Chapter 1

Literature survey

(A) - Literature survey

1.1 General introduction.

The rabies virus belongs to the *Lyssavirus* genus, a group of bullet shaped viruses that have a nearly worldwide distribution. These viruses are responsible for causing a fatal encephalitis, which results in the deaths of thousands of people each year (WHO, 1992; Swanepoel, 1994; Rupprecht *et al.*, 2002). It is maintained in two overlapping epidemiological cycles, one that is limited to domestic dogs, and the other that occurs in terrestrial and chiropteran wildlife (Swanepoel, 1994). The virus is transmitted by the bite of an infected animal, and all warm-blooded animals tested so far are susceptible to the disease (Hemachudha, 2002).

Although effective and safe vaccines have been available for decades for the pre- and post-exposure prophylaxis of rabies in humans, and the control of dog and wildlife rabies, the brunt of the disease continues to be borne by the worlds poor. The disease remains a problem in the developing world due to a combination of unfavorable factors, which includes the limited funding which is often available for protracted control campaigns, as well as the unavailability of safe and inexpensive anti-rabies biologics (Rupprecht *et al.*, 2002). This problem is compounded by underdeveloped public health infrastructures, due to wars and poverty, as well as the presence of diseases such as AIDS, tuberculosis and malaria, which represent competing public health priorities in the developing world (Rupprecht *et al.*, 2002). Control is further complicated by virus specific factors, such as the virus' high adaptability, and thus its ability to establish infectious cycles in whatever host species are ecologically capable of sustaining it (Wunner, 2002; Sabeta *et al.*, 2003). Indeed rabies continues to be maintained by wildlife reservoirs, even in the developed world, where the disease has been brought under control in dogs through vaccination (Smith, 1996).

It is thus apparent that significant challenges are posed to public health and veterinary authorities charged with the elimination of the disease in third world countries. In these circumstances, the development and utilization of enhanced methods of disease surveillance becomes essential, in order to provide information to relevant authorities for the development and evaluation of feasible control strategies (Bourhy *et al.*, 1995). In this

respect, molecular epidemiological studies have become increasingly more important for understanding the basic factors that govern the epidemiology of rabies within the different affected regions of the world (Bourhy *et al.*, 1995).

This project was specifically focused on determining the molecular epidemiology of rabies in the KwaZulu Natal province of South Africa. The purpose of section A of the literature review was to give an overview of the history, classification, molecular biology and epidemiology of rabies in the world, as well as in southern Africa. Section B focuses specifically on a discussion of the techniques and methodology employed when conducting a molecular epidemiological study, while section C presents an overview of the different aims of this project. Chapter 2 will focus on the particulars of the epidemic in KwaZulu Natal, a description of the experimental techniques which were employed, as well as the results which were obtained from these experiments. In chapter 3 the major experimental findings of the project will be discussed in further detail, with the emphasis being placed on the implications of these findings in the maintenance, persistence and control of the disease. The main conclusions that can be drawn from the results will be summarized in chapter 4.

1.2 History of rabies

Rabies is one of the oldest infectious diseases known to man, with a history that can be traced back for thousands of years. The presence of rabies was documented in Mesopotamian, Egyptian, Chinese, Greek and Roman texts, with the characteristic symptoms and inevitable outcome of the disease, ensuring its notoriety, even among the major epidemics that plagued the citizens of the ancient world (Wilkinson, 1988; Swanepoel, 1994; Steele and Fernandez, 1991; Wilkinson, 2002). The Greeks referred to the disease as "Lyssa", meaning madness, whereas the word "rabies" comes from the Latin word "*rabere* ", that has an earlier origin which may be derived from an old Sancrit word, "*rhabdas*", which means violence (Wilkinson, 1988; Swanepoel, 1994; Steele and Fernandez, 1991; Wilkinson, 2002).

The first reference to rabies appears in the legal document, the *Eshunna* code in the 23rd century B.C., in which it was indicated to what degree the owner of a rabid dog (*Canis familiaris*) was financially liable, when either a free man or a slave was bitten (Wilkinson, 1988; Swanepoel, 1994; Steele and Fernandez, 1991; Wilkinson, 2002). The Greek

philosopher Aristotle was 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, while initial attempts at prophylaxis was described by the Roman doctor Celsus, as early as the first century A.D (Wilkinson, 1988; Steele and Fernandez, 1991; Wilkinson, 2002). Speculations as to the source of the disease varied throughout the ages, and it was widely argued that it was spontaneously generated due to stress associated with heat, thirst, sexual frustration and celestial events (Wilkinson, 1988; Steele and Fernandez, 1991; Wilkinson, 2002). In 1546, Giralamo Fricastro suggested the possibility of a disease "contagion" associated with the saliva of an infected dog, which would eventually become widely believed to be generated as a consequence of putrefaction (Wilkinson, 2002). The discovery of Antony Von Leeuwenhoek's (1632-1723) "animalculae" and the refutal of the spontaneous generation theory by Francesco Redi in 1688 laid this theory to rest (Wilkinson, 1988; Wilkinson, 2002).

In 1793, a surgeon by the name of William Hunter proposed the possibility of performing transmission studies on the etiological agent of rabies, so that the disease's transmission pathway and clinical progression could be monitored (Wilkinson, 1988, Wilkinson, 2002). A German by the name of Georg Gottfried Zinke, put Hunter's suggestion into practice in 1813, by successfully transmitting rabies to a wide range of animal hosts (Wilkinson, 1988; Wilkinson, 2002). This paved the way for Pierre-Victor Galtier's (1879) development of the rabbit as an experimental animal model, and the historical work performed by Louis Pasteur and Colleagues on the development of the first rabies vaccine (Wilkinson, 1988; Swanepoel, 1994; Steele and Fernandez, 1991; Wilkinson, 2002).

Pasteur and Colleagues applied Edward Jenner's principle of vaccination, and successfully developed vaccines against anthrax and fowl cholera (Swanepoel, 1994). In a similar manner they reasoned, it would be possible to develop a vaccine against rabies that would be able to induce immunity after an exposure, before the disease manifested in its victim (Wilkinson, 1988; Swanepoel, 1994; Steele and Fernandez, 1991; Wilkinson, 2002). With their discovery of the virus' neurotropic character in 1881, and the development of a fixed laboratory strain with a reproducible incubation period, they laid the groundwork for the development of the first neural tissue vaccine (Wilkinson, 1988; Steele and Fernandez, 1991; Swanepoel, 1994; Wilkinson, 2002). It consisted of infected rabbit spinal cord, desiccated for successively longer days over potassium hydroxide. Following extensive testing in animal models this material was successfully used in

humans in 1885 (Wilkinson, 1988; Steele and Fernandez, 1991; Swanepoel, 1994; Wilkinson, 2002). The neural tissue vaccine found widespread application, and established the principle of post exposure prophylaxis (PEP) against rabies (Wilkinson, 1988; Swanepoel, 1994; Steele and Fernandez, 1991; Wilkinson, 2002).

Although a vaccine against rabies had been developed, the fundamental question as to the nature of the pathogen still remained. In 1903, Paul Remlinger illustrated the filterable nature of the disease agent, consigning it to the newly described class of microorganisms called viruses (Wilkinson, 1988; Swanepoel, 1994; Steele and Fernandez, 1991; Wilkinson, 2002). In the same year, Adelchi Negri described the neuronal viral cytoplasmic inclusion bodies that still bear his name, incorrectly identifying them as Protozoa, and claiming that they are the causative agent of the disease (Wilkinson, 1988; Swanepoel, 1994; Steele and Fernandez, 1991; Wilkinson, 2002). Even though Negri was wrong as to the etiological agent of rabies, Negri bodies would continue to play an important role in the diagnosis of the disease for decades to come (Wilkinson, 1988; Swanepoel, 1994; Steele and Fernandez, 1991; Wilkinson, 2002). In 1936 ultrafiltration studies indicated that the rabies virus particle had a diameter of 100-150 µm, a tolerable approximation of the presently accepted dimensions of 180 x 80 nm (Wilkinson, 1988). The bullet shaped morphology of the virus was however only revealed when Matsumoto, Davies *et al.*, and Atanasiu *et al.* first described it by using electron microscopy in 1963 (Steele and Fernandez, 1991).

Other notable developments of the 20th century include Fermi's improvement of the Pasteur vaccine by treating infected nerve tissue with phenol in 1908, and the eventual development of an inactivated neural tissue vaccine by Semple in 1911 (Wilkinson, 1988; Swanepoel, 1994; Steele and Fernandez, 1991; Wilkinson, 2002). Although effective, inactivated nerve tissue vaccines carried the risk of causing neurologic complications due to an autoimmune reaction against myelin, which is present within these preparations (Swanepoel, 1994, Warrell and Warrell, 1994). This led to the development of vaccines prepared from virus grown in suckling mouse brain, as well as in embyonated chicken and duck eggs (Swanepoel, 1994). Successful serial subculture of rabies in cell culture, first reported in 1958, has subsequently led to the development of highly effacious vaccines with a reduced or absent antigen load, such as the purified chick embryo cell (PCEC) and human diploid cell vaccines (HDCV) in use today (Swanepoel, 1994; Warrell and Warell, 2004). The inclusion of human rabies immunoglobin (HRIG) as part of the post-exposure

vaccine regimen in the 1950s, has further increased the effectiveness of rabies PEP, allowing for the neutralization of the virus at the site of the inoculation, before it can gain access to the central nervous system (CNS) (Cabasso *et al.*, 1971; Warrell and Warrell, 1994).

Prior to the 1950s, 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 superceded by the development of the direct/indirect fluorescent antibody test (DFA, IFA), first described by Goldwasser and Kissling in 1958, which is based on the detection of rabies virus antigen on touch impression brain smears using a fluorescently labeled antibody (Ab) conjugate (Campbell and Barton, 1988; Steele and Fernandez, 1991; Swanepoel, 1994; Rupprecht, *et al.*, 2002). The advent of reverse transcription polymerase chain reaction (RT-PCR) assays in the 1980s, have provided a further useful adjunct to rabies virus diagnosis, but remains limited in general usage due to the requirement for expensive lab equipment and trained personnel (Rupprecht *et al.*, 2002).

Early literature reflected the growing opinion that rabies could be controlled by quarantine, muzzling, and restriction of dog movement, as well as through the destruction of strays (Wilkinson, 1988; Swanepoel, 1994). Elimination of rabies through such measures was only achieved in a very few instances, where the insulatory nature of the country concerned, as well as a lack of wild reservoirs, favored the control of the disease (Swanepoel, 1994). Control was first achieved in the Scandinavian countries of Norway, Denmark and Sweden in 1826, and in Britain in 1902, and 1922, following a reintroduction of the disease (Swanepoel, 1994). Elimination of rabies through mass vaccination was achieved in 1956 in Japan, while Taiwan, Singapore, Malaysia, Switzerland, France and Portugal are among the few other countries that have managed to eradicate the disease (Steele and Fernandez, 1991; Swanepoel, 1994).

The advent of monoclonal antibody techniques (MAb) opened new doors in the study of the taxonomy and epidemiology of rabies and rabies related viruses. Wiktor *et al.* in 1978 were the first to describe the antigenic variation that is present within viral proteins, a finding which would eventually be applied to the classification as well as diagnosis of Lyssaviruses (Wiktor *et al.*, 1978; Steele and Fernandez, 1991). The development of MAb

further proved useful in studies of the epidemiology of the disease by allowing for the discrimination between viral variants which were obtained from different geographic regions or host species (Steele and Fernandez, 1991; Dietzschold *et al.*, 1988). The development of PCR, automated nucleotide sequencing, as well as computer algorithms and hardware required to analyze sequence data, has largely superceded the use of MAb typing as a means for strain discrimination in molecular epidemiology (Nadin Davis, 2000; Smith and Seidel, 1994; Bourhy *et al.*, 1995; Hungnes *et al.*, 2000; Smith, 2002).

1.3. Classification of the *Lyssavirus* genus

The rabies virus belongs to the *Lyssavirus* genus, which together with the genus *Vesiculovirus, Ephemerovirus, Novirhabdovirus*, *Cytorhabdovirus* and *Nucleorhabdovirus* make up the family *Rhabdoviridae* (ICTV, 2005)*.* Together with the families *Paramyxoa*, *Filo* and *Bornaviridea*, the family *Rhabdoviridae* belongs to the order *Mononegavirales* whose members are characterized by a single stranded, non-segmented, negative-sense RNA genome (Swanepoel, 1994; Levy and Owens, 1994; Rupprecht *et al.*, 2002; ICTV, 2005).

The family *Rhabdoviridae* has a wide host range, and include rod shaped viruses that have been isolated from vertebrates, invertebrates and plants (Tordo and Poch, 1988; Rupprecht *et al.*, 2002). Rhabdoviruses infecting animals were initially classified into the *Lyssavirus* (types species, rabies) and *Vesiculovirus* (type species, vesiculo stomatis virus) genera, based on seroneutralization of the N protein (Tordo and Poch, 1988; Nadin Davis, 2000). The genus *Lyssavirus* was then further subdivided into four serotypes, based on monoclonal antibody (MAb) studies, which were again directed against the N protein (WHO, 1992).

More recently the *Lyssavirus* genus has been reclassified into six genotypes by Bourhy *et al.*, 1993, based on percentage nucleoprotein amino acid (aa) similarity (Bourhy *et al.*, 1993). The six genotypes identified by Bourhy *et a*l., 1993, consist of the classical rabies virus (genotype 1), which has except for a few protected islands a worldwide distribution, as well as the rabies related viruses, consisting of genotypes 2 to 6 (LBV, MOKV, DUVV, EBLV-1, EBLV-2), that have been isolated from Africa and Eurasia (Bourhy *et al.*, 1993; Wunner et al., 2002). Viruses recently isolated in 1996 from Australian flying foxes and insectivorous bats (ABLV), belong to a seventh genotype, which is most closely related to

genotype 1 (Fraser *et al.*, 1996; Gould *et al.*, 1998; Halpin *et al.*, 1999; Nadin-Davis *et al.*, 2002; Wunner *et al.*, 2002; Guyatt *et al.*, 2003). Table 1.1. presents an abridged version of the classification of the *Mononegavirales* order as presented by the International Committee for the Taxonomy of Viruses (ICTV, 2005). The genotype designation has recently been changed to a generic "type species" designation by the ICTV (ICTV, 2005). The original genotype designation (WHO) will however be utilized throughout the rest of this project to refer to the viral type species, since the original genotype designation is still widely used throughout the literature.

Current classification scheme of the Lyssavirus genus.

Genotype/Serotype 1: Rabies virus (RABV).

Genotype 1 includes numerous fixed and vaccinal strains which were derived from the Pasteur virus, a viral strain which was originally isolated from a cow by Louis Pasteur in 1882. This genotype also includes the majority of isolates of terrestrial origin, as well as viruses isolated from insectivorous bats from North America, and heamatophagous bats from Latin America (King and Crick, 1988; Swanepoel, 1994; WHO, 1992).

The major reservoir of genotype 1 (RABV) throughout the world is the domestic dog (*Canis familiaris*) (Childs, 2002). The genetic and antigenic variation of classical rabies, circulating in the world-wide dog reservoir is low, and it has been found that terrestrial virus isolates from regions such as Latin America, Africa, Asia, Europe and the U.S.A. are very similar to each other (Bourhy *et al.*, 1995; Smith and Seidel, 1994; Childs, 2002). It is speculated that the low degree of antigenic and genetic variation among these geographically diverse isolates, is probably due to the importation of dog rabies into these countries as a result of European colonization (1800-1900) (see section 1.9.2) (Bourhy *et al.*, 1995; Smith and Seidel, 1994; Childs, 2002). These closely related groups of genotype 1 viruses are generally referred to as the Cosmopolitan strain of the rabies virus (Badrane and Tordo, 2001) . In many geographic regions the cosmopolitan lineage exists in concert with autochthonous genotype 1 variants, which are presumed to have been present before the importation of the cosmopolitan lineage. These include a variant which circulate in the Artic region among artic and red foxes, two variants which circulate among dogs and mongoose in sub-Saharan Africa, three variants which have spread among

Asian dogs and several strains which have spread within both chiropteran hosts and terrestrial carnivores (skunks and racoons) in the Americas (Badrane and Tordo, 2001).

Genotype/Serotype 2: Lagos Bat virus (LBV).

LBV was first isolated from frugivorous *Eidelon helvum* bats from Lagos Island, Nigeria, in 1956, and has there after been isolated from a number of different bat species as well as domestic cats and dogs (Shope, 1982; King and Crick, 1988 King *et al.*, 1994; Childs; 2002). It is the only known lyssavirus that has not been associated with rabies in humans, even though it has been shown to be pathogenic for both mice and vervet monkeys when inoculated intracerebrally (King *et al.*, 1994). The virus was first isolated from KwaZulu Natal in the 1980s from a number of *Epomophorus wahlbergi* bats. During this period there was an increase in public awareness of the danger of rabies in the province, due to an epidemic that had been raging among dogs since the late 1970s (Van der Merwe, 1982; Shope, 1982; Swanepoel, 1994; King *et al.*, 1994; Childs, 2002). Abnormally behaving and sick bats were thus submitted at a higher frequency to diagnostic laboratories, allowing for the isolation, and identification of the virus. Since then further isolations have been made from additional *E. wahlbergi* bats, as well as from a domestic dog and a marsh mongoose (*Atilax paludonsis*) (Swanepoel, 1994; King *et al.*, 1994; Markotter, University of Pretoria, 2005, unpublished results).

Genotype/Serotype 3: Mokola Virus (MOKV).

Mokola virus has a wide distribution throughout Africa, and has been associated with clinical disease in humans as well as animals (Kemp *et al.*, 1971; Swanepoel, 1994; King and Crick, 1988; King *et al.*, 1994). It was first isolated from *Crocudura* sp. shrews, captured close to the Mokola forest, near Ibadan, Nigeria in 1968 (Kemp *et al.*, 1971; Shope, 1982; Swanepoel, 1994; King and Crick, 1988; King *et al.*, 1994). Since then isolations have been made from hosts including humans, shrews, rodents as well as from domestic cats and dogs (Kemp *et al.*, 1971; Familusi *et al.*, 1972; Swanepoel *et al*, 1994; King and Crick, 1988; King *et al.*, 1994). It remains the only lyssavirus that has not been isolated from bats, and although it has been reported that shrews and rodents may represent potential reservoir hosts, uncertainty as to its reservoir species remains (Swanepoel, 1994). The Mokola virus also appears to be unique among the lyssaviruses, as it has been shown to have the ability to replicate in insect cells (King and Crick, 1988).

The first isolation of Mokola virus in South Africa was made from a cat near Uhmlanga Rocks in Durban, South Africa in 1971 (Swanepoel, 1994). It was however only identified as Mokola virus in the 1980s after an outbreak of the virus in 6 cats and one previously vaccinated dog in the Bulawayo area of Zimbabwe, between the periods of 1980–1982 (Shope, 1982; Foggin, 1982; Foggin, 1983). A further five isolations were made from cats, from regions surrounding East London and Pietermaritzburg in the period between 1995-1998 (Von Teichman *et al.*, 1998). Mokola virus has an unusual disease presentation in cats when compared to a classical rabies virus infection, in that the animals only display aggressive symptoms when handled (Von Teichman *et al.*, 1998).

Genotype/Serotype 4: Duvenhage virus (DUVV).

Only three isolates of Duvenhage virus are known – all made from southern Africa. The virus was first isolated from a man in 1970 in the Warmbath district, 100 km north of Pretoria, South Africa. This man succumbed to a rabies like disease after having been bitten five weeks prior by an unidentified insectivorous bat species, possibly *Miniopterus schreiberssi (*Van der Merwe, 1982; Meridith *et al.*, 1971; Shope, 1982; King *et al.*, 1994). Duvenhage virus was again isolated from an unidentified bat in 1981, caught by a cat in Louis Trichardt, South Africa, while a third isolation was made from a slit-faced *Nycteris thebaica* bat caught in a survey in Bulawayo, Zimbabwe in 1986 (Van der Merwe, 1982; Schneider *et al.*, 1985; King, 1993).

Serotype/Genotype 5 and 6: European Bat Lyssavirus 1 and 2 (EBLV-1 and EBLV-2).

EBLV-1 and EBLV-2 occur in insectivorous bats throughout Europe and Russia. EBLV-2 has also been isolated from bats from the United Kingdom, although isolations of EBLV-1 have thus far not been made from this region. The viruses were initially classified into serotype 4 together with the Duvenhage virus, and it was thought that these viruses were introduced into Europe as a result of bat migration, or transport on ships from Africa (Shope, 1982; King, 1993; King *et al.*, 1994; Schneider *et al.*, 1994; Childs, 2002). It only became clear that DUVV, EBLV-1 and EBLV-2 represented distinct disease entities, after monoclonal antibody studies and phylogenetic analysis of the N gene, distinguished between them (King, 1993; Bourhy *et al.*, 1993). Although, rabies in bats has been recognized in Europe since 1954, EBLV-1 was only isolated in 1985, when a large number of infected bats, mostly *Eptesicus serotinus*, were reported from Denmark and Germany

(Childs, 2002). In the same year EBLV-2 was first isolated from a Swiss bat biologist who died of rabies in Finland, after he had been bitten by a *Myotis daubentoni* bat (Childs, 2002).

Serotype/Genotype 7: Australian Bat Lyssavirus (ABLV).

Australian Bat Lyssavirus was first isolated from a flying fox (*Pteropus alecto*) in New South Wales in Australia, during a survey for equine morbillivirus in 1996 (Fraser *et al.*, 1996 Gould *et al.*, 1998; Halpin *et al.*, 1999). Since then, this virus has been shown by phylogenetic analysis of the G gene, to consist of two subtypes, of which the first type has been isolated from flying fox species (*P. poliocephalus, P, scapulatus*, and *P. conspicullatus*), and the second type has been isolated from an insectivorous bat (*Saccolaimus flaviventris*) (Guyatt *et al.*, 2003). The virus has a wide distribution on the eastern coast of Australia, and antibodies to ABLV have been found in an additional six bat species from as far a field as the Philippines, even though no viable virus could be isolated (Arguin *et al.*, 2002). ABLV was first isolated from humans in November 1996, from a bat rehabilitator who had been bitten and scratched by a large bat, five weeks before the onset of rabies like symptoms (Childs, 2002). A second person succumbed of the disease in 1998, after an unusually long incubation period of 2 years, after she had received a bite from a flying fox (Childs, 2002).

Unclassified rabies related viruses: West Caucasian Bat Virus (WCBV), Irkut, Aravan, Khujand.

Aravan, Khujand, West Caucasian Bat virus (WCBV) and Irkut.

The Aravan, Khujand, WCBV and Irkut viruses were recently isolated from bats in Eurasia (Kuzmin *et al.*, 1992; Arai *et al.*, 2003; Botvinkin *et al.*, 2003). The Aravan virus was isolated from the Lesser Mouse-eared Bat (*Myotis blythi*) in the Osh region of Kyrghyztan in 1991, while the Khujand virus was isolated from a Whiskered Bat (*Myotis mystacinus*) near the town of Khujand in Northern Tajikistan in 2001. The WCBV and Irkut viruses were isolated from a Schreibers' bat (*Miniopterus schreibersi*), and a Greater Tube-nosed bat (*Murina leucogaster*), from the Caucasus and Eastern Siberia in 2002, respectively. Based on results of the phylogenetic analyses of different gene regions of the viral genome, as well inter and intragenotypic percentage identity values, it has been proposed

that these viruses should be regarded as new lyssavirus genotypes (Arai *et al.*, 2003; Botvinkin *et al.*, 2003; Kuzmin *et al.*, 2003; Kuzmin *et al.*, (b) unpublished results).

Phylogroup designation for genotypes within the Lyssavirus genus.

The Lyssavirus genus has recently been reclassified into two phylogroups by Badrane *et al.*, 2001, based on viral phylogeny, pathogenicity and immunogenicity (Badrane *et al.*, 2001). Phylogroup 1 comprises genotypes 1, 4, 5, 6 and 7, while phylogroup 2 consists of genotypes 2 and 3 (Badrane *et al.*, 2001). The phylogroups differ from each other in terms of pathogenicity, in that phylogroup 1 is pathogenic for mice by both the intacerebral and intramuscular route, while phylogroup 2 is only pathogenic by the intracerebral route (Badrane *et al.*, 2001). These differences can be ascribed to the absence of arginine (or lysine) in position 333, in antigenic domain III of the glycoprotein, which has been implicated as an important determinant of virulence, it being replaced with aspartic acid in phylogroup 2 (Dietzschold *et al.*, 1983; Coulon *et al.*, 1993; Badrane *et al.*, a2001). With respects to immunogenicity it was found that serum antibodies produced when mice were immunized with the G protein of the PV (genotype 1) and the Mokola virus could cross neutralize their respective phylogroup members, but not vice versa (Badrane *et al.*, 2001). This has important implications for vaccine development, as vaccines prepared from genotype 1 vaccine strains, may not be able to cross protect against infection with a member of phylogroup 2 (Shope, 1982; Badrane *et al.*, 2001; Nel, 2005).

