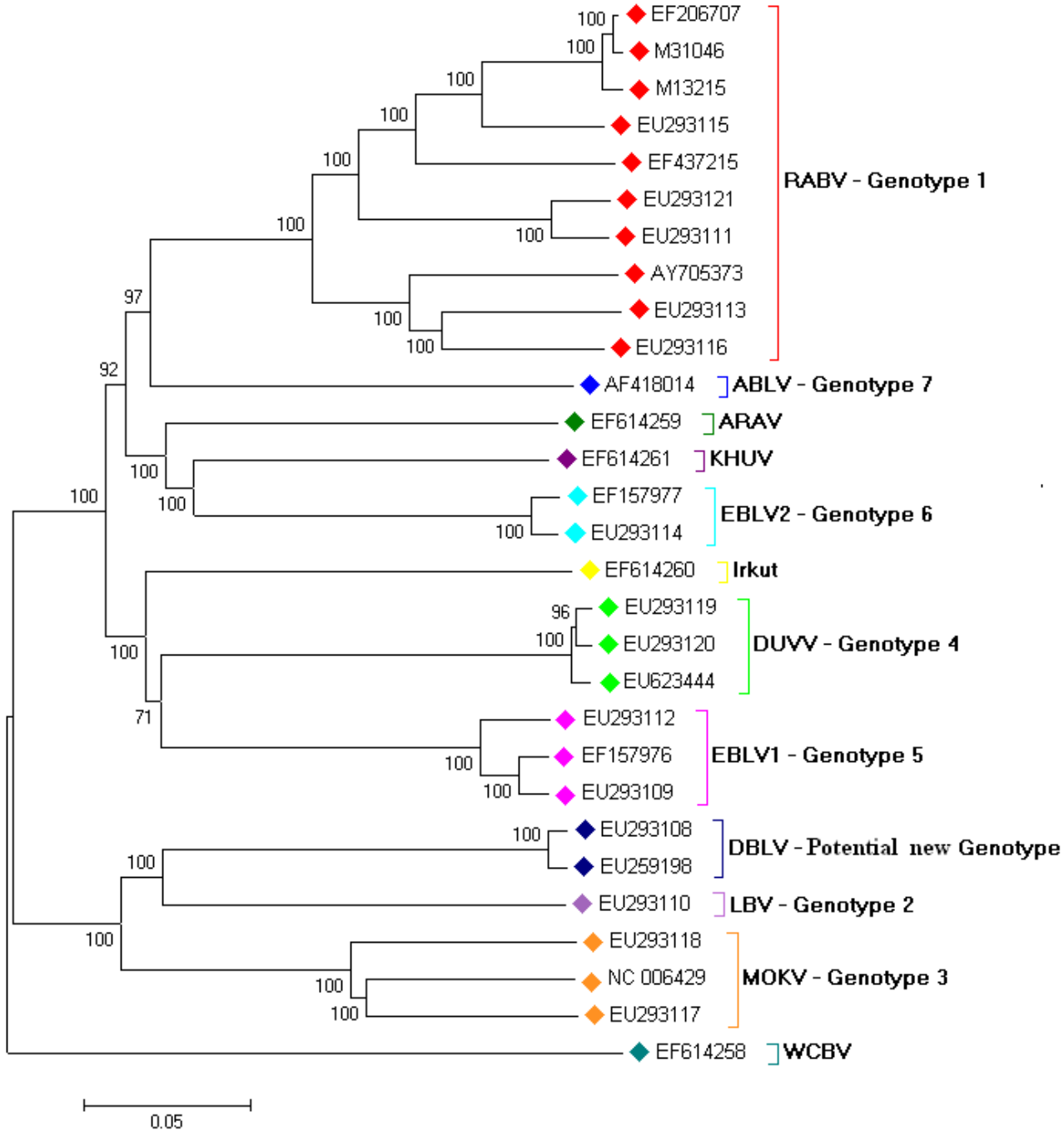


Figure 3.11 Neighbour joining tree of the full genome sequences of the seven lyssaviruses genotypes as well as the putative lyssavirus genotypes. GenBank accession numbers are indicated for each isolate. Bootstrap values are indicated at the nodes and branch lengths are drawn to scale.

3.3.7.2 Polymerase protein

Neighbour joining phylogenetic trees based on the full length gene were constructed for both the nucleotide (Figure 3.12a) and deduced amino acid sequences (Figure 3.12b) of the polymerase protein. A set of 29 complete L gene sequences of all the lyssavirus representatives (gt 1-7, Irkut, Aravan, Khujand and WCBV) was generated. Members of the lyssaviruses, clustered together in their respective genotypes, the grouping however corresponded to that obtained in M gene analysis (Section 2.3.4.3) where at the nucleotide level ABLV clustered with the EBLV2 group and not with RABV its closest relative. Bootstrap support for the EBLV2 and EBLV1 clusters (73% and 84% respectively) was relatively low when compared to the remaining clusters which had bootstrap support values of 93% and higher. Amino acid based tree topologies are however in agreement with the current classification of lyssaviruses as described in previous studies (Kuzmin *et al.,* 2003; Kuzmin *et al.,* 2005).

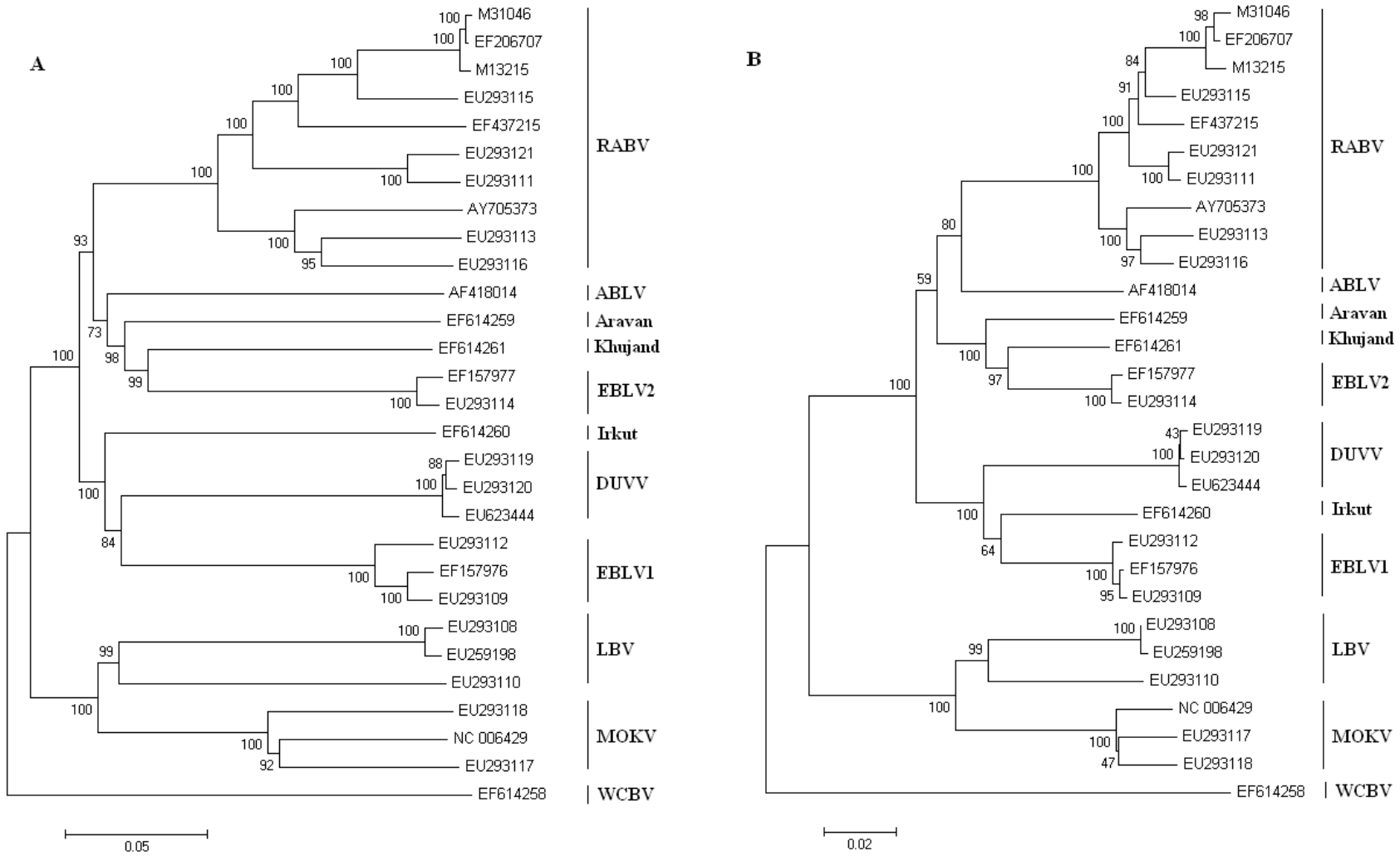


Figure 3.12 Neighbour joining tree of the L gene nucleotide (A) and deduced amino acid (B) sequences of the seven lyssaviruses genotypes as well as the putative lyssavirus genotypes. GenBank accession numbers are indicated for each isolate. Bootstrap values are indicated at the nodes and branch lengths are drawn to scale.

3.4 Discussion

The full genome sequence and genomic properties of the DUVVSA06 isolate were determined to further our understanding of the relationship between the Duvenhage virus isolates and the other lyssavirus representatives. The DUVVSA06 isolate is one of the few lyssavirus genomes to have an odd number of nucleotides as has previously been found for Khujand, LBV and several strains of RABV (Kuzmin *et al*., 2008b). It is as yet unclear whether either an odd or an even number of nucleotides has any critical importance in viral life history. The intergenic regions for N-P, P-M, M-G were typical for phylogroup I lyssaviruses (Marston *et al.,* 2007) the G-L IGS was however an atypical, 36 nt. This may be an indication that DUVV is more ancient than other lyssaviruses, as it was suggested by Kuzmin *et al.,* (2008b) that viruses with longer IGSs are more ancestral and that evolution in this genus would lead to shortening of the IGSs. Analysis of antigenic epitopes present on the N and G genes of the three DUVV isolates, showed them to be identical in these regions, DUVV was most similar to EBLV1 at all but one site; N site III (374-383) where DUVV and the other African lyssaviruses LBV and MOKV shared greater similarity. These results reiterate the close relationship between DUVV and EBLV1.

Several authors have suggested the L protein to be an ideal target for phylogenetic comparisons because of its significant conservation and neutral evolution (Le Mercier *et al.,* 1997; Warrilow *et al.,* 2002). Our analysis showed the DUVV isolates to be organized into the six conserved blocks (I-VI) with functional motifs previously detected in the polymerase of Mononegavirales (Poch *et al.,* 1990). Phylogenetic analysis however showed non conformant grouping of ABLV and EBLV2 at the nucleotide level, as was seen previously for the M gene (Section 2.3.4.3), these results are not in agreement with those found by Kuzmin *et al*., (2008b), this study was however based on only partial L gene sequences. Intragenotypic/intergenotypic ratios indicated that the L gene may be considered as an adequate candidate for lyssavirus classification.

Genomic terminal sequences were determined using the RNA circularization method of Kuzmin et al., (2008b). Full length terminal sequences saw the incorporation of five nucleotides at the site of ligation. Analysis of 70 clones showed a great deal of variation in the size of fragments obtained for the genomic termini, this truncation of fragments was also observed by Kuzmin *et al.,* (2008b). All variation was observed at the 3'end, where the shortest sequence proved to be missing 113 nucleotides. This degradation may have been as

a result of the presence of RNases and DNases (www.ambion.com/techlib/tn/91/9113.html), whether endogenous or environmental. With the large amount of 3' degradation observed, many non specific RNAs would have been present at the time of ligation. The ability of T4 RNA ligase to anneal any fragment with a 5' phosphoryl-terminated nucleic acid donor and a 3'hydroxyl-terminated nucleic acid receptor, could easily have led to our observed insertion. Due to the great level of conservancy among lyssaviruses termini, these base pairs where not considered to be part of either the 5' or 3' end, resulting in the observation that all DUVV isolates have the conserved 11 nucleotides that are complementary at opposing termini in most lyssaviruses (Marston *et al.,* 2007; Delmas *et al.,* 2008; Kuzmin *et al.,* 2008b).

This study was the first to analyze the relationship between the seven lyssavirus (RABV, LBV, MOKV, DUVV, EBLV1, EBLV2 and ABLV) and the putative lyssavirus genotypes (Irkut, Aravan, Khujand and WCBV) at the full genome level. Analysis of 29 genomes revealed the separation of the lyssaviruses into three major groups, previously described as phylogroups (Kuzmin *et al.,* 2005) and 12 component branches, representing the seven lyssavirus genotypes, four putative genotypes and the proposed new genotype within LBV (Delmas *et al.,* 2008; Markotter *et al.,* 2008a). These groupings are consistent with our analysis of the N, P and G genes, which were all in keeping with previous studies by Kuzmin *et al.,* (2003; 2005). The bootstrap support values for the full genome analysis were much higher than for any other gene, indicating that full genomes may be best for lyssavirus classification, as was suggested by Delmas *et al.,* (2008).

Percentage identity values obtained for the full genomes demonstrated DUVV to have the highest intragenotypic value (99%). Such high levels of conservancy have also been observed for EBLV, and could be linked to adaptation of these viruses to a particular host species (Davis *et al.,* 2005; Marston *et al.,* 2007). The limited number of isolates available for DUVV may have led to over estimation of these values. As previously described in Chapter 2, partial analysis of N gene (Section 2.3.3) showed the southern Africa isolates (used in this study) to be very similar to one another (97.7-100%), the east Africa isolate however, deviated from these by 11%. Thus accurate interpretation can only be achieved by the incorporation of full gene/genome sequences of all current Duvenhage virus isolates as well as the discovery and inclusion of new DUVV isolates.

Analysis of antigenic epitopes present on the N and G proteins showed all DUVV isolates to be conserved, with EBVL1 only differing from DUVV isolates at three sites; antigenic site:

III (N), I (G) and antigenic site III (G). For the isolates studied the binding site for the cytoplasmic light chain of dynein, LC8 (aa 143-148) of the P protein was found to be conserved between DUVV and EBLV1, with the exception of EBLV1 isolate EF157976. These observations reiterate the close relationship between these two lyssaviruses.

Our studies have shown that not all lyssavirus genes are equally adept to phylogenetic analysis as was previously suggested by Wu *et al.,* (2007). The variation observed in individual gene analyses, and the strong support shown for full genome analysis, leads us to believe that full genomes should be used for lyssavirus classification so as to avoid the potential bias of individual gene analyses.

CHAPTER 4
Conclusion

At the outset of this study beginning 2006, there were only three isolates of Duvenhage virus. These were the index human isolate and two isolates from insectivorous bats, the species of which has only once been positively identified (*Nycteris thebaica*). With the limited sequence information focusing primarily on only two of these isolates, as such very little was known about the molecular epidemiology of this African lyssavirus. Then early in the study a second human case (fourth DUVV isolate) was reported and this isolate became the focus point of our full genome analysis. At this time, DUVV had only been isolated from southern Africa (South Africa and Zimbabwe). Fascinatingly, in late 2007, the virus was isolated from a Dutch tourist who had been on vacation, in the east African country of Kenya. Analysis of all five DUVV isolates, which could only be based on partial gene analysis, saw phylogenetic and percentage identity analyses split the isolates into two disinct lineages. These two linegeas appear to have split earlier than those of the EBLVs and seem to be divided based on a geographical seperation that occurred some time ago.