The first four genotypes have been found to be in accordance with the initial serological classification, while monoclonal antibody studies (King *et al.*, 1993) and genotypic classification (Bourhy *et al.*, 1993) could clearly distinguish between DUVV, EBLV-1 and EBLV-2, that had previously been classified under serotype 4. Figure 1.1. demonstrates a UPGMA phylogenetic tree that was constructed from the alignment of stretches of the amino acid sequence obtained from the N gene of the classical, as well as rabies related viruses, illustrating the genetic relationships which are present between the members of the *Lyssavirus* genus (Smith, 2002). Within the tree, 7 distinct genetic clusters, comprising the seven genotypes can be distinguished, with genotype 7 being most closely related to, and genotype 2 and 3 being most distantly related to genotype 1. Genotype 4, 5 and 6 also belong to clearly separable branches on the tree, illustrating the increased sensitivity of genetic typing over serological classification for taxonomic studies. Phylogroup divisions

are also indicated, with MOKV and LBV comprising phylogroup 2, while ABLV, RABV, EBLV-2, EBLV-1 and DUVV comprise phylogroup 1 (Bourhy *et al.*, 1993; Smith, 2002).
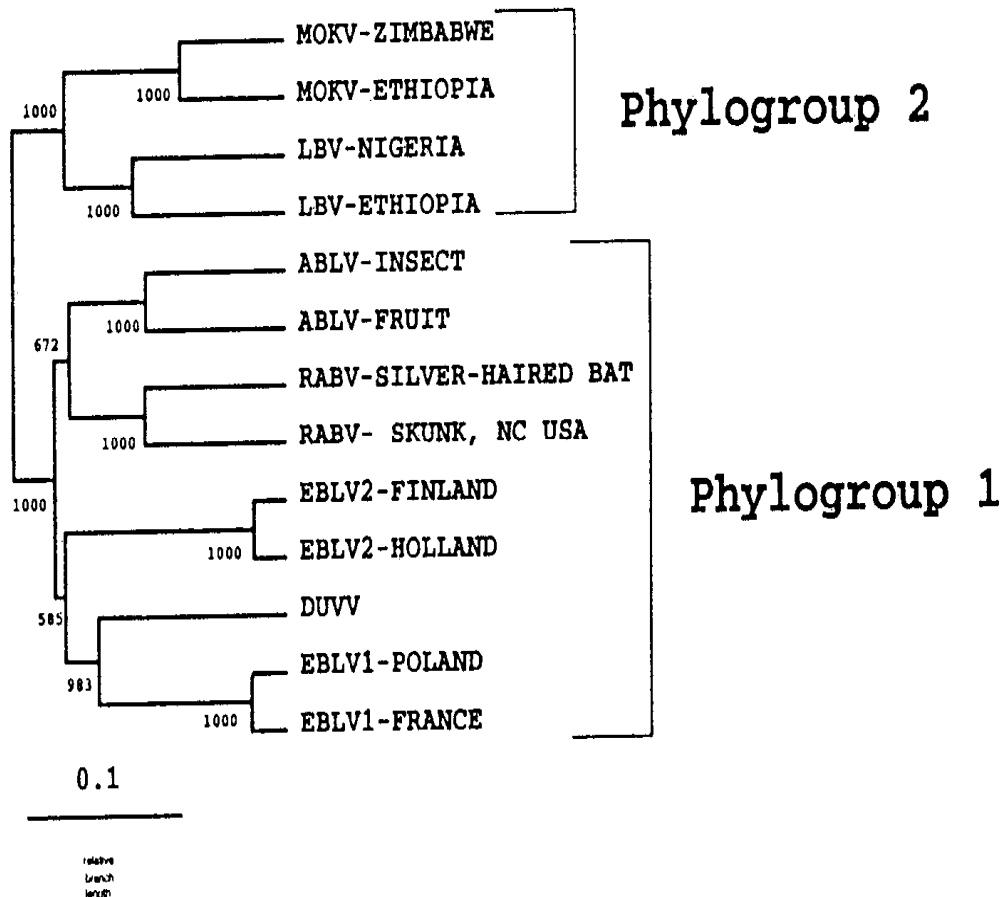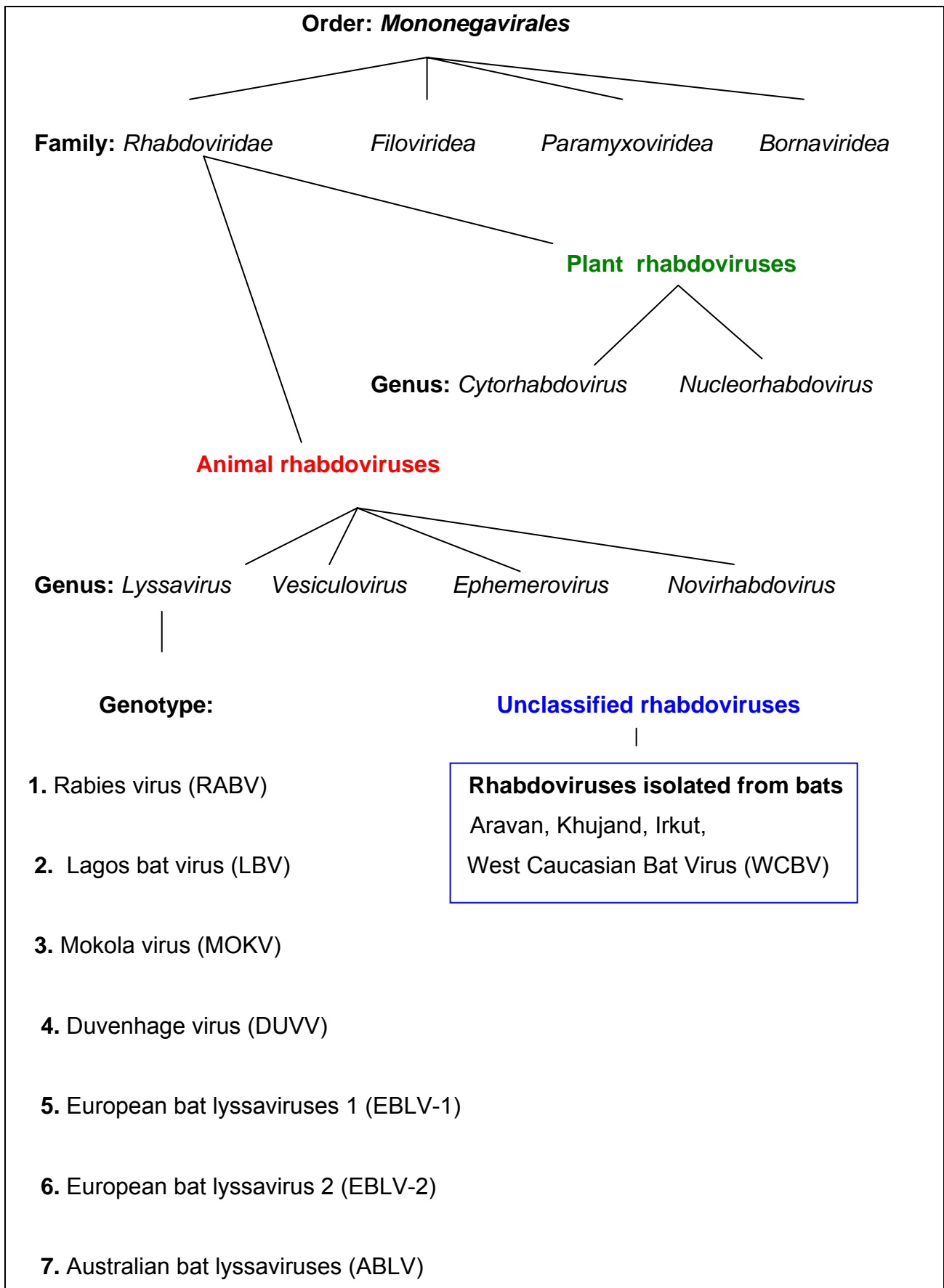

Figure 1.1. UPGMA phylogenetic tree constructed using the translated amino acid sequence of a 1350 nt. region of the N gene, for a representative sample of the seven lyssavirus genotypes. Numbered nodes indicate the number of times each node was reproduced in a 1000 bootstrap resamplings of the data (Smith *et al.*, 2002: reprinted with permission from Elsevier Science).

Table 1.1. Abridged classification of the *Mononegavirales* order (ICTV, 2005).

Order: *Mononegavirales*

Family: *Rhabdoviridae*   *Filoviridea*   *Paramyxoviridea*   *Bornaviridea*

Plant rhabdoviruses

Genus: *Cytorhabdovirus*   *Nucleorhabdovirus*

Animal rhabdoviruses

Genus: *Lyssavirus*   *Vesiculovirus*   *Ephemerovirus*   *Novirhabdovirus*

Genotype: **Unclassified rhabdoviruses**

1. Rabies virus (RABV)

Rhabdoviruses isolated from bats
Aravan, Khujand, Irkut,
West Caucasian Bat Virus (WCBV)

2. Lagos bat virus (LBV)

3. Mokola virus (MOKV)

4. Duvenhage virus (DUVV)

5. European bat lyssaviruses 1 (EBLV-1)

6. European bat lyssavirus 2 (EBLV-2)

7. Australian bat lyssaviruses (ABLV)

1.4 Molecular biology of the RABV

The vesicular stomatitis virus (VSV) is considered to be the prototype virus of the *Rhabdoviridae* family (Tordo and Poch, 1988; Banjeree, 1987). Most of the data generated on rhabdovirus architecture, mode of replication, as well as structural and functional features are based on research which was conducted on this virus (Tordo and Poch, 1988; Banjeree, 1987). VSV and RABV however differ in certain characteristics, such as their degree of neurotropism, the degree of inhibition of host cell macromolecular synthesis, and the speed at which the viral replication cycle is completed (Tordo and Poch, 1988).

1.4.1 Structure of the RABV

The RABV virion is a bullet shaped particle, 180nm in length (130 to 200nm), with an average diameter of 75nm (60 to 110nm) (Tordo and Poch, 1988; Wunner, 1991; Wunner, 2002). It consists of a helical ribonucleocapsid (RNP) that is composed of a negative-sense single stranded viral RNA genome, bound by molecules of nucleo (N), phospho (P) and polymerase (L) protein (Wunner, 1991; Swanepoel, 1994). The nucleocapsid contains 30 to 35 coils and is approximately 4.2 to 4.6 µm in length (Wunner, 1991; Swanepoel, 1994). The RNP is surrounded by a viral envelope consisting of a layer of matrix protein, which is in turn surrounded by a lipid membrane which was derived from the host cell membrane during viral budding (Wunner, 1991; Swanepoel, 1994; Wunner, 2002). Imbedded within the lipid membrane are numerous 10 nm long peplomeric spikes, consisting of trimers of glycoprotein (G), which have previously been described as stalks with hollow knobs at their distal ends (Murphy and Harrison, 1971; Wunner, 1991; Swanepoel, 1994). The viral particle is also typically hemispherical at one end, and planar at the other, the latter end budding from the host cell last (Tordo and Poch, 1988; Wunner,1991; Swanepoel, 1994; Wunner, 2002). Figure 1.2. presents a schematic presentation of a RABV particle (Tordo and Poch, 1988)

RABV can also generate defective interfering particles (DI) that vary in their size and replicative ability from standard rabies virions. DI are easily generated in a cell sub-culture of undiluted RABV, and vary in size from approximately one third, to two thirds the length of a standard rabies virus particle (Tordo and Poch, 1988: Wunner, 1991; Swanepoel, 1994; Wunner, 2002). Except for their reduced length, they usually exhibit identical

structural features to standard viral particles, however due to a truncated genome, they cannot replicate on their own, and are dependant on the co-infection of the host cell with a homologous parental virus for their propagation (Wunner, 1991; Swanepoel, 1994; Wunner, 2002).
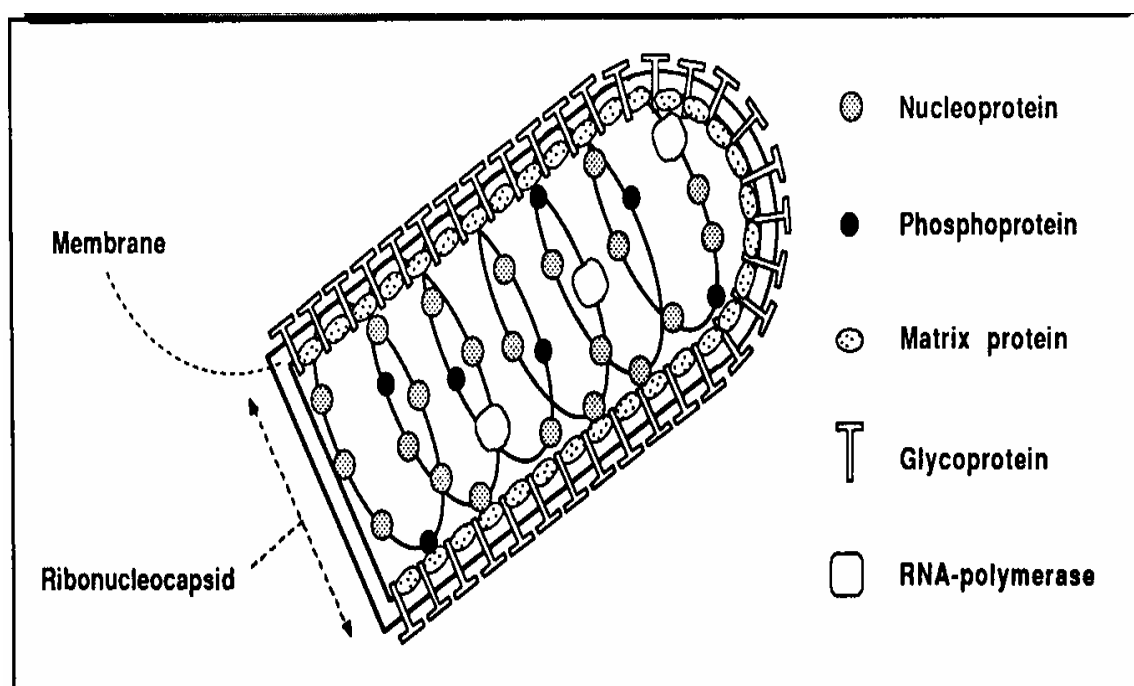

Figure 1.2. Schematic presentation of the rabies virus particle (Tordo and Poch, 1988). The coiled ribonucleocapsid (RNP) composed of the nucleoprotein (N), the phosphoprotein (P), the polymerase (L) and the single stranded negative sense RNA genome, is located on the inside of the particle. A layer of matrix protein surrounds the RNP, which is in turn surrounded by a lipid envelope which was derived from the host cell membrane during viral budding. A large number of peplomeric spikes, consisting of trimers of glycoprotein, are imbedded within the lipid membrane, and project from the outside surface of the virion (Reprinted with permission from Alice Essenpreis Springer).

1.4.2 Organization of the RABV genome

The complete nucleotide sequences of four lyssaviruses have been reported to date. These include the complete genomic sequence of the Pasteur virus (PV), the Street Alabama Dufferin (SAD)–B19 strain, the ABLV, and MOKV genome (Tordo *et al.*, 1986, 1988; Bourhy *et al.*, 1989, 1993; Conzelmann *et al.*, 1990; Le Mercier *et al.*, 1997; Wunner, 2002). Additional complete genomic sequences are also available from Genbank

for the Nishigara, HEP-Flurry, RC-HL, NI-CE and SHBRV-18 strains (NCBI). The RABV contains an unsegmented, negative sense RNA genome, 11 932 nucleotides (nt.) in length, with a molecular weight of approximately 4.6×10^6 Dalton (Da.) (Tordo and Poch, 1988; Wunner, 1991; Wunner, 2002). Its genome structure as well as its gene organisation, is similar to other negative stranded RNA viruses of the *Mononegavirales* order, and in particular to members of the *Rhabdoviridae* family (Wunner, 2002). From the 3' to the 5' end, the genome is transcribed into a short leader RNA, the nucleoprotein (N), the phosphoprotein (NS, P), the matrixprotein (M), the glycoprotein (G) and the viral RNA polymerase (L) (Wunner, 1991; Wunner, 2002). 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 *et al.*, 1993). The matrix and glycoprotein encoding gene regions of certain vaccine strains are an exception to the rule, in that they contain two consecutive stop signals from which long and short mRNAs are alternatively produced (Tordo *et al.*, 1993).

Non-coding intergenic regions separate the genes from each other, and except for the highly variable G-L intergenic region (423 nt.), are typically between 2 – 23 nt. in length (Wunner, 1991; Wunner, 2002). The viral genome also contains additional external signals, the 3' polymerisation promoter (58 nt.), which is recognised by the viral RNA polymerase in order to initiate transcription and replication, and a 5' end encapsidation promoter (70 nt.), that is recognised by the first molecule of nucleoprotein during the encapsidation of the viral genome (Wunner, 1991; Tordo *et al.*, 1993; Wunner, 2002). Conservation of these promoters at both extremities of the genome, results in an inverted complementarity of approximately 11 nucleotides at the genome ends (Tordo *et al.*, 1993; Wunner, 2002). Figure 1.3. demonstrates the order and length of the genes on the RABV genome (Tordo *et al.*, 1986; Levy *et al.*, 1994).

Figure 1.3. (page 42). Schematic presentation illustrating the order and length (nt.) of the genes (3'-5') on the rabies virus genome (Tordo *et al.*, 1986; Levy *et al.*, 1994). Prefixes on the figure are as follows: I, leader sequence, N, nucleoprotein, P, phosphoprotein, M, Matrix protein, G, glycoprotein and L, RNA polymerase. Numbers below the genes and intergenic regions, indicate their approximate nucleotide sequence lengths.

Rabies virus RNA genome

Nucleotides of intergenic regions
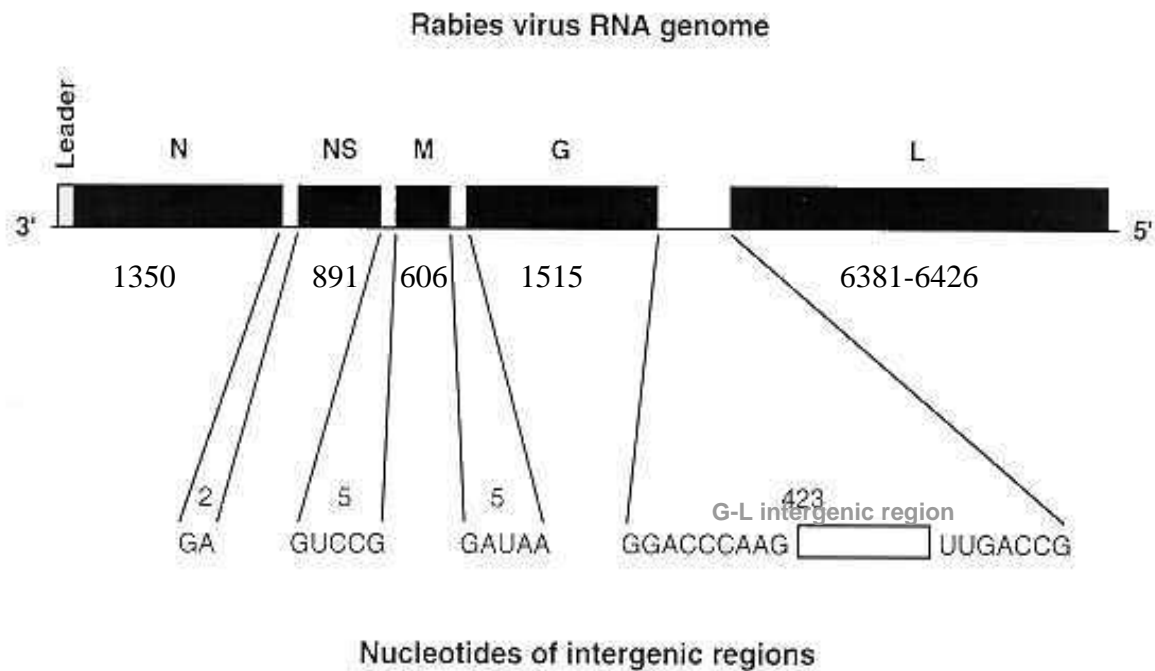
1.4.3.1 Coding regions

(a) The nucleoprotein (N)

The nucleoprotein is the major component of the internal viral nucleocapsid, and is tightly associated with the viral RNA genome (Tordo and Poch, 1988). It is a phosphorylated protein, 450 aa in length, with an approximate molecular weight of 57,000 Da. (Tordo and Poch, 1988; Wunner; 1991; Wunner, 2002). It is typically found inside the cell cytoplasm as a homologous N-N or heterologous N-P complexes, or bound to the viral RNA as part of the viral RNP (Wunner, 1991; Wunner, 2002).

The major function of the N protein is to encapsidate the viral RNA genome and viral replicative intermediates (RI), in order to prevent their digestion by endogenous cellular ribonucleases (Tordo and Poch, 1988; Wunner, 2002). It has also been proposed that the N protein regulates viral transcription and replication by its ability to promote read through of viral transcription termination signals, when present at a sufficient concentration (Tordo and Poch, 1988; Wunner, 2002). The absolute requirement for the N proteins' RNA binding ability is reflected by the fact that the N proetin of RABV contains a segmented amino acid homology with the N protein of VSV, illustrating the conservation of the domains required for RNA binding among these members of the *Rhabdoviridae* family (Tordo *et al.*, 1988).

Under electron microscopy (EM) the N protein appears to form bi-lobed structures that are closely packed along the viral RNA at intervals of 9 to 11 nt. (Wunner, 2002). It contains an RNA binding site within the N terminal domain between 298 and 352 aa, and after binding typically undergoes a conformational change that exposes serine 389, which is subsequently phosphorylated (Wunner, 2002). Phosphorylation at this residue is unique to the rabies virus, in contrast to the N protein of the VSV, and it has been proposed that this may directly influence the altered rates of transcription and replication observed between these viruses (Wunner, 2002).

In terms of its immunogenic and antigenic structure and function, the N protein represents the second most extensively analyzed protein to date (Wunner, 2002). This interest stems from its ability to protect animals against a peripheral rabies virus challenge in the absence of the production of virus neutralizing antibodies (VNAb), and its ability to prime the immune system with a resultant increase of VNAb following the subsequent inoculation of a target animal with an inactivated rabies vaccine (Wunner, 2002). 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 it has been shown to be able to act as a super antigen in humans (Dietzschold *et al.*, 1987; Fu *et al.*, 1994; Lafon *et al.*, 1992; Kawai and Morimoto, 1994; Wunner, 2002).

(b) The phosphoprotein (P, NS, M1).

The RABV phosphoprotein is a phosphorylated protein, 297 aa in length, with a molecular weight of between 38 000 – 41 000 Da. (Tordo and Poch, 1988; Wunner, 1991; Wunner, 2002). Among its many functions, is an ability to act as a chaperone of nascent N protein, binding to its carboxyl terminal extremity, and preventing its self-aggregation and non-specific binding to the viral RNA genome (Wunner, 2002). It has further been shown that it specifically directs genome encapsidation with nascent N protein as part of the N-P complex, and that it is an essential component of the viral RNA polymerase complex (P-L), in which it acts as a cofactor during transcription and replication (Wunner, 2002). As cofactor it has a duel function, simultaneously stabilizing the viral RNA polymerase and placing it on the viral genomic template, a function that the L protein is unable to achieve on its own (Wunner, 2002).

The phosphoprotein in the virion, and viral infected cells, is present in a number of different forms that differ from each other in terms of the degree of phoshorylation, as well as their length, and which can be distinguished from each other by an altered mobility on SDS-PAGE (Wunner, 2002). The P protein is furthermore also able to oligomerize with itself, with the oligomeric forms existing in equilibrium with the monomeric species in the cell cytoplasm (Wunner, 2002). The P protein contains a highly conserved hydrophilic core, and is phosphorylated on its N terminal domain at serine 63 and 64. Phosphorylation is carried out by two cellular kinases, the first of which is named *rabies virus protein kinase*, which is the predominant kinase packaged inside the virion, and isomers of a second less abundant kinase, named *protein kinase C* (Wunner, 2002).

In accordance with its multifunctional role as part of N-P and N-L complexes, the P proteins' co-function is mediated by domains which specifically interact with the N and L proteins. By using deletion mutant analysis, two independent N binding sites on the P protein have been mapped to the carboxyl terminal domain (1-30 aa) and the amino terminal domain (69-177 aa) respectively, while the L binding site has been mapped within the first 19 aa of the carboxyl terminal domain (Wunner, 2002). The P protein also contains a site on the amino terminal domain (138-173 aa) that allows it to bind to dynein LC8. Dynein LC8 is a component of the myosin V complex, a microtubule-associated motor protein complex that is implicated in the actin based transport of endoplasmic reticulum (ER) vesicles in brain neurons (Wunner, 2002). As such it has been speculated that the interaction of the P protein with the myosin V complex, may facilitate the spread of uncoated RNPs from neuron to neuron (Wunner, 2002). This hypothesis has however been weakened, by the demonstratiion that the G protein is required for transsynaptic spread of RABV *in vivo* (Etessami *et al.*, 2000).

(c) The Matrix protein (M)

The matrix protein is the smallest and most abundantly synthesized of the virion proteins. It is a 202 aa palmitoylated protein, with an approximate molecular weight of 25 000 Da, and is similar to the M protein found in VSV and Paramyxoviruses (Poch and Tordo, 1988; Wunner, 1991; Wunner, 2002). It contains an N terminal domain (35-38 aa), rich in proline and charged amino acid residues, as well as a central hydrophilic core (89-107 aa) (Wunner, 2002). The former domain is responsible for inhibition of host cell RNA

synthesis, a function that appears to correlate with the cell rounding cytopathic effects observed in cell culture for VSV, and to a lesser degree in RABV infected cells, while the latter domain appears to mediate the interaction of the M protein with the membrane lipids (Wunner, 2002).

The M protein plays an important role during the viral replication cycle, and acts as stimulatory factor for viral replication, an inhibiting factor of cellular RNA transcription, and a facilitator for viral maturation and budding (Wunner, 2002; Finke *et al.*, 2003). After inhibition of cellular RNA transcription, it forms a sheath-like structure around the viral nucleocapsid, condensing it into a tightly coiled ribonucleocapsid-M complex (Wunner, 2002). At the same time it facilitates the interaction of the assembled RNP complex with the cytoplasmic domain of membrane imbedded glycoprotein at the cell surface, in order to facilitate viral budding (Wunner, 2002).

Based on the results of negative staining immuno-electron microscopy and preliminary skeleton liposome binding studies, an alternative model for rhabdovirus molecular morphology has been proposed. In this model, instead of the matrix protein forming an external sheath around the RNP, the M protein forms a central core around which the RNP is wound (Coll, 1995). The M protein is then attached to either the G and/or the lipid membrane, but only at the ends of the viral skeleton (Coll, 1995). It has been suggested that both viral structures could exist in a dynamic equilibrium, depending on external factors such as pH. (Coll, 1995).

Although the precise mechanism of viral budding is not known, it appears to be mediated by the same N terminal rich proline motif (PY motif) responsible for the inhibition of host cell RNA synthesis (Wunner, 2002). The PY motif is able to interact with a double tryptophan domain (28-40 aa long) which is found in a range of cellular proteins and which is involved among other functions in cytoskeleton formation, signal transduction and gene regulation (Wunner, 2002). As such it has been speculated that the RABV M proteins interaction with cellular proteins containing these motifs, might play a role in viral budding (Wunner, 2002).

Finally it has also been found that the matrix protein plays an important role in the induction of cellular apoptosis through a TRAIL (Tumor Necrosis Factor Related Apoptosis Inducing Ligand) dependant pathway involving caspase 8, 3 and 6 (Kassis *et al.*, 2004).

The matrix protein does not directly interact with the TRAIL-death receptors on the cell surface, but may instead facilitate the activation of the apoptopic transduction pathway through the stabilization of the oligomerization, or the post translational modification of the TRAIL protein (Kassis *et al.*, 2004).

The Glycoprotein (G)

The glycoprotein is the best studied of all the rabies virus proteins due to its importance in the induction of the host immune response, and thus its importance for vaccine preparation (Tordo and Poch, 1988; Kawai and Morimoto, 1994; Coll, 1995; Wunner, 2002). The mature glycoprotein is a 505 aa glycosylated protein, with an approximate molecular weight of 65 000 Da (Wunner, 2002). From its amino to carboxyl terminal regions its nascent form contains a 19 aa signal peptide domain, a 439 aa ecto-domain, a 22 aa transmembrane domain and a 44 aa cytoplasmic domain (Tordo and Poch, 1988; Coll, 1995; Wunner, 2002). The function of these domains can be summarized as follows:

The hydrophobic signal peptide domain - provides the signal responsible for the insertion of the G protein into the membranes of the rough endoplasmic reticulum-Golgi-plasma membrane pathway (Tordo and Poch, 1988; Wunner, 2002). It is cleaved off from the N terminal region after the protein has been inserted into the Golgi apparatus membrane, resulting in the formation of the 505 aa mature glycoprotein (Tordo and Poch, 1988; Wunner, 2002).

The ectodomain - forms the trimeric peplomeric spikes, which extends to the outside of the mature viral particle. It is responsible for viral binding to host cell receptors, and fusion of the viral particle with the endosomal membrane (Tordo and Poch, 1988; Wunner, 2002). Consequently it also plays an important role in determining virulence (Tordo and Poch, 1988; Wunner, 2002). The ectodomain also plays a critical role in the host immune response as it contains the major antigenic sites responsible for eliciting VNAb, and as it is a major target for virus specific Th and cytotoxic T cells (Tordo and Poch, 1988; Wunner, 2002).

The hydrophobic domain – is responsible for anchoring the glycoprotein to the viral envelope (Tordo and Poch, 1988; Wunner, 2002).

The cytoplasmic domain - interacts with the M protein of the viral skeleton particle in order to complete the assembly of the virus during budding (Tordo and Poch, 1988; Wunner, 2002). The efficiency of viral budding seems to be influenced by the interaction of the RNP-M complex with the cytoplasmic domain of the envelope G protein, suggesting that a concerted action between these proteins is necessary for efficient virion recovery (Tordo and Poch, 1988; Wunner, 2002).

In terms of the proteins immunogenicity it contains at least 8 antigenic sites (I-IV, "a", and G1), located on the ectodomain of different RABV strains (Wunner, 2002). Sites I, III, VI and "a" have been mapped to aa located at positions 231, 330-338, 264, and 342, respectively, while site II is a discontinuous antigenic site involving two separate stretches of aa in position 34-42 and 198-200, and which are presumably linked by a disulfide bridge (Wunner, 2002). Antigenic site III is of special interest as it contains the residue Arg 333 (or Lys-333) which plays a critical role in viral pathogenesis. Virus variants containing glutamine, (Gln), isoleucine (Ile), glycine (Gly), methionine (Met), or serine (Ser) substituted for Arg-333 in the G protein, express a phenotype with a reduced pathogenicity when compared to the wild type virus when inoculated intracerebrally into adult immunocompetent mice (Dietzschold *et al.*, 1983; Coulon *et al.*, 1993; Badrane *et al.*, 2001; Wunner, 2002). It appears that the presence of Arg-333 affects the ability and rate at which the virus is able to spread from cell to cell, as well as what neuronal pathway the virus takes to reach the CNS (Wunner, 2002).

(e) The RNA polymerase (L) protein

The L protein is the largest polypeptide encoded by the RABV genome, it being between 2127 and 2142 aa in length, and comprising approximately half of the genomes encoding potential (Wunner, 1995; Tordo and Poch, 1988; Wunner, 2002). It is the main catalytic component of the viral polymerase complex, and together with the P protein, is responsible for the majority of enzymatic activities during viral transcription and replication (Wunner, 2002). Not only is it responsible for RNA synthesis, but it is also responsible for the cotranscriptional modification of primary mRNA transcripts which includes 5' capping, methylation and 3' polyadenyaltion (Wunner, 2002)

Based on the comparison of L-nucleotide sequences between members of the *Mononegavirales,* a number of conserved sequence domains have been identified

(Wunner, 2002). Four of these motifs labeled A through D, present between aa 530-1177 in the RABV L, and between aa 532-1201 in the VSV L, represent regions of the highest similarity (Wunner, 2002). It is maintained in same linear arrangement in all viral RNA and DNA polymerases, and is thought to constitute the polymerase module (Wunner, 2002). A tri-amino acid core sequence named GDN (standing for glycine, aspartic acid and asparagine), present within motif C, is extremely conserved throughout the *Mononegavirales* order, and together with specific aa sequences downstream of the core sequence is crucial in the maintenance of nucleotide polymerization activity (Wunner, 2002). In addition, at least two other sequences between aa residue 754-778 and 1332-1351 in the VSV L protein have been identified as consensus sites for the binding and utilization of ATP. These ATP binding sites are required for the L enzymatic activity involved in (1) transcription, which requires the binding of substrate dNTPs, (2) polyadenylation and (3) protein kinase activity involved in the specific phosphorylation of the P protein during transcriptional activation (Wunner, 2002).

1.4.3.2 Non-coding regions

(a) Leader RNA (Le)

The first RNA to be transcribed from the viral genome is the short leader RNA that is between 55 to 58 nt. in length (Tordo and Poch, 1988; Wunner, 2002). It is complementary to the 3' end of the viral genome, and is produced at the highest molecular ratio of all the RNA transcripts (Tordo and Poch, 1988; Wunner, 2002). It is neither translated, capped nor polyadenylated and has been proposed to functions as a inhibitor of host cell macromolecular synthesis (Kurilla and Keene, 1983; Wunner, 2002) and that it regulates transcription of the viral genes in a start-stop manner by its ability to interact with viral promoter elements (Tordo and Poch, 1988; Wunner, 2002). Further it has also been proposed that it acts as a decision point in the switch between the transcriptive and replicative mode of the viral genome by its ability to bind viral nucleoprotein (Tordo and Poch, 1988; Wunner, 2002).

24

(b) Intergenic sequence

Intergenic sequences are defined as lying between the 3' end of one mRNA encoding gene region on the viral genome and the 5' start of the following one (Tordo *et al*, 1986; Tordo and Poch, 1988). The intergenic regions of the RABV all start with guanosine, and are 2, 5, 5 and 423 nt. in length between the N-P, P-M, M-G and G-L regions respectively (Tordo *et al.*, 1986). Of special interest is the unusually long G-L intergenic region, as it has been found to contain either one, or two, transcription termination and polyadenylation signals (TTP) for the G gene, among certain laboratory adapted strains and wildtype street virus. (Tordo *et al.*, 1986; Tordo and Poch, 1988; Ravkov et al., 1995). In viral strains, which contain two TTP signals, the first is located 70 nt. downstream from the stop signal of the G gene (position 4974), while the second sequence is located 25 nucleotides upstream from the L gene transcription initiation signal (Tordo *et al.*, 1986; Ravkov *et al.*, 1995). (Wunner, 2002). In the PV and Evelyn-Rokitnicki-Abelseth (ERA) strains both polyadenylation and transcription termination signals are present, while only the downstream signal is present among geographically diverse wildtype street viruses and laboratory adapted strains such as the Flury high egg passage (HEP), CVS, Pitman Moore (PM), the Nishigara strain and the MOKV (Wunner, 2002). The absence of the first TTP signal has also been reported for rabies virus isolates from South Africa and Zimbabwe (Sabeta *et al.*, 2003).

The presence of both TTP signals among the PV and the ERA strain raised the possibility that the G-L intergenic region represented a remnant protein open reading frame (ORF), in an intermediate stage of Rhabdovirus evolution (pseudogene) (Tordo *et al.*, 1986). This hypothesis was based on the considerations that the flanking sequences might have resulted from the degeneration of the transcription consensus signals, and that the G-L intergenic region is the sole large genomic region of the viral genome which is extensively blocked in all reading frames (Tordo *et al.*, 1986). Further evidence for the pseudogene nature of this genomic region is provided by the fact that a sixth gene, corresponding to the G-L intergenic region in a fish rhabdovirus (infectious hematopoietic necrosis virus, IHNV), encodes a protein of unknown function (NV, non-virion protein), and that paramyxoviruses, which also form part of the *Mononegavirales* order, encode a second glycoprotein, whose encoding gene is located in an corresponding section of the viral genome as the G-L intergenic region(Tordo *et al.*, 1986; Tordo and Poch, 1988).

A conflicting view as to the pseudogene nature of the G-L intergenic region was presented by Ravkov *et al.*, in 1995. By analyzing the nucleotide sequences of a number of geographically diverse street viruses and laboratory adapted strains, the author came to the conclusion that the G-L intergenic region does not represent a vestigal gene, but that instead, the rabies virus genome encodes a G mRNA with a long 3' non-coding sequence. This conclusion was based on the findings that the presence of the double TTP signals appeared to be confined to a specific vaccine lineage, and that the upstream signal appeared to be absent from all the street viruses which were included in the analysis (Ravkov *et al.*, 1995). Furthermore phylogenetic analysis of the sequence data indicated that none of the laboratory adapted strains which contained both these signals were ancestral to the viruses which contained only the downstream signal, precluding the possibility of the presence of both signal motifs as being representative of the rabies virus in its ancestral form (Ravkov *et al.*, 1995). The author proposed that the additional upstream TTP signal may have been incorporated due to polymerase stuttering during genome replication and that this signal together with the presence of the 3' non-coding trailer may be involved in the down regulation of the expression of the L protein (Ravkov *et al.*, 1995).

1.5 Genetic variability of the Lyssavirus genome

The nucleotide sequence determination, and the comparison of the complete genomic sequence of the PV strain (genotype 1) and the most divergent member of the Lyssavirus genus, the MOKV (genotype 3), have allowed for the determination of the relative conservation that exists between the different genes contained within the *Lyssavirus* genus genome (Bourhy *et al.*, a1993, b1993; Nadin-Davis, 2000). This has led to the development of oligonucleotide primers that bind to conserved docking sites, flanking the informative regions of the RABV genome, allowing for the development of PCRs for diagnostic and molecular epidemiological purposes (Sacramento *et al.*, 1991)

Figure 1.4. demonstrates a similarity profile which was generated by comparing the complete genome nucleotide sequence of the PV strain and the MOKV, illustrating the highly variable degrees of homology that exists between the five genes of the *Lyssavirus* genome (Bourhy *et al.*, a1993; Nadin-Davis, 2000). The N gene is well conserved (77%), followed by the M (70.8%), G (54.2%) and P genes (45.9%) (Bourhy *et al.*, 1995; Nadin-Davis, 2000). The G gene is divided into discrete regions, exhibiting high (ectodomain)

26

and low conservation (signal, transmembrane and cytoplasmic domains). Of these regions the G cytoplasmic domain is the least conserved (Nadin-Davis, 2000). The most variable region of the lyssavirus genome is the non-coding G-L intergenic region, which does not encode a gene product, and which is thus exempt from selective pressures, allowing it to essentially mutate at random (Bourhy *et al.*, 1995; Nadin-Davis, 2000). Interestingly enough it has been shown that the rate of mutation is not uniform throughout the G-L intergenic region, with clearly distinguishable regions of high and low variation being present (Nel *et al.*, 1993). The reason for the variable rates of mutation in the different regions of the pseudogene is however currently unknown. The L gene is the most conserved of all the lyssavirus genes, due to a requirement for the conservation of the domains which encode the structural features of the protein, which are essential for its catalytic functions (Bourhy *et al.*, 1995; Nadin-Davis, 2000).

The degree of homology within the different regions of the Lyssavirus genome affects the suitability of these regions for diagnostic, taxonomic and epidemiological studies. For diagnostic and taxonomic studies, a well-conserved region of the genome proves most suitable, while for epidemiological studies the choice of the genomic region targeted depends on the degree of strain discrimination that is required for a particular study (Nadin-Davis, 2000; Bourhy *et al.*, 1993, 1995). The N protein has been widely applied to diagnostic and taxonomic studies, while the N, P, G and G-L intergenic regions have been used for strain discrimination at a global and regional level (Nadin-Davis, 2000). The highly variable G-L intergenic region is particularly suited for studies requiring sensitive strain discrimination, even down to a local level (Nadin-Davis, 2000).

The RABV has been shown to evolve without recombination, and the evolutionary relationships which can thus be constructed from any given gene region thus usually correspond (Nadin-Davis, 2000). Furthermore phylogenetic trees constructed using shorter stretches of sequence usually correspond to trees which were constructed from longer gene regions. The trees constructed from the shorter sequences however, usually have lower bootstrap confidence values with respects to the reconstructed phylogeny, due to the fact that fewer informative nucleotide sites are available from the respective sequence alignments (Nadin-Davis, 2000).
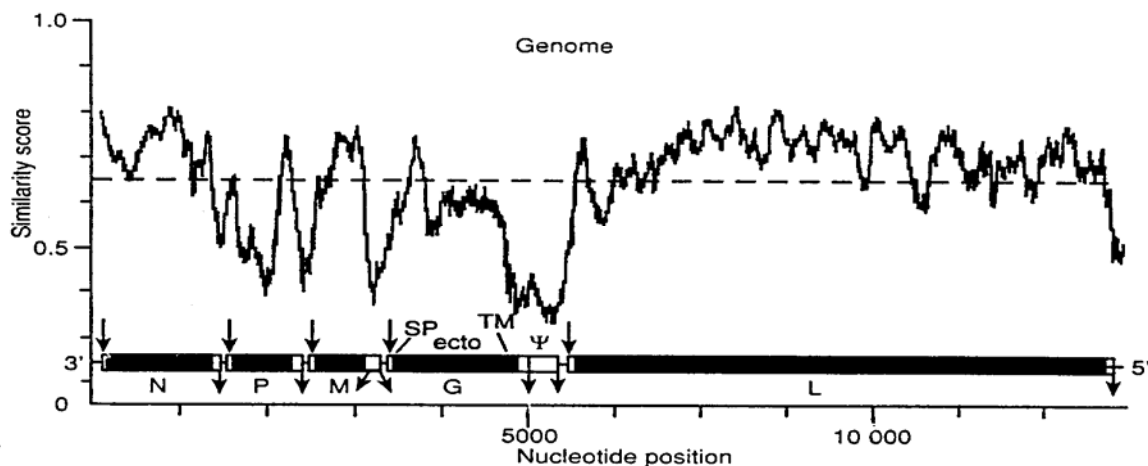
Figure 1.4. Similarity profile generated by comparing the complete genomic nucleotide sequence of the PV (GT1) and MOKV (GT3) genomes (Bourhy *et al.*, 1993; Bourhy et al., 1995; Nadin-Davis, 2000). Shaded areas indicate the mean similarity values for each gene, while the hatched line indicates the overall similarity value between the two genomes. Black bars at the bottom correspond to coding regions, while arrows below and above the bar indicate the transcription initiation and stop signals, respectively. The location of the signal peptide (SP), ecto and transmembrane domains of the G protein, as well as the G-L intergenic region (Ψ), are also indicated.

1.6 Heterogeneous nature of RNA genomes: Quasispecies concept

The quasispecies concept originally developed by Eigen and Colleagues (1979), has been shown to accurately describe the composition of an RNA virus population at the molecular level (Eigen and Schuster, 1979; Domingo, 1997). It presupposes that RNA viruses, including the RABV, exists as a range of mutants within an infected individual, with individual viral genomes differing by one or more nucleotides from the average consensus sequence of the entire viral population (Domingo, 1997). As a result these quasispecies populations have a significant adaptive potential, as they allow for the rapid selection of mutants with the highest fitness, in any new environmental condition (Morimoto *et al.*, 1998).

The rate of mutation in RNA viruses (10^{-3} to 10^{-5} bp substitutions/replication cycle) are several magnitudes higher than in DNA containing organisms ($\leq 10^{-7}$ bp substitutions/replication cycle), due to the fact that the RNA polymerase lacks a proofreading/repair function, and is thus unable to carry out post replicative repairs on erroneously incorporated bases in a newly replicated viral genome (Levy *et al.*, 1994; Domingo, 1997). This coupled with a short generation time for, and the production of a

large number of progeny virions, results in the development of a mixture of related but distinct RNA genomes, in which heterogeneity increases rapidly with infection cycles over time (Smith *et al.*, 1997). The fixation of mutations in a viral population, will depend on selective pressures which are functioning in the regions of the genome these mutations occur in. For example, negative selection in a coding region will tend to result in the conservation of the consensus sequence, while positive selection will promote evolution and the fixation of new mutations in a viral population (Smith *et al.*, 1997). Mutations in non-coding regions are not under selective pressure, and will thus be incorporated at a rate equal to the error rate of the RNA polymerase (Smith *et al.*, 1997). Additional factors that may also affect the diversity of viral populations, include genetic bottlenecks in which only a few particles are transferred to a new host, or where selection occurs for viral variants with a specific cell tropism or the ability to evade the host' immune system (Smith *et al.*, 1997).

The genetic variation generated in this way, forms the basis for the comprehensive differences that allows viruses to be taxonomically sorted into different families, genera and species, as well as for the minor differences that allows for the grouping of viruses into serotypes and genotypes within a viral genus (Wunner, 2002). In the RABV, the large amount of genetic instability that exists within the viral quasispecies provides the driving force for the evolution of new viral variants, with distinctive phenotypes or unique virus-host relationships (Wunner, 2002) (see section 1.12). For example, in any specific ecological zone a single dominant host (maintenance host) will be responsible for maintaining a variant which induces specific characteristics in its host that will ensure its maximal transmission. Examples of such virus induced properties include the induction of appropriate biting behavior during a virally influenced clinical survival period, timed for example, to allow the transmission of the virus to conspecifics (biotype). Typically transmission of these host specific variants to other species results in a dead end infection, due to for example the failure of the virus to induce the appropriate biting behavior, or the failure of the virion to shed efficiently from the spillover hosts' salivary glands. However, occasionally spillover infections with host adapted variants are successful, resulting in the establishment of viable infectious cycles in a new host species. This is due to the quasispecies nature of the RABV genome, which ensures that there will be a sufficient pool of genetically variable viruses available, from which a better-fit variant can be selected (Bingham, 2005).

1.7 Intracellular replication cycle of rabies virus

The sequence of events in the rabies virus replication cycle *in vivo*, can be divided into three phases:

- **The early phase** - includes virus attachment to surface receptors on susceptible host cells, entry of viral particles into the cell by the direct fusion of the virus to either the plasma or endosomal membrane, uncoating of the virus particles, and the release of the helical RNP into the cell cytoplasm (Wunner, 2002).

- **The middle phase** - includes the transcription and translation of viral mRNAs, the replication of the viral genome, and the post-translational modification of viral proteins (Wunner, 2002).

- **The late phase** - includes the assembly of progeny virions at the plasma membrane of the host cell, and the release (budding) of the completed particles from the surface of the infected cell into the surrounding cellular environment (Wunner, 2002).

Early-Phase Events.

RABV infection starts with the virions' attachment to a target cell surface, and uptake, most likely by a receptor or cellular receptor unit (CRU), which permits the virus entry into susceptible cells in culture (*in vitro*), or specific target cells at the site of inoculation (*in vivo*) (Wunner, 2002). Several studies using various cell culture systems have implicated lipids, gangliosides, carbohydrates, and proteins of the plasma membrane in rabies virus binding to cell culture, while others have focused on specific CRU's *in vivo* that appear to correlate with the defined tropism of the virus (Tsiang, 1988; Wunner, 2002). CRU's that have been implicated in the attachment of viral particles to muscle cells include the acethycholine receptor (AChR), while the neural cell adhesion molecule (NCAM) CD56 and the low affinity neurotrophin (NT) receptor p75NTR have also been implicated as possible neural attachment CRUs (Tsiang, 1988; Wunner, 2002).

After the virus had bonded to the cell receptor, it is internalised into the cell via receptor-mediated endocytosis (Wunner, 2002). After internalisation, the viral G protein mediates the fusion of the viral envelope with the endosomal membrane, during a low pH-

dependant fusion process, mediated by the low pH environment of the endosome (pH 6.2-6.3) as well as a low pH fusion domain thought to lie between aa 102-179, thereby resulting in the release of the naked RNP into the cell cytoplasm (Wunner, 2002).

Middle-Phase events.

After the RNP is released into the cell cytoplasm, the tightly coiled nucleocapsid of the RNP structure relaxes, to form a loosely coiled helix (Wunner, 2002). The transcription process in virus replication is carried out on the viral RNA-N complex, by the virion associated RNA polymerase complex (Wunner, 2002). The virion associated polymerase either starts transcription at the 3' end of the genome, or it starts at the next mRNA start site, downstream of the position in which the RNA polymerase was "frozen" to the viral genome, during progeny virus assembly in a previous infected cell (Wunner, 2002). During transcription, six RNA species are produced in a sequential order, from the 3' to the 5' end of the genome, of which five are eventually translated into viral proteins (Wunner, 2002).

The amount of protein produced from each gene is strongly influenced by its position on the genome (Tordo and Poch, 1988; Wunner, 2002). As the polymerase moves from the 3' end to the 5' end of the genome, it pauses at intergenic regions preceding each gene before continuing the transcription of downstream mRNAs (Tordo and Poch, 1988). It is estimated that during these pauses an approximate 20-30% of the polymerase complex dissociates from the viral genome (Wunner, 2002). As a result fewer polymerase molecules remain associated with the genome RNA-N template after each gene junction, resulting in a decreased rate of transcription and translation for genes from the 3' end to the 5' end of the genome (Wunner, 2002).

Two sequences flank each gene, and mark where initiation and termination of transcription occurs (Wunner, 2002). The sequence 5'-AACA-3' at the 3' end of each gene identifies the start site of transcription, while the sequence $3'\text{-AC(U)}_{7-8}\text{-}5'$ at the 5' end of the gene acts as termination and polyadenylation signal for transcription in the RABV and MOKV genome (Wunner, 2002). During synthesis of nascent mRNAs, the L protein caps the 5' end of each mRNA by attaching a 7-methyl guanosine (5'-m^{7}Guppp) to the 5' nucleotide of the newly synthesized strand. When the viral polymerase complex reaches the string of 7 to 8 U in the genome at the end of each gene, it reiteratively copies

31

the U's, resulting in the formation of a poly A tail of variable length (up to 200 A's) at the 3' end of each mRNA (Wunner, 2002). The mRNA, which has been capped and polyadenylated, is then released, and the transcriptase after skipping the next intergenic sequence, restarts transcription at the next downstream transcription initiation signal (Wunner, 2002).

The protein synthesis machinery of the host cell is responsible for the translation of viral mRNAs (Wunner, 2002). The N-, P-, M-, and L-mRNAs are translated by free ribosomes in the cell cytoplasm, whereas the G-mRNA is translated by the membrane bound ribosomes of the endoplasmic reticulum (ER) (Wunner, 2002). During the translation of the G-mRNAs, the nascent G protein is inserted into the lumen of the ER, where post-translational modification, including the formation of disulfide bonds, folding by molecular chaperones, core glycosylation and N-glycan processing takes place (Wunner, 2002). Furthermore, monomers of G protein associate with each other in the ER lumen, to form the homotrimeric spikes which are eventually incorporated in the viral envelope (Wunner, 2002). The addition of N-linked carbohydrate chains takes place in the Golgi apparatus as the G protein trimers are transported to the cell surface. This is accomplished by the sequential removal and addition of monosaccharides by glycosidases and glycosyltranferase (Wunner, 2002).

Viral replication becomes the dominant event following transcription in the late phase of viral growth (Wunner, 2002). During this process the polymerase no longer recognises intergenic signals that cause transcription initiation and termination at gene junctions. This results in the synthesis of only complementary full length (+) sense viral RNA genomes (RI), that act as templates for the synthesis of new full length progeny genomes (Wunner, 2002). It is unknown what the precise mechanism is that switches the viral replicative cycle from transcriptive to replicative mode, although it has been speculated that this may be directly related to the concentration of N protein in the cell. In this mechanism, N encapsidation of the Le+ RNA transcripts either prevents termination of transcription at the Le-N junction, or prevents the Le+ transcripts from initiating the transcription of individual down stream mRNAs (Wunner, 2002).

Late-Phase events.

Once a sufficient pool of viral progeny genomes and N, P and L protein have accumulated in infected cells, rabies virus nucleocapsids are formed, and virus assembly begins and continues as long as the cell is metabolically active (Wunner, 2002). This process appears to be dependent on the relative concentrations of N and P protein (ratio 2:1) in the cell cytoplasm, which allows for the formation of heterologous N-P complexes (Wunner, 1991; Wunner, 2002).