In analyzing the complete N, P M, G and L gene and protein sequences, it was found that these genes are not all equal for phylogenetic analysis as previously suggested by Wu *et al*., (2007). On analysis of the potential role of individual genes in genotype classification, variation between the genes was again observed, with the N, P and G genes showing the most promising results. Although the N gene was found to be most consistent for lyssavirus genotype classification, the criteria upon which genotype distinction is based need to be revised. Intergenotypic and intragenotypic identities already show overlap at the amino acid level and as more lyssavirus isolates are discovered, the diversity of this genus will continue to expand and challenge current genotype classification criteria. Our study suggests that full genomes should be used for genotype classification so as to avoid the potential bias of individual gene analysis. However, to further investigate the taxonomy of the lyssaviruses, effort should be made to obtain more sequence data of all lyssavirus representatives, to better determine the diversity of this genus and to allow decisions regarding taxonomy to be based on this increased information.

The first full genome phylogenetic analysis of all the lyssavirus and putative lyssavirus genotypes was conducted in this study. Results obtained were in agreement with previous studies on the lyssavirus genotypes as described by Kuzmin *et al.,* (2003; 2005). Our analysis however showed much stronger support for the division of the lyssaviruses into the three previously defined phylogroups (Kuzmin *et al.,* 2005) and 12 clusters representative of the various lyssavirus genotypes. Analysis of the DUVV virus genomes showed them to

follow the lyssavirus 3' N, P, M, G, L 5' gene order, have 11 conserved nucleotides that are complementary at the genomic termini and have conserved transcription and termination signals. Our study showed DUVV to have an unusually long G-L IGS, which is suggestive of DUVV being a more ancestral virus within the *Lyssavirus* genus. Investigation of antigenic epitopes showed all DUVV isolates to be conserved, with very high levels of similarity being found between DUVV and EBLV1. Sites of pathogenicity were identical between these two viruses. The six conserved blocks of the L gene also showed DUVV and EBLV1 to be most similar to one another, these results emphasising the close relationship between the two lyssavirus genotypes.

On examination of the evolutionary forces acting on DUVV and its closest relative EBLV1, it was found that DUVV, as with all other lyssaviruses, is subject to strong selective constraints against amino acid change. Evolution between these two viruses may have been influenced by a population bottleneck which occurred during speciation. Nucleotide substitution patterns were indicative of viruses at an early stage of divergence. Although different insectivorous bat species have been associated with DUVV and EBLV1 many of these species are known to co-colonize roosts, where close contact allows for the spread of RNA viruses between species with relative ease. The ecology of these bat species and their propensity to migrate, increases their potential role in the spread and evolution of these viruses. Additional surveillance among bat species in Africa is needed to establish more information about the distribution, prevalence, genetic diversity and host species associated with DUVV so that informed decisions can be made regarding the potential threat of these viruses to public health.

APPENDIX

Appendix A

Sequence similarity between isolates was determined using the distance estimation program of MEGA 3.1 (Kumar *et al.* 2004). Genetic distances were calculated for both the nucleotide and deduced amino acid sequences of the N, P, M, G and L genes using the p-distance model (Nei and Gojobori, 1986). Full genome distances were also determined.

i) Full genomes nucleotide sequence similarity

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
	1	#M13215																													
	2	#M31046	98.58																												
	3	#EU293121	84.18	84.02																											
	4	#EU293111	84.15	84.02	96.41																										
RABV	5	#EU293115	92.06	91.96	85.15	85.15																									
	6	#EU293113	80.82	80.77	81.4	81.37	81.7																								
	7	#EU293116	81.89	81.94	82.26	82.22	82.73	89.62																							
	8	#AY705373	81.27	81.26	81.77	81.77	81.81	86.24	88.85																						
	9	#EF437215	87.7	87.59	84.84	84.66	88.27	81.6	82.41	81.76																					
	10	#EF206707	98.88	99.53	84.18	84.21	92.13	80.88	82.06	81.38	87.77																				
ABLV	11	#AF418014	72.77	72.87	73.35	73.37	72.99	73.47	73.41	73.16	73.59	72.91																			
	12	#EU623444	69.85	69.96	70.21	70.27	70.37	69.78	70.06	69.7	69.66	70.04	70.66																		
DUVV	13	#EU293119	69.76	69.9	70.15	70.25	70.33	69.81	69.99	69.67	69.59	69.95	70.57	98.82																	
	14	#EU293120	69.78	69.89	70.1	70.21	70.35	69.89	70.04	69.71	69.6	69.94	70.7	98.62	98.99																
	15	#EF157976	70.51	70.6	70.52	70.81	70.85	70.54	70.33	70.48	70.53	70.55	72.07	75.31	75.34	75.35															
EBLV1	16	#EU293112	70.38	70.48	70.61	70.87	70.88	70.55	70.34	70.32	70.75	70.45	71.79	75.23	75.3	75.25	95.96														
	17	#EU293109	70.44	70.5	70.57	70.79	70.81	70.65	70.41	70.6	70.62	70.45	72.06	75.33	75.31	75.33	98.1	95.62													
EBLV2	18	#EF157977	71.29	71.37	72.13	72.25	71.9	71.78	72.14	71.63	71.91	71.46	74.14	71.95	71.94	71.95	72.99	73.12	72.98												
	19	#EU293114	71.5	71.59	72.21	72.27	72.22	71.82	72.25	71.78	71.96	71.66	74.42	71.74	71.73	71.73	73.03	73.15	72.93	98.25											
Irkut	20	#EF614260	69.97	70	70.09	70.22	70.49	70.41	70.69	70.24	70.17	70.05	71.29	73.22	73.17	73.29	75.54	75.33	75.55	72.23	72.32										
Khujand	21	#EF614261	71.83	71.9	71.85	72.01	72.4	71.42	71.85	71.5	71.63	71.92	74.72	72.12	72	72.08	73.51	73.48	73.51	78.35	78.16	72.88									
LBV	22	#EU293108	65.64	65.74	66	66.03	66.21	65.96	66.16	65.86	65.8	65.83	66.51	66.27	66.23	66.23	66.93	66.85	66.76	65.91	65.79	66.45	66.48								
Aravan	23	#EF614259	71.38	71.44	71.54	71.72	71.84	71.59	71.79	71.34	71.46	71.5	73.65	72.07	72.07	72.14	74.16	74.03	74.09	76.07	76.07	72.79	76.74	66.8							
LBV	24	#EU259198	65.72	65.85	66.11	66.15	66.34	65.94	66.13	65.82	65.85	65.94	66.67	66.23	66.21	66.21	67.01	66.84	66.79	65.88	65.73	66.55	66.41	98.83	66.79						
	25	#EU293110	65.76	65.84	66.03	66.12	66.05	66.39	66.32	66.13	65.85	65.9	66.52	66.19	66.18	66.21	66.64	66.57	66.5	66.65	66.69	66.79	66.09	75.71	66.86	75.63					
	26	#NC_006429	65.87	66.03	65.84	65.91	66.27	65.92	66.12	66.03	65.88	66.03	66.19	66.21	66.2	66.19	66.03	66.11	66.16	65.96	66	66.54	65.71	73.31	66.25	73.22	72.57				
MOKV	27	#EU293117	65.84	66.05	65.99	66.08	66.28	65.59	65.63	65.69	65.77	66.06	66.05	65.58	65.63	65.64	65.79	65.91	65.83	66.4	66.48	66.28	65.79	73.47	66.21	73.36	72.48	87.33			
	28	#EU293118	65.87	66.03	65.97	66.05	66.27	65.88	65.61	65.74	65.6	66.05	66.08	65.67	65.59	65.65	66.08	66.21	66.23	66.31	66.52	66.78	65.45	72.75	66.45	72.81	72.56	86.38	86.32		
WCBV	29	#EF614258	63.92	63.97	64.16	64.21	64.49	63.7	64.16	64.26	63.91	64.07	64.3	64.64	64.79	64.69	64.43	64.43	64.53	64.27	64.25	64.16	64.35	64.65	64.53	64.59	64.74	64.32	64.07	64.19	

ii) Genetic distances for the nucleoprotein nt 8-406

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
EBLV2	1	#AY863407																		
	2	#EF157976	71.61																	
	3	#AY863375	71.61	99.5																
EBLV1	4	#AY863380	71.86	99.25	99.75															
	5	#AY863397	72.11	94.97	95.48	95.23														
	6	#AY863383	72.61	94.47	94.97	94.72	96.48													
	7	#EF157977	97.24	71.11	71.11	71.36	71.36	72.36												
EBLV2	8	#AY863405	96.73	70.6	70.6	70.85	71.36	71.61	97.49											
	9	#AY863406	95.48	73.62	73.62	73.37	73.37	74.62	95.73	95.23										
	10	#DUVV_kenya	73.12	76.88	77.14	77.39	76.38	75.13	71.61	72.36	72.36									
	11	#AY062080_DUVVZIM86	74.12	77.14	76.88	77.14	77.14	76.63	72.61	73.37	73.87	89.2								
DUVV	12	#EU623437_DUVVSA71	73.87	77.89	77.64	77.89	77.89	77.39	72.36	73.12	73.62	89.7	98.99							
	13	#EU623438_DUVVSA81	74.12	77.14	76.88	77.14	77.14	76.63	72.61	73.37	73.87	89.2	100	98.99						
	14	#EU623444_DUVVSA06	73.87	77.64	77.39	77.64	77.14	76.63	72.36	73.12	73.62	88.94	97.74	98.74	97.74					
Khujand	15	#AY262024_Khujand	78.14	71.86	71.61	71.86	71.61	71.86	79.4	77.64	77.64	73.62	72.11	72.61	72.11	72.11				
Aravan	16	#AY262023_Aravan	76.63	75.38	75.38	75.13	75.13	75.13	76.63	76.13	76.63	74.62	73.12	73.37	73.12	73.12	77.14			
Irkut	17	#AY333112_Irkut	76.38	76.88	76.63	76.88	76.38	76.63	75.63	75.38	75.13	77.39	77.64	77.64	77.64	78.14	73.62	71.86		
RABV	18	#M13215_RABV_PV	70.1	69.1	68.84	69.1	69.85	69.35	70.1	71.36	70.6	70.1	70.1	70.1	70.1	69.35	72.61	70.6	73.12	

iii) Nucleoprotein nucleotide sequence similarity

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
DUVV	1	#EU623438_DUVVSA81																			
	2	#EU623444_DUVVSA06	98.74																		
	3	#EU623437_DUVVSA71	99.26	99.19																	
EBLV2	4	#EF157977	76.05	75.98	75.91																
	5	#AY863405	76.27	76.35	76.27	97.93															
	6	#AY863406	75.91	76.13	76.05	95.57	95.64														
	7	#AY863407	75.83	75.91	75.83	96.53	96.53	95.79													
EBLV1	8	#EF157976	79.23	79.38	79.31	76.42	76.64	77.46	76.35												
	9	#AY863375	78.94	79.08	79.01	76.05	76.27	77.09	75.98	99.26											
	10	#AY863380	79.08	79.23	79.16	76.35	76.57	77.24	76.27	99.41	99.11										
	11	#AY863397	79.45	79.6	79.82	76.79	77.31	77.68	76.57	96.01	95.79	96.01									
	12	#AY863383	79.31	79.45	79.53	76.79	77.09	77.75	76.64	96.01	95.79	96.01	98								
Khujand	13	#AY262024_Khujand	75.91	75.98	76.13	79.75	79.31	79.16	79.08	77.9	77.46	77.83	77.38	77.16							
Aravan	14	#AY262023_Aravan	77.9	78.05	78.2	76.87	76.79	77.24	77.01	78.2	77.9	78.12	78.2	77.97	79.08						
Irkut	15	#AY333112_Irkut	77.83	78.12	78.05	77.61	77.53	76.72	77.53	78.79	78.34	78.86	78.64	78.42	76.2	76.35					
RABV	16	#M13215_RABV_PV	74.65	74.35	74.5	74.21	74.94	74.35	74.5	75.02	74.87	75.24	75.09	75.17	75.46	75.09	74.5				
	17	#EF206707_RABV_ERA	74.8	74.65	74.8	74.58	75.31	74.58	74.72	75.39	75.24	75.61	75.46	75.54	75.39	75.39	74.72	99.11			
	18	#M31046_RABV_SAD_B19	74.65	74.5	74.65	74.58	75.31	74.58	74.72	75.39	75.24	75.61	75.46	75.68	75.24	75.54	74.65	98.97	99.7		
ABLV	19	#AF418014_ABLV	77.16	76.87	77.09	76.27	76.5	76.94	76.5	77.53	77.31	77.46	77.24	77.16	78.05	76.57	75.98	78.49	78.64	78.79	

iv) Nucleoprotein amino acid sequence similarity

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
DUVV	1	#EU623438_DUVVSA81,																			
	2	#EU623444_DUVVSA06,	99.56																		
	3	#EU623437_DUVVSA71,	100	99.56																	
EBLV2	4	#EF157977,	86.44	86.44	86.44																
	5	#AY863405,	86.22	86.22	86.22	99.33															
	6	#AY863406,	86.22	86.22	86.22	98	97.78														
	7	#AY863407,	86.44	86.44	86.44	99.56	99.33	98													
EBLV1	8	#EF157976,	92.89	92.44	92.89	87.78	87.56	88	87.78												
	9	#AY863375,	92	91.56	92	86.89	86.67	87.11	86.89	99.11											
	10	#AY863380,	92.89	92.44	92.89	87.78	87.56	88	87.78	100	99.11										
	11	#AY863397,	93.33	92.89	93.33	88.22	88	88	88.22	99.33	98.44	99.33									
	12	#AY863383,	93.11	92.67	93.11	88	87.78	87.78	88	99.11	98.22	99.11	99.78								
Khujand	13	#AY262024_Khujand,	89.33	88.89	89.33	91.33	90.67	90.44	90.89	89.78	88.89	89.78	90.22	90							
Aravan	14	#AY262023_Aravan,	91.78	91.33	91.78	88.89	88.67	88.89	88.89	92	91.33	92	92	91.78	92.67						
Irkut	15	#AY333112_Irkut,	90.44	90.22	90.44	86.89	86.67	86.22	86.89	92.22	91.33	92.22	92.67	92.44	88	90.67					
RABV	16	#M13215_RABV_PV,	87.56	87.33	87.56	87.33	86.89	86.89	87.11	88.22	87.33	88.22	88.22	88	90.22	89.33	86.89				
ABLV	17	#AF418014_ABLV,	90	89.56	90	88	87.78	88.44	88	90	89.11	90	90	89.78	92.22	92.44	87.78	92			
RABV	18	#EF206707_RABV_ERA,	88	87.78	88	87.78	87.33	87.33	87.56	88.67	87.78	88.67	88.67	88.44	90.67	89.78	87.33	99.11	92.44		
	19	#M31046_RABV_SAD_B19,	88	87.78	88	87.78	87.33	87.33	87.56	88.67	87.78	88.67	88.67	88.44	90.67	89.78	87.33	99.11	92.44	99.56	