During encapsidation, the N protein recognises a 5' cis acting encapsidation signal that acts as nucleation signal for the interaction of soluble N protein with the viral genomes and antigenomes (Wunner, 2002). After recognition of this signal by the N protein, encapsidation proceeds rapidly from the 5' to the 3' direction on the RNAs, independent of the viral RNA sequence (Wunner, 2002). The precise mechanism by which the L protein is incorporated in the viral RNP is unknown, although it is thought that the RNA-N-P complex mediates its binding to the viral genome (Wunner, 2002).

After the assembly of the viral RNP is completed, it starts migrating towards the cell surface. While the RNP moves towards the cell surface it binds to soluble M protein, which forms a skeleton like structure around the RNP (Wunner, 2002). The M protein localises the RNP at the cell surface, where it subsequently interacts with the cytoplasmic domain of the membrane embedded G proteins, resulting in the acquisition of the viral envelope, and budding of the virus, through the host cell plasma membrane (Wunner, 2002).
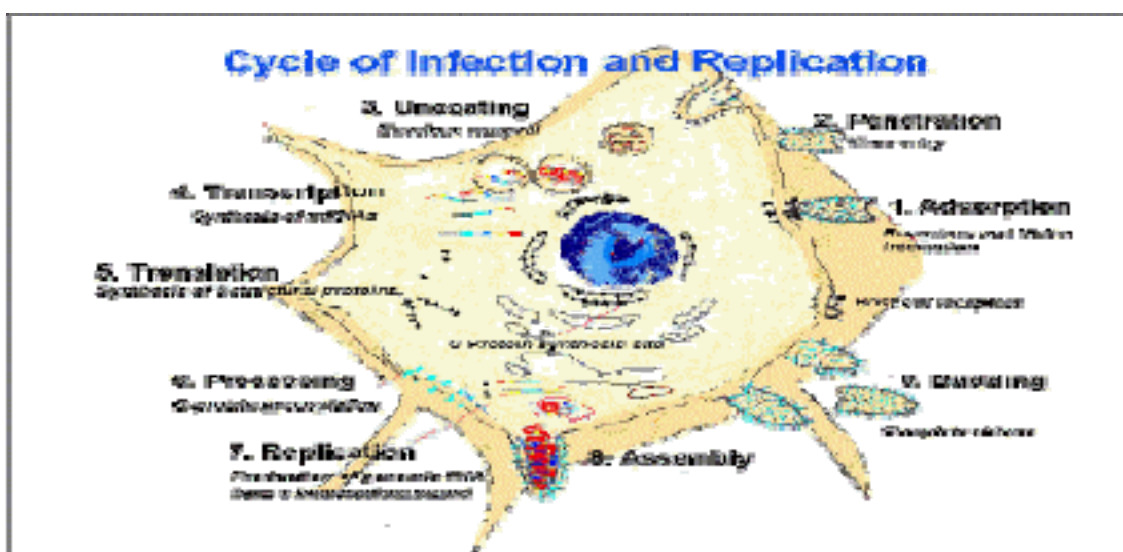

Figure 1.5. Infection and replication cycle of the rabies virus.

1.8 Rabies in humans: Clinical presentation and pathogenesis

Rabies is an acute incurable encephalitis, which can be caused by all members of the Lyssavirus genus, although classical rabies virus (RABV) accounts for most cases worldwide (Hemachudha *et al.*, 2002). All warm blooded mammals tested so far are susceptible to RABV infection, which is usually transmitted by the exposure of wounds or cuts on the skin, to virus-laden saliva, most commonly inoculated through a bite from an infected carnivore (Warrell and Warrell, 2004, Hemachudha *et al.*, 2002). Non-bite transmissions are highly unusual, but have been documented to occur in the literature. These include transmission of high concentrations of aerosolized RABV to laboratory workers and spelunkers in bat-infested caves, transmission through the transplant of infected human organs, as well as by oral transmission through the ingestion of contaminated breastmilk (Afshar, 1979).

Classic rabies is almost always fatal once symptoms appear and is associated with true RABV (genotype 1) infection (Hemachudha *et al.*, 2002) Survival of rabies is extremely rare, and only half a dozen cases are currently accepted as authentic. Most of these patients received some type of rabies PEP treatment, but in most cases recovery was associated with residual neurological complications (Krebs *et al.*, 1995; Hemachudha, 2002; Warrell and Warrell, 2004). Only one patient, a girl from Wisconsin in the U.S.A who was bitten by a bat in 2004, survived rabies without any pre or post exposure prophylaxis. Unfortunately no RABV could be isolated from saliva or nuchal skin biopsies, and therefore the identification of the virus variant which was responsible for this infection was not possible (Willoughby *et al.*, 2004).

Rabies mortality in humans after bites from rabid dogs, varies from 38% to 57%, and depends on the severity and location of the wound, as well as the presumed concentration of RABV in the saliva (Hemachudha *et al.*, 2002). Bites on the head, face, neck and hand, particularly with bleeding, carries the greatest risk, and is generally associated with short incubation periods (Hemachudha *et al.*, 2002). After inoculation of the virus laden-saliva, the virus can either replicate locally in muscle cells before gaining access to the CNS via neuromuscular spindles or muscle motor end plates, or alternatively can attach directly to sensory nerve endings (Swanepoel, 1994; Warrell and Warrell, 2004). Having gained entry to peripheral nerves, it travels via retrograde axoplasmic transport to the brain, where massive replication of RABV ensues within neurons (Warrell and Warrell, 2004). A

combination of features explains the apparent lack of an antibody response during the initial stages of the disease. Firstly viremia has never been seen in any stage of the disease, whiles secondly the sequestration of the virus in nervous tissue shields the virus from the hosts' immune system (Krebs *et al.*, 1995). Within the CNS the virus can directly spread from cell to cell across synaptic junctions, or it can spread along neuronal pathways via intra-axonal transport (Swanepoel, 1994; Warrell and Warrell, 2004). Following infection of the CNS, there begins a reverse spread of the virus along peripheral nerves to sites throughout the body, including the salivary glands, where it may be shed in the saliva, with, before or after the development of clinical signs (Krebs *et al.*, 1995; Fekadu, 1993; Warrell and Warrell, 2004). It has been reported that apparently healthy dogs and bats are able to excrete RABV in their saliva, raising the possibility of a possible carrier state that may play a role in the transmission and perpetuation of the disease (Irvin, 1970; Fekadu, 1993; Swanepoel, 1994). Such a possible carrier state has possibly been demonstrated for spotted hyenas (*Crocuta crocuta*) in the Serengeti. East *et al.*, demonstrated by PCR that rabies virus RNA was present in the saliva, of a small percentage of apparently healthy hyenas (East *et al*, 2001). The findings from the PCR amplification of the virus RNA, could however not be confirmed by virus isolation from the saliva (East *et al.* 2001).
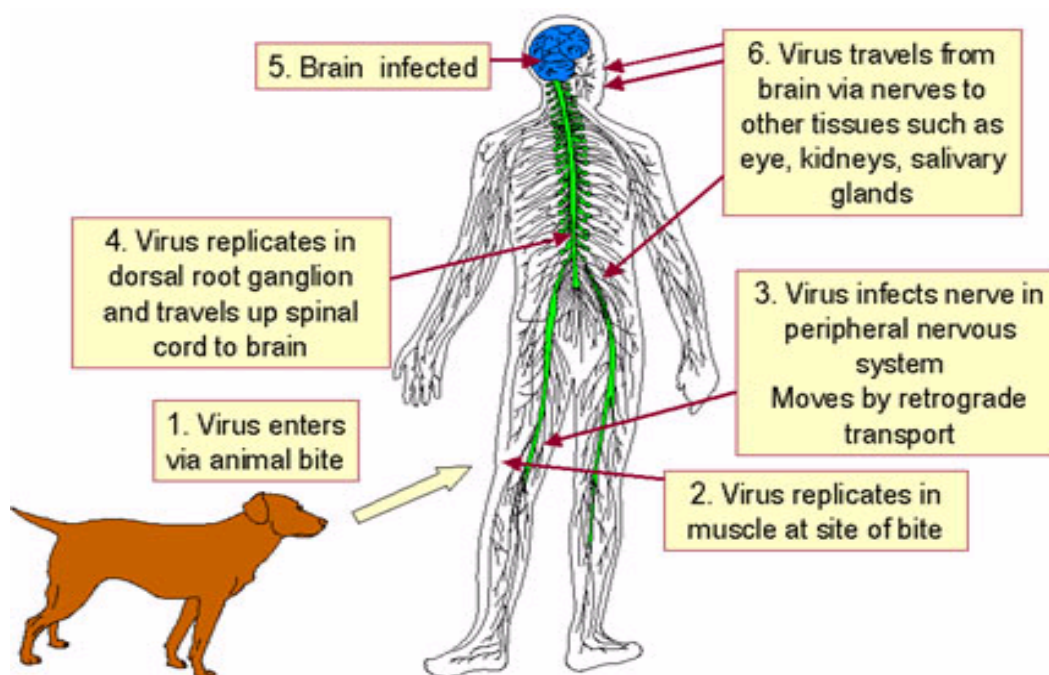
Figure 1.6. Route of spread of the RABV, from a peripheral inoculation site, to the CNS, and from there to other sites throughout the body (http://pathmicro.med.sc.edu/virol/rabies.htm).

Clinical features of the RABV infection consist of five stages: the incubation period, the prodrome, the acute neurological phase, coma and death (Hemachudha *et al.*, 2002). The acute or neurological phase of classical rabies can be divided into the encephalitic and paralytic forms of the disease, that differ from each other in term of the diseases' clinical presentation (Hemachudha *et al.*, 2002). The encephalitic form of the disease corresponds to the aggressive phase seen in dogs, in which animals often bite without provocation, while the paralytic form of the disease in humans corresponds to the so called "dumb" form of the disease in dogs, in which paralytic symptoms dominate (Hemachudha *et al.*, 2002).

Incubation period.

The symptomless incubation period of the RABV infection in humans is highly variable, with the most common period varying from one to two months (Hemachudha *et al.*, 2002). Incubation periods shorter than 1 week, and longer than 6 years have however been demonstrated (Hemachudha *et al,* 2002).

Prodrome.

The prodrome stage begins when the virus moves from the peripheral to dorsal root ganglia, and from there on into the CNS (Hemachudha *et al.*, 2002). These developments mark the end of the incubation period, and most patients die within the next two weeks (Hemachudha *et al.*, 2002). Symptoms at this phase are non-specific, but include fever, headache, malaise and non-specific gastrointestinal symptoms (Hemachudha *et al.*, 2002). Other symptoms, including irritability, depression, anxiety and insomnia, may also be present (Hemachudha *et al.*, 2002). Sensory symptoms at the bite site, which has usually healed over, are commonly experienced, and may include intense pruritis (itching), paraesthesia (non specific sensory symptoms) or pain (Bishop *et al.*, 2003). Prodromal symptoms last a few days, but are generally not present for more than a week (Hemachudha *et al.*, 2002).

Acute neurological phase.

The acute neurological phase is characterized by the appearance of signs in which neurological and behavioral symptoms dominate (Hemachudha *et al.*, 2002; Bishop *et al.*,

2003). Encephalitic rabies is found in two thirds of patients with classical rabies, while the rest present with the paralytic form of the disease (Hemachudha *et al.*, 2002). Periods to morbidity vary between the paralytic and encephalitic form, with patients with encephalitic rabies usually succumbing within a week from the onset of the disease. The average survival period for patients presenting paralytic rabies can however be as long as two weeks (Hemachudha *et al.*, 2002).

It has been proposed that different rabies virus variants, associated with particular geographic regions or vector species, may be responsible for the altered clinical presentation observed between paralytic and encephalitic rabies. Hemachuda *et al.*, 2003 investigated whether strain specific differences are associated with the altered clinical presentations, by analyzing the complete nucleotide sequence of the G, N and P protein from viruses obtained from humans and dogs who had exhibited either encephalitic or paralytic symptoms during their clinical illness (Hemachudha *et al.*, 2003). The authors found that there was no single nucleotide or amino acid substitution pattern among these genes that could account for the altered disease progressions (Hemachudha *et al.*, 2003). The authors hypothesized that an immune response might be responsible for the altered disease presentations, since it has been observed that patients with an intact T cell immunity to RABV, with high serum IL-2 and IL-6r receptors, appeared to manifest encephalitic rabies, in contrast to patients lacking this response who demonstrate paralytic disease (Hemachudha *et al.*, 2003). This hypothesis is partially supported by the description of a patient with encephalitic rabies, who developed paralysis of all limbs, following the administration of human rabies immunoglobin (HRIG) (Hemacudha *et al.*, 2003).

Encephalitic (furious) rabies.

The earliest features presented during the furious form of the disease include hyperactivity and convulsive seizures, aggravated by thirst, fear, light, noise and other stimuli (Hemachudha *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 (Hemachudha *et al.*, 2002). 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 (Hemachudha *et al.*, 2002). Phobic spasms resulting from the

contraction of the pharynx, larynx, or the diaphragm and accessory inspiratory muscles, can be elicited by offering the patient a drink of water (hydrophobia), or by blowing or fanning air on the face or chest wall (aerophobia) (Swanepoel, 1994; Hemacudha *et al.*, 2002). Inspiratory spasms are similar to phobic spasms but occur spontaneously without stimulation (Hemachudha *et al.*, 2002). Autonomic signs that may be present, include hypersalivation, fixed, dilated or constricted pupils, anisocoria (unequal pupil size), generalized piloerection (goose bumps), neurogenic pulmonary oedema (fluid buildup in intracranial cavities), excessive sweating, priapism (prolonged erection in males) and spontaneous ejaculation (Hemacudha *et al.*, 2002). Death may supervene abruptly after 10 days, or the patient may die from aphyxiation following the onset of paralysis (Swanepoel, 1994; Hemachudha *et al.*, 2002).

Paralytic (dumb) rabies.

Paralytic rabies differs from the encephalitic form of the disease, in that paralytic symptoms dominates throughout the disease progression, with a marked absence of an agitated phase as seen in the encephalitic form of the disease (Hemachudha *et al.*, 2002; Bishop *et al.*, 2003). It is extremely difficult to diagnose, owning to the fact that cardinal symptoms are less prominent than in encephalitic rabies, and only appear late in the viral infection (Hemachudha *et al.*, 2002). Phobic spasms are observed in only approximately half of patients, while inspiratory spasms are less prominent than in encephalitic rabies, due to the fact that there is a development of muscular weakness (Hemachudha *et al.*, 2002).   Muscular weakness usually starts at the bitten limb and soon spreads to encompass all limbs. Death usually results from aphyxiation, due to the paralysis of the bulbar and respiratory muscles (Swanepoel, 1994; Hemachudha *et al.*, 2002).

1.9 Epidemiology of rabies

Rabies is a zoonotic disease that is maintained in two inter-related ecological cycles. One that is primarily limited to domestic dogs and cats (*Felis catus*) (urban rabies), and the other which occurs in wildlife vectors (sylvatic rabies), and that varies according to different regions of the world (Swanepoel, 1994; Krebs *et al.*, 1995; Childs, 2002). Humans and livestock do not play a role in the maintenance of rabies, and infections are usually a result of incidental spillover from the primary reservoir hosts (Krebs *et al.*, 1995; Childs, 2002). The elimination of rabies is complicated in areas where urban and sylvatic

cycles exist in concert, since either may act as a potential source for the re-introduction of the disease (Krebs *et al.*, 1995).

Members of the canid, viverrid and chirpoteran species are the most important vectors of rabies, with dogs being the main source of infections for humans (Irvin, 1970; Rupprecht *et al.*, 2002). The role that rodents play in the epidemiology of rabies is not clearly understood, although they may be of importance in localized areas (Irvin, 1970). Bats act as vectors of classical rabies (GT1) in the Americas, but also act as the principal vector for most of the rabies-related viruses (GT3 remains to be excluded) in Eurasia and Africa (McColl *et al.*, 2000). These observations led to the hypothesis that many of the lyssaviruses were originally viruses of bats, of which some were eventually transmitted to terrestrial carnivores (McColl *et al.*, 2000). The importance of bats as reservoirs of rabies and rabies related viruses cannot be emphasized enough, since some species are highly mobile and are responsible for spreading the disease throughout their migratory ranges (McColl *et al.*, 2000). Figure 1.6. illustrates the global distribution of the primary host species responsible for the maintenance and dissemination of rabies and the rabies related viruses (Rupprecht *et al.*, 2002).
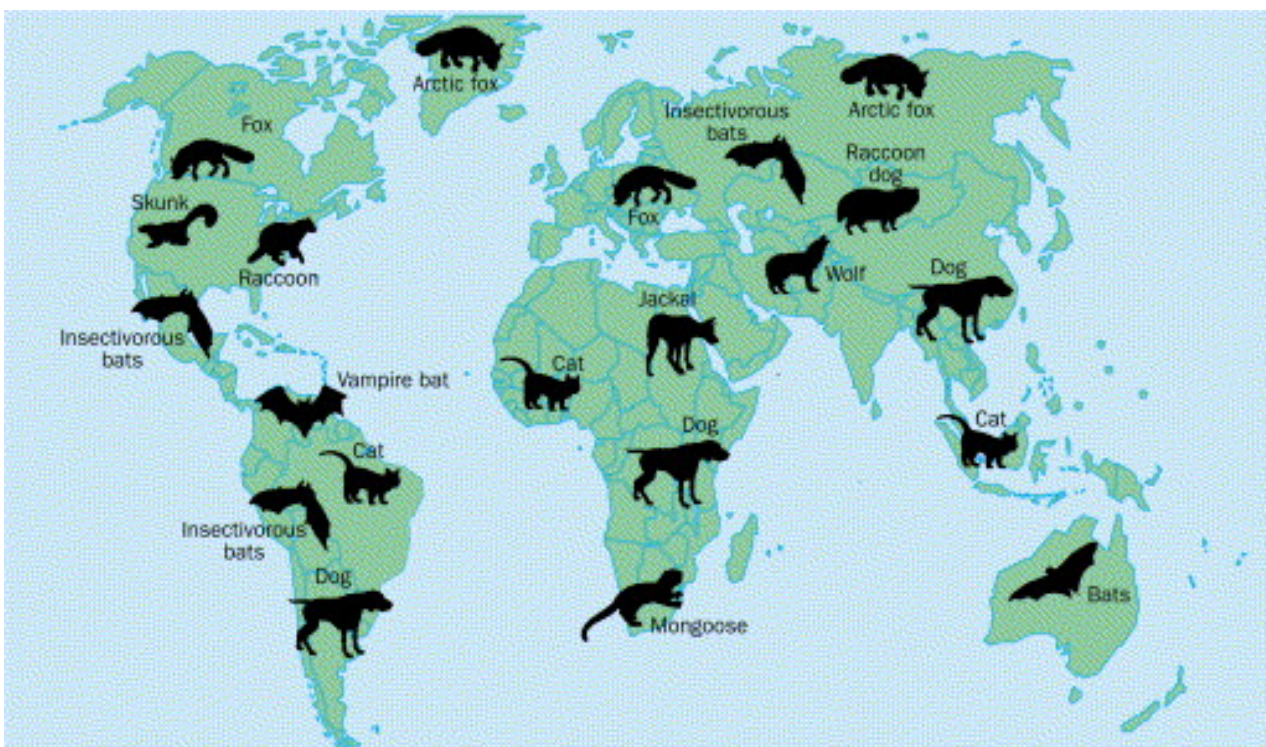

Figure 1.7. Global distribution of the primary host species responsible for the maintenance of rabies and rabies related viruses (Rupprecht *et al.*, 2002: reprinted with permission from the Lancet Neurology).

1.9.1 Worldwide distribution of the RABV

North America and Canada.

Rabies (genotype 1) is maintained in North America primarily in sylvatic cycles. This is due to widespread vaccination campaigns that were first implemented in the 1950s, and which had by the 1960s resulted in the disease being eliminated among domestic pets (Finnegan *et al.*, 2002).

In North America antigenic and genetic characterization has proven useful for identifying the major reservoir hosts for different wildlife variants of the RABV (Dietzschold *et al.*, 1988; Smith *et al.*, 1995; Krebs *et al.*, 1995; Smith, 1996). A variant is maintained by raccoons (*Procyon lotor*) along the eastern seaboard, with the recent translocation of animals from the southeastern states being responsible for the emergence of the disease among these animals, in the mid-Atlantic and northeastern states (Smith, 1996; Smith *et al.*, 1995). A variant associated with a long-standing reservoir in red (*Vulpes vulpes*) and arctic foxes (*Alopex lagopus*) has been documented in Alaska, and spread during the 1950s to affect foxes in the northern, eastern and southern regions of Canada, especially in the regions surrounding Ontario (Smith, 1996; Smith *et al.*, 1995). Three different variants are maintained by the striped skunk (*Mephitis mephitis*) of which one is found in long standing reservoirs in California, the northern and southern central states of the USA as well as the Canadian prairie province, while the remaining two variants have been identified in long standing gray fox (*Urocyon cinereorgenteus*) reservoirs in Arizona and Texas, respectively (Smith, 1996; Smith *et al.*, 1995). A recently recognized outbreak of rabies in coyotes (*Canis latrans*) in south Texas, is due to spillover infection from a long-standing dog reservoir at the Texas-Mexico border (Smith, 1996; Smith *et al.*, 1995).

Overlying the disease in terrestrial animals are multiple independent reservoirs of rabies in several species of insectivorous bats (Smith, 1996). Two rabies virus variants associated with the rarely encountered silver haired bat (*Lasionycteris noctivagans*) and the eastern pipestrelle (*Pipistreiilus subflavus*), have been implicated as causing most of the indigenously acquired cryptic rabies cases in humans since 1980 (Smith, 1996; Messenger *et al.*, 2002). This led investigators to hypothesize that these variants may have a higher likelihood of establishing an infection after a superficial contact (Messenger *et al.*, 2002; Morimoto *et al.*, 1996). This hypothesis was partially supported by a series of

40

experiments in which it was shown that the silver haired bat variant was able to grow to higher titers at a lower temperature, in epithelial and muscle cells, when compared to rabies virus isolates of terrestrial origin (Messenger *et al.*, 2002; Morimoto *et al.*, 1996).

Central and South America.

Rabies in central and south America is maintained in sylvatic cycles by vampire bats, and in urban cycles by domestic dogs. The common vampire bat (*Desmodus rotundus*) has been associated with outbreaks of rabies in humans throughout Latin America, with affected regions including the Island of Trinidad, Brazil, Peru, Venezuela, Argentina and Mexico (Smith and Baer, 1988; McColl *et al.*, 2000; Childs, 2002). This vector species is also responsible for the occurrence of bovine paralytic rabies among cattle throughout the ranching areas of South America, which leads to huge annual economic losses (Van der Merwe, 1982; Smith and Baer, 1988; Childs, 2002). Because of this economic burden, vampire bats have been the target of major control efforts, which have focused on the population reduction of vampire bats, by the topical wound or systemic treatment of cattle with anticoagulants (Childs, 2002). The mongoose (*Herpestes auropunctatus*), which was imported from India to combat rats in sugar cane fields, represents another sylvatic reservoir of rabies in Cuba, Puerto Rico, Grenada, the Dominican Republic and other Caribbean islands (Smith and Baer, 1988).

Dog rabies remains a serious problem in Latin America, although vaccination campaigns instituted by the Pan American Health Organization and the WHO Regional Offices for the Americas during the last two decades, has resulted in a significant reduction of the number of human cases (Chomel, 1993; Swanepoel, 1994; WHO, 2004). Here genetic characterization has proved useful for distinguishing between isolates which were obtained from urban and sylvatic cycles. In one such study by Tordo *et al.*, 1993, partial sequencing of the viral G protein could distinguish between overlapping cycles, one of which occurred in cats and dogs, and the other that occurred in vampire bats and bovines in Sao Paulo, Brazil and Mexico City (Tordo *et al.*, 1993).

Europe, Eurasia, and the Arctic.

The most important reservoir for rabies in Europe is the red fox (Blancou, 1988; Chomel, 1993; Swanepoel, 1994; Krebs *et al.*, 1995; Childs, 2002; Woldehiwet, 2002). The current

epizootic raging among these animals started in 1935, in a focus of infection among dogs in Poland, from where the disease entered, and spread throughout the European red fox population (Swanepoel, 1994; Child, 2002; Finnegan *et al.*, 2002; Woldehiwet, 2002). During the Second World War the disease spread progressively westwards and southwards, at a rate of 30-60 km per annum, until most of the countries of northern Europe were affected (Swanepoel, 1994; Child, 2002; Finnegan *et al.*, 2002; Woldehiwet, 2002). Today, vulpine rabies affects primarily the eastern regions of Europe, having been brought under control through the implementation of successful oral vaccination campaigns in the Western regions of the continent (Blancou, 1988; WHO, 2004). The Racoon dog (*Nyctereutes procynoides*) acts as another principal vector of rabies in the more northerly regions of Eurasia (Blancou, 1988; Chomel, 1993; Swanepoel, 1994; Krebs, 1995). The arctic (*Alopex lagopus*) and red fox maintains another rabies virus variant in the Arctic Circle (Blancou, 1988; Swanepoel, 1994). Antigenic and phylogenetic characterization has proved useful for distinguishing among the Arctic fox strain, and other RABV variants that occurs among foxes in more southerly latitudes (Chomel, 1993; Smith *et al.*, 1995; Smith, 1996).

The dog is the major vector of rabies over the length of the Asiatic continent, with wolves, mongoose and jackal representing sylvatic vectors of the disease (Blancou, 1988, Swanepoel, 1994). Several thousands of human cases are recorded each year from India, Pakistan, Indonesia, Thailand and Vietnam, with fewer cases being reported from Bangladesh, Burma, Iran, Iraq and the Philippines (Swanepoel, 1994). India has the highest dog associated human rabies fatality rate in the world, with estimates ranging from 15 000 to 25 000 cases per annum (Swanepoel, 1994).

Africa and Madagascar.

Rabies has probably been present on the African continent since ancient times, but due to a lack of written records, the presence and distribution of the disease before European colonization cannot be ascertained (Blancou, 1988). Urban rabies predominates throughout Africa, with the domestic dog being the principal victim, disseminator and reservoir of the disease (Blancou, 1988). Dogs are also the only known vectors of the disease in Madagascar, with the occasional cases occurring in wildlife, being the result of spillover infections (Blancou, 1988).

North Africa, which includes the countries of Algeria, Morocco, Tunisia and Egypt, has been endemic for canine rabies for hundreds of years (Blancou, 1988; Swanepoel, 1994). Here, several hundreds of cases among dogs are diagnosed per annum, with occasional cases occurring among cats and humans (Blancou, 1988; Swanepoel, 1994). Rabies has however been brought under control in the countries of Tunisia and Libya, due to rabies-control programs that have been implemented there since the 1980s (Blancou, 1988; Swanepoel, 1994). In countries that span the Sahara desert, and which are arid and sparsely populated, rabies occurs in scattered foci, and is primarily disseminated by dogs belonging to nomads or refugees, who are fleeing war or drought (Blancou, 1988; Swanepoel, 1994). Occasional cases are also recognized among camels following outbreaks of the disease among dogs, with hyenas (*Crocuta crocuta*) and jackal representing the wildlife species that are most often diagnosed as rabid (Blancou, 1988; Swanepoel, 1994).

In West African countries such in Nigeria, Senegal, Niger, Benin, Guinea, Ivory Coast, Ghana, Central African Republic and Cameroon, rabies primarily affects dogs, with occasional cases occurring in domestic livestock, humans and wildlife (Blancou, 1988; Swanepoel, 1994). "Mad dog disease" or "*oulo fato*", a non-fatal form of rabies was first reported from West Africa at the beginning of the 20th century (Blancou, 1988; Swanepoel, 1994). Viral strains capable of causing non-fatal chronic infections have since been isolated from Ethiopia in the 1950s and 1970s respectively, and have been shown to differ in their pathogenic properties from classical rabies (Blancou, 1988; Swanepoel, 1994).

In East Africa, rabies outbreaks have been recognized in Kenya and Uganda since the 1900s, but have been known to be present by indigenous inhabitants since long before the arrival of Europeans (Swanepoel, 1994). A serious outbreak of rabies was first reported from Tanzania and Zambia in 1936 and 1956, respectively, and spread northeastwards and northwestwards from Zambia during the 1960s to produce a severe epidemic in the northeastern regions of Tanzania in the 1970s (Swanepoel, 1994). During the 1980s the epidemic spread throughout Kenya, where it remains endemic to this day (Swanepoel, 1994).

Rabies in Southern Africa represents a unique situation with respect to the rest of Africa, in that it is the only part of the continent in which sylvatic cycles among yellow mongoose (*Cynictis penicillata*) and kudu antelope (*Tragelaphus strepsiceros*) exists independently

from the cycles among canids (Meredith, 1982; Blancou, 1988; Swanepoel, 1994; King *et al.*, 1994; Woldehiwet, 2002; Bishop *et al.*, 2003). Other vectors of the disease include suricates (*Suricata suricata*) and ground squirrels (*Xerus inauris*) in the mongoose rabies endemic regions of South Africa, while the bat eared fox, black-backed jackal and side striped jackal (*Otocyon megalotis*, *Canis mesomalas*, *Canis adustus*) are the principal wildlife canid hosts in the western and northern regions of South Africa, as well as in Namibia, Zambia and Zimbabwe (Meredith, 1982; Blancou, 1994; Swanepoel, 1994; King *et al.*, 1994; Bishop *et al.*, 2003). Wild cats (*Felis* sp.) and genets (*Genneta* sp.) represent supplementary vectors of rabies in southern Namibia and the Northern Cape, while dog rabies has become an ever-increasing problem in the South African province of KwaZulu Natal (Blancou, 1988; Swanepoel, 1994; Bishop, 2003).

1.9.2 Worldwide molecular epidemiological studies on the RABV

The development of PCR and nucleotide sequencing applied to the characterization of Lyssaviruses, stimulated worldwide research on the taxonomy, epidemiology and evolution of the rabies and rabies related viruses. Sacramento *et al.*, 1992 were the first to describe the evolutionary processes that shape viral diversity for RABV isolates from foxes in France (Sacramento *et al.*, 1992). By analyzing the nucleotide sequence of the G-L intergenic region for isolates from proximal and distant foci of infection, they demonstrated that the viruses showed high sequence homogeneity, which increased with the proximity of the geographical origin of the isolates from each other (Sacramento et al., 2002). A number of variants associated with small enzootic foci could also be identified, which were present independently from their host species of origin. This suggested that a slow evolution of the virus occurred in parallel with the spatio-temporal progression of the disease (Sacramento *et al.*, 1992). These findings were confirmed by passaging wildlife isolates in mice for a few times, with the results that no mutations in the consensus sequence could be demonstrated (Bourhy *et al.*, 1995).

Studies on the host origin of Lyssaviruses.

It has been speculated that lyssaviruses evolved from insect rhabdoviruses which were first transmitted to insectivorous bats in the distant past (Shope, 1982; Badrane and Tordo, 2001). Evidence for this argument is supplied by the fact that members from most of the genera in the *Rhabdoviridae* family, (ex. Kotokan and Obodhiang viruses) have

been isolated from insects (King and Crick, 1988). In addition the Mokola virus, which was first isolated from an insectivorous shrew has been demonstrated to be able to replicate in inoculated *Aedes aegypti* mosquitoes (Shope, 1982; King and Crick, 1988). Phylogenetic and molecular clock analysis of sequence data obtained from the G gene from a global assembly of carnivoran and chiropteran rabies as well as the rabies related viruses, has dated the presumed common insect virus ancestor of the lyssavirus genotypes, as being present approximately 7 080 to 11 631 years ago (Tordo and Badrane, 2001).

The study by Tordo and Badrane, 2001 also addressed the hypothesis that suggested that the epizootic of genotype 1 RABV in terrestrial carnivores was due to the successful host switching of the virus, from chirpotera to terrestrial carnivores in the distant past. In the study it was found that viral isolates belonging to genotype 1 all shared a common ancestor, and consisted of variants that could be divided into the worldwide cosmopolitan lineage, the racoon lineage in the USA, as well as the indigenous strain that circulates among insectivorous and hematophagous bats in North and South America (Tordo and Badrane, 2001). Phylogenetic analysis suggested that at least two spillover events were responsible for establishing genotype 1 RABV in terrestrial carnivore reservoirs. The first of these spillovers occurred in North America and gave rise to the racoon and side striped skunk lineages, while the second spillover event occurred in an unknown geographic region and gave rise to the cosmopolitan lineage (Tordo and Badrane, 2001). These lineages were shown by molecular clock analysis to have emerged from 888-1,459 year ago, a finding that was surprising in lieu of the fact that the occurrence of rabies in domestic carnivores had been documented from ancient Mesopotamia since the 23rd century B.C (Tordo and Badrane, 2001). This discrepancy was explained by acknowledging that the carnivoran strain that circulated in ancient Mesopotamia, may have been part of a different epizootic, that came to a halt when the virus went extinct, due to either viral, historical or environmental factors (Tordo and Badrane, 2001).

Characterization of the global diversity of genotype 1 Lyssaviruses.

A number of molecular epidemiological studies targeting the N, G and P genes have been conducted in order to investigate the genetic variation that is present within genotype 1 RABV reservoirs throughout the world. This has led to the delineation of the principal host species, as well as geographic range, of groups of viral variants.

Kissi *et al.*, (1993) investigated the genetic diversity that is present within the N gene (400 bp) of a large global collection of over 70, serotype 1 lyssavirus isolates (Kissi *et al.*, 1995). Similarity levels of > 83.3% (nt.) and > 92% (aa.), supported the classification of these isolates into genotype 1, and at least 11 phylogenetic lineages associated with particular host species, geographic regions and/or vaccine types could be distinguished (Bourhy *et al.*, 1993; Kissi *et al.*, 1995). These included nine phylogenetically distinct groups isolated from Africa (Africa 1-3), Asia (Thailand), the Arctic, Europe/Middle East, Latin America (1-2) and two groups of vaccinal strains (AVO1, CVS and PV, SAD-B19) (Nadin-Davis, 2000; Bourhy *et al.*, 1993; Bourhy *et al.*, 1995; Kissi *et al.*, 1995). Africa group 1 was isolated from dogs and wildlife from throughout Africa. It could be sub-divided into Africa 1a, isolated from North and East Africa, and group 1b that circulates in Southern Africa. Africa 2 was isolated from dogs from several central and eastern African countries, while Africa 3 was isolated primarily from mongoose from South Africa. Group 2 and 3 were unrelated to other globally identified lyssavirus groups, suggesting that they are indigenous to the African continent. A single isolate from a human bitten by a dog in Egypt segregated distinctly from the other African groups, raising the speculation that it may represent a fourth African lineage (Nadin-Davis, 2000; Bourhy *et al.*, 1993; Bourhy *et al.*, 1995; Kissi *et al.*, 1995). It has further also been speculated that Africa group 1a, 1b, 2 and 3 have probably arisen independently on the Africa continent from different progenitor viruses. A close relationship between Africa 1 and the Europe Middle East group however, suggested that they probably shared a recent common ancestor, an observation that was supported by historical data on the emergence of the disease (Nadin-Davis, 2000; Bourhy *et al.*, 1993; Bourhy *et al.*, 1995; Kissi *et al.*, 1995).

Smith *et al.*, (1992) conducted a similar study on a global assembly of 87 serotype 1 viral isolates, by targeting a 200-bp region on 3' terminal of the N gene, but focused primarily on isolates from the Americas and Asia. The authors defined three categories of isolates, based on quantitative measures of sequence divergence. Closely related samples from a single outbreak showed <5% divergence, while isolates that were historically, politically or geographically related showed a divergence of <10%. The third category, consisting of isolates sharing less than 10% homology, consisted of viruses that were only distantly related (Smith *et al.*, 1992). Using this divisioning system the authors identified seven major groups of viral isolates circulating in worlwide animal reservoirs (<10% divergence) (Smith *et al.*, 1992). Group I was isolated from Africa, Europe and America and was subdivided into two groups e.g. group Ia and Ib. Group Ia was, with the exclusion of

laboratory adapted strains, isolated from the Americas, and included isolates from the Texas grey fox, dog isolates from Mexico and the Mexico-Texas border, as well as dog isolates from Latin America. Group Ib contained specimens isolated from dogs and mongoose from the USA, Central and North America, South America, the Caribbean and Central Africa (Smith *et al.*, 1992). The Asian isolates formed a highly heterogeneous group of viruses and consisted of groups II, III, IV and Va, b that were isolated from dogs from countries including Thailand, the Philippines, Indonesia and Pakistan (Smith *et al.*, 1992). Group VI was isolated from humans bitten by dogs in Nigeria, while group VII was isolated from heamatophagous bats from Latin America (Smith *et al.*, 1992).

Similar findings to those of Kissi *et al.*, and Tordo *et al.*, have also been demonstrated by an investigation of a 500 nt. region on the N terminal ectodomain of the G gene, from a global assembly of 47 genotype 1 isolates (Tordo *et al.*, 1993). It was found that conservation values of 81.5% (nt.) and 88.5% (aa.) defined the maximum intragenotypic diversity for this region of the viral genome for the isolates studied (Smith *et al.*, 1992; Nadin-Davis, 2000).

The diversity for the genotype 1 lyssaviruses as well as the rabies related viruses was further investigated by using the P gene, and showed similar findings to studies conducted using the N and G gene (Nadin-Davis *et al.*, 2002). Intragenotypic variation for a 935 nt. region of the P gene for the global assembly of 128 isolates, showed that the diversity for the region analyzed was $\geq$73.5% (nt.) while intergenotypic identity values between the RABV and the other genotypes ranged between 69.5% (ABLV) to 56% (MOKV) (Nadin-Davis, 2000; Nadin-Davis *et al.*, 2002). The strong sampling of American isolates in this study further allowed for the demonstration of the presence of the indigenous genotype 1 strains, which differ from USA isolates of the cosmopolitan lineage, as previously demonstrated by Tordo and Badrane, 2001. (Badrane and Tordo, 2001; Nadin-Davis *et al.*, 2002) The distinctiveness of these strains for the P region analyzed, and the demonstration of the close phylogenetic relationship between isolates from terrestrial and chiropteran hosts, provided further support for the hypothesis that presupposed the chiropteran origins of RABV in terrestrial carnivores (Nadin-Davis *et al.*, 2002).

The observation of the close phylogenetic relationship between RABV variants of the cosmopolitan lineage from diverse geographical locations from throughout the world has been explained by the transportation of dogs that were incubating the disease, from

Europe to other parts of the world. This led to the introduction of the disease into new geographic regions, where the virus was subsequently transmitted to new wildlife reservoirs. The virus persisted in these reservoirs and over time evolved into distinct strains (Childs, 2002; Smith *et al.*, 1992; Bourhy *et al.*, 1993; Bourhy *et al.*, 1995; Kissi *et al.*, 1995; Childs, 2002). The highly heterogeneous nature of Asian isolates however, suggested that the disease was introduced by a number of independent rabies epizootics long before European colonization had even started (Smith *et al.*, 1992).

1.9.3 Distribution of rabies in South Africa

A wide diversity of animal species is involved in the maintenance of rabies within discrete geographical regions within South Africa. These include members of the *Canidae* family such as the bat eared fox in the northern and northwestern Cape region, the black backed jackal in the northern Limpopo and northwestern provinces, as well as the dog in KwaZulu Natal. An endemic form of rabies is associated with the family *Herpestidae*, and is primarily maintained by the yellow mongoose over the central plateau region (Meredith, 1982; Swanepoel, 1994; King *et al.*, 1994; Bishop *et al.*, 2003). Figure 1.7., demonstrates the geographical distribution of the primary host species responsible for rabies endemicity in South Africa (Bishop *et al.*, 2003).
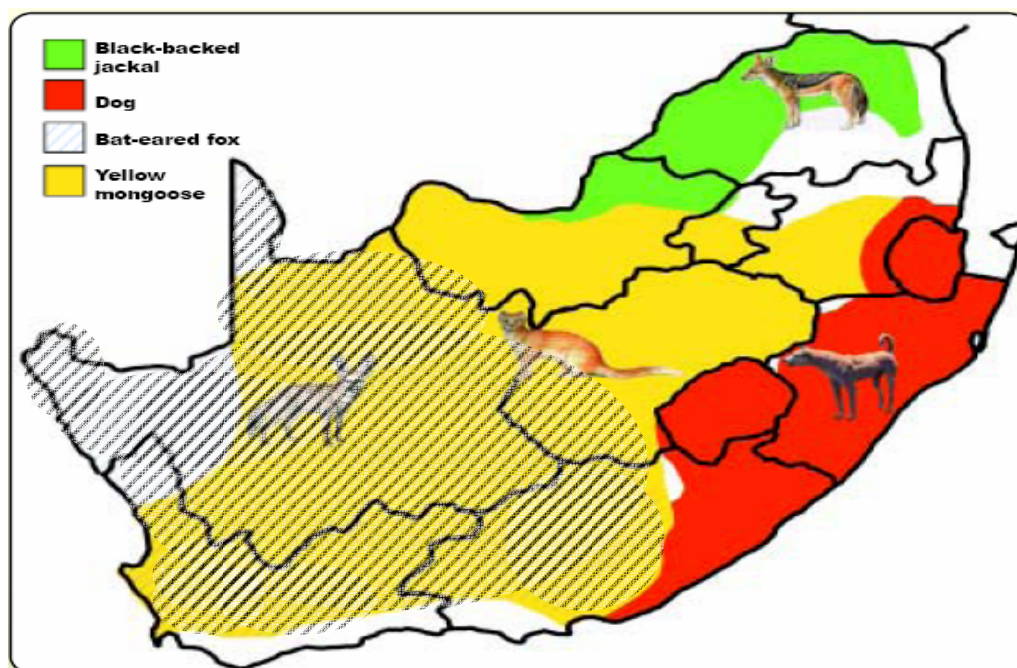

Figure 1.8. The distribution of the primary host species responsible for rabies endemicity in South Africa (Bishop *et al.*, 2003).

Canine rabies.

Canine rabies has been reported in South Africa from as early as the 18th century. The first case was confirmed in Port Elizabeth in the Eastern Cape in 1892, following an outbreak believed to have been initiated by the importation of an infected dog from Britain one-year prior (King *et al.*, 1988; Swanepoel *et al.*, 1993; Bishop *et al.*, 2003). It was brought under control by 1894, through muzzling and control of dog movement, as well as the destruction of strays, and affected 90 dogs, 7 cats and a few cattle with no reported spillover to wildlife (Swanepoel *et al.*, 1993; Bishop *et al.*, 2003). An outbreak in dogs was also recognized in southwestern Zimbabwe in 1902, following an introduction of the disease from western Zambia, where it had been known to be rampant since 1901 (Swanepoel *et al.*, 1993). It was brought under control in 1913, and apparently did not affect wildlife or extend southwards into South Africa (Swanepoel *et al.*, 1993).

Canine rabies became endemic in South Africa as a consequence of a dog-associated epizootic that started in the border region between Namibia and Angola in 1947. It spread throughout Namibia and Botswana, reaching the northern Limpopo province of South Africa in 1950 (King *et al.*, 1988; Swanepoel *et al.*, 1993; Bishop *et al.*, 2003). It presumably did not spread further southwards into the dry and sparsely populated region of the southwestern parts of the Limpopo province, but instead took an eastward and northward course, entering the eastern Mpumulanga Lowveld region and southwestern Zimbabwe by June and August 1950, respectively (King *et al.*, 1988; Swanepoel *et al.*, 1993; Bishop *et al.*, 2003). By 1952 it had entered the densely populated southern Maputo district of Mozambique, where it became endemic in dogs, and from where it spread southwards along the eastern coastal belt of South Africa, entering Swaziland in 1954, and appearing in the northern districts of KwaZulu Natal (KZN) in 1961 (King *et al.*, 1988; Swanepoel *et al.*, 1993; Bishop *et al.*, 2003). Except for unconfirmed report of rabies in the province in the early nineteenth century, KZN had up till then been free of the disease (Swanepoel, 1994). The Kruger National Park and other game reserves in the KZN province have with the exception of occasional intrusions of rabid domestic dogs from Mozambique, also remained free of rabies (Bishop, 2003)

The introduction of rabies into the northern regions of KwaZulu Natal initiated an epidemic of unprecedented intensity in dogs, with the high population densities associated with rural areas providing favorable conditions for the dissemination of the disease (Swanepoel,

1994). It was brought under control in 1968 through the vaccination and prohibition of transport of unvaccinated animals, even though the vaccination coverage which was achieved during this time period (1961-1968) in dogs, never rose to above 41%. Higher vaccination coverage was however achieved by targeted vaccination campaigns in the most severely affected regions of the province (Swanepoel, 1994).

Rabies appeared again in Natal in 1976, at a time when refugees were fleeing from Mozambique, following that countries assumption of independence from Portugal (Meredith, 1982; King *et al.*, 1988; Swanepoel, 1994 Bishop *et al.*, 2003). During the intervening time period between the first and second epidemic there had been an increase in the density of human populations in informal settlements, as people moved from rural to urban areas. Consequently there was also an increase in the mobility of human populations, which according to Swanepoel, 1994, could have been responsible for the appearance of the disease in geographically widely separate locations throughout KZN. It spread throughout the province, extending as far inland, as Lesotho and the Freestate, and as far southwards as the Eastern Cape, which it entered in 1987 (Swanepoel, 1994). Even though vaccination coverage of 59% was achieved for dog populations in the period between 1980-1981, the epidemic has proven to be intractable, and despite concerted efforts by the Directorate of Veterinary services it has as of yet not been brought under control (Bishop *et al.*, 2003). Reasons for this are varied, but include a subsequent restructuring of the veterinary authority, an inability to control dog movement, and the inadequate vaccination coverage which exists in the densely populated regions of the province (Bishop *et al.*, 2003).   Full details on the particulars of the second epidemic will be given in section 2.2.

The spread of canine rabies led to the emergence of the disease in wildlife throughout Namibia, Zimbabwe, Botswana and South Africa. In South Africa, sporadic cases of rabies among the bat eared fox populations (*Otocyon megalotis*) in the dry and semi arid regions of the northern and north western Cape province, was recorded from 1955 onwards, with an increase in disease incidence being noticed in the 1970s, and a westward spread along the coast being recognized by the 1980s (Swanepoel, 1994). A unique outbreak among kudu (*Tragelaphus strepsiceros*) antelope, believed to have been initiated by black backed jackal (*Canis mesomelas*), was also recorded in central Namibia from 1977 to 1985, with the infection apparently spreading among the antelope by the oral route, independently from the usual vector species (Hassel, 1982; King *et al.*, 1988; Swanepoel,

1994). Rabies among black backed jackal and cattle emerged as a problem in the ranching areas of the northern Limpopo province in the early 1950s (Swanepoel, 1994). Initial attempts at control were made by the distribution of poisoned meat baits, and over 3900 jackals were poisoned from 1951 to 1956 (Swanepoel, 1994; Bishop *et al.*, 2003). The campaign however provided only temporary and localized control, and the disease re-emerged among this vector species in the mid 1970s, following a possible reintroduction into the Messina district from Zimbabwe across the Limpopo River (Swanepoel, 1994). Since then rabies has remained endemic among jackal populations within this region of the country (Bishop *et al.*, 2003). Black-backed jackal are potentially efficient vectors of the disease, as they secrete a large amount of virus in their saliva, and as they reach high population densities in commercial crop and ranching areas (Bingham and Foggin, 1993). Some controversy however still exists as to whether jackal are able to maintain the disease independently of rabies cycles in domestic dogs (Bruckner et al., 1978; Cummings, 1982; King *et al.*, 1988). Jackal rabies furthermore also represents a significant danger to endangered African wildlife. A point in case, is the documentation of rabies in an African wild dog (*Lycaon pictus*) pack in the Madikwe Game Reserve in the Northwestern Province of South Africa (Hofmeyer *et al.*, 2000). Sequence analysis indicated that the rabies variant that decimated this pack was indistinguishable from a variant that circulated locally among black backed jackal (Hofmeyer *et al.*, 2000).

Mongoose rabies.

Anecdotal evidence for the presence of an indigenous form of rabies associated with viverrids date back to 1885 (Swanepoel, 1994). In particular, there was a belief in the eastern and northern Cape that the bite from a genet led to a fatal, rabies like disease (Swanepoel, 1994). Cluver *et al.*, (1927) reported on 11 unconfirmed cases that had occurred in between the period of 1916 to 1927 in the Mpumalanga, the Freestate and the northern Cape province, following bites from yellow mongoose (*Cynictis penicillata*), genets (*Genetta genetta*) and domestic dogs (Cluver *et al.*, 1927). The presence of rabies in viverrids was however only confirmed in 1928, in two children who had succumbed to the disease in Wolmaransstad, in western Mpumalanga, after they had been bitten by an apparently tame mongoose (Meredith, 1982; King *et al.*, 1988; Swanepoel, 1994; Bishop, 2003).

It soon become apparent that endemic rabies in South Africa was more widely distributed than was previously suspected, with cases being recognized from throughout the country. Initially it was though that these outbreaks were due to an extension of the dog-associated outbreaks that had occurred in the Eastern Cape (1892-1894) and Zimbabwe (1902-1923) (Swanepoel, 1994). However, additional surveillance indicated that this was not the case, as endemic rabies was confluent over the central plateau region, in accordance with the distribution of the yellow mongoose (Swanepoel, 1994). Attempts to control the disease in yellow mongoose populations started as early on as the 1930s and was focused on the population reduction of yellow mongoose through the gassing of burrows with cyanogas and later phosphine (Swanepoel, 1994). Approximately 50 000 to a 160 000 hectares were treated annually, until the 1970s (1974-1976), when heavy rainfalls restricted the access of control teams to the affected regions (Swanepoel, 1994).

Control measures based on the population reduction of mongoose however, proved largely ineffectual, and rabies has occurred in epidemic proportions among this vector species over the interior plateau since the 1970s (Swanepoel, 1994; Bishop *et al.*, 2003). In this region incidental infections occur in cattle and a variety of other animals, including mustelids, felids and canids (Swanepoel, 1994).

1.9.4 Molecular epidemiology of rabies in South Africa

Historically it had been recognized that the disease maintained by viverrids differed from the disease in canids, in that although the disease was cross-transmissible between these host species, there was no real tendency for it to spread among them (Swanepoel, 1994). This suggested that the viruses in these hosts had undergone a high level of species adaptation, resulting in the development of so called biotypes of the virus, which have unique differences in terms of antigenicity and pathogenicity (Swanepoel, 1994). King *et al.*, demonstrated the distinction, which exists between the biotypes maintained by the South African viverrid and canid hosts, by reacting viral isolates against a panel of 80 MAb directed against the nucleoprotein (King *et al.*, 1988, 1993). It was found that the viruses corresponded to two reaction patterns, one chiefly confined to isolates from canid hosts, and the other to isolates from viverrids. Further it was also apparent that while the reaction pattern was the same for all the canid isolates, substantial heterogeneity existed within the viverrid pattern (King *et al.*, 1988, 1993). Nel *et al.*, and Von Teichman *et al.* verified these findings by phylogenetic analysis of the nucleoprotein gene and non-coding pseudogene,

by demonstrating that the canid viruses were closely related to each other and European vaccine strains, but distant from the viverrid viruses, which formed a separable and phylogenetically diverse group (Nel *et al.*, 1993; Von Teichman, 1995).

Jaftha *et al.*, and Nel *et al.*, expanded on the work conducted by Von Teichman, by investigating the diversity that exists within the mongoose biotype of the RABV. They demonstrated that the viverrid isolates formed a number of phylogenetically distinct groups that clustered according to the geographical region these viruses were isolated from (Jaftha *et al.*, 1997; Nel *et al.*, 1993, 2005). This suggested that these viral groups had been evolving independently within different subsets of the yellow mongoose population for a prolonged period of time, a hypothesis which was supported partially by a study in which mitochondrial DNA analysis was used to differentiate between mongoose subpopulations (Taylor *et al.*, 1993; Jaftha *et al.*, 1997; Nel *et al.*, 1993, 2005).