v) Phosphoprotein nucleotide sequence similarity

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
DUVV	1	#EU623444_DUVVSA06															
	2	#EU623439_DUVVSA81	98.99														
	3	#EU623436_DUVVSA71	99.1	98.99													
EBLV2	4	#AF049121	65.85	65.17	65.73												
EBLV1	5	#EF157976	70.1	70.55	70.21	67.97											
	6	#AF049113	69.76	70.21	69.88	67.75	99.22										
	7	#AF049117	69.76	70.21	69.88	67.64	98.54	98.66									
EBLV2	8	#EF157977	64.61	63.94	64.5	95.52	68.09	67.86	67.86								
RABV	9	#M13215	63.72	63.94	64.05	66.63	62.37	62.15	62.37	66.18							
Aravan	10	#AY262023_Aravan	66.29	66.52	66.07	72.56	70.21	69.99	70.55	72.23	67.19						
Irkut	11	#AY333112_Irkut	67.3	67.75	67.53	66.41	71.78	71.44	71.56	66.41	64.39	65.85					
Khujand	12	#AY262024_Khujand	65.51	65.4	65.4	74.13	67.75	67.64	67.53	73.46	69.32	74.58	66.74				
RABV	13	#EF206707	64.28	64.5	64.61	66.97	62.15	61.93	62.04	66.52	98.99	66.85	64.73	69.76			
	14	#M31046	64.17	64.39	64.5	67.08	62.26	62.04	62.04	66.63	98.54	66.97	64.73	69.88	99.33		
ABLV	15	#AF418014_ABLV	61.48	61.81	61.81	68.42	64.28	64.05	63.72	68.09	69.32	69.2	63.16	68.76	69.09	69.2	

vi) Phosphoprotein amino acid sequence similarity

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
DUVV	1	#EU623444_DUVVSA06															
	2	#EU623439_DUVVSA81	100														
	3	#EU623436_DUVVSA71	98.99	98.99													
EBLV2	4	#AF049121	59.6	59.6	59.93												
EBLV1	5	#EF157976	71.38	71.38	71.38	67.68											
	6	#AF049113	71.72	71.72	71.72	67.34	98.65										
	7	#AF049117	72.39	72.39	72.39	68.01	98.99	98.99									
EBLV2	8	#EF157977	58.59	58.59	58.92	98.32	67	66.67	67.34								
RABV	9	#M13215	58.25	58.25	58.59	68.01	63.3	62.96	63.3	68.01							
Aravan	10	#AY262023_Aravan	62.96	62.96	62.96	78.45	72.39	72.05	72.73	77.44	67.34						
Irkut	11	#AY333112_Irkut	67	67	67	67	70.71	71.04	71.38	66.33	64.98	67					
Khujand	12	#AY262024_Khujand	62.96	62.96	62.96	78.79	68.69	68.69	69.02	77.78	69.7	80.13	68.69				
RABV	13	#EF206707	58.25	58.25	58.59	68.69	62.63	62.29	62.63	68.69	97.98	67.68	64.65	69.7			
	14	#M31046	57.91	57.91	58.25	69.02	62.96	62.63	62.96	69.02	97.31	67.68	64.65	70.03	98.65		
ABLV	15	#AF418014_ABLV	58.59	58.59	58.59	71.38	62.63	62.63	62.96	71.38	75.08	70.03	64.31	72.05	74.41	74.41	

vii) Matrix protein nucleotide sequence similarity

			1	2	3	4	5	6	7	8	9	10	11	12
DUVV	1	#EU623440_DUVVSA71												
	2	#EU623441_DUVVSA81	99.34											
	3	#EU623444_DUVVSA06	98.52	98.52										
EBLV2	4	#EF157977	74.84	75.16	74.67									
EBLV1	5	#EF157976	80.1	80.26	80.26	77.63								
RABV	6	#EF206707	75.49	75.82	76.15	76.15	76.81							
	7	#M13215	74.84	75.16	75.49	75.99	76.64	97.86						
	8	#M31046	75	75.33	75.33	75	76.48	98.36	96.22					
ABLV	9	#AF418014_ABLV	76.64	77.14	77.14	79.44	76.81	75.82	75	74.84				
Aravan	10	#AY262023_Aravan	80.43	80.59	80.92	80.76	79.11	75.99	75.66	75.33	81.09			
Irkut	11	#AY333112_Irkut	78.29	78.62	78.29	76.81	79.61	73.03	72.2	72.53	75.66	78.95		
Khujand	12	#AY262024_Khujand	77.3	77.3	76.97	78.95	78.95	75.16	74.01	74.84	79.61	80.92	79.77	

viii) Matrix protein amino acid sequence similarity

			1	2	3	4	5	6	7	8	9	10	11	12
DUVV	1	#EU623440_DUVVSA71												
	2	#EU623441_DUVVSA81	99.5											
	3	#EU623444_DUVVSA06	99.5											
EBLV2	4	#EF157977	85.57	86.07	86.07									
EBLV1	5	#EF157976	92.54	93.03	93.03	85.07								
RABV	6	#EF206707	82.09	82.59	82.59	81.09	81.59							
	7	#M13215	81.59	82.09	82.09	81.09	80.6	94.53						
	8	#M31046	80.1	80.6	80.6	79.1	80.1	96.52	91.04					
ABLV	9	#AF418014_ABLV	86.07	86.57	86.57	88.56	87.06	88.56	87.06	86.57				
Aravan	10	#AY262023_Aravan	91.54	92.04	92.04	91.04	90.55	83.08	81.59	81.59	92.04			
Irkut	11	#AY333112_Irkut	92.04	92.54	92.54	86.57	93.03	82.59	81.09	81.09	88.56	92.54		
Khujand	12	#AY262024_Khujand	88.56	89.05	89.05	90.55	89.05	83.58	82.59	81.59	92.54	96.52	91.04	

ix) Glycoprotein nucleotide sequence similarity

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
DUVV	1	#EU623444_DUVVSA06																				
	2	#EU623443_DUVVSA81	98.66																			
	3	#EU623442_DUVVSA71	98.79	98.85																		
EBLV1	4	#EF157976	73.15	73.09	73.41																	
EBLV2	5	#AY863343	72.19	72.26	71.81	73.6																
EBLV1	6	#AY863318	72.96	72.9	73.15	99.55	73.6															
	7	#AY863321	73.09	73.02	73.28	99.36	73.47	99.43														
	8	#AY863330	72.58	72.51	72.77	96.43	73.6	96.49	96.36													
	9	#AY863335	72.64	72.77	72.7	95.98	73.85	96.05	95.85	98.21												
EBLV2	10	#EF157977	72.58	72.64	72.19	72.83	96.81	72.83	72.7	72.83	72.96											
	11	#AY863346	72.26	72.32	71.87	73.02	96.43	73.02	72.9	73.02	73.21	98.98										
	12	#AF298145	71.68	71.75	71.3	72.58	95.54	72.58	72.45	72.58	72.77	98.09	99.11									
	13	#AY863345	72.77	72.83	72.39	73.02	95.28	73.02	72.9	72.83	73.15	95.15	94.77	94.13								
RABV	14	#M13215	65.69	65.37	65.37	67.54	69.96	67.41	67.67	67.03	66.65	69.96	69.83	69.39	69.9							
	15	#EF206707	65.37	65.05	65.05	67.22	69.77	67.09	67.35	66.77	66.39	69.77	69.64	69.2	69.71	98.34						
	16	#M31046	65.37	65.05	65.05	67.22	69.77	67.09	67.35	66.77	66.39	69.77	69.64	69.2	69.71	98.09	99.74					
ABLV	17	#AF418014_ABLV	68.24	68.62	68.3	70.47	72.39	70.34	70.66	70.15	69.9	72.32	72.45	71.87	73.09	71.11	71.24	71.24				
Aravan	18	#AY262023_Aravan	70.15	70.22	70.34	73.47	75.45	73.21	73.34	73.66	73.6	75.32	75.7	75	75.06	68.88	69.01	69.13	71.43			
Irkut	19	#AY333112_Irkut	69.9	70.28	70.03	73.66	71.3	73.66	73.72	73.47	73.98	70.85	71.11	70.41	70.92	66.26	65.94	66.01	68.69	70.15		
Khujand	20	#AY262024_Khujand	71.75	71.56	71.62	72	78.76	71.94	71.94	71.68	71.87	78.64	78.64	78.06	78.7	70.15	70.09	70.09	73.34	74.55	70.03	

x) Glycoprotein amino acid sequence similarity

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
DUVV	1	#EU623444_DUVVSA06																				
	2	#EU623443_DUVVSA81	99.43																			
	3	#EU623442_DUVVSA71	99.81	99.23																		
EBLV1	4	#EF157976	79.89	79.89	79.89																	
EBLV2	5	#AY863343	79.31	79.69	79.12	80.27																
EBLV1	6	#AY863318	79.69	79.69	79.5	99.43	80.08															
	7	#AY863321	80.08	80.08	79.89	99.81	80.46	99.62														
	8	#AY863330	80.08	80.08	79.89	98.47	80.65	98.28	98.66													
	9	#AY863335	80.08	80.27	79.89	98.47	80.84	98.28	98.66	99.62												
EBLV2	10	#EF157977	79.69	80.08	79.5	80.27	99.23	80.08	80.46	80.65	80.84											
	11	#AY863346	79.5	79.89	79.31	80.27	99.04	80.08	80.46	80.65	80.84	99.81										
	12	#AF298145	78.16	78.54	77.97	78.93	97.51	78.74	79.12	79.31	79.5	98.28	98.47									
	13	#AY863345	79.31	79.69	79.12	80.46	97.89	80.27	80.65	80.84	81.03	97.51	97.32	96.55								
RABV	14	#M13215	69.73	69.73	69.54	70.69	74.14	70.5	70.88	70.88	70.88	74.52	74.33	72.99	73.18							
	15	#EF206707	69.16	69.35	68.97	70.5	73.95	70.31	70.69	70.69	70.69	74.33	74.14	72.8	72.99	97.13						
	16	#M31046	69.35	69.54	69.16	70.5	74.14	70.31	70.69	70.69	70.69	74.52	74.33	72.99	73.18	96.55	99.43					
ABLV	17	#AF418014_ABLV	74.14	74.33	73.95	76.44	80.27	76.25	76.63	76.63	76.44	80.46	80.27	78.93	80.08	76.05	76.63	76.82				
Aravan	18	#AY262023_Aravan	79.31	79.69	79.12	83.14	84.87	82.95	83.33	83.72	83.91	85.25	85.25	83.91	84.1	73.75	73.56	73.75	79.5			
Irkut	19	#AY333112_Irkut	75.48	76.05	75.29	80.84	77.01	80.65	81.03	80.84	81.03	76.82	76.82	75.48	76.63	67.43	67.43	67.82	73.56	77.59		
Khujand	20	#AY262024_Khujand	77.78	77.97	77.59	79.31	87.36	79.12	79.5	79.12	79.31	87.74	87.55	86.21	87.16	76.82	76.63	76.82	82.38	83.72	76.63	

xi) Polymerase protein nucleotide sequence similarity

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
	1 #M13215																													
	2 #M31046	99.08																												
	3 #EU293121	84.77	84.53																											
	4 #EU293111	84.53	84.3	96.3																										
RABV	5 #EU293115	92.43	92.4	85.82	85.4																									
	6 #EU293113	81.78	81.71	82.39	82.33	82.33																								
	7 #EU293116	82.87	82.89	83.41	83.27	83.56	90.36																							
	8 #AY705373	82.36	82.28	82.84	82.94	82.56	86.98	89.6																						
	9 #EF437215	88.12	87.98	85.4	85.08	88.23	82.45	83.36	82.69																					
	10 #EF206707	99.25	99.61	84.69	84.47	92.46	81.84	83.05	82.4	88.12																				
ABLV	11 #AF418014	74.33	74.33	75.21	75.24	74.32	75.6	75.24	74.95	75.32	74.4																			
	12 #EU623444	72.89	73	73.6	73.55	73.44	73.46	73.38	73.17	72.88	73.06	73.27																		
DUVV	13 #EU293119	72.89	73.02	73.63	73.69	73.5	73.46	73.36	73.24	72.92	73.06	73.14	98.84																	
	14 #EU293120	73	73.11	73.64	73.69	73.57	73.55	73.44	73.28	72.94	73.16	73.32	98.79	99.11																
	15 #EF157976	73.55	73.69	73.39	73.75	73.93	73.61	73.27	73.55	73.21	73.68	74.4	77.01	77.06	77.11															
EBLV1	16 #EU293112	73.53	73.71	73.57	73.77	74.1	73.68	73.32	73.38	73.52	73.71	74.3	76.94	77	77.01	96														
	17 #EU293109	73.57	73.64	73.57	73.79	73.99	73.93	73.52	73.85	73.5	73.64	74.4	77.11	77.08	77.17	98.15	95.69													
EBLV2	18 #EF157977	73.85	73.9	74.62	74.69	74.29	75.02	74.87	74.46	74.29	73.99	76.64	74.46	74.65	74.58	74.71	74.87	74.84												
	19 #EU293114	74.18	74.24	74.76	74.69	74.71	75.05	75.1	74.57	74.38	74.32	77.04	74.18	74.37	74.32	74.73	74.76	74.76	98.21											
Irkut	20 #EF614260	73.41	73.35	73.39	73.43	74.02	74.02	74.02	73.85	73.66	73.47	74.33	75.48	75.35	75.49	77.19	77.01	77.12	74.54	74.55										
Khujand	21 #EF614261	74.44	74.41	74.43	74.65	75.05	74.08	74.38	74.07	74.16	74.46	76.76	74.73	74.62	74.9	75.45	75.51	75.37	79.65	79.6	75.74									
LBV	22 #EU293108	69.74	69.68	70.37	70.42	70.18	70.32	70.46	70.2	69.92	69.84	70.64	70.76	70.71	70.82	71.15	71.17	70.93	70.93	70.76	70.73	70.93								
Aravan	23 #EF614259	74.13	74.08	74.15	74.16	74.48	74.35	74.16	74.02	74.15	74.22	75.99	74.07	74.08	74.27	76.01	75.9	75.78	77.73	77.59	75.59	78.03	71.2							
	24 #EU259198	69.87	69.84	70.57	70.64	70.37	70.31	70.37	70.1	69.99	69.99	70.86	70.79	70.78	70.84	71.29	71.22	70.98	70.93	70.76	70.81	70.86	98.73	71.18						
LBV	25 #EU293110	69.73	69.57	70.17	70.17	69.98	70.84	70.39	70.42	69.55	69.73	70.28	70.59	70.46	70.51	70.86	70.76	70.76	70.78	70.84	70.98	70.39	77.11	71.18	76.94					
	26 #NC_006429	70.51	70.43	69.98	70.18	70.76	70.64	70.62	70.39	70.43	70.57	70.43	70.9	71.01	71.11	70.42	70.5	70.75	70.67	70.81	71.26	70.15	75.63	70.82	75.43	75.24				
MOKV	27 #EU293117	70.04	70.03	69.96	69.98	70.2	69.84	69.77	69.9	69.85	70.17	70.26	70.23	70.23	70.37	69.87	69.98	70.04	70.57	70.67	70.48	69.96	75.78	70.51	75.62	74.66	87.65			
	28 #EU293118	70.18	70.09	70.29	70.23	70.45	70.86	70.32	70.18	69.74	70.21	70.46	70.13	70.07	70.29	70.17	70.24	70.39	70.68	70.97	71.14	69.82	74.99	70.84	75.01	75.35	86.95	86.84		
WCBV	29 #EF614258	67.31	67.28	67.78	67.72	67.96	67.72	67.8	67.77	67.47	67.42	68.02	68.25	68.25	68.36	68.52	68.52	68.58	67.78	67.99	68.13	68.36	68.3	68.25	68.29	68.47	67.94	67.25	67.74	