Sabeta *et al.*, 2003, investigated the genetic diversity which is present between viral isolates of the canid biotype, which were obtained from the principal canid host species from Zimbabwe and Southern Africa (Sabeta *et al*, 2003). Phylogenetic analysis of a region encompassing the cytoplasmic domain of the glycoprotein and the non-coding pseudogene, indicated that these viruses were highly related to each other, but distinct from the PV (Sabeta *et al.*, 2003). Minor phylogenetic branching between viral isolates based on their geographic origin and host species of isolation could be demonstrated, but no general distinction could be made between them (Sabeta *et al.*, 2003). These findings supported historical data on the emergence and spread of the canid epizootic throughout Southern Africa by supporting the hypothesis of a common progenitor for this biotype in Sub-Saharan Africa (Swanepoel, 1994; Bishop *et al.*, 2003; Sabeta *et al.*, 2003). Furthermore the study also demonstrated that canid biotype viruses were highly adaptable, as they could enter and establish rabies cycles in whatever canid host species were ecologically capable of sustaining them (Sabeta *et al.*, 2003).

Monoclonal and antibody studies as well as phylogenetic analysis has further proved useful in demonstrating that spillover infections of the viverrid biotype occurred to canid species, while the vice versa event was first demonstrated by phylogenetic analysis of the glycoprotein cytoplasmic domain and the G-L intergenic region in a study by Jacobs *et al.*, (1997) (King *et al.*, 1993; Nel *et al.*, 1993; Von Teichman, 1995, Jacobs *et al.*, 1997).

1.13   Prevention and control of rabies

<u>Control of dog rabies.</u>

Dogs represent the most dangerous reservoir of rabies in the developing world, including South Africa (Fekadu, 1993; Swanepoel, 1994; Bishop *et al.*, 2003). This is reflected by the fact that the majority of human PEP treatments and human rabies cases occur in developing countries, and are associated with bites from infected dogs (WHO, 1992). According to the WHO (1992), human rabies in the developing world can reach 30 000 - 50 000 cases per year, of which 99% can be ascribed to rabies transmission from dogs (WHO, 1992). Furthermore, the majority of the 4 million PEP treatments given per annum (90% of the total), are given to people who live in, and become exposed to rabies in areas of canine rabies endemicity (WHO, 1992). Weighted against these factors, the impetus to control rabies among dogs remains the single most important factor in preventing the transmission of the disease to humans. A case in point was the decrease that was observed in human rabies cases following the inception and implementation of urban rabies control programs by the PAHO/AMRO in some south American countries in 1983 (Larghi *et al.*, 1988; Debbie, 1988; WHO, 1992).

During the last half of the 20th century rabies has become an ever-increasing problem in dogs in third world countries, and especially in countries of Sub-Saharan Africa. Here the rabies situation appears to be aggravated by the increase of the dog population due to the increased urbanization, density and mobility of human populations (WHO, 1992). In these regions concerted and comprehensive schemes must be implemented in order to eliminate the disease (WHO, 1992). At least three different elements have been recognized as forming part of rabies control strategies in dogs and other domestic animals. These include the requirement for epidemiological surveillance, immunization and dog control (WHO, 1992; Childs, 2002).

Epidemiological surveillance forms the cornerstone on which rabies control programs are planned, organized and implemented (WHO, 1992). Surveillance data necessary to control rabies include a requirement for information on the distribution and occurrence of the disease, and specifically for urban rabies, the degree of ownership, confinement level, number, longevity, reproductive turnover rate and the age and sex structure of dogs in an affected region (Chomel, 1992; WHO, 1992; Swanepoel, 1994; Bishop *et al.*, 2003). This

provides valuable information, which influences the manner and frequency at which vaccination campaigns will be conducted (Swanepoel, 1994). It should also be determined whether wildlife reservoirs exists in concert with the disease in dogs, and whether dog rabies can be controlled independently from sylvatic cycles (Swanepoel, 1994).

The first step in controlling dog rabies is to declare the area where it occurs as a rabies endemic region, thus permitting the enforcement of control measures (Swanepoel, 1994). Vaccination of dogs in the affected area then becomes compulsory, and dogs may not be removed from the region without a movement permit or proof of vaccination (Swanepoel, 1994). In developing countries, among the higher income groups, enforcement of such measures are easy to achieve, while the government is usually responsible for the implementation of vaccination campaigns for the dogs of the lower income segment of the population (Swanepoel, 1994).

The WHO recommends that dogs be vaccinated annually with an inactivated vaccine, and that a vaccination coverage target of at least 75% of the dog population in an affected community should be reached within one month from the start of a campaign (WHO, 1992). Depending on local circumstances, different basic approaches can be used to achieve this, including house-to-house visits, fixed vaccination posts, or mobile clinics (WHO, 1992). Count is kept of the number of dogs vaccinated, and dogs are typically marked at some of the locations where the vaccinations were performed (Swanepoel, 1994). Vaccination coverage is then either evaluated by comparing estimates of the dog population with the number of dogs vaccinated, by comparing the ratio of marked to unmarked dogs, or by conducting Ab surveys (Swanepoel, 1994). The success of the campaign is then judged by the continued monitoring of the occurrence of the disease in the affected region. It is usually assumed that complete control has been achieved, if the target region remains rabies free for at least two years (Swanepoel, 1994). Once control has been achieved, it may be necessary to continue vaccinating animals in a cordon, bordering areas from which the disease may be re-introduced (Swanepoel, 1994; Bishop et al., 2003). Vaccination of farm herbivores in dog rabies affected regions is usually made optional, and is performed at the discretion and expense of the owner (Swanepoel, 1994).

An alternative to parenteral vaccination may be the use of live attenuated oral vaccines in mass vaccination campaigns (Bogel, 2002). This may facilitate the control of rabies in areas where dogs are difficult to vaccinate by injection, and contains the added advantage

in that dogs do not need to be captured or handled (Bogel, 2002). Unfortunately the concentrations of virus in these baits need to be substantially higher than in conventional wildlife vaccines, in order to elicit a protective response (Bogel, 2002). These vaccines thus carry an additional risk to non-target animals that may inadvertently take up the baits. Moreover, these vaccines are given close to, or even inside human habitation, which leads to even more stringent requirements for oral vaccines and their application (Bogel, 2002).

The WHO does not recommend the removal of strays, or the destruction of unvaccinated dogs, as means to decrease dog population densities towards the control of dog rabies (Macdonald, 1982; WHO, 1992; Smith, 1996). The high turnover rate of dogs, as well as the recuperative reproductive rates following dog removal, necessitates that nearly 50-80% of individuals have to be removed each year for there to be any sustained effect on the dog population (Swanepoel, 1994, WHO, 1992). This coupled with the increase in dog intraspecific contact due to a disturbance of the population structure (i.e. reestablishment of social hierarchies by males), the high implementation cost, as well as the resistance which may occur from some local communities to dog removal, rather proves counter productive to rabies control (WHO, 1992; Swanepoel, 1994). Instead it has been proposed that reducing the carrying capacity of neighborhoods through proper garbage disposal, as well as the sterilization of animals, can be as an effective means for reducing dog numbers in urban environments (Swanepoel, 1994; Finnegan *et al.*, 2002).

Import and quarantine regulations constitute another important facet of rabies control (Swanepoel, 1994; Childs, 2002). Countries like Britain, which is free of terrestrial rabies, have previously required that imported dogs and cats be vaccinated and kept in quarantine for at least six months (Finnegan *et al.*, 2002; Woldehiwet, 2002). These measures have recently been amended (2001) by the introduction of a pet travel scheme, that allows for the importation of animals without the accompanying quarantine period, provided that they came from countries and territories that satisfy certain conditions (Finnegan *et al.*, 2002; Woldehiwet, 2002). These criteria include the implantation of a radio frequency identification device (microchip), that animals three months old have been vaccinated with an inactivated rabies vaccine, and that at least six months before importation blood samples be tested for Ab titers against rabies (Finnegan *et al.*, 2002; Woldehiwet, 2002).

Control of wildlife rabies .

Up till recently, wildlife rabies control programs have focused on efforts to decrease the population levels of susceptible wildlife reservoirs, in order to lower the population density below the threshold that is required for the maintenance of the disease (Irvin, 1970; Winkler and Bogel, 1992). Traditional population reduction techniques included hunting, trapping, poisoning and den-gassing have proved largely ineffectual, and have been discontinued due to ecological, economic, and ethical reasons (Rupprecht *et al.*, 1995; Rupprecht *et al.*, 2002; Winkler and Bogel, 1992). Vector populations in areas where control programs are applied, are often rapidly replenished either by the influx of new animals from neighboring areas, or by the increased fecundity of the remaining animals in the target region (Swanepoel, 1994). This necessitates that animals be removed on a continued basis for there to be any effect on rabies incidence (Swanepoel, 1994). Furthermore, as public interest in wildlife conservation has risen, programs based on wholesale slaughter have become increasingly unacceptable (Winkler and Bogel, 1992). Removal of carnivores from their ecological niches may in addition, also have a significant ecological impact, such as for example where there is an increase in population size of pest species (e.g. rodents) that are normally kept in check by predation (WHO, 1992). Wildlife rabies control programs have consequently moved away from population reduction as a control technique, and today rather focuses on oral vaccination through the distribution of vaccine-laden baits.

1.11 Vaccines

Significant advances have been made in the preparation and delivery of human and animal vaccines, since the development of the original neural tissue vaccine by Pasteur. Today a variety of safe, stable, and highly immunogenic inactivated cell culture and veterinary vaccines are available, that can be used to vaccinate humans and animals either before, or after an exposure to RABV (Swanepoel, 1994). Vaccination of domestic animals are essential to prevent rabies infections in humans, since these animals provide the major transmission pathway linking rabies virus cycles in wildlife, to the disease in humans (Krebs *et al.*, 1995).

Today, post-exposure prophylaxis with modern cell culture vaccines can virtually guarantee complete protection following a RABV exposure, provided that it is applied in a correct and timely manner, and is combined with the correct wound treatment and recommended regimens of HRIG (WHO, 1992). Pre-exposure vaccination is recommended for individuals at occupational risk, such as laboratory staff working with RABV, veterinarians, animal handlers, wildlife officers, and other individuals who are living in, or are travelling to, areas where rabies is endemic (WHO, 1992).

In recent decades it has been demonstrated that wildlife can also be vaccinated against rabies by using oral vaccines delivered through edible baits (Krebs *et al.,* 1995). The goals of these vaccinations are to create immune barriers that can prevent or slow the dissemination of the disease (Krebs *et al.,* 1995). Oral vaccines that have successfully been used in the field include a number of attenuated as well as a genetically engineered viral vaccines (WHO, 1992). Earlier field trials of these vaccines proved favourable, and the subsequent mass vaccination of red foxes in Europe and Canada has changed the geographical distribution and incidence of rabies among this and other vector species (Krebs *et al.,* 1995).

Five attenuated vaccines, all of which are mutants of the original Street Alabama Dufferin strain (SAD), have thus far been approved for use in the field vaccination of foxes (Canada and Europe) and raccoon dogs (Finland) (WHO, 1992; Swanepoel, 1994; Pastoret and Brochier, 1999; Finnegan *et al.,* 2002; Woldehiwet, 2002). These include the ERA, SAD-Bern, SAD-B19, and VNukovo-32 strains (WHO, 1992). The SAG-2 oral vaccine strain represents an improvement in safety and genetical stability over the SAG-1 vaccine strain, and was selected from SAD-Bern in a two-step selection process using anti-G MAb. Arg-333 in antigenic domain 3 of the glycoprotein was first mutated to lysine, and then to glutamic acid, resulting in a avirulent vaccine strain with a two-nucleotide difference at position 333, to any of the six possible arginine codon triplets (Schumacher *et al.,* 1993). These mutations provide an additional safety mechanism, since two point mutations instead of one (SAG-1) would be required for the reversion of this strain to virulence (Schumacher *et al.,* 1993). SAG-2 has proven to be apathogenic to target and non-target species, and provides protection against a lethal RABV challenge when taken orally by dogs and foxes (WHO, 1992; Woldehiwet, 2002).

A vaccinia recombinant (VRG), expressing the RABV G-protein has been extensively used to immunize foxes in Belgium and France, and in limited trials in raccoons in America (WHO, 1992; Swanepoel, 1994; Rupprecht *et al.*, 1995; Pastoret and Brochier, 1999; Finnegan *et al.*, 2002). It was developed by inserting a cDNA encoding the G protein of the ERA vaccine strain, into the thymidine kinase gene of the vaccinia Copenhagen strain, a gene that is required by the poxvirus to infect nervous tissue (WHO, 1992; Woldehiwet, 2002). Advantages of using a vaccinia vector as recombinant rabies vaccine is that it has a low cost of production, is easy to administrate, is highly stable and has extensively been used in the past for vaccine production (Ferguson, 1982). Additional recombinants that may be suitable vaccine candidates are those created with racoon pox, canary pox, adenovirus, attenuated Salmonella and BCG (bacille Calmette-Guerin) vectors (WHO, 1992). Neither live attenuated, nor recombinant vaccines are suitable for human use, due to the safety aspects associated with their application (WHO, 1992).

The immunization of susceptible hosts with plasmid DNA vaccines is another potentially promising way of conferring protection against a RABV infection. DNA vaccines have several advantages over conventional killed, live attenuated and recombinant vaccines, in that they are highly stable, there is no possibility of reversion to virulence, and that there is no risk of contamination with adventitious agents, particularly 'live' virus during vaccine preparation (Woldehiwet, 2002). Furthermore, by incorporating multiple viral genes into the vaccine design strategy, they may expand the range of lyssavirus cross reactivity. Mice immunized with DNA vaccines have been shown to develop VNAb and virus specific cellular responses, and to be protected from a lethal virus challenge (Woldehiwet, 2002). Similar results have also been reported in dogs and cats (Woldehiwet, 2002). Disadvantages include the fact that the primary inoculation may need to be followed up by a booster dose for there to be any efficacy, and that they may lack the rapid kinetics required for post exposure prophylaxis in humans (Rupprecht *et al.*, 2002).

Another interesting development has been the demonstration that rabies viral antigens can be expressed in plants, offering the possibility of effective, inexpensive and safe vaccine production and delivery. Plant viruses such as the tobacco mosaic virus (TMV) and tomato bushy stunt virus (TBSV) have successfully been used for the expression of foreign antigens, and have been used for the production of prototypes of plant derived, genetically manufactured vaccines (Briggs *et al.*, 2002). Rabies virus antigen expressed in plant systems have been shown to retain their immunogenic properties, and have

protected mice against a lethal RABV challenge when instilled either orally, or through intraperitoneal injection (Briggs *et al.*, 2002). Although plant viruses represent a promising option for the future control of rabies, substantial problems with practicality and yield however, remain (Rupprecht *et al.*, 2002).

(B) – Themes in molecular epidemiology

1.12 What is molecular epidemiology?

Molecular epidemiology is a scientific field of study in which some measurable molecular characteristic (marker), is used to track the diversity and routes of dissemination of a pathogen within and between host populations. This is done with the express aim of identifying the factors that influence the incidence, distribution and spread of a disease, towards the aim of devising strategies which can bring it under control (See internet references). In lieu of this, molecular epidemiological studies have become a prequisite for providing information to public health authorities for the development and evaluation of possible control strategies (Bourhy *et al.*, 1995).

In RABV the diversity, which occurs between viral populations, associated with discrete geographical regions or host species, are generated as a result of the occurrence of quasispecies (Smith, 2002). Normally the occurrences of these quasispecies are unnoticeable, because the mutations are distributed around a relatively constant consensus sequence (Smith, 2002). Genetic drift of the consensus sequence however occurs when transmission is constrained by a genetic bottleneck, which limits passage to a subset of the viral population (Smith, 2002). This may either be due to the selection of viral variants within a new host, which contains some beneficial mutations promoting its spread, or it may be due the accumulation of neutral mutations that were randomly selected from the viral population during its transmission from host to host (Smith, 2002). In the course of time, variants that circulate within specific geographic boundaries or host species will thus develop unique genetic or antigenic markers, that can be characterized through molecular techniques, in order to distinguish these viral populations from each other (Smith, 2002).

Initially RABV strain discrimination was based on the characterization of the antigenic variation that is contained within the viral proteins by using panels of MAb. This technique provided an advantage in that it was a rapid and an inexpensive method for the large-scale surveillance for both rabies related, as well as common RABV variants (Smith, 2002). It however carried the disadvantage in that strain discrimination was based on the recognition of a few MAbs to a limited number of conformational epitopic sites, which are critical to the structural-functional aspects of the proteins (Smith, 2002). Unfortunately the variations between these epitopic sites are limited, and it was thus not always possible to distinguish between closely related variants and even genotypes (Bourhy *et al.*, 1993; Smith, 2002). MAb typing techniques have thus been supplanted by more sensitive genetic typing techniques that characterize viral variants at the amino acid and nucleotide sequence level. Molecular sequence data allows for the quantitative measurement of relatedness between strains, and is used in the construction of phylogenetic trees that allow for the visualization of the evolutionary relationship between taxa (Smith and Seidel, 1993; Bourhy *et al.*, 1995; Smith, 2002)

1.13 Methods used to generate sequence data

1.13.1 Polymerase chain reaction (PCR)

The discovery of PCR by Kary Mullis (1989) has revolutionized the field of molecular biology (Mullis and Faloona, 1987). PCR allows for the rapid amplification of a target DNA sequence using a heat resistant DNA polymerase enzyme (*Taq*), obtained from a thermophilic bacterium (*Thermus aquaticus*), together with a pair of sequence specific oligonucleotide primers. Successive cycles of heating and cooling, allows for the interaction of the primers with the DNA template, resulting in the priming and extension of new daughter strands, in a semi conservative fashion, on the pre-existing DNA template (Mullis and Faloona, 1987; Sakai *et al.*, 1988).

The advent of PCR, coupled with the improvements in sequence methodology, has resulted in the exponential increase in the rate of the accumulation of sequence data from rabies and rabies related viruses over the past decade (Nadin-Davis, 2000). With the incorporation of a reverse transcription step, RT-PCR can generate a sequence specific amplicon from any lyssavirus, directly from total RNA preparations made from infected brain tissue (Nadin-Davis, 2000; David *et al.*, 2002). Amplicons can then either be directly

sequenced, or subjected to other analytical techniques, such as restriction fragment length polymorphism (RFLP), without the need for time consuming strain adaptation to cell culture, or cloning activities, that had previously restricted molecular epidemiological analysis to a few laboratory-adapted strains (Nadin Davis, 2000).

Sacramento *et al.*, 1991 were the first to describe the use of the PCR technique as an alternative method, for the diagnosis and molecular epidemiology of RABV and the rabies related viruses. Their work initiated most of the molecular epidemiological studies on Lyssaviruses, and has been expanded on in the literature to include the development of strain specific PCRs that can distinguish between different RABV variants and lyssavirus genotypes (Rodney *et al.*, 1997; Nadin-Davis, 1998; Nel *et al.*, 1998).

It should be noted that a typical *Taq* polymerase used in PCR has a mutation rate in the order of 1 mismatch per 10 kb. However, this is unlikely to affect a generated consensus sequence, unless it occurs during the first few rounds of amplification (Hugnes *et al.*, 2000).

1.13.2 Automated fluorescent sequencing

Automated fluorescent sequencing is based on the dideoxynucleotide chain termination method, first introduced by Sanger over 25 years ago (Sanger *et al.*, 1977; Old and Primrose, 1994). It involves the incorporation of fluorecently labeled dideoxynucleotides (ddNTPs) into an extending DNA chain, in a cycle reaction similar to PCR. The reaction differs from PCR, in that only one primer is utilized, and that a mixture consisting of an optimized ratio of ddNTPs to dNTPs is used as monomers for chain synthesis. Incorporation of ddNTPs into an extending DNA chain results in chain termination, due to the replacement of the 3' hydroxyl group on the ddNTPs sugar base with a hydrogen atom. This results in the generation of a mixture of a nested set of fragments during the cycle reaction, each of which terminates at a different base, and each of which is labeled with a uniquely colored fluorescent dye. The labeled fragments are then either separated by SDS-PAGE, or through capillary electrophoresis, after which a laser visualizes the fluorescent dye labels on the sequence fragments. The emitted light is fed into a spectrograph that separates the light according to wavelengths, and from there into a charged-coupled device or CCD camera, that generates a signal which can be interpreted

by appropriate sequencing software on a personal computer (Sanger *et al.*, 1977; Old and Primrose, 1994).

1.14   Methods used to infer phylogenetic relationships

Phylogenetic trees are mathematical structure used to model the evolutionary history of a group of organisms, by using stretches of thier genomic sequence (Li and Graur, 2000; Vandamme, 2003). The actual pattern of historical relationships are referred to as the phylogeny of the sequences under study, and it is the objective of phylogenetic tree construction methods to reconstruct the true phylogeny among a set of organisms by using either their nucleotide or amino acid sequence data (Hungnes *et al.*, 2000; Li and Graur, 2000; Smith; 2002; Baldauf, 2003; Vandamme, 2003).

The foundations for the field of molecular phylogenetics were established in the 1950s. Two events instigated these developments and included the discovery of the structure of DNA by Watson and Crick (1953), and the first publication of a comparison between amino acid sequences from the insulin gene from different domestic livestock species (Li and Graur, 2000). By the 1960s the increase in the amounts of amino acid sequence data had led the development of models of evolution, which are essential for accurately recovering evolutionary information from sequence data (Li and Graur, 2000). The three most important assumptions introduced in the 1960s was that sequence evolution was determined primarily by random events, that by observing only the number of differences between sequences the nucleotide divergence could be underestimated, and that genes may evolve at a constant rate among lineages, allowing evolutionary events to be placed within a time frame (i. e. molecular clock) (Li and Graur, 2000).

Sarich and Wilson (1967) in an investigation of the evolutionary relationships between humans and primates, were the first to demonstrate the use of gene sequence data for the reconstruction of phylogenetic relationships (Li and Graur, 2000). Since then there has been an increased understanding of how sequences evolve, coupled with the development of ever increasingly more sophisticated methods for recovering evolutionary information from sequence data. These developments ,combined with the increased availability of nucleotide and amino acid sequence data, has led to field of molecular systematics coming into its own over the last two decades (Li and Graur, 2000).

Phylogenetic reconstruction can be divided into five essential steps which include the (1) construction of a sequence dataset, (2) the compilation of the gathered sequences into a multiple alignment, (3) the identification of a nucleotide substitution model by which the sequences may have evolved, (4) the actual tree building itself, and finally (5) the evaluation of the confidence levels which exist with respect to the reconstructed tree topology (Internet reference). Table 1.2. presents some terms which are commonly used in the field of phylogenetics (Li and Graur, 2000)

Table 1.2. Some common terms used in the field of phylogenetics (Adapted from Li and Graur, 2000)

Bifurcation	The graphical representation in a phylogenetic tree of an evolutionary speciation event whereby an ancestral taxon splits into two.
Branches	The relationships among the units in terms of descent and ancestry.
Branch length	Represents the number of changes that have occurred in a branch.
External nodes	Represent the currently existing taxonomic units under comparison, are referred to as **operational taxonomic units (OTUs).**
Internal nodes	Represent ancestral units.
Inferred trees	A tree that is obtained using a certain set of data and a certain method of tree construction. May or may not be identical with the true tree.
Nodes	Represents the taxonomic units, eg. species, populations, individuals or genes.
Phylogenetic tree	Evolutionary relationships among a group of organisms illustrated by means of a graph composed of nodes and branches, in which only one branch connects any two adjacent nodes.
Rooted tree	In such a tree there exists a particular node, called the root from which a unique path leads to any other node. The direction of each path corresponds to evolutionary time, and the root is the common ancestor of all the OTUs under study.
Scaled branches	The length of the branch is proportional to the number of changes.
Topology	The branching pattern of a tree.
True tree	The one tree that can be built with a given number of OTUs representing the true evolutionary history.
Unrooted tree	A tree that only specifies the relationships among the OTUs and does not define the evolutionary path.
Unscaled branches	The length of the branch is not proportional to the number of changes.

14.1 Multiple sequence alignment

Multiple sequence alignment which are carried out on a microcomputer, is a prequisite for most phylogenetic analyses (Higgins, 2003). Phylogenetic inference techniques require that homologous residues (i.e. residues that descend from a common ancestor) be compared, in order to determine rates or patterns of change in DNA or protein sequences (Higgins, 2003). The best way to achieve this is to align the sequences, so that homologous positions are aligned in columns within a multiple sequence alignment. This is easy to achieve if all the sequences are of the same length, but unfortunately the introduction of deletions and insertions (indels) of nucleotides which complicate sequence alignment ,is a hallmark of viral evolution (Higgins, 2003).

In order to obtain an optimal alignment between two nucleotide sequences, a scoring system has been devised that computes the "cost" of an alignment. The alignment that maximizes the alignment score then represents the optimal alignment for any two sequences under study (Higgins, 2003). A simple measurement of sequence identity can be calculated by counting the number of matches between nucleotides, divided by the number of aligned positions (excluding gaps) within a sequence alignment (Higgins, 2003). Alignments with more matches between nucleotides will thus result in a higher alignment score, when compared to alignments in which the sequences have been sub-optimally aligned (Higgins, 2003). Unfortunately this method can result in nonsensical alignments in which an excess number of gaps can be introduced to maximize the alignment score (Higgins, 2003). To overcome this problem gap penalties (GP) have been introduced that penalize the alignment score, thereby limiting the insertion of gaps to optimize the alignment (Higgins, 2003). Furthermore, separate gap penalties are also introduced for the insertion and extension of gaps, due to the fact that the insertion or deletion of a nucleotide is in general evolutionary more "expensive" than the cost required to extend a gap (Baldauf, 2003). A typical formula for calculating gap penalties that takes this into account, is given by the formula $GP = g + hl$, where l represent the length of the introduced gap, h the gap extension penalty, and g the gap opening penalty (Higgins, 2003). Finally, it is also customary to make the insertion of gaps at the end of a nucleotide sequence alignment free, in order to prevent the unnecessary insertion of gaps at the termini of sequence alignments, in which sequences of unequal length have been aligned (Higgins, 2003).