xii) Polymerase protein deduced amino acid sequence similarity

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
	1 #M13215																													
	2 #M31046	98.68																												
	3 #EU293121	95.67	95.44																											
	4 #EU293111	95.86	95.63	99.15																										
RABV	5 #EU293115	96.85	96.71	97.04	97.18																									
	6 #EU293113	93.7	93.56	94.87	94.97	95.01																								
	7 #EU293116	94.07	94.03	95.67	95.58	95.58	97.55																							
	8 #AY705373	93.7	93.65	95.06	95.06	95.25	96.24	96.99																						
	9 #EF437215	96.24	96	96.85	96.99	97.7	95.06	95.63	95.11																					
	10 #EF206707	99.01	99.29	95.81	96	96.99	93.89	94.36	93.89	96.28																				
ABLV	11 #AF418014	88.1	88.05	89.18	89.28	89.37	89.23	89.7	89.09	88.95	88.43																			
	12 #EU623444	84.34	84.38	84.9	85.14	85.23	84.67	84.81	84.85	84.95	84.57	85.61																		
DUVV	13 #EU293119	84.34	84.38	84.85	85.09	85.18	84.67	84.9	84.9	84.81	84.52	85.56	99.53																	
	14 #EU293120	84.38	84.43	84.9	85.14	85.23	84.81	84.9	84.95	84.95	84.57	85.65	99.62	99.62																
	15 #EF157976	85.84	85.79	86.36	86.45	86.92	86.22	86.55	86.27	86.45	85.94	87.58	91.11	91.16	91.2															
EBLV1	16 #EU293112	85.84	85.79	86.41	86.45	86.92	86.27	86.45	86.17	86.55	85.94	87.63	91.2	91.11	91.25	99.39														
	17 #EU293109	85.75	85.65	86.27	86.36	86.83	86.22	86.45	86.12	86.45	85.84	87.58	90.97	90.92	91.06	99.62	99.29													
EBLV2	18 #EF157977	86.88	86.78	87.91	88.19	88.05	87.54	87.91	87.39	87.68	87.06	90.73	86.74	86.69	86.78	89.04	89.18	88.99												
	19 #EU293114	86.83	86.74	87.86	88.01	88.05	87.54	87.86	87.35	87.72	87.02	91.02	86.78	86.74	86.83	89.13	89.28	89.09	99.39											
Irkut	20 #EF614260	85.42	85.32	86.27	86.55	86.41	86.36	86.55	86.08	85.94	85.56	87.25	89.23	89.32	89.46	92.76	92.76	92.71	88.66	88.66										
Khujand	21 #EF614261	87.3	87.25	88.15	88.29	88.29	87.63	87.82	87.54	88.01	87.54	91.58	87.44	87.44	87.54	89.79	89.89	89.7	93.89	93.98	89.42									
LBV	22 #EU293108	79.73	79.82	80.2	80.29	80.34	80.24	80.86	80.15	80.15	80.06	80.43	80.39	80.34	80.39	81.61	81.61	81.42	80.95	80.95	81.61	81.19								
Aravan	23 #EF614259	87.16	87.16	87.86	88.15	88.24	87.96	88.1	87.39	88.15	87.44	90.08	87.54	87.44	87.63	90.26	90.4	90.17	92.24	92.38	89.75	93.13	81.66							
	24 #EU259198	79.54	79.63	80.01	80.1	80.15	80.06	80.67	79.92	79.96	79.87	80.24	80.24	80.2	80.24	81.37	81.37	81.19	80.81	80.81	81.37	81	99.72	81.47						
LBV	25 #EU293110	79.59	79.63	80.06	80.1	80.29	80.24	80.53	80.15	79.82	79.68	80.43	80.29	80.24	80.2	81.7	81.7	81.51	81.28	81.42	81.51	81	91.39	81.47	91.16					
	26 #NC_006429	78.93	78.93	79.16	79.26	79.49	79.4	79.77	79.26	79.26	79.16	79.73	79.49	79.49	79.59	80.81	80.81	80.67	80.24	80.43	81	80.06	89.09	80.01	88.85	88.33				
MOKV	27 #EU293117	78.6	78.6	79.02	79.21	79.16	79.16	79.59	79.16	79.07	78.83	79.68	79.26	79.26	79.35	80.9	80.95	80.76	80.43	80.57	80.95	80.01	88.57	80.01	88.33	88.19	96.66			
	28 #EU293118	78.83	78.88	79.02	79.21	79.4	79.44	79.82	79.3	79.07	79.12	79.49	78.93	78.93	79.02	80.53	80.57	80.39	80.29	80.43	80.86	79.92	88.85	80.01	88.62	88.57	96.66	96.66		
WCBV	29 #EF614258	74.74	74.88	75.31	75.21	75.49	75.12	75.35	75.26	75.16	75.02	76.58	75.59	75.54	75.49	76.39	76.43	76.34	76.48	76.62	76.15	76.67	76.2	76.39	76.01	76.53	75.54	75.59	75.82	

Appendix B

Origin of isolates used in molecular epidemiology analysis

VIRUS CODE	GENOTYPE	SPECIES ISOLATED FROM	YEAR OF ISOLATION	GEOGRAPHIC LOCATION	REFERENCE/ SOURCE	GENBANK ACCESSION NUMBER
RAVMMGN	1	Rabies virus, laboratory strain Pasteur			Tordo *et al.*, 1988	M13215 (Full genome)
ERA	1	Rabies virus, laboratory strain Evelyn-Rokitnicki-Abelseth			Unpublished	EF206707
8743THA	1	*Homo sapiens*	1983	Thailand	Delmas *et al.*, 2008	EU293121 (Full genome)
8764THA	1	*Homo sapiens*	1983	Thailand	Delmas *et al.*, 2008	EU293111 (Full genome)
9147FRA	1	Fox	1991	France	Delmas *et al.*, 2008	EU293115 (Full genome)
9001FRA	1	Dog bitten by bat	1990	France	Delmas *et al.*, 2008	EU293113 (Full genome)
9704ARG	1	*Tadarida brasilliensis*	1997	Argentina	Delmas *et al.*, 2008	EU293116 (Full genome)
SHBRV-18	1	*Lasionycteris noctivagans*	1983	USA	Faber *et al.*, 2004	AY705373 (Full genome)
NNV-RAB-H	1	*Homo sapiens*	2006	India	Unpublished	EF437215 (Full genome)
SADB19	1	Rabies virus, laboratory strain Street Alabama Dufferin B-19			Conzelmann *et al.*, 1990	M31046 (Full genome)
8619NGA	2	*Eidolon helvum*	1956	Nigeria	Delmas *et al.*, 2008	EU293110 (Full genome)
0406SEN	2	*Eidolon helvum*	1985	Senegal	Delmas *et al.*, 2008	EU293108 (Full genome)
KE131	2	*Eidolon helvum*	2007	Kenya	Kuzmin *et al.*, 2008a	EU259198 (Full genome)
MOKV	3	Cat	1981	Zimbabwe	Le Mercier *et al.*, 1997	NC_006429 (Full genome)
86100CAM	3	Shrew	1974	Cameroon	Delmas *et al.*, 2008	EU239117 (Full genome)
86101RCA	3	Rodent	1981	Central African Republic	Delmas *et al.*, 2008	EU293118 (Full genome)
DUVVSA06	4	*Homo sapiens*	2006	South Africa	This study	EU623444 (Full genome)
DUVVSA81 (94286SA)	4	Bat	1981	South Africa	This study	EU623438(N) EU623439(P) EU623441(M) EU623443(G)
					Delmas *et al.*, 2008	EU293120 (Full genome)
DUVVSA71 (86132SA)	4	*Homo sapiens*	1971	South Africa	This study	EU623436(P) EU623437(N) EU623440(M) EU623443(G)
					Delmas *et al.*, 2008	EU293119 (Full genome)

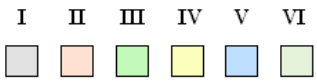
VIRUS CODE	GENOTYPE	SPECIES ISOLATED FROM	YEAR OF ISOLATION	GEOGRAPHIC LOCATION	REFERENCE/ SOURCE	GENBANK ACCESSION NUMBER
DUVVKenya	4	*Homo sapiens*	2007	Kenya	van Thiel *et al*., 2008	Received from Dr. M. Schutten
RV131	4	*Nycteris thebiaca*	1986	Zimbabwe	Johnson *et al*., 2002	AY062080(N)
86132SA	4	*Homo sapiens*	1971	South Africa	Davis *et al*., 2005	AY996323(N) AY996321(G)
					Nadin-Davis *et al*., 2001	AF049115(P)
					Delmas *et al.,* 2008	EU293119 (Full genome)
94286SA	4	*Miniopterus schreibersii*	1981	South Africa	Davis *et al*., 2005	AY996324(N) AY996322(G)
					Nadin-Davis *et al*., 2001	AF049120(P)
					Delmas *et al.,* 2008	EU293120 (Full genome)
02010DEN	5	*Eptesicus serotinus*	1995	Denmark	Davis *et al*., 2005	AY863375(N) AY863318(G)
02016DEN	5	Sheep	2002	Denmark	Davis *et al*., 2005	AY863380(N) AY863321(G)
V002	5	*Eptesicus serotinus*	1986	Denmark	Nadin-Davis *et al*., 2001	AF049113(P)
V023	5	*Eptesicus serotinus*	1986	Denmark	Nadin-Davis *et al*., 2001	AF049117(P)
9367HOL	5	*Eptesicus serotinus*	1992	Netherlands	Davis *et al*., 2005	AY863383(N) AY863335(G)
0002FRA	5	*Eptesicus serotinus*	2000	France	Davis *et al*., 2005	AY863397(N) AY863330(G)
9395GER	5	*Eptesicus serotinus*	1968	Germany	Marston *et al*., 2007	EF157976 (Full genome)
8918FRA	5	*Eptesicus serotinus*	1989	France	Delmas *et al.,* 2008	EU293112 (Full genome)
03002FRA	5	*Eptesicus serotinus*	2003	France	Delmas *et al.,* 2008	EU293109 (Full genome)
9018HOL	6	*Myotis dasycneme*	1986	Holland	Badrane *et al*., 2001	AF298145(G) RVU22847(N)
					Delmas *et al.,* 2008	EU293114 (Full genome)
9367HOL	5	*Eptesicus serotinus*	1992	Netherlands	Davis *et al*., 2005	AY863383(N) AY863335(G)
94112HOL	6	*Myotis dasycneme*	1989	Netherlands	Davis *et al*., 2005	AY863405(N) AY863346(G)

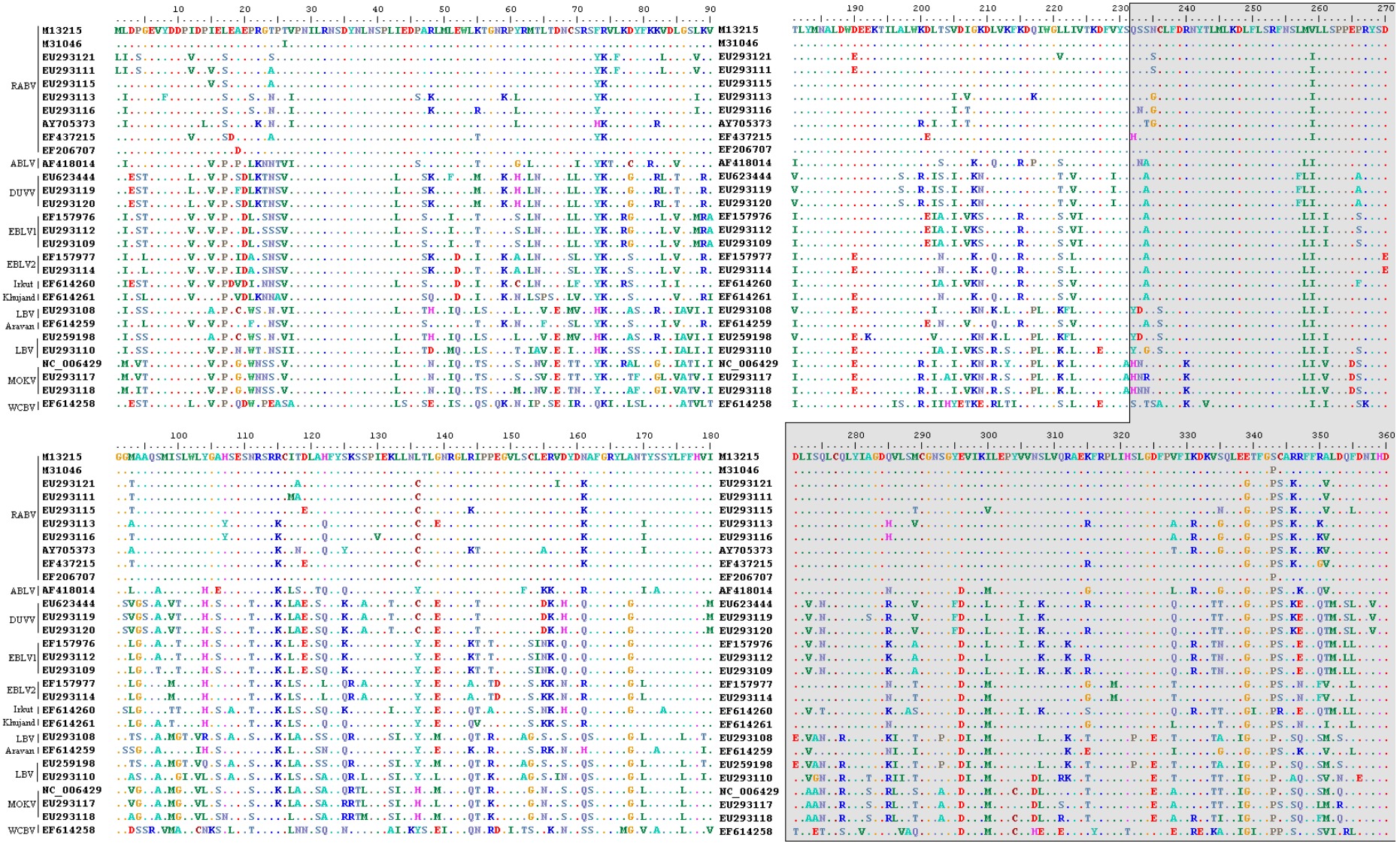
VIRUS CODE	GENOTYPE	SPECIES ISOLATED FROM	YEAR OF ISOLATION	GEOGRAPHIC LOCATION	REFERENCE/ SOURCE	GENBANK ACCESSION NUMBER
9007FIN	6	*Homo sapiens*	1986	Finland	Davis *et al*., 2005	AY863406(N) AY863345(G)
9337SWI	6	*Myotis dasycneme*	1993	Switzerland	Davis *et al*., 2005	AY863407(N) AY863343(G)
V286	6	*Myotis daubentonii*	1992	Switzerland	Nadin-Davis *et al*., 2001	AF049121(P)
RV1333	6	*Homo sapiens*	2002	United Kingdom	Marston *et al*. 2007	EF157977 (Full genome)
ABLh	7	*Homo sapiens*	1998	Australia	Warrilow *et al*., 2002	AF418014 (Full genome)
Irkut		*Murina leucogaster*	2002	Russia	Kuzmin *et al*., 2005	AY333112 (N-G)
					Kuzmin *et al*., 2008b	EF614260 (Full genome)
West Caucasian bat virus		*Miniopterus schreibersi*	2002	Russia	Kuzmin *et al*., 2005	AY333113 (N-G)
					Kuzmin *et al*., 2008b	EF614258 (Full genome)
Khujand		*Myotis daubentonii*	2001	Tajikistan	Kuzmin *et al*., 2003	AY262024 (N-G)
					Kuzmin *et al.,* 2008b	EF614261 (Full genome)
Aravan		*Myotis blythi*	1991	Kyrgyzstan	Kuzmin *et al*., 2003	AY262023 (N-G)
					Kuzmin et al., 2008b	EF614259 (Full genome)

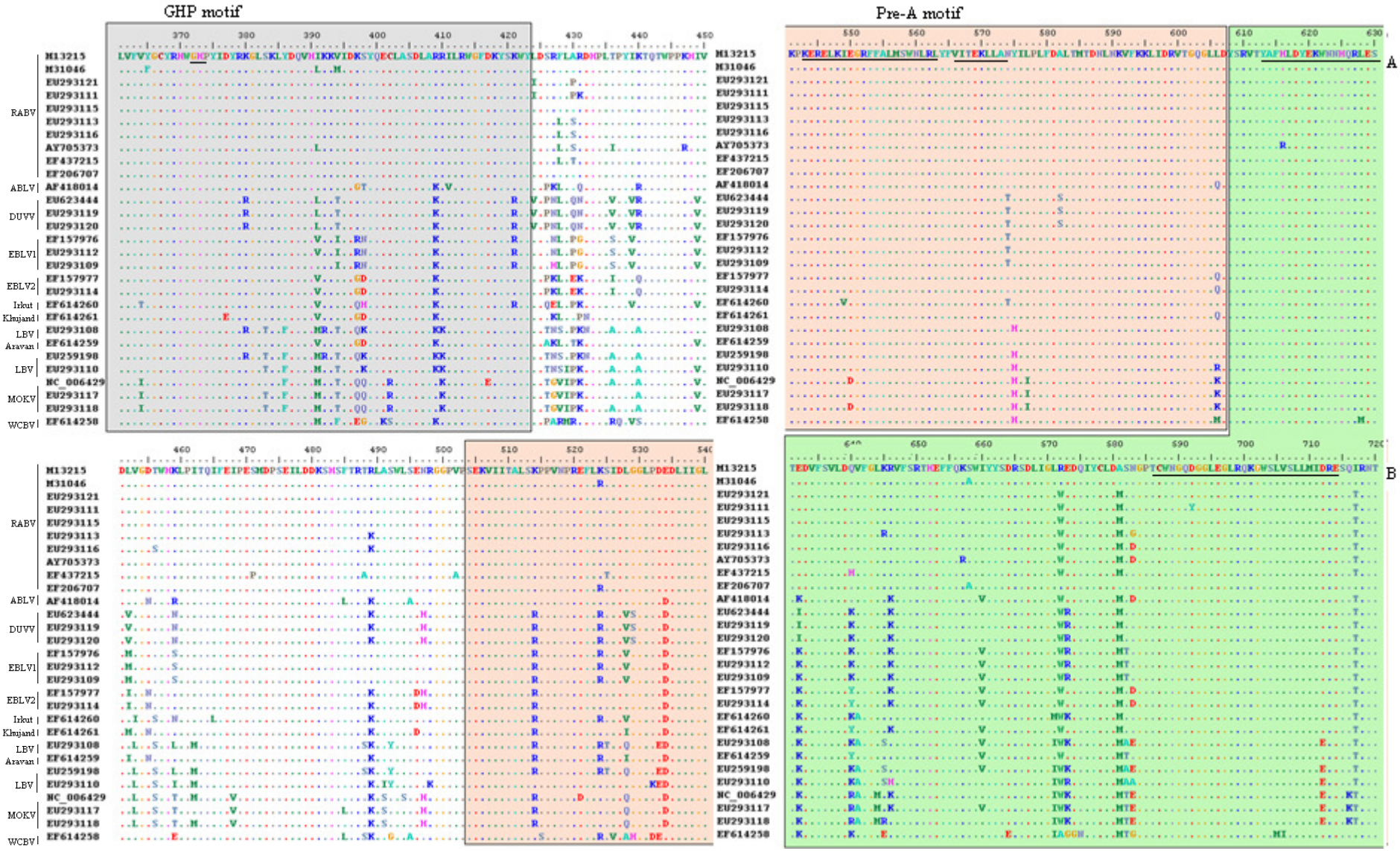
Appendix C

Domains in the L gene were investigated through multiple alignments which were carried out using the ClustalW subroutine (Thompson *et al*., 1994), which forms part of the Bioedit program. Dots represent identity to PV; hyphens are gaps for optimal alignment.

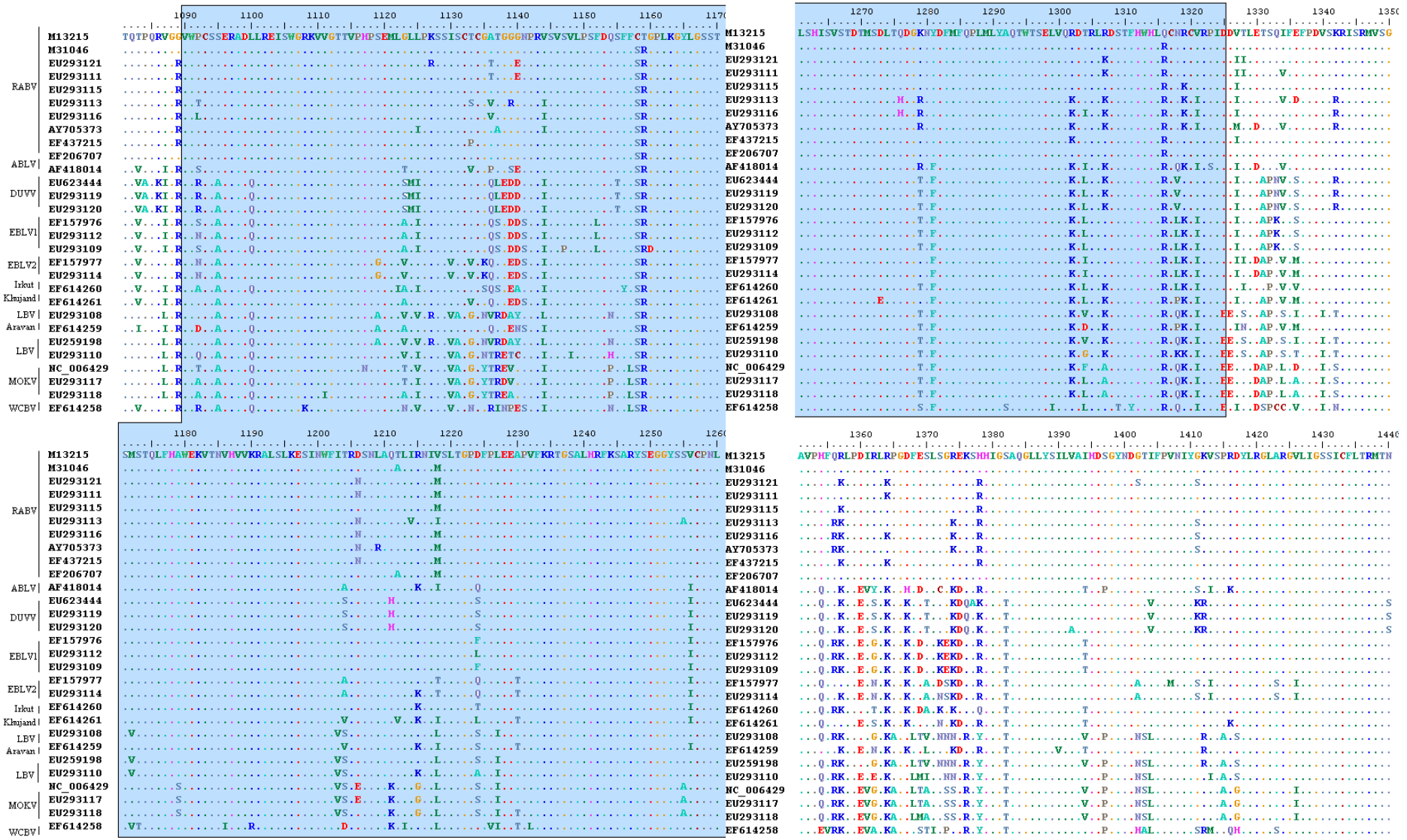
The six conserved domains (I-VI) are boxed (Poch *et al*., 1990).

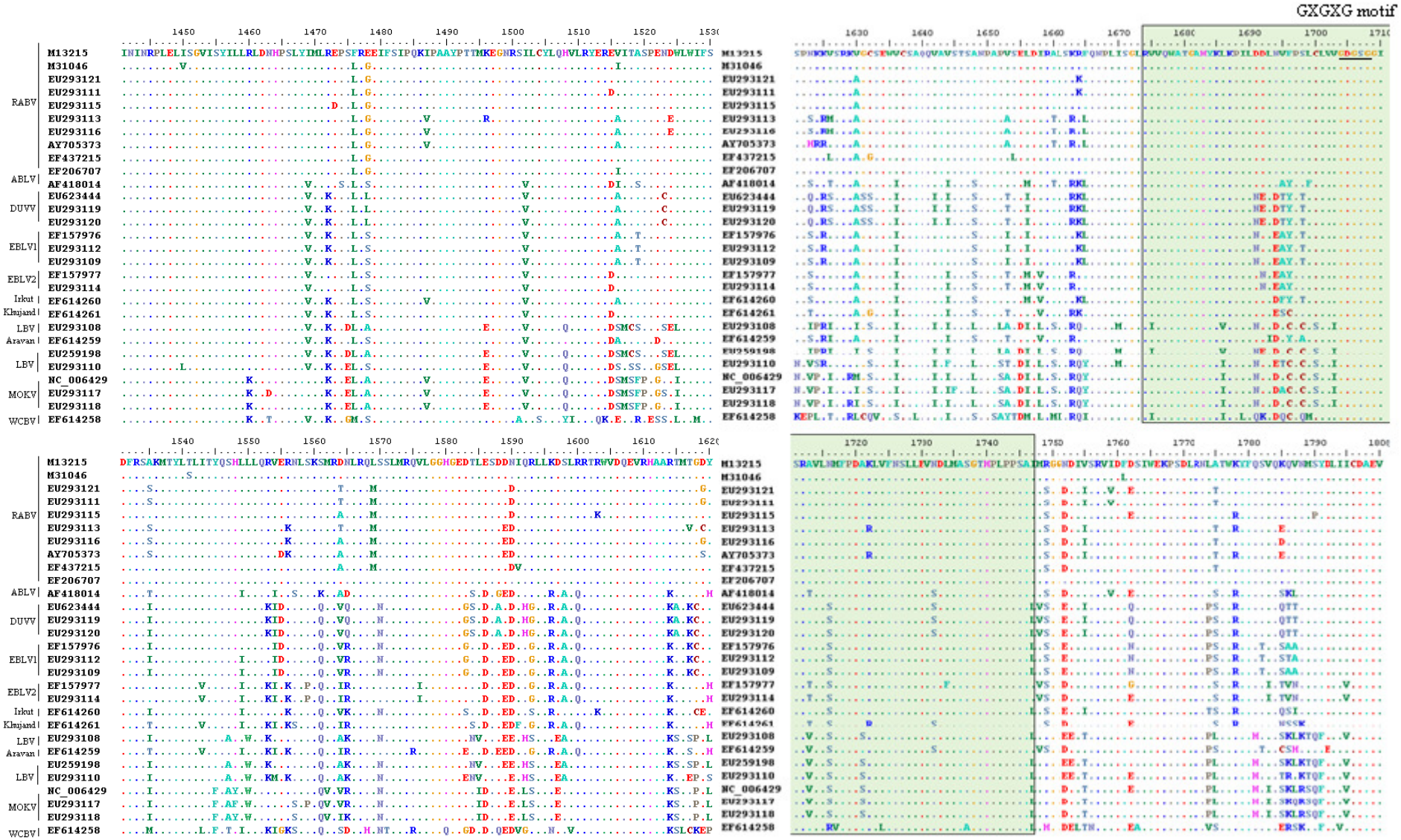


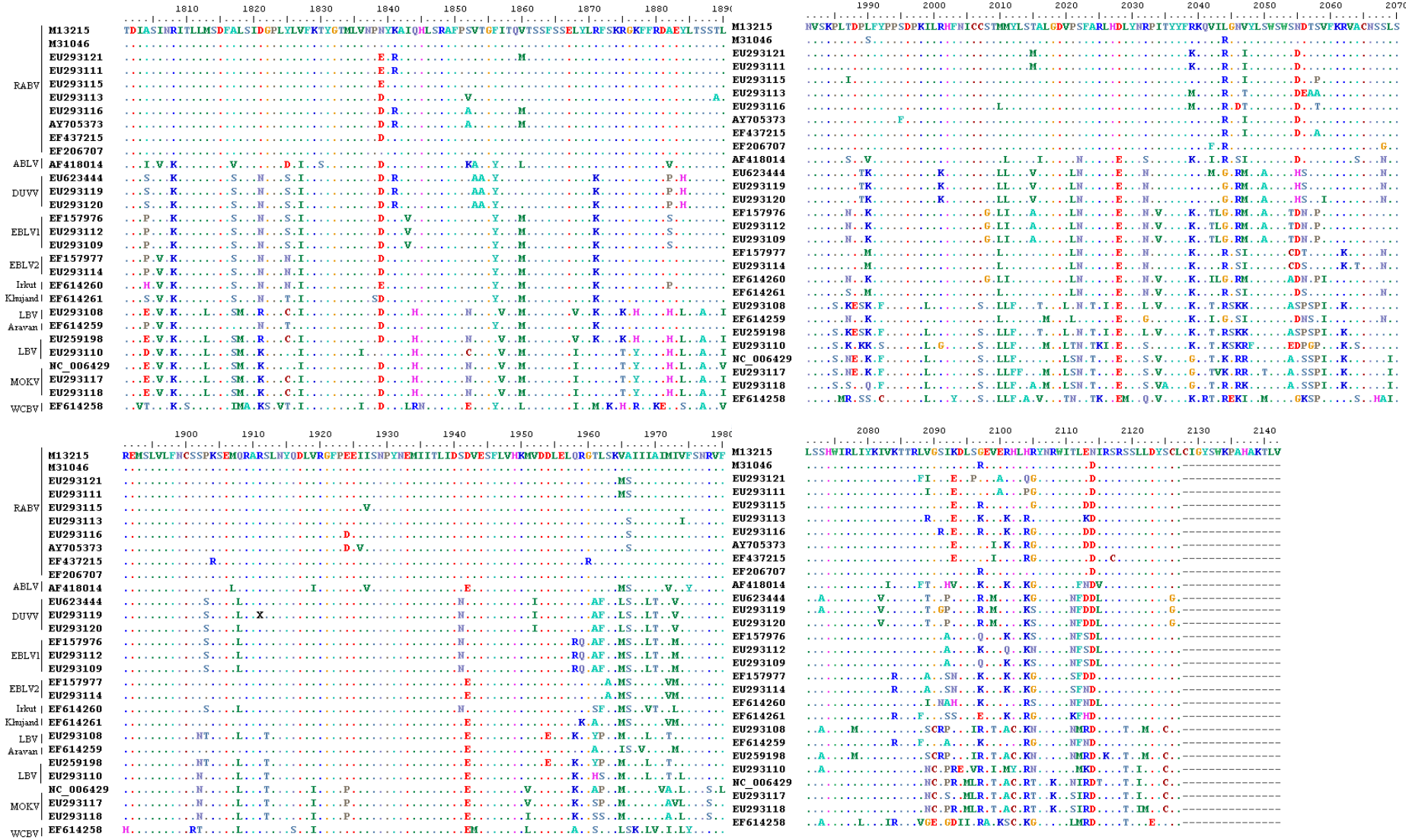



GHP motif
Pre-A motif
A
B
M13215
M31046
EU293121
EU293111
EU293115
EU293113
EU293116
AY705373
EF437215
EF206707
AF418014
EU623444
EU293119
EU293120
EF157976
EU293112
EU293109
EF157977
EU293114
EF614260
EF614261
EU293108
EF614259
EU259198
EU293110
NC_006429
EU293117
EU293118
EF614258
RABV
ABLV
DUVV
EBLV1
EBLV2
Irkut
Khujand
LBV
Aravan
MOKV
WCBV