Amino acid sequences are aligned in a similar manner to nucleotide sequences except that scoring matrixes have been devised that take into account the biochemical properties of the amino acids, and the tendency for their substitutions to remain relatively conserved, to amino acids with similar chemical or physical properties (Higgins, 2003). Several classes of scoring matrixes have been devised which include tfor example the PAM and BIOSSUM matrixes. These matrixes are tailor made for amino acid alignments, depending on the percentage of nucleotide sequence divergence which is present between the set of sequences to be aligned (Higgins, 2003).

More than two sequences are typically aligned by progressive multiple sequence alignment algorithms. These algorithms progressively add more similar sequences to an alignment, followed by more dissimilar ones (Baldauf, 2000; Higgins, 2003). This process is usually conducted by the construction of crude guide tree (typically a neighbourhood-joining tree) which is reconstructed from similarity scores from a pairwise alignment of all the sequences to be included in the alignment (Baldauf, 2000). The alignment is then progressively built up by following the order of the relationships between the sequences as presented in the guide tree (Baldauf, 2000). The insertion of a gap is maintained throughout the alignment, as sequences are progressively added. This is because it is assumed that the best information on gap placement will be found among the most similar sequences (Baldauf, 2000).

1.14.2 Methods used in tree construction

Methods used in the construction of phylogenetic trees can be divided according to two criteria. The first of these is whether the tree construction method uses discrete character states (discrete character methods), or a distance matrix of pairwise dissimilarities between sequences (distance methods) in the tree construction algorithm (Vandamme, 2003). The second criteria classifies tree construction methods according to whether the method clusters sequences (OTUs) in a step wise manner (stepwise clustering methods), or whether it considers all theoretical trees that can be constructed from the given dataset (exhaustive search methods) (Vandamme, 2003). Table 1.3. summarizes some of the methods most commonly used in phylogenetic analysis as well the strategies these methods are based on (Vandamme, 2003).

Table 1.3. Some construction methods used in phylogenetic analysis, and the strategies they are based on (Vandamme, 2003).

	Exhaustive search	**Stepwise clustering**
Character state	Maximum parsimony	
	Maximum likelihood	
Distance matrix		UPGMA
		Neighbourhood- joining (NJ)

Distance methods.

In distance methods a tree is constructed from evolutionary distances contained in a distance matrix. The distance matrix contains information on the dissimilarity between sequences (roughly percentage nucleotide sequence difference), and is calculated from a pairwise comparison of all the OTUs which were included in the analyses (Felsenstein, 1984; Felsenstein, 1988; Hungnes, 2000; Vandamme, 2003). The amount of dissimilarity between the sequences provides a measurement of the evolutionary relatedness between the OTUs, which can then be interpreted and visualized by phylogenetic tree construction methods (Felsenstein, 1988; Hungnes *et al.,* 2000; Vandamme, 2003).

For highly divergent nucleotide sequences, distances based on the percentage divergence can give an underestimation of the true genetic distance between the OTUs being studied. This is because more than one mutation could have taken place consecutively (multiple hits) at a specific position within a sequence, as it became more divergent from its ancestral sequence (Hungnes *et al.*, 2000; Baldauf, 2003; Strimmer and Haeseler, 2003). Distance methods take multiple hits into account when calculating a distance matrix, by applying mathematical models that correct the percentage difference between the analyzed sequences. The corrected distances are referred to as the "evolutionary distance", and are always larger than the percentage values calculated directly by a comparison between the sequences (Vandamme, 2003).

Two different models that are commonly used to correct distance matrixes include the Jukes-Cantor's one- parameter, and Kimura's two-parameter model (Jukes and Cantor, 1969; Kimura, 1980; Hungnes, 2000; Strimmer and Haeseler, 2003). The Jukes cantor model corrects the distance matrix by assuming that there is no bias in the change of the

direction in the nucleotides, but does take into account that a base can mutate back to its initial position (Hungnes *et al.*, 2000). Kimura's two-parameter model corrects distances by assuming that there is an unequal rate between transition (purine to purine, pyrimidine to pyrimidine) and transversion mutations (purine to pyrimidine or vice versa), with transition mutations occurring more frequently than transversion mutations (Hungnes *et al.*, 2000).

Character state methods.

Character state methods use the discrete characters of nucleotide or amino acid sequences directly for the construction of phylogenetic trees. For nucleotide and amino acid sequences, every position in an alignment is considered to be a "character", while the type of amino acid or nucleotide base at a specific position is referred to as the "state" of that character (Vandamme, 2003). In these methods the states of characters within each position within a multiple alignment is examined separately, and independently from each other, and a tree is constructed which best accommodates all of this information (Baldauf, 2003; Vandamme, 2003).

Stepwise clustering methods.

Stepwise clustering methods results in the construction of a single tree by sequentially adding OTUs to a growing tree. It starts by clustering the most closely related OTUs together, which are then taken as a new OTU. The next most closely related OTU is then added to the newly created OTU, and the process is repeated until all but two of the OTUs have been clustered (Vandamme, 2003).

Exhaustive search methods.

Exhaustive search methods are methods that examine every possible tree topology that can be constructed from a given dataset. It assigns probability scores to reconstructed trees, based on how well the data fits these trees under a specific model of evolution. It then selects the best tree among the reconstructed trees by using an optimality criterion (determined by the method of tree construction) (Li and Graur, 2000; Vandamme, 2003). The best tree (or trees) will then be those that maximize the probability score for the optimality criterion (Li and Graur, 2000; Vandamme, 2003).

The number of possible trees grows quickly as the numbers of OTUs in an analysis are increased, and these methods are typically not applicable to more than 10 taxa (Vandamme, 2003). Strategies (heuristic, branch and bound, close neighbor interchange (CNI) method etc.) have however been developed which attempt to search "tree space" for the optimal tree, without examining every possible tree topology (Hungnes *et al.*, 2000; Swofford and Sullivan, 2003). These strategies however carry the disadvantage in that they cannot guarantee that the best tree had been examined during the analysis (Hungnes *et al.*, 2000).

1.14.2.1 Distance methods

Unweighted pair group method with arithmetic mean (UPGMA).

The UPGMA method is one of the simplest methods used in tree construction. It assumes that the OTUs under study evolved by a molecular clock (i.e. that the rate of mutation is constant in all the OTUs) and the branches on the tree are thus ultrameric (Hungnes *et al.*, 2000; Vandamme, 2003). UPGMA uses a clustering algorithm in which sequences with the smallest genetic distances between them are sequentially added to a growing tree. Its starts by constructing a distance matrix for a pairwise comparison between all the OTUs under study. It then clusters the first two OTUs together with the smallest genetic distance between them, and treats these OTUs as a new single OTU. The average genetic distance between the new OTU and every other OTU is then recalculated and a new distance matrix is constructed. The process is then repeated until all but two of the OTUs have been clustered (Vandamme, 2003). The UPGMA method can lead to the construction of incorrect tree topologies when the rates of evolution between lineages are uneven. When this happens it is better to utilize other methods such as the neighbourhood joining method, which is based on the use of the true (additive), rather than average distances (ultrameric distances) between the OTUs, during the tree construction process.

The neighbourhood-joining method (NJ).

In the neighbourhood joining method an unrooted tree is constructed, in which the distances between any pair of OTUs, represents the number of mutation that has occurred in the evolutionary pathway connecting them since their divergence from a common ancestor (Saitou and Nei, 1987; Studier and Keppler, 1988; Van de Peer, 2003). The

method does thus not assume that the sequences being considered evolved by a molecular clock, and the distances on the reconstructed tree thus reflect the true (additive distances) rather than the average distances between the OTUs (ultrameric distances) (Saitou and Nei, 1987; Studier and Keppler, 1988; Van de Peer, 2003).

The algorithm starts by constructing a corrected distance matrix between the sequences being analyzed, and represents the OTUs as a star tree in which one node connects an OTU to every other OTU. It then calculates the netto divergence from every OTU to every other OTU by adding the distances between them, and summarizes these values in a new distance matrix. It creates a new bifurcating node connecting the OTUs with the smallest distance between them, and calculates the branch lengths between the newly created node to the newly joined OTUs. This is then followed by the calculation of the distance between the newly created node to every other unclustered terminal OTU, after which the distances are again summarized in a distance matrix. This process is then repeated until all the OTUs have been clustered (Saitou and Nei, 1987; Studier and Keppler, 1988; Van de Peer, 2003).

1.14.2.2 Discrete character methods

Maximum parsimony method (MP).

The objective of the maximum parsimony method is to examine every possible tree topology that can be constructed from a given dataset, in an attempt to find the tree (or trees) which requires the least amount of evolutionary changes required to explain the pattern of nucleotide substitutions in the sequence data (Fitch, 1971; Hungnes *et al.*, 2000; Swofford and Sullivan, 2003). The method examines the nucleotide sequence data directly, and only parsimony informative characters (contains at least two types of states at a character position and at least two of them occur with a minimum frequency of two) from four or more aligned sequences are used during the analysis (Bourhy et al., 1995; Li and Graur, 2000).

The ancestral sequence is inferred separately at each character site for a given topology, under the assumption that substitutions can occur in all directions. The smallest number of nucleotide or amino acid changes is then computed that will explain the entire evolutionary process. This computation is done for all potentially correct tree topologies, and the

topologies which require the least number of substitutions are then chosen as the most parsimonious trees (Li and Graur, 2000; Swofford and Sullivan, 2003).

Maximum likelihood method (ML).

The maximum likelihood method uses a likelihood function to identify the optimal tree from all possible tree topologies that can be constructed from a dataset (Vandamme, 2003). The objective of ML is to determine what the probability is of observing your data when given a hypothesis. It then chooses among different competing hypothesis the one that maximizes the likelihood of observing the data (Li and Graur, 2000). In phylogenetics terms, the tree (hypothesis) that has the highest likelihood of all possible tree topologies, is taken as the correct tree, because it maximizes the probability of observing the data (Li and Graur, 2000).

The tree is constructed from the nucleotide sequence data under the assumption of a model of molecular evolution (nucleotide substitution model) (Hungnes et al., 2000; Li and Graur, 2000). The total likelihood for a given tree is then calculated by determining the likelihood scores for each column in the sequence alignment, and by multiplying the individual likelihoods for all the columns within a sequence alignment, with each other (Li and Graur, 2000). This is evaluated for all possible tree topologies, and the tree that maximizes the likelihood values for observing a specific set of data under the chosen model of evolution, is then taken as the best tree (Hungnes *et al.*, 2000; Li and Graur, 2000).

1.15 Bootstrap analysis

The bootstrap analysis is statistical technique used to generate confidence limits on phylogenies, where the underlying sample distribution of OTUs which are used in the construction of a phylogenetic tree is unknown (Felsenstein, 1985; Van de Peer, 2003). It is based on the random re-sampling of columns from a multiple sequence alignment, until a dataset of equal length to the original has been constructed (pseudoreplicate dataset). For each pseudoreplicate dataset included in an analysis, a phylogenetic tree is constructed, and the frequencies with which particular groupings (present in the original phylogenetic tree) are reproduced is then computed. The proportion, at which a particular group was reproduced from the total number of pseudoreplicate datasets, is then taken as

the statistical support for that particular grouping (Baldauf, 2003; Van de Peer, 2003). In general 200-2000 resamplings are recommended for an analyses, while bootstrap support values of 70-75% for branches and groups are regarded as being statistically significant (Baldauf, 2003; Van de Peer, 2003). Two approaches can be used to demonstrate the bootstrap values on a phylogenetic tree. The first of these involves the summary of the bootstrap analyses in majority rule consensus tree (discrete character methods), while in the second approach the bootstrap values are superimposed on the tree obtained from the original sequence alignment (distance methods) (Van de Peer, 2003).

1.16 Molecular clock analysis

A molecular clock presupposes that the rate of nucleotide substitution within a gene among lineages accumulates linearly over time as they diverge from each other (i.e. the rates of nucleotide substitution remains constant over time) (Vandamme, 2003). This has important implications, since if the rate of nucleotide substitution is known it can be used to date the unknown divergence times between any two lineages (Vandamme, 2003). Sequences however do not always evolve by a molecular clock, since selective pressures that promote or restrict evolution at particular gene loci differ among different sites. In order to date divergence events, a dataset thus needs to be tested on whether the molecular clock hypothesis holds true (Posada, 2003). One method by which this can be achieved is by conducting a relative rate (Tajimas) test, in which the amount of divergence between any two sequences is compared to the divergence that has occurred in an outgroup (Tajima, 1993; Posada, 2003). Under a perfect molecular clock the distances between the first sequence and the outgroup will be equal to the distance between the second sequence and the outgroup, with the distances between the two sequences and the outgroup being equal. This hypothesis will hold true for a perfect molecular clock, irrespective of the substitution model chosen, and whether or not the substitution rate varies at the particular site analyzed (Tajima, 1993). If this hypothesis is rejected for a set of sequences however, it can be assumed that they did not evolve by a molecular clock.

1.17 Advantages and disadvantages of tree construction methods

The choice of method used to reconstruct the evolutionary relationship between taxa is influenced by a number of factors that include the number of taxa, the length of sequences analysed, as well as the type of questions that are being investigated. Tree construction methods may, or may not, result in reconstruction of the true evolutionary tree, and it is thus advised that different methods be used in parralel to evaluate the correctness of the tree topology which was obtained by any chosen method (Bourhy *et al.*, 1995). Table 1.4. summarises some of the advantages and disadvantages of the different tree construction methods as discussed above (Bourhy *et al.*, 1995).

Table 1.4. Advantages and disadvantages of a few tree construction methods (Bourhy *et al.*, 1995)**.**

Tree construction method	Advantages	Disadvantages
UPGMA	Fast , can be used for a large number of taxa.	Construction not based on the sequences themselves
Neighbourhood-joining (NJ)	Fast, can be used for a large number of taxa, branch lengths are proportional to amount of genetic change between lineages	Construction not based on the sequences themselves
Maximum parsimony (MP)	Efficient, considers the nucleotide sequence data itself. Allows reconstruction of ancestral sequences.	Computationally expensive and only parsimony informative characters are analysed. Thus limited to a few taxa from which longer sequence segments have been obtained.
Maximum likelihood (ML)	Very efficient, considers nucleotide sequence data itself. Allows reconstruction of ancestral sequences.	Computationally very expensive and thus limited to small datasets.

(C) - Aims of this study

The domestic dog is recognized as the principal disseminator and maintenance host for rabies in the South African province of KwaZulu Natal (KZN), and two separate epidemics, believed to have originated in Mozambique, have occurred among this vector species during the late 1960s and middle 1970s, respectively (Swanepoel, 1994; Bishop *et al.*, 2003). The second epidemic has as of yet not been brought under control, and is responsible for the majority of human cases that are reported annually from South Africa (Swanepoel, 1994; Bishop *et al.*, 2003). In order to contribute to the understanding of the molecular epidemiology of the disease in KZN, as well as South Africa as a whole, we undertook a molecular phylogenetic analysis of a representative panel of viral isolates obtained from the province during the calendar year 2003, and characterized these isolates phylogenetically.

Previous studies conducted in southern Africa, included a limited number of isolates from KZN, but were primarily aimed at illustrating the distinction which exists between the canid and viverrid biotypes (Nel *et al.*, 1993; Von Teichman, 1995), or were focused on determining the genetic diversity of the canid biotype over a broader geographical region (Sabeta *et al.*, 2003). In this study the emphasis was placed on determining the regional genetic variation of viral isolates which were specifically obtained from the different magisterial districts of KZN. The analysis was based on a portion of the carboxyl terminal domain of the glycoprotein (cytoplasmic domain) and the G-L intergenic region, which are the most variable part of the RABV genome, and which would thus prove suitable for distinguishing between closely related strains (Nadin-Davis, 2000). These regions have previously been shown to be useful markers for determining the genetic variation for rabies virus isolates from France, South Africa, Zimbabwe and Columbia (Sacramento *et al.*, 1992; Nel *et al.*, 1993; Von Teichman, 1995; Jacobs *et al.*, 1997; Sabeta *et al.*, 2003; Paez *et al.*, 2003).

This study also aimed at answering broader questions pertaining to the origin and spread of the epidemic, into and throughout the province, and to place this information in a historical context. Furthermore the study also attempted to answer questions relating to whether single or multiple introduction events were responsible for the establishment of rabies within the province, and to determine the possible role, if any, that wildlife plays in the maintenance of the disease. Furthermore it was decided to illustrate the utility of

phylogenetic analysis for retracing case histories of humans infections from the province, in order to demonstrate how the methods applied in this study may assist health professionals in procuring this type of information.

The epidemiological situation of rabies in the province was subsequently re-evaluated based on the molecular data which was generated during this investigation, and suggestions were made on how current control strategies within the province might be improved. The sequences which were generated during the course of this project will be submitted to Genbank, and will undoubtedly play an important role in future surveillance efforts targeted to elucidating the spread of the virus in regions throughout the KwaZulu Natal province.

The following techniques were applied during the course of this project:

1. Extraction of RNA from infected brain material.
2. cDNA synthesis from RNA using reverse transcription (RT).
3. Polymerase chain reaction (PCR) of the glycoprotein cytoplasmic domain and G-L intergenic region, by using gene specific oligonucleotide primers.
4. Purification of amplified DNA products from electrophoresis gels.
5. Nucleotide sequencing of purified amplicons using automated fluorescent sequencing.
6. Multiple sequence alignment and phylogenetic analysis of the obtained sequence data.

Chapter 2

Sequence analysis of rabies virus isolates obtained from KwaZulu Natal

2.1 Summary

Chapter 2 provides a discussion of the current epidemiological situation of rabies in KwaZulu Natal, as well as the materials and methods which were implemented during the course of this study. The result which were obtained from the phylogenetic analysis which was conducted on the KwaZulu Natal isolate and reference sequence data is also described. The implications of the reconstructed phylogeny with respects to the epidemiology of rabies within the KwaZulu Natal province, will however only be discussed in detail in chapter 3.

2.2 Introduction

The aim of this investigation was to establish the molecular epidemiology of rabies in the KwaZulu Natal (KZN) province of South Africa, based on a molecular genetic sequence analysis of a representative sample of virus isolates from the single calendar year, 2003. Although being one of South Africa's most populous provinces, KwaZulu Natal is relatively small, comprising only approximately 7% of the total land area of the Republic. It is about 600 km long and about 200 km wide and extends from the borders with Swaziland and Mozambique in the north, to the Eastern Cape in the south. Inland it is bound by the Kingdom of Lesotho, the Freestate, Swaziland and the Mpumalanga province (www.nationmaster.com/Demographics-of-South-Africa). The province can be divided into three geographical regions: the lowland region along the coast of the Indian Ocean, plains in the central region, and two mountain ranges, the Drakensberg and Lebombo Mountains in the west and north, respectively (www.nationmaster.com/Demographics-of-South-Africa)). Rainfall occurs mainly through the summer and varies from 400mm to 1000mm per annum. Most of the rivers in the province have their origin in the west and flow eastwards towards the Indian Ocean (www.nationmaster.com/Demographics-of-South-Africa)). The geographic location of the KZN province within South Africa is indicated in figure 2.1.

Figure 2.1. Provinces of South Africa

Agriculture is extremely diverse within KZN, as is reflected by the provinces diverse topography and climatic conditions. Warm and humid weather prevails along the coast, allowing for the production of tropical fruit and sugar cane, whereas high rainfall areas and cooler temperatures in the central regions are ideal for crop and timber production Inland in the northeastern region of the province is the Lowveld, a subtropical region that is limited in cropping potential, but is well suited for beef ranching and game farming. Main industrial centres are located around coastal cities such as Pietermaritzburg, Durban, Richards Bay and Port Shepstone. Major highways connect these centres to each other, and includes among others the N2 highway which runs along the eastern coastal belt, and that connects Richards Bay, Durban and Port Shepstone to Cape town via East London and Port Elizabeth (Internet reference). The N3 highway runs through the central and northwestern districts of the province, and connects Durban to Johannesburg. The coastal region north of Durban is referred to as the "North Coast", whereas the region south of Durban is known as the "South Coast"(http://agriculture.kzntl.gov.za.dae.index.aspx?ID=4, www.nationmaster.com/Demographics-of-South-Africa).

The KZN province, like the 8 other provinces of the Republic of South Africa, is subdivided into 51 independently managed governmental units known as magisterial districts (Internet reference). The ethnic composition of the provinces' population is diverse and consists of approximately 79% black, 10% white, 9% coloured and 3% Asian (Internet reference). The black population is concentrated within the former homelands, as well as in informal settlements surrounding towns and cities, at or near the agricultural, industrial and recreational hubs of the province (http://agriculture.kzntl.gov.za.dae.index.aspx?ID=4, www.nationmaster.com/Demographics-of-South-Africa).

Prior investigations using MAb typing and phylogenetic analysis have indicated that rabies viruses associated with infections in dogs in KZN, are of the canid biotype (King et al., 1993; Nel *et al*., 1993; Von Teichman *et al*., 1995; Bishop *et al*., 2003). These viruses form a discrete group that are closely related to each other, but distinct from the cosmopolitan PV (Nel *et al*., 1993; Von Teichman *et al*., 1995). The number of rabies cases in the region typically amounts to between 121 to 409 animal, and 3 to 22 human cases per annum (1992-2003), with children being the principal victims of the disease (Randles, 2003). Human patients who succumb to the disease in the province invariably receive little to no post exposure prophylaxis, highlighting the lack of education of the public on the dangers of rabies (NICD annual report, 2003).

The number of human cases reported each year appears to closely correlate with the number of infected dogs, implicating the importance of this vector as the principal transmission host of rabies to humans in the province (Swanepoel, 1994). During periods in which a high number of cases are reported among dogs, similar increases in infection rates are also observed among humans and other animal species. The inverse is noted following the implementation of vaccination campaigns that reduce the incidence of dog rabies (see figure 2.1 and 2.2). The number of infected dogs also show peaks every 3 to 4 years, with a strong seasonal bias being present (Randles, 2003; Swanepoel, 1994). The highest numbers of cases are recorded in the late and early spring months, although the precise reason for this has as of yet not been determined (Swanepoel, 1994). This trend in dogs is unexpected in light of the fact that these fluctuations are typically only observed among wildlife species (ex. *C. mesomelas*) (Swanepoel, 1994) For wildlife the 3 to 4 year movement averages are indicative of fluctuating vector populations as a result of the decimation and recovery of susceptible host populations following a rabies outbreak, while the seasonal trend in the number of cases correlates with the seasonal availability of food

and water, or the dispersal of new progeny (Swanepoel, 1994; Bishop *et al.*, 2003). Dogs which are largely dependent on humans for their sustenance, would thus not be expected to show a similar seasonal trend. It could however be expected that rabies cases in dogs that have become feral, may show a such a fluctuating trend (Bateman, 2005).

Although it has been reported that feral dog packs do not play a role in the epidemiology of the disease in KZN, this situation is rapidly changing due to the HIV/AIDS pandemic that is currently raging throughout the province (Bateman, 2005). High mortality rates associated with the disease, has resulted in an increase in the percentage of free roaming unowned dogs, that form feral packs, and contribute to the spread of rabies (Bateman, 2005). Veterinary workers estimate that a 50% increase in feral dogs has occurred in some country villages, while in informal settlements close to industrial areas like Richards Bay and Empangeni, the figure may be even higher (Bateman, 2005).

 Regionally the highest numbers of rabies cases are observed along the coastal regions, especially in the coastal magisterial districts surrounding the major industrial hubs of Durban, Richards bay and Port Shepstone (Randles, 2003). With the exception of magisterial districts in the northeastern regions of the province, fewer cases are reported inland, possibly reflecting the lower density of the human population which is present in the internal regions of the province (Randles, 2003; Bishop *et al.*, 2003).

Figure 2.2. presents a graph illustrating the dog and human rabies trends per annum, for the periods from 1992 to 2002. The annual number of dogs vaccinated during this period is presented in figure 2.3 (Randles, 2003). In the periods from 1990 to 1995, an increase in the number of dog, human and animal infections were observed. According to Randles (2003), this increase may likely be related to social and political changes that had occurred during that time period. The time period preceding and following the change of government after the fall of apartheid (1994), was characterised by political instability and factional infighting that exacerbated the difficulties associated with obtaining sufficient vaccination coverage for all the rabies affected regions of the province (Randles, 2003; Bishop *et al.*, 2003). The fall of apartheid further provided new freedom of movement to people, and the subsequent migration of people from rural to urban areas coupled with the parallel increase in the dog population, could be another factor that contributed to the higher number of infections observed (Randles, 2003; Bishop *et al.*, 2003).