C
D
M13215
M31046
EU293121
EU293111
EU293115
EU293113
EU293116
AY705373
EF437215
EF206707
AF418014
EU623444
EU293119
EU293120
EF157976
EU293112
EU293109
EF157977
EU293114
EF614260
EF614261
EU293108
EF614259
EU259198
EU293110
NC_006429
EU293117
EU293118
EF614258
RABV
ABLV
DUVV
EBLV1
EBLV2
Irkut
Khujand
LBV
Aravan
MOKV
WCBV


RABV
ABLV
DUVV
EBLV1
EBLV2
Irkut
Khujand
LBV
Aravan
MOKV
WCBV
M13215
M31046
EU293121
EU293111
EU293115
EU293113
EU293116
AY705373
EF437215
EF206707
AF418014
EU623444
EU293119
EU293120
EF157976
EU293112
EU293109
EF157977
EU293114
EF614260
EF614261
EU293108
EF614259
EU259198
EU293110
NC_006429
EU293117
EU293118
EF614258


GXGXG motif
RABV
ABLV
DUVV
EBLV1
EBLV2
Irkut
Khujand
LBV
Aravan
MOKV
WCBV


1810 1820 1830 1840 1850 1860 1870 1880 1890
1990 2000 2010 2020 2030 2040 2050 2060 2070
1900 1910 1920 1930 1940 1950 1960 1970 1980
2080 2090 2100 2110 2120 2130 2140
RABV
ABLV
DUVV
EBLV1
EBLV2
Irkut
Khujand
LBV
Aravan
MOKV
WCBV
M13215
M31046
EU293121
EU293111
EU293115
EU293113
EU293116
AY705373
EF437215
EF206707
AF418014
EU623444
EU293119
EU293120
EF157976
EU293112
EU293109
EF157977
EU293114
EF614260
EF614261
EU293108
EF614259
EU259198
EU293110
NC_006429
EU293117
EU293118
EF614258

REFERENCES

Allworth A., Murray K. and Morgan J. (1996). A human case of encephalitis due to a lyssavirus recently identified in fruit bats. *Comm. Dis. Intell.* 20: 504

Amengual B., Whitby J.E., King A., Serra Cobo J. and Bourhy H. (1997). Evolution of European bat lyssaviruses. *J Gen. Virol*. 78: 2319-2328

Amengual B., Bourhy H., Lopez-Roig M. and Serra-Cobo J. (2007). Temporal dynamics of European bat lyssavirus type 1 and survival of *Myotis myotis* bats in natural colonies. *PLoS ONE* 2, e566. Published online 2007 June 27. doi: 10.1371/journal.pone.0000566

Arai Y.T., Kuzmin I.V., Kameoka Y. and Botvinkin A.D. (2003). New lyssavirus genotype from the Lesser Mouse-eared Bat (*Myotis blythi*), Kyrgyzstan. *Emerg. Infect. Dis*. 9(3): 333-337

Auewarakul P. (2005). Composition bias and genome polarity of RNA viruses. *Virus Res*. 109: 33-37

Badrane H. and Tordo N. (2001). Host switching in Lyssavirus history from the Chiroptera to the Carnivora Orders. *J. Virol.* 17: 8096-8104

Badrane H., Bahloul C., Perrin P. and Tordo N. (2001). Evidence of two lyssavirus phylogroups with distinct pathogenicity and immunogenicity. *J Virol.* 75: 3268-3276

Baer G.M. (1994). Rabies – an historical perspective. *Infect. Agents and Dis.* 3: 168-180

Bahloul C., Jacob Y., Tordo N., Perrin P. (1998). DNA-based immunization for exploring the enlargement of immunological cross-reactivity against the lyssaviruses. *Vaccine* 16(4): 417-425.

Benmansour A., Leblois H., Coulon P. (1991). Antigenicity of rabies virus glycoprotein. *J Virol.* 65:4198-4203

Bingham J., Foggin C.M., Wandeler A.I., Hill F.W. (1999). The epidemiology of rabies in Zimbabwe. 1. Rabies in dogs (Canis familiaris). *Onderstepoort J Vet. Res*. 66(1): 1-10

Bingham J. (2005). Canine rabies ecology in southern Africa. *Emerg. Infect. Dis*. 11: 1337-1342

Bishop G.C., Durrheim D.N., Kloeck P.E., Godlonton J.D., Bingham J. and Speare J. (2003). *Rabies: Guide for the medical, veterinary and allied professions*. Rabies advisory group, South African Department of Agriculture and Health.

Bizri A.R., Azar A., Salam N. and Mokhbat J. (2000). Human rabies in Lebanon: lessons for control. *Epidemiol. Infect*. 125: 175-179

Blumberg B.M., Giorgi C and Kolakofsky D. (1983). N protein of vesicular stomatitis virus selectively encapsidates leader RNA *in vitro. Cell* 32: 559-567

Botvinkin A.D., Poleschuk E.M., Kuzmin I.V., Borisova I.T., Gazaryan S.V. Yager P. and Rupprecht C.E. (2003). Novel lyssavirus isolated from bats in Russia. *Emer. Infect. Dis.* 9(12): 1623-1625

Boulger L.R. and Porterfield J.S. (1958). Isolation of a virus from Nigerian fruit bats. *Trans. R. Soc. Trop. Med. Hyg.* 52: 421-424

Bourhy H., Sureau P. and Tordo N. (1990). From rabies to rabies-related viruses. *Vet. Microbiol.* 23: 115-128

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: 2419-26

Bourhy H., Kissi B. and Tordo N. (1993). Molecular diversity of the lyssavirus genome. *Virology* 194: 70-81

Brookes S.M., Parsons G., Johnson N., McElhinney L.M. and Fooks A.R. (2001). Rabies human diploid cell vaccine elicits cross neutralizing and cross protecting immune responses against European and Australian bat lyssaviruses. *Vaccine* 23: 4101-4109

Campbell, J.B. and Barton L.D. (1958). In, *Rabies* J.B. Campbell and K.M. Charlton (Eds), Kluwer Academic, Boston, pp. 223-241

Chanda P.K. and Banerjee A.K. (1979). Two distinct populations of vesicular stomatitis virus ribonucleoprotein cores with differential sensitivities to micrococcocal nuclease. *Biochem. Biophys. Res. Commun*. 91: 1337-1345

Coetzee P and Nel L.H. (2007). Emerging epidemic dog rabies in coastal South Africa: a molecular epidemiological analysis. *Virus Res*. 126(1-2): 186-95.

Coetzee P., Weyer J., Paweska J.T., Burt F.J., Markotter W. and Nel L.H. (2007). Use of a molecular epidemiological database to track human rabies case histories in South Africa. *Epidemiol Infect*. 136(9): 1270-1276

Cohen C., Sartorius B., Sabeta C., Zulu G., Paweska J., Mogoswane M. *et al.,* (2007). Epidemiology and viral molecular characterization of re-emerging rabies, South Africa. *Emerg Infect Dis*. 13(12): 1879-1886

Conzelmann K.K., Cox J.H., Schneider L.G. and Thiel H.J. (1990). Molecular cloning and complete sequence of the attenuated rabies virus SAD B19. *Virol*. 175: 485-499

Corbet G.B. (1991). Chiroptera. In, *The handbook of British mammals*. Harris S (Ed). Blackwell scientific publications, Oxford, United Kingdom. pp 583

Corbet G.B. (1992). In, *The Mammals of the IndoMalayan Region: A Systematic Review*. Corbet G.B. and Hill J.E. (Eds). Oxford University Press, Oxford. pp 1-418

Coulon P., Lafay F. and Flamand A. (1993). Rabies virus antigenicity: An overview. *Ondersteport J Vet. Res*. 60: 271-275

Crick J., Tignor G.H. and Moreno K. (1982). A new isolate of Lagos bat virus from the Republic of South Africa. *Trans. R. Soc. Trop. Med. Hyg*. 76: 211-213

Crowley J.C., Dowling P.C., Mennona J., Silverman J.I., Schuback D., Cook S.D., and Blumberg B.M. (1988) Sequence variability and function of measles virus 3' and 5' ends and intercistronic regions. *Virology* 164: 498-506

Davis P.L., Holmes E.C., Larrous F., Van der Poel W.H., Tjornehoj K., Alonso W.J. and Bourhy H. (2005). Phylogeography, population dynamics and molecular evolution of European bat lyssaviruses. *J Virol*. 79: 10487-10497

De B.P., Das T. and Banerjee A.K. (1997). Role of cellular kinases in the gene expression of nonsegmented negative strand RNA viruses. *Biol. Chem*. 378: 489-493

Delmas O., Holmes C.E., Talbi C., Larrous F., Dacheux L., Bouchier C. and Bourhy H. (2008). Genome diversity and evolution of the lyssaviruses. *PLoS ONE* 3(4): e2057

Dietzschold B., Wunner W.H., Wiktor T.J., Lopes A.D., Lafon, A.D., Smith C.L. and Koprowski H. (1983). Characterization of an antigenetic determinant of the glycoprotein that correlates with pathogenicity of rabies virus. *Proc. Natl. Acad. Sci. USA* 80: 70-74

Dietzschold B., Wiktor T.J., Trojanowski J.Q., Macfarlan R.I., Wunner W.H., Torres-Anjel M.J. and Koprowski H. (1985). Differences in cell to cell spread of pathogenic and apathogenic rabies virus *in vivo* and *in vitro*. *J Virol*. 56: 12-18

Dietzschold B., Tollis B., Rupprecht C.E., Cellis E. and Koprowski H. (1987). Antigenic variation in rabies and rabies-related viruses: Cross protection independent of glycoprotein-mediated virus-neutralizing antibody. *J Infect. Dis.* 56: 815-822

Dietzschold B., Rupprecht C.E., Tollis M., Lafon M., Mattei J., Wiktor T.J. and Koprowski H. (1988) Antigenic diversity of the glycoprotein and nucleocapsid proteins of rabies and rabies-related viruses: Implications for epidemiology and control of rabies. *Rev. Infect. Dis.* 10: 8785-98.

Dietzschold B., Gore M., Marchadier D., Niu H.S., Bunschoten H.S., Otvos L., Wunner W.H., Ertl H.J.C., Osterhous A.D.M.E. and Koprowski H. (1990). Structural and immunological characterization of a linear virus-neutralizing epitope of the rabies virus glycoprotein and its possible use in a synthetic vaccine. *J Virol.*64: 3804-3809.

Emerson S.U. (1982) Reconstitution studies detect a single polymerase entry site on the vesicular stomatitis virus genome. *Cell* 31: 635-642

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

Faber M., Pulmanausahakul R., Nagao K., Prosniak M., Rice A.B., Koprowski H., Schnell M.J. and Dietzschold B. (2004). Identification of viral genomic elements responsible for rabies neuroinvasivness. *Proc. Natl. Acad. Sci. USA* 101(46): 16328-16352

Familusi J.B., Osunkoya B.O., Moore D.L., Kemp G.E. and Fably A. (1972). A fatal human infection with Mokola virus. *Am. J. of Trop. Med. and Hyg*. 21: 959-963

Fekadu M., Shaddock J.H., Sanderlin D.W. and Smith J.S. (1988). Efficacy of rabies vaccines against Duvenhage virus isolated from European house bats (Eptesicus serotinus), classic rabies and rabies-related viruses. *Vaccine* 6(6): 533-539

Finnegan C.J., Brookes S.M., Johnson N.J., Smith J., Mansfield K.L., Keene V.L., McElhinney L.M. and Fooks A.R. (2002). Rabies in North America and Europe. *J R Soc Med*. 95: 9-13.

Fitch W.M. (1995). Uses for evolutionary trees. *Phil. Trans. R. Soc. London*. 349: 93-102

Flamand A., Wiktor T.J. and Koprowski H. (1980a). Use of hybridoma monoclonal antibodies in the detection of antigenic differences between rabies and rabies-related virus proteins. I. The nucleocapsid protein. *J Gen. Virol*. 48: 97-104

Flamand A., Wiktor T.J. and Koprowski H. (1980b). Use of hybridoma monoclonal antibodies in the detection of antigenic differences between rabies and rabies-related virus proteins. II. The glycoprotein. *J Gen. Virol*. 48: 105-109

Foggin C.M. (1983). Mokola virus infection in cats and a dog in Zimbabwe. *Vet. Rec*. 113: 115

Foggin C.M. (1988). Rabies and rabies-related viruses in Zimbabwe: historical, virological and ecological aspects. *[Doctoral dissertation]*. Harare: University if Zimbabwe.

Fooks A.R., Brookes S.M., Johnson N., McElhinney L.M. and Hutson A.M. (2003) European bat lyssaviruses: an emerging zoonosis. *Epidemiol. Infect.* 131: 1029-1039.

Franka R., Johnson N., Müller T., Vos A., Neubert L., Freuling C., Rupprecht C.E. and Fooks A.R. (2008). Susceptibility of North American big brown bats (Eptesicus fuscus) to infection with European bat lyssavirus type 1. *J Gen. Virol*. 89: 1998–2010

Fraser G.C., Hooper P.T., Lunt R.A., Gould A.R., Gleeson L.J., Hyatt A.D., Rusel G.M. and Kattenbeldt J.A. (1996). Encephalitis caused by fruit bats in Australia. *Emerg. Infec. Dis.* 2: 327-331

Fu Z.F., Wunner W.H and Dietzschold B. (1994). Imunnoprotection against rabies virus nucleoprotein. In, *Lyssaviruses*. C.E. Rupprecht, B. Dietzschold and H. Koprowski (Eds), Springer-verlag, Berlin, Heidelberg. pp. 145-160

Ghatak S., Banerjee R., Agarwak R.K. and Kapoor K.N. (2000). Zoonoses and bats: a look from human health viewpoint. *J Comm. Dis*. 32: 40-48

Goldwasser R.A. and Kissling R.E. (1958). Fluorescent antibody staining of street and fixed rabies virus antigens. *Proc Soc Exp Biol Med*. 98: 219

Goto H., Minamoto N., Ito H., Ito N., Sugiyama M., Kinjo T. and Kawai A. (2000). Mapping of epitopes and structural analysis of antigenic sites in the nucleoprotein of rabies virus. *J Gen. Virol*. 81: 119-127

Gould A.R., Hyatt A.D., Lunt R., Kattenbelt J.A., Hengstberger S. and Blacksell S.D. (1998). Characterization of a novel lyssavirus isolated from Pteropid bats in Australia. *Virus Res*. 54: 165-187

Gould A.R., Kattenbelt J.A., Hyatt A.B., Gumley S.G. and Lunt R.A. (2002). Characterization of an Australian bat lyssavirus variant isolated from an insectivorous bat. *Virus Res*. 89(1): 1-28

Gray P. A., Fenton M. B. and van Cakenberghe V. (1999). *Nyteris thebaica* – In: *Mammalian species*. 612: 1-8

Guyatt K.J., Twin J., Davis P., Holmes E.C., Smith G.A., Smith I.L., Mackenzie J.S., and Young P.L. (2003). A molecular epidemiological study of Australian bat lyssavirus. *J Gen. Virol*. 84: 485-496

Hall T.A. (1999). Bioedit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Res.* 41: 95-98

Hanada K., Suzuki Y. and Gojobori T. (2004). A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Mol. Biol. Evol*. 21: 1074-1080

Hanahan, D., Jessee, J. and Bloom, FR. (1991). Plasmid transformation of E.coli and other bacteria. *Methods in Enzymology* 204: 63-114.

Hanlon C.A., Kuzmin I., Blanton J.D., Weldon W.C., Managan J.S. and Rupprecht C.E. (2005). Efficacy of rabies biological against new lyssaviruses from Eurasia. *Virus Res*. 111(1): 44-54

Hannah J., Carney I., Smith G., Tannenberg A., Deverill J., Botha J., Serafin I., Harrower B., Fitzpatrick P., and Searle J. (2000). Australian bat Lyssavirus infection: a second human case, with a long incubation period. *Med. J Aust.* 172: 597-599

Harcourt B.H., Tamin A., Halpin K., Ksiazek T.G., Rollin P.E., Bellini W. and Rota P.A. (2001). Molecular characterisation of the polymerase gene and genomic termini of nipah virus. *Virology* 287: 192-201

Harty R.N., Brown M. E., McGettigan J.P., Wang G., Jayakar H.R., Huibregtse J.M., Whitt M.A. and Schnell M.J. (2001). Rhabdoviruses and the cellular ubiquitin-proteasome system: a budding interaction. *J Virol.* 75: 10623-10629

Hassel R.H. (1982). Incidence of Rabies in Kudu in South West Africa/Namibia. *S. Afr. J Sci.* 78: 418-421

Haydon D.T., Randall D.A., Matthews L., Knobel D.L., Tallents L.A., Gravenor M.B., Williams S.D., Pollinger J.P., Cleaveland S., Woolhouse M.E., Sillero-Zubiri C., Marino J., Macdonald D.W. and Laurenson M.K. (2006). Low-coverage vaccination strategies for the conservation of endangered species. *Nature* 443: 692-695.

Hemachuda T., Laothamatas J. and Rupprecht C.E. (2002). Human rabies: a disease of complex neuropathogenic mechanisms and diagnostic challenges. *Lancet Neurology*. 1: 101-109

Hofmeyer M., Hofmeyer D., Nel L. and Bingham J. (2004). A second outbreak of rabies in African wild dogs (*Lycaon pictus*) in Madikwe Game Reserve, South Africa, demonstrating the efficacy of vaccination against natural rabies challenge. *Anim. Conserv.* 7: 193-198

Holmala K. and Kauhala K. (2006). Ecology of wildlife rabies in Europe. *Mammal Review* 36: 17

Holmes E.C., Woelk C.H., Kassis R. and Bourhy H. (2002). Genetic constraints and the adaptive evolution of rabies virus. *Virology* 292: 247-257

Hooper P.T., Lunt R.A., Gould A.R., Samaratunga H., Hyatt A.D., Gleeson L.J. Rodwell B.J. Rupprecht C.E., Smith J.S. and Murray P.K. (1997). A new lyssavirus – the first endemic rabies-related virus recognized in Australia. *Bull. Inst. Pastuer*. 95: 209-218

Hubschle O.J. (1988). Rabies in the kudu antelope (*Tragelaphus strepsiceros*). *Rev Infect Dis*. 11: 629-633.

Hughes G.J., Smith J.S., Hanlon C.A., Rupprecht C.E. (2004). Evaluation of a TaqMan PCR assay to detect rabies virus RNA: influence of sequence variation and application to quantification of viral loads. *J Clin. Microbiol*. 42(1): 299-306

Hughes G.J., Orciari L.A. and Rupprecht C.E. (2005). Evolutionary timescale of rabies virus adaptation to North American bats inferred from the substitution rate of the nucleoprotein gene. *J. Gen. Virol*. 86: 1467-1474

Institute Pasteur. Isolations of Lagos bat virus in West Africa. Internal reports of Centre Collaborateur OMS de Reference et Recherche Pour les Arbovirus. Dakar (Senegal). 1985. Available from http://www.pasteur.fr/recherche/

banques/CRORA/bibref/br01600.htm and http://www.pasteur.fr/recherche/banques/CRORA/bibref/br01610.

Isaac C.L. and Keene J.D. (1982). RNA polymerase-associated interactions near template promoter sequences of defective interfering particles of vesicular stomatitis virus. *J Virol*. 43: 241-249

Jayakar H.R., Murti K.G. and Whitt M.A. (2000). Mutations in the PPPY motif of vesicular stomatitis virus matrix protein reduce virus budding by inhibiting a late step in virion release. *J Virol.* 74: 9818-9827

Jenkins G.M., Rambuat A., Pybus O.G. and Holmes E.C. (2002). Rates of molecular evolution in RNA viruses: quantitative phylogenetic analysis. *J Mol. Evol*. 54: 152-161

Johnson N., McElhinney L.M., Smith J., Lowings P. and Fooks A.R. (2002). Phylogenetic comparison of the genus Lyssavirus using distal coding sequences of the glycoprotein and nucleoprotein genes. *Arch. Virol*. 147: 2111-2123

Johnson N., Black C. Smith J., Un H., McElhinney L.M., Aylan O. and Fooks A.R. (2003). Rabies emergence amongst foxes in Turkey. *J Wildl. Dis.* 39: 262-270

Jukes T.H. (1987). Transitions, transversions, and the molecular evolutionary clock. *J Mol. Evol.* 26: 87-98

Karatas A., Yigit N., Colak E., Kankilic T. (2003). Contribution to *Rousettus aegyptiacus* (Mammalia: Chiroptera) from Turkey. *Folia. Zool*. 52(2): 137-142

Kassis R., Larrous F., Estaquier J. and Bourhy H. (2004). Lyssavirus matrix protein induces apoptosis by TRAIL-dependent mechanism involving caspase-8 Activation. *J Virol.* 78 (12): 6543-6555

Kawai A. (1977). Transcriptase activity associated with rabies virion. *J Virol*. 24: 826-835

Kawai A. and Morimoto K. (1994). Functional aspects of lyssavirus proteins. In, *Lyssaviruses*.C.E. Rupprecht, B. Dietzschold and H. Koprowski (Eds). Springer-Verlag, Berlin, Heidelberg, pp.27-42

Keene J.D., Schubert M and Lazzarinin R.A. (1979). Terminal sequences of vesticular stomatitis virus RNA are both complementary and conserved. *J Virol*. 32: 16-174

Keene J.D., Thornton B.J. and Emerson S.U. (1981). Sequence specific contracts between the RNA polymerase of vesicular stomatitis virus and the leader RNA gene. *Proc. Natl. Acad. Sci. U.S.A*. 78: 6191-6195

Kemp G.E., Causey O.R., Moore A.K. and Fabiyi A. (1971). Mokola virus. *Am. J. of Trop. Med. and Hyg*. 21(3): 356-359

Kimura M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol. Evol*. 15: 111-120

King A. and Crick J. (1988). Rabies-related viruses. In, *Rabies*. Campbell J.B and Charlton K.M. (Eds). Kluwer Academic Publishers, Boston. pp. 177-200

King A., Davies P. and Lawrie A. (1990). The rabies viruses of bats. *Vet. Microbiol.* 23: 165-74

King A.A., Meredith C.D. and Thomson G.R. (1994). The biology of Southern African lyssavirus variants. In, *Lyssaviruses.* Rupprecht C.E., Dietzschold B. and Koprowski H. (Eds). Springer-Verlag, Berlin, Heidelberg. pp. 267-295

Kissi B., Tordo N. and Bourhy H. (1995). Genetic polymorphism in the rabies virus nucleoprotein. *Virology* 209: 526-537

Koopman K. (1994). Chiroptera: Systematics. In, *Mammalia.* Walter de Gruyter, Berlin. pp 217

Koprowski H. (1996).The mouse inoculation test. In: *Laboratory techniques in rabies*. Meslin F.X., Kaplan M.M. and Koprowski H. (Eds). World Health Organization, Geneva. pp 80-87

Krebs W.J., Mark L. and Childs E.C. (1995). Rabies – Epidemiology, prevention and future research. *J Mammalogy*. 76(3): 681-694

Krebs J.W., Smith J.S., Rupprecht C.E. and Childs J.E. (1999). Rabies surveillance in the United states during 1998. *The J of the Vet. Med. Ass*. 215: 1786-1798

Kucera P., Dolivo M., Coulon P. and Flamand A. (1985). Pathways of the early propagation of virulent and avirulent rabies strains from the eye to the brain. *J Virol*. 55: 158-162

Kumar S., Tamura K. and Nei M. (2004). Mega3: integrated software for molecular evolutionary genetic analysis and sequence alignment. *Brief Bioinform.* 5: 150-163

Kurilla M.G., Cabradilla C.D., Holloway B.P. and Keene J.D. (1984). Nucleotide sequence and host La protein interactions of rabies virus leader RNA. *J Virol*. 50: 773-778

Kuzmin I.V., Orciari L.A., Yohko T.A. Smith J.S. Hanlon C.A., Kameoka Y. and Rupprecht C.E. (2003). Bat lyssaviruses (Aravan and Khujand) from central Asia: Phylogenetic relationships according to the N, P and G gene sequences. *Virus Res*. 97: 65-79

Kuzmin I.V., Hughes G.J., Botvinkin A.D., Orciari L.A. and Rupprecht C.E. (2005). Phylogenetic relationships of Irkut and West Caucasian bat viruses within the *Lyssavirus* genus and suggested quantitative criteria based on the N gene sequence for lyssaviruses genotype definition. *Virus Res.* 111: 28-43

Kuzmin I.V., Niezgoda M., Franka R., Agwanda B., Markotter W., Beagley J.C., Urazova O.Y., Breiman R.F. and Rupprecht C.E. (2008a). Lagos bat virus in Kenya. *J Clin. Micro*. 46(4): 1451-1461

Kuzmin I.V., Wu X., Tordo N. and Rupprecht C.E. (2008b) Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions. *Virus Res.* 136: 81-90

Lafon M., Wiktor T.J. and Macfarlan R.I. (1983) Antigenic sites on the CVS rabies virus glycoprotein: Analysis with monoclonal antibodies. *J Gen. Virol*. 64: 843-851.

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

Lafon M., Herzog M. and Sureau P. (1986). Human rabies vaccines induce neutralizing antibodies against the European bat rabies virus (Duvenhage). *The Lancet*. 30: 515

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 J., Lafage M., Martinez-arends A., Ramirez R., Vuilier F., Charron D., Looteau V. and Scott-Algara D. (1992). Evidence for a viral super antigen in humans. *Nature* 358: 507-510

Le Mercier P., Jacob Y. and Tordo N. (1997). The complete Mokola virus genome sequence: structure of the RNA-dependent RNA polymerase. *J Gen. Virol*. 78: 1571-1576

Lo K.W, Naisbitt S., Fan J.S., Sheng M. and Zhang M. (2001). The 8-kDa dynein light chain binds to its target via a conserved (K/R)XTQT motif. *J Biol. Chem.* 276: 14059-14066

Lumio J., Hillbom M., Roine R., Ketonen L., Haltia M., Valle M., Neuvonen E. and Lahdevirta J. (1986). Human rabies of bat origin in Europe. *The Lancet.* 8477: 378

Ma J., Jones G., Zhang B., Shen J., Metzner W., Zhang L. and Liang B. (2003). Dietary analysis confirms that Ricketts big footed bat (*Myotis ricketti*) is a piscivore. *J Zool*. 261: 245-248

Mackensie J.S, Field H.E. and Guyatt K.J. (2003). Managing emerging diseases borne by fruit bats (flying foxes) with particular reference to henipaviruses and Australian bat lyssavirus. *J Appl. Microbiol*. 94: 59-69

Mansfield K.L., Johnson N. and Fooks A.R. (2004). Identification of a conserved linear epitope at the N terminus of the rabies virus glycoprotein. *J Gen. Virol*. 85: 3279–3283

Mansfield K., McElhinney L., Hübschle O., Mettler F., Sabeta C., Nel L.H. and Fooks A.R. (2006). A molecular epidemiological study of rabies epizoonotics in kudu (*Tragelaphus strepsiceros*) in Namibia. *BMC Vet. Res.* Published online 2006 January 13. doi: 10.1186/1746-6148-2-2

Markotter W., Kuzmin I., Rupprecht C.E., Randles J., Sabeta C.T., Wandeler A.I. And Nel L.H. (2006a) Isolation of Lagos bat virus from water mongoose. *Emerg. Infect. Dis.* 12(2): 1913-1918

Markotter W., Randles J., Rupprecht C.E., Sabeta C.T., Wandeler A.I., Taylor P.J and Nel L.H. (2006 b). Recent Lagos bat virus isolations from bats (suborder *Megachiroptera*) in South Africa. *Emerg. Infect. Dis*.12: 504-506

Markotter W., Kuzmin I., Rupprecht C.E. and Nel L.H. (2008a). Phylogeny of Lagos Bat Virus: Challenge for lyssavirus taxonomy. *Virus Res*. 135(1): 10-21

Markotter W., Van Eeden C., Kuzmin I., Rupprecht C.E., Paweska J.T., Swanepoel R., Fooks A.R., Sabeta C.T., Cliquet F. and Nel L.H. (2008b). Epidemiology and pathogenicity of African bat lyssaviruses. *Dev. Biol. Basel. Krager* 131: 972-977

Marston D.A., McElhinney L.M., Johnson N., Müller T., Conzelmann K.K., Tordo N. and Fooks A.R. (2007). *J Gen. Virol*. 88: 1302-1314

Martinez-Burnes J., Lopez A., Medellin J., Haines D., Loza E. and Martinez M. (1997). An outbreak of vampire bat-transmitted rabies in cattle in north eastern Mexico. *Can Vet J.* 38: 175-177

McColl B.J., Epstein J.H., Neill A.S., Heel K., Field H., Barret J., Smith G.A., Selvey L.A., Rodwell B. and Lunt R. (2000). Potential exposure to Australian bat lyssavirus, Queensland, 1996-1999. *Emerg. Infect. Dis*. 6: 259-264

Mebatsion T., Cox J.H., and Frost J.W. (1992). Isolation and characterization of 115 street rabies virus isolates from Ethiopia by using monoclonal antibodies: identification of 2 isolates as Mokola and Lagos bat viruses. *J Infect. Dis.* 166: 972-977

Mebatsion T., Weiland F. and Conzelmann K.K. (1999). Matrix protein of rabies virus is responsible for the assembly and budding of bullet-shaped particles and interacts with the transmemebrane spike glycoprotein G. *J Virol.* 73: 242-250

Mebatsion T. (2001). Extensive attenuation of rabies virus by simultaneously modifying the dynein light chain binding site in the P protein and replacing Arg333 in the G protein. *J Virol.* 75: 11496-11502

Meredith C.D., Rossouw A.P., and Van Praag Koch H. (1971). An unusual case of human rabies thought to be of chiropteran origin. *S. Afr. Med. J.* 45: 767-769

Minamoto N., Tanaka H., Hishida M., Goto H., Ito H., Naruse S., Yamamoto K., Sugiyama M., Kinjo T., Mannen K. and Mifune K. (1994). Linear and conformation-dependent antigenic sites on the nucleoprotein of rabies virus. *Microbiol. Immunol*. 38: 449-455

Montano Hirose J.A., Bourhy H., and Lafon M. (1990). A reduced panel of anti-nucleocapsid monoclonal antibodies for bat rabies virus identification in Europe. *Res. Virol.* 141: 571-581

Morzunov S.P., Winton J.R. and Nichol S.T. (1995). The complete genome structure and phylogenetic relationship of infectious hematopoietic necrosis virus. *Virus Res*. 38: 175-192

Müller T., Cox J., Peter W., Scafer R., Bodamer P., Wulle U., Burow J. and Müller W. (2001). Infection of a stone marten with European bat lyssavirus (EBLV1). *Rabies Bulletin Europe* 25(3): 9-11

Nadin-Davis S.A. (2000) S.A. Nadin-Davis, Rabies and rabies-related viruses. In: R.C.A.
Nadin-Davis S.A., Abdel-Malik M., Armstrong J. and Wandeler A.I. (2002). Lyssavirus P gene characterization provides insights into the phylogeny of the genus and identifies structural similarities and diversity within the encoded phosphoprotein. *Virology* 298: 286-305

Nadin-Davis S.A., Huang W., Armstrong J., Casey G.A., Bahloul C., Tordo N. and Wandeler A.I. (2001). Antigenic and genetic divergence of rabies viruses from bat species indigenous to Canada. *Virus Res*. 74: 139-156

Nadin-Davis S.A., Simani S., Amstrong J., Fayaz A. and Wandeler A. (2003). Molecular And antigenic characterization of rabies viruses from Iran identifies variants with distinct epidemiological origins. *Epidemiol. Infect*. 131: 777-790

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

Nel L.H., Jacobs J., Jaftha J., von Teichman B. and Bingham J. (2000). New cases of Mokola virus infection in South Africa: A genotypic comparison of southern African virus isolates. *Virus genes*. 20:2 103-106

Nel L.H and Markotter W. (2007). Lyssaviruses. *Critical Rev. in Microbiol.* 33: 1-24

Nel L.H. and Rupprecht C.E. (2007). Emergence of lyssaviruses in the old world: The case of Africa. In, *Current topics in microbiology and immunology: wildlife and emerging diseases: The biology, circumstances and consequences of cross-species transmission*. Heidelberg, Springer Berlin. pp. 161-193

Nel L.H., Thomson G.R. and Von Teichman B.F. (1993). Molecular epidemiology of rabies in South Africa, *Onderstepoort J Vet. Sci.* 60: 301–306

Nichol S.T. and Holland J.J. (1987) Genome RNA terminus conservation and diversity among vesiculoviruses. *J Virol*. 61: 200-205

Niezgoda M., Hanlon C.A. and Rupprecht C.E. (2002). Animal Rabies. In, *Rabies*. Jackson A.C. and Wunner W.H. (Eds), Elsevier Science, London. pp. 163-207.