The veterinary services from KwaZulu and Natal were previously divided into two separate veterinary authorities (Randles, 2003; Bishop *et al.*, 2003). These authorities were restructured into a single new veterinary authority in the early 1990s. This led to the increased co-operation between these authorities in campaigns targeted to eliminating rabies, possibly explaining the drop in cases that was observed in the time period from 1995-1997 (Randles, 2003; Bishop *et al.*, 2003). An upward trend in dog, human and animal infections was however again noted from 1997 onward. This could be ascribed to an outbreak of Foot and Mouth disease that had occurred in the province during that time period, and which refocused the efforts of veterinary staff to the detriment of rabies control (Randles, 2003). This is illustrated in figure 2.1. and 2.2 by the observed decrease in the number of vaccinated dogs, and the subsequent increase in the number of rabies cases in the affected species, from 1997 onward.
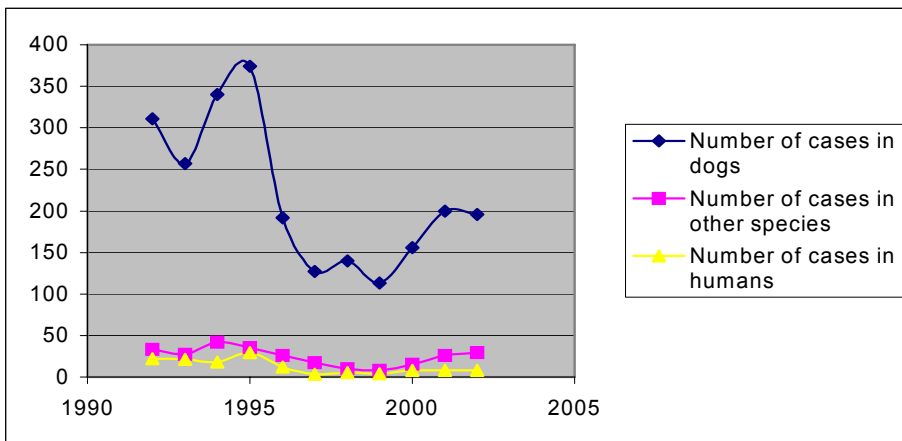

Figure 2.1. Rabies virus trends for KwaZulu Natal between 1992-2002 (Randles, 2003).
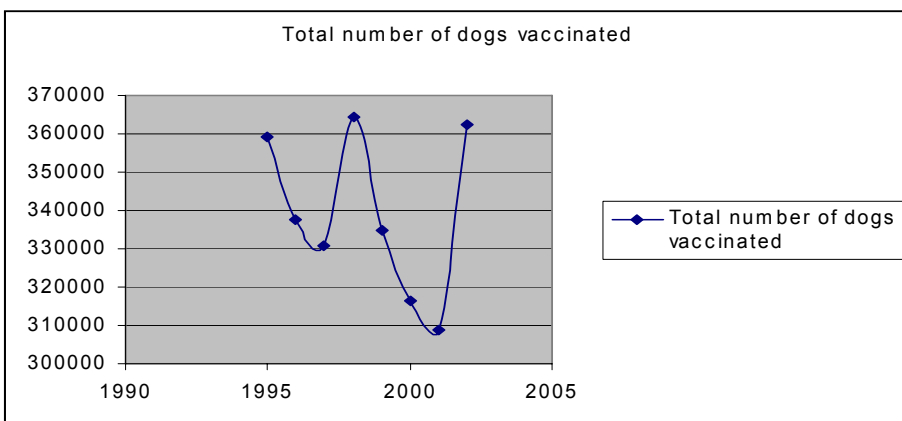
Figure 2.2. Total number of dogs vaccinated in KwaZulu Natal for the time period between 1992-2002 (Randles, 2003).

81

Rabies outbreaks typically appear to form localized outbreaks on map, that varies over time. This creates the impression of a situation in which outbreaks are brought under control by targeted vaccination campaigns, after which new outbreaks flare up in geographically distant locations (Randles, 2003). Although dogs represent the main vector of rabies in KZN, incidental spillover to domestic livestock also occurs. Table 2.1 presents the total number of cases of rabies diagnosed in dogs, humans, domestic livestock in KZN for the year 2003 (Randles, 2003). No other wildlife isolates were submitted during thisthis time period. Table 2.2 gives a summary of the total number of infections that had occurred in KZN in 2003, by magisterial district of isolation (Randles, 2003).

Table 2.1. Total number of infections in domestic dogs, livestock, humans and wildlife for the year 2003 (Randles, 2003).

Species	Number of cases	% of cases
Canine	203	83
Bovine	16	6.5
Caprine	6	2.5
Feline	4	1.6
Equine	3	1.2
Ovine	1	0.4
Human	11	4.5
	Total number of cases	Total percentage
	244	100

Table 2.2. (page 107) Summary of the number of cases diagnosed per magisterial district for the calendar year 2003. Red zones have been defined as districts in which more than ten cases have been diagnosed from a single calendar year, while amber and green zones represent districts in which 1 to 10, or no cases had been diagnosed during the same time period (Randles, 2003)

Red Zone district (>10 cases)	
District	Number of cases (dogs)
Lower Umfolozi	40 (36)
Vryheid	29 (27)
Hlabisa	21 (19)
Port Shepstone	14 (11)
Umzinto	13 (9)
Mtuzini	10 (10)
Amber Zone district (1-10 cases)	
District	Number of cases (dogs)
Ixopo	9 (7)
Lower Tugela	9 (7)
PR/P	9 (6)
Inanda	8 (8)
Ubombo	7 (5)
Alfred	6 (5)
Eshowe	6 (6)
Ndwedwe	6 (5)
Nongoma	6 (4)
Pinetown	6 (6)
Durban	5 (3)
Mhlabatini	5 (4)
Umvoti	5 (4)
KNA	4 (2)
Camper Down	3 (3)
Msinga	3 (2)
Ngotshe	3 (2)
Ingwavuma	2 (2)
Mount Currie	2 (1)
Mtonlaneni	2 (2)
Nqutu	2 (2)
Pietermaritzburg	2 (1)
Mapumolo	1 (1)
New Hanover	1 (1)
Green Zone district – No (?) cases	
District	District
Babanango	Impendle
Bergville	Klip Rivier
Chatsworth	Kranskop
Dannhauser	Lions River
Dundee	Mooi Rivier
Estcourte	New Castle
Richmond	Underberg
Umbumbulu	Utrecht
Umlazi	Weenen
Glencoe	Polela

Different methods and strategies have been applied in an attempt to control rabies in KZN and South Africa. For urban outbreaks caused by the mongoose biotype, it is customary to vaccinate all dogs and cats in the affected street as well as neighbouring blocks (Bishop *et al.*, 2003). For an outbreak of the mongoose biotype on farms and smallholdings, all animals on the property as well as surrounding properties are also vaccinated (Bishop *et al.*, 2003). For outbreaks associated with the canid biotype, attempts are made to

83

vaccinate all cats and dogs in a radius of 25-km surrounding the primary outbreak area (Bishop *et al.*, 2003).

Vaccination strategies that have been applied with some success in the province include large scale, cordon, ring, house to house, central point and oral vaccination (Bishop *et al.*, 2003). The aim of synchronised large-scale vaccinations is to achieve 70% vaccination coverage in the shortest time possible, in order to break the transmission chain of the epidemic (Bishop *et al.*, 2003). Cordon vaccinations are performed to prevent the spread of rabies from affected to rabies free regions, and is done for a cordon 20 to 30 km wide on regions bordering areas from which the disease may be reintroduced (Bishop *et al.*, 2003). Cordon vaccination is typically performed along international borders to prevent the introduction of rabies from neighbouring countries, as well as along regions bordering game parks in order to prevent the spread of the disease into wildlife (Bishop *et al.*, 2003). Ring vaccinations are performed in a 20 to 30 km radius (70% coverage) surrounding a primary outbreak, and are performed to reduce the chance of the transport of an unvaccinated animal which may be incubating the disease, to a distant geographic location (Bishop *et al.*, 2003). Door to door vaccinations have proven to be particularly useful for achieving higher vaccination coverage than which is usually achievable by mass vaccination. It has been demonstrated for example that a vaccination coverage of 94% (6498 dogs) for dogs in 30 villages stretching over 594 square kilometre region in the Mpumalanga province, was achievable by 20 veterinary officials in just 7 days (Bishop *et al.*, 2003). Central point vaccination on the other hand, results in a lower vaccination coverage of between 20-80%, depending on the good advertising and planning of the vaccination campaign (Bishop *et al.*, 2003).

Of note is the demonstration that dogs in South Africa may also be vaccinated using oral vaccines. Baits containing SAG2 attenuated vaccine, specifically formulated for dogs, was offered to 755 animals in a study in South Africa by Bishop, 1999 (unpublished results). A large number of dogs (75%) accepted these baits, illustrating the possible utility of this strategy for obtaining a sufficient vaccination coverage among the dog populations of KZN. Oral vaccines have however not as of yet been licensed for use in dogs in South Africa (Bishop *et al.*, 2003). A euthanasia campaign targeting the large number of free roaming unowned dogs was also introduced from 2003, in several municipalities throughout the province. This campaign has resulted in the direct and dramatic drop of rabies cases in the targeted regions (Bateman, 2005).

2.3 Materials and methods

2.3.1 Rabies virus isolates from KwaZulu Natal

Brain material from suspect rabies cases in KZN from the calendar year 2003 were submitted to the Allerton Veterinary Reference Laboratory, and were isolated at the Rabies Unit of the Onderstepoort Veterinary Institute (OVI). A further four human isolates from 2003, and one isolate from 2002, were obtained from, and isolated at the National Institute of Communicable Diseases (NICD). All samples were diagnosed as rabies positive using immunofluorescent staining techniques, and were stored at -20ºC in 50% glycerol-phosphate buffered saline (PBS) solution without further passage in mice.

Rabies cases are not evenly distributed throughout the KZN province, but are primarily localised in the more densely populated coastal regions. It was thus decided to use the magisterial district divisioning system in our determination of a representative virus sampling. We selected where possible, approximately half of rabies virus isolates from each magisterial district, from where rabies cases were reported from during the year 2003. Even though many of these isolates were likely to be very similar to each other or even identical, given their isolation over a short period of time, the strategy would ensure sufficient representation from all of the affected areas of the province. The location of these isolates on a map of the province was indicated by using the grid referencing system, as implemented on the isolate submission forms received from the Allerton Veterinary Reference Laboratory.

2.3.2 Preparation of RNA

Total RNA was extracted from infected brain material using the Trizol^{TM} reagent (Promega), based on the method originally developed by Sacchi and Chomzynski (1987). In short 50-100 ng of brain material was added to 1 ml Trizol, and homogenized using a syringe. The mixture was incubated for 5 minutes at room temperature (15-30ºC) in order to permit the dissociation of nucleoprotein complexes, after which 0.2 ml of chloroform was added in order to facilitate the separation between the organic and aqueous phases. After a further 3 minute incubation (15-30ºC), the mixture was centrifuged at 13 000 g for 10 minutes (15-30ºC), after which the RNA containing aqueous phase was transferred to a new Eppendorf tube. The RNA was subsequently precipitated for 10 minutes (15-30ºC)

using 0.5 ml of 100% isopropyl alcohol, and recovered by centrifugation at 13 000 g for 15 minutes (15-30ºC). The RNA pellet was washed with 1 ml of 75% ethanol, allowed to dry, and dissolved in between 30-50 µl nuclease free water (Eppendorf). Dilutions were carried out so that the final RNA concentration was approx. 1µg/ul, as estimated by UV absorbance at 260nm. The RNA preparations were then stored at –70ºC until further use.

2.3.3 Primer selection

The oligonucleotide primers which were used in this investigation were designed to anneal to conserved regions flanking the glycoprotein cytoplasmic domain and G-L intergenic region, based on a comparison between the Pasteur and divergent Mokola virus genomes (Sacramento *et al.*, 1991). The G (+) primer anneals to the glycoprotein gene upstream from the transmembrane domain at nucleotide position 4665 - 4687, while the L (-) primer anneals to the amino terminal region of the L gene at nucleotide position 5520 – 5543, relative to the published PV genome (Tordo *et al.*, 1986). A summary of these oligonucleotides are given in table 2.3., while a schematic presentations of the area of the genome amplified, as well as the primer annealing positions and orientations are supplied in figure 2.4. The primers were ordered from Integrated DNA Technologies (IDT), diluted to 10 pMol using nuclease free water (Eppendorf), and used to amplify a viral specific product (ca. 879 bp).

2.3.4 Reverse transcription

Between 0.5-3.9 µg (2.5µl) of total extracted RNA was denatured together with 40 pMol of G (+) primer at 70ºC for 5 minutes. The mixture was subsequently cooled on ice for 2 minutes followed by reverse transcription at 42ºC for 60 minutes in a 12.5 µl mixture, containing 0.5 µl (100 U) of Murine Moloney Leukemia Virus reverse transcriptase (M-MLV, USB[TM]), 2.5 µl RT buffer (1X = 50 mM Tris-HCL; pH 8.6; 75 mM KCl; 3 mM $MgCl_2$, 10mM DTT), 1.25 µl of a 10 mM mixture of all four deoxyribonucleotide triphosphates (dNTPs), and 0.5 µl (20 U) of RNasin[®] ribonuclease inhibitor (Promega).

2.3.5 Polymerase chain reaction (PCR)

PCR was carried out in a 50 µl reaction containing 1 µl of the undiluted RNA/cDNA hybrid; 1 µl of both the G (+) and L (-) primer (10 pMol); 5 µl Taq 10x buffer (50mM KCl, 10mM Tris-HCL pH 9.0; 0.1% Triton X-100); 3 µl MgCl$_2$ (25 mM), 1 µl of a 10 mM mixture of all four deoxyribonucleotide triphosphates (dNTPs) and 0.2 µl (5 U/µl) of *Taq* DNA polymerase (Promega). The PCR was performed in a programmable Perkin-Elmer thermocycler (GeneAmp PCR 2400) using a profile that involved an initial denaturation for 2 minutes at 94ºC, 30 cycles of 94ºC for 2 minutes; 45ºC for 90 seconds; 72ºC for 2 minutes and a final elongation step at 72ºC for 7 minutes, using conditions similar to those described by Sacramento *et al.*, 1991 and Sabeta *et al.*, 2003.

PCR products were visualized by agarose gel electrophoresis. A small amount of the PCR mixture (5 µl) was added to 1µl (5X) loading buffer (0.25% bromophenol blue; 40% sucrose) and electrophoresed at a 120V in a 1% ethidium bromide stained agarose gel in 1X SB buffer (10 mM NaOH adjusted to pH 8.5 with boric acid), using a Labnet power station 300 plus electrophoresis cell (Labnet International Inc.). Amplified DNA was subsequently visualized by UV fluorescence, and the size was estimated with a 100bp DNA weight marker (Promega).
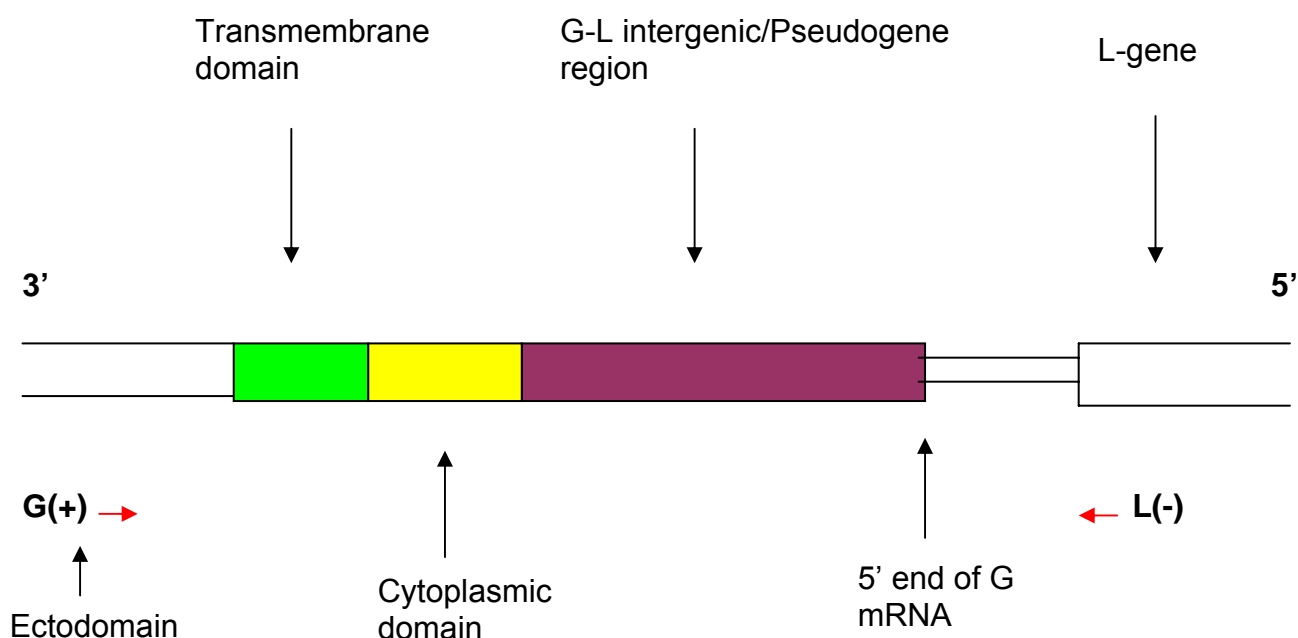

Figure 2.4. Schematic presentation of the annealing positions and orientations of the primers which were used to amplify the glycoprotein cytoplasmic domain and G-L intergenic region (Sacramento *et al.*, 1991).

Table 2.3. Annealing positions and nucleotide sequences of the primers utilized to amplify the glycoprotein cytoplasmic domain and G-L intergenic region. Primer annealing positions are given according to the numbering of the PV genome (Sacramento *et al.*, 1991, Tordo *et al.*, 1986).

Primer	Primer annealing position (nt.)	Nucleotide sequence of primer 5'-3'
G(+)	4665 – 4687	GACTTGGGTCTCCCGAACTGGGG
L(-)	5520 – 5543	CAAAGGAGAAGTTGAGATTGTAGTC

2.3.6 Purification of PCR products

Electrophoresed PCR products (20 µl in a double well) were excised from the agarose gel at the minimum UV exposure time, and purified using a Wizard® SV gel purification kit (Promega) according to the manufacturers instructions. In brief, 10 µl of membrane binding solution was added per 10 mg of excised gel and incubated at 50-65ºC for 10 minutes until the gel slice had completely melted. The DNA was bound to the silica membrane contained within the spin basket, by centrifuging the mixture through the column at 13000 g for 1 minute. The silica bound-DNA was subsequently washed three times with membrane wash solution, and eluted into 15 µl of nuclease free water (Eppendorf). The purified DNA product was then quantified (1 µl) by using agarose gel electrophoresis

2.3.7 Nucleotide sequencing

Nucleotide sequencing was performed by using automated fluorescent sequencing. In short, a 100 ng of purified PCR product (1-4 µl) was added to a reaction mixture containing 1 µl (10 pMol) G+ primer; 4 µl BigDye TerminatorReady reaction premix (PE Applied Biosystems, V 3.1); 2 µl BigDye sequencing buffer (5X) and nuclease free water (Eppendorf) to a final volume of 20 µl. The reaction was processed in a programmable Perkin-Elmer thermocycler (GeneAmp PCR 2400) using a profile that consisted of an initial denaturation step at 94ºC for 1 minute, 25 cycles of 96ºC for 10 seconds; 50ºC for 5 seconds and 60ºC for 4 minutes. After the reaction was completed, unincorporated ddNTPs were removed by precipitating the reaction in a 100µl 60% ethanol for 15

minutes. The extension products were recovered by centrifugation at 13 000 g for 30 minutes and washed with 250 µl 75% ethanol. After being allowed to dry on the bench for 15 minutes, the completed reaction was submitted to a commercial facility for resolution on an ABI 377 DNA sequencer (PE Applied Biosystems). A distance matrix for a pairwise comparison between all the KwaZulu Natal isolates and the PV was constructed by using the MEGA3 software. Distance correction as implemented by Kimura's two-parameter method was used to obtain the evolutionary distances between the sequences, or groups of sequences (Kimura, 1980; version 2.1, Kumar S., Tamura K., Jakobsen I. B and Nei M, 1993-2001; Sabeta *et al.*, 2003).

2.3.8 Phylogenetic analysis.

Phylogenetic analysis of the KZN RABV isolates was based on an alignment of a 591 nucleotide sequence which included the cytoplasmic domain of the glycoprotein gene, and the major part of the G-L intergenic region. It has previously been shown that similar tree topologies are obtained when using either of the two regions individually (Von Teichman, 1995), and it was thus decided to combine these regions during the analysis, as previously demonstrated by Sabeta *et al.*, 2003.

Various different combinations of isolates, as well as a number of sequences from the public domain, were included in the analysis, in order to provide phylogenetic reference points or clarity. Two phylogenetic trees were constructed in order to address the different aspects of the epidemic in the province. The first tree included viral sequences which were obtained from domestic dogs, humans, and incidental spillover hosts from KZN, as well as sequences which were obtained from dogs and wildlife (*C. mesomelas, O. megalotis*) from the other rabies endemic regions of South Africa and Zimbabwe. This analysis was conducted in order to demonstrate the evolutionary relationships which exist between viral isolates which were obtained from the province, and rabies viral isolates which were obtained from the different rabies affected geographical regions and host species present in southern Africa. The inclusion of dog and wildlife sequences from Zimbabwe, was furthermore also an essential step in creating the phylogeny used to reconstruct the spread of the virus, from the northern regions of South Africa and bordering countries, along the eastern coastal belt of southern Africa, southwards into KZN. This tree was further also used to demonstrate the relationship of a viral group (subfamily B), suspected of being maintained simultaneously in jackal and dogs in the

northern and northeastern regions of the KZN province, with wildlife isolates which were obtained from elsewhere in South Africa and Zimbabwe. Ideally this tree would have been constructed by the inclusion of additional sequence data from Mozambique, Swaziland and Lesotho. No studies describing the genetic variation of isolates from these regions have however been conducted to date, and consequently sequence data from viral variants circulating in these regions were not available for inclusion in this analysis.

The second tree included isolates from KZN only, and was constructed in order to describe the regional genetic variation that exists specifically between the viral isolates from this province. The description of the distribution of viral variants from the different magisterial districts of KZN, as well as the order in which these isolates clustered on the reconstructed tree, was required for elucidating the possible spread of rabies into and throughout the province, and further allowed us to make suggestions as to the factors that play a role in the maintance and persistence of the disease. This tree also allowed us to demonstrate the relationships which existed between a number of human cases which were diagnosed from the province during the time period between 2002-2003, and the variants that circulate in the regions of the province where these exposures had presumably occurred in. In this way were able to demonstrate the utility of the methodology which was employed during the course of this study, to the reconstruction of human case histories from the province.

In each case the obtained sequences were trimmed using the Bioedit sequence alignment editor (V 7.0, Tom Hall, Isis Pharmaceuticals, Inc 1997-2004), and alignments were carried out by using the ClustalW subroutine which is incorporated into this software. (Thompson *et al.*, 1994; V 7.0, Tom Hall, Isis Pharmaceuticals, Inc 1997-2004). The alignments used to construct the distance matrixes for the different trees, are presented in appendix A and B, respectively. Calculation of genetic distances and phylogenetic tree construction, was carried out by using the MEGA3 software (version 3.1, Kumar, Tamura and Nei 2004). Genetic distances were calculated between pairs of sequences by using the Kimura's two-parameter method (Kimura, 1980; Sabeta *et al.*, 2003), and based on these distances, a Neighbourhood-joining tree (NJ) was constructed using the methods of Saitou and Nei (1987). Positions with gaps were excluded from the analysis. The branching order of the tree was evaluated by using bootstrap analysis of a 1000 pseudoreplicate datasets, with a random seed generator number of 64238 (Swofford,

1993). Bootstrap values of more than 70% were generally regarded as providing evidence for a phylogenetic grouping.

Results were validated by maximum parsimony as implemented in MEGA3 (Li and Graur, 2000). Ambiguous sites were again deleted from the alignment, and tree space was searched for the most parsimonious trees by using the close neighbourhood interchange (CNI) search. The graphical output for the 50% majority rule consensus trees was obtained by using the MEGA3 tree explorer (version 3.1, Kumar, Tamura and Nei 2004). The trees which were obtained from the MP analysis are presented in appendix C and D, respectively. For the construction of the MP trees as presented in appendix C and D, a 123, and 44 parsimonious character sites could be identified from the sequence alignments, respectively.

2.4 Results

2.4.1, 2.4.2 Rabies virus isolates, RNA extraction, cDNA synthesis and PCR.

By using the magisterial divisioning system, a final panel of 128 virus isolates were selected for further processing, the details of which are supplied in table 2.4. and figure 2.9., respectively. RNA extractions, cDNA synthesis and PCRs were performed as described in section 2.3.2-2.3.5. The G-L primer set yielded a virus specific product approx. 850 nt. in size, for each of the isolates processed (results not shown).

2.4.3 Purification of PCR amplicons and nucleotide sequence determination

Purification of PCR amplicons and nucleotide sequencing was conducted as described in section 2.3.6 and 2.3.7 respectively. Sequence files were retrieved and manually inspected using the sequence editing option of the sequence alignment editor, BioEdit (V 7.0, Tom Hall, Isis Pharmaceuticals, Inc 1997-2004), in order to confirm that the obtained bases were accurately called, and to manually correct them where necessary. Nucleotide sequencing in general yielded approximately 600-700 bases of nucleotide sequence data for all of the isolates processed. Obtained sequences were trimmed using the Bioedit software to give a final sequence fragment of 591 nt. encompassing most of the G protein cytoplasmic domain and the major part of the G-L intergenic region. A distance matrix was constructed from a pairwise comparison of nucleotide sequences which were obtained

from the isolates from KZN as well as for the reference sequences (results not show). For the 591 bp region analyzed, the KZN canid isolates showed an intrinsic homology of 98.9%, indicating that these viruses were very closely related to each other. The KZN isolates differed on average by 18.1% from the corresponding sequence of the PV and by 29.6 % from a phylogenetically typical isolate of the mongoose biotype (m669.90). These values are similar to those calculated by Von Teichman *et al.*, 1995, which characterized the divergence between isolates of the canid biotype, the PV, and the South African mongoose biotype (Von Teichman *et al.*, 1995). In addition, the KZN canid viruses showed an average nucleotide sequence divergence of 3% and 3.5% from a phylogenetically typical jackal virus (j101) and bat eared fox isolate (o03.179), respectively.

Table 2.4. KwaZulu Natal rabies virus isolates analyzed during the course of this study.

Virus nr.	Allerton_ref nr.	Species	Country/province	Magisterial district	Year of isolation	Grid reference number (fig 2.9.)
1	03/83	Bovine/Cow	South Africa/KwaZulu Natal	Alfred	2003	G17
2	03/89	Canid/Dog	South Africa/KwaZulu Natal	Pongola	2003	L4
3	03/103	Bovine/Cow	South Africa/KwaZulu Natal	Pongola	2003	L4
4	03/105	Canid/Dog	South Africa/KwaZulu Natal	Eshowe