Nowak R. (1999). Chiroptera: Bats. In, *Walkers mammals of the world*. The John Hopkins University press. Baltimore, Maryland. pp 253-414

O'Shea T. J., Shankar V., Bowen R. A., Rupprecht C. E. and Wimsatt J. H. (2003). Do bats acquire immunity to rabies? Evidence from the field. *Bat Research News* 44: 161

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

Paweska J.T., Blumberg L.H., Liebenberg C., Hewlett R.H., Grobbelaar A.A., Leman P.A., Croft J.E., Nel L.H., Nutt L. and Swanepoel R. (2006) Fatal human infection with rabies-related Duvenhage virus, South Africa. *Emerg. Infect. Dis.* 12: 1965-66

Pazos F., Helmer-Citterrich M., Ausiello G. andValencia A. (1997). Correlated mutations contain information about protein-protein interaction. *J Mol. Biol*. 271: 511-523

Peaz A., Nunez C., Garcia C., and Boshell J. (2003). Molecular epidemiology of rabies epizoonotics in Colombia: evidence for human and dog rabies associated with bats. *J Gen. Virol*. 84: 795-802

Picard-Meyer E., Barrat J., Wasniewski M., Wandeler A., Nadin-Davis S., Lownings J.P., Fooks A.R., McElhinney L., Bruyire V. and Cliquet F. (2004). Epidemiology of rabid bats in France, 1989 to 2002. *Vet. Rec*. 155: 774-777

Poch O., Blumberg B.M., Bougueleret L. and Tordo N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative strand RNA viruses: theoretical assignment of functional domains. *J Gen. Virol.* 71: 1153-1162

Poisson N., Real E., Gaudin Y., Vaney M., King S., Jacob Y., Tordo N. and Blondel D. (2001). Molecular basis for the interaction between rabies virus phosphoprotein P and the dynein light chain LC8: Dissociation of dynein-binding properties and transcriptional functionality of P. *J Gen Virol.* 82: 2691-2696

Prehaud C., Coulon P., Lafay F., Thiers C. and Flamand A. (1988). Antigenic site II of the rabies virus glycoprotein: Structure and role in viral virulence. *J Virol.* 62: 1-7

Presetnik P. (2004). Bat species and conservation issues in the castle na Gorickem (NE Slovenia). *Mammalia* 68(4): 427-435

Ridley M. (2004). Natural selection and random drift. In, *Evolution*. Ridley M. (Ed). Blackwell Publishing. Oxford, United Kingdom. pp 184-185

Ronsholt L., Sorensen K.J., Bruschke C.J.M., Wellenberg G.J., van Oirschot J.T., Johnstonbe P., Whitby J.E. and Bourhy H. (1998). Clinically silent rabies infection in (zoo) bats. *Veterinary record.* 142: 519-520

Ronsholt L. (2002). A new case of European bat lyssavirus (EBL) infection in Danish sheep. *Rabies Bulletin Europe* 2/02

Rosatte R.C., Power M.J., Donovan D., Davies J.C., Allan M., Bachmann P., Stevenson B., Wandeler A. and Muldoon F. (2007). Elimination of arctic variant rabies in red foxes, Metropolitan Toronto. *Emerg. Infect. Dis.* 13: 25-27

Rupprecht C.E., Hanlon C.A. and Hemachudaha T. (2002). Rabies re-examined. *Lancet Infect. Dis.* 2: 327-343

Sabeta C.T., Markotter W., Mohale D.K., Shumba W., Wandeler A.I. and Nel L.H. (2007a). Mokola virus in domestic mammals, South Africa. *Emerg. Infec. Dis*. 13(9): 1371-1373

Sabeta C.T., Mansfield K.L., McElhinney L.M., Fooks A.R. AND Nel L.H. (2007b). Molecular epidemiology of rabies in bat-eared foxes (Otocyon megalotis) in South Africa. *Virus Res*. 129(1): 1-10

Sacramento D., Bourhy H and Tordo N. (1991). PCR technique as an alternative method for diagnosis and molecular epidemiology of rabies virus, *Mol. Cell. Probes* 5: 229–240

Saitou N. and Nei M. (1987). The neighbour-joining method: A new method for constructing Phylogenetic trees. *Mol. Biol. and Evol.* 4: 406-425

Sato G., Kobayashi Y., Motizuki N., Hirano S., Itou T., Cunha E.M.S., Ito F.H. and Sakai T. (2008). A unique substitution at position 333 on the glycoprotein of rabies virus street strains isolated from non-hematophagus bats in Brazil. *Virus genes*. DOI 10.1007/s11262-008-0290-5

Schneider L.G., Dietzschold B., Dierks R.E., Matthaeus W., Enzmann P.J. and Strohmaier K. (1973). Rabies group-specific ribonucleoprotein antigen and a test system for grouping and typing of rhabdoviruses. *J Virol.*11: 748-755.

Schneider L.G. (1982). Antigenic variants of rabies virus. *Comp. Immunol. Microbiol. Infect. Dis.* 5: 101-107.

Seif I., Coulon P., Rollon P.E. and Flamand A. (1985). Rabies virulence effect on pathogenicity and sequence characterization of rabies virus mutations affecting antigenic site III of the glycoprotein. *J Virol* 53: 926-934

Serra-Cobo J., Amengual B., Abellan C. and Bourhy H. (2002). European bat lyssavirus infection in Spanish bat populations. *Emer. Infect.* Dis. 8(4): 413-420

Shioda T., Iwasaki K. and Shibuta H. (1986). Determination of the complete nucleotide sequence of the sendai virus genome RNA and the predicted amino acid sequences of the FHN and L proteins. *Nucleic Acids Res*. 14: 1545-1563

Shope R.E. (1982). Rabies-related viruses. *The Yale J. of Biol. and Med*. 55: 271-275

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

Smith J. S., Orciari L.A. and Yager P.A. (1995). Molecular epidemiology of rabies in the United States. *Seminars in Virology* 6: 387-400

Smith J.S. (1996). New aspects of rabies with emphasis on epidemiology, diagnosis and prevention of the disease in the United States. *Clin. Microbiol. Rev*. 9: 166-176

Steele H.S. and Fernandez P.J. (1991). History of rabies and global aspects. In, *The natural history of rabies*. Baer G.M. (Ed) CRC press, Boca Raton, Florida. pp 72-97

Swanepoel R. (2004). In: *Infectious diseases of livestock: with special reference to southern Africa.* 2nd edition. Coetzer J.A.W. and Tustin R.C. (Eds). Oxford University Press, Cape town, South Africa. pp 1123-1182

Takayama-Ito M., Inoue K., Shoji Y., Inoue S., Iijima T., Sakai T., Kurane I. and Morimoto K. (2006a). A highly attenuated rabies virus HEP-Flurry strain reverts to virulent by a single amino acid substitution to arginine at position 333 in the glycoprotein. *Virus Res*. 119: 208-215

Takayama-Ito M., Ito N., Yamada K., Sugiyama M. and Minamoto N. (2006b). Multiple amino acids in the glycoprotein of rabies virus are responsible for pathogenicity in adult mice. *Virus Res.* 115: 169-175

Thompson J.D., Higgins D.G. and Gibson T.J. (1994). CLUSTAL W: improving the Sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680

Tignor G.H and Shope R.E. (1972). Vaccination and challenge of mice with viruses of the rabies serogroup. *J Infect. Dis*. 125: 322-324

Tignor G.H., Murphy F.A., Clark H.F., Shope R.E., Madore P., Bauer S.P., Buckley S.M. and Meredith C.D. (1977). Duvenhage virus: Morphological, biochemical, histopathological and antigenic relationships to the rabies serogroup. *J Gen. Virol.* 37: 595-611

Tordo N. and Poch O. (1988). Structure of rabies virus. In *Rabies.* Campbell J.B. and Charlton K.M. (Eds). Kluwer Academic Publishers, Boston. pp. 25-45

Tordo N. and Kouknetzoff A. (1993). The rabies virus genome: an overview. *Onderstepoort J Vet. Res.* 60: 263-269

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

Tordo N., Poch O., Ermine A., Keith G. and Rougeon F. (1988). Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative strand RNA viruses. *Virol.* 165: 565-576

Tordo N., Benmansour A., Calisher C., et al., (2005). Rhabdoviridae. In, *Virus Taxonomy: The Classification and Nomenclature of Viruses*. The Eight Report of the International Committee on Taxonomy of Viruses. Fauquet C.M., Mayo M.A., Maniloff J., Desselberger U. and Ball L.A. (Eds). Elsevier Academic Press, London. pp. 623-644

Truffereau C., Lebloish., Benejean J., Coulon P., Lafay E. and Flamand A. (1989). Arginine or lysine at position 333 of ERA and CVS glycoprotein is necessary for rabies virulence in adult mice. *Virology* 172: 206-212

Van der Merwe M. (1982). Bats as vectors of rabies. *S. Afr. J. Sci.* 78: 421-422

Van Thiel P., van den Hoek J., Eftimov F., Tepaske R., Zaaljer H., Spanjaard L., de Boer H., van Doornun G., Schutten M., Osterhaus A. and Kager P. (2008). Fatal case of human rabies (Duvenhage virus) from a bat in Kenya: the Netherlands, December 2007. *Euro. Surveill.* 13(2). pii 8007

W.H.O. expert consultation on rabies. (2005). Technical Report Series 931, World Health Organization, Geneva
Warrell M.J. and Warrell D.A. (2004). Rabies and other lyssavirus diseases. *The Lancet.* 363(9413): 959-969

Warrell M. and Warrell D. (2004) Rabies and other lyssavirus diseases. *The Lancet*. 363:959-969.

Warrilow D., Smith I.L., Harrower B. and Smith G.A. (2002). Sequence analysis of an isolate from a fatal human infection of Australian bat lyssavirus. *Virol.* 297: 109-119

Warrilow D. (2005). Australian bat lyssavirus: A recently discovered new Rhabdovirus. *Curr. Top. Microbiol. Immunol*. 292: 25-44

Wiktor T.J. and Koprowski H. (1978). Monoclonal antibodies against rabies virus produced by somatic cell hybridization: detection of antigenic variants. *Proc. Natl. Acad. Sci. USA* 75: 3938-3942

Wilkinson L. (1988). Understanding the nature of rabies: An historical perspective. In, *Rabies*. Campbell J.B. and Charlton K.M. (Eds). Kluwer Academic, Boston. pp. 1-23

Wilkinson L. (2002). History. In, *Rabies*. A. C. and W. H. Jackson (Eds), Elsevier Science, London pp. 1-21

Willoughby R.E., Tieves K.S., Hoffman G.M., Ghanayem N.S., Amlie-Lefond C.M., Schwabe M.J., Chusid M.J. and Rupprecht C.E. (2005). Survival after treatment of rabies with induction of a coma. *NEJM*. 352: 2508-2514.

Woelk C.H. and Holmes E.C. (2002). Reduced positive selection in vector borne RNA viruses. *Mol. Biol. Evol*. 19: 2333-2336

Wu X. et al., (2007). Are all lyssavirus gene equal for phylogenetic analysis? *Virus Res*. 129(2): 91-103

Wunner H.W. (1991). Chemical composition and molecular structure of rabies viruses. In, *The natural history of rabies*. Baer G.M. (Ed) CRC Press, 2nd ed. pp. 27-65

Wunner H.W. (2002). Rabies virus. In, *Rabies*. A.C. Jackson and W.H. Wunner (Eds), Elsevier Science, London. pp. 23-77

Yakobson B.A., David D. and Aldomy F. (2004). Rabies in Israel and Jordan. In, *Historical perspective of rabies in Europe and the Mediterranean Basin*. King A.A., Fooks A.R., Aubert M. and Wandeler A. I. (Eds). World Health Organization for animal health (OIE) in conjunction with the WHO, Geneva, Paris. pp. 171-183

Yan X., Mohankumar P.S., Dietzschold B., Schnell M.J. and Fu Z.F. (2002). The rabies virus glycoprotein determines the distribution of different rabies virus strains in the brain. *J Neurovirol*. 8: 345-352

Yang J., Hooper D.C., Wunner W.H. Koprowski H., Dietzschold B. and Fu Z.F. (1998). The specificity of rabies virus RNA encapsidation by nucleoprotein. *Virol*. 242: 107-117

Yang J., Koprowski H., Dietzschold B. and Fu Z.F. (1999). Phosphorylation of rabies virus nucleoprotein regulates viral RNA transcription and replication by modulating leader RNA encapsidation. *J Virol.* 73: 1661-1664

Yang Z. and Yoder A.D. (1998). Estimation of the transition/transversion rate bias and species sampling. *J Mol. Evol*. 48: 274-283

Internet references

1. http://www.cdc.gov/ncidod/dvrd/rabies/professional/publications/DFA_diagnosis

   CDC. Centres for disease control and prevention. Rabies

2. http://www.ambion.com/techlib/tn/91/9113.html

   Applied Biosystems: Ten ways to improve your RNA isolation

COMMUNICATIONS

C. van Eeden, W. Markotter, L.H. Nel. Molecular epidemiology and genetic characterization of a rabies-related virus, Duvenhage virus. Molecular and Cell Biology Group Symposium (MCBG). Pretoria, South Africa. 17 October 2007.

C. van Eeden, W. Markotter, L.H. Nel. Genetic characterization of rabies-related Duvenhage virus. 9th Meeting of the Southern and Eastern African Rabies Group (SEARG). Centre for In-service and Continuing Education, Botswana College of Agriculture, Gaborone, Botswana. 25 - 28 August 2008.

PUBLICATIONS

Markotter W., Van Eeden C., Kuzmin I., Rupprecht C.E., Paweska J.T., Swanepoel R., Fooks A.R., Sabeta C.T., Cliquet F. and Nel L.H. 2008. Epidemiology and pathogenicity of African bat lyssaviruses. *Dev. Biol. Basel. Krager* 131: 972-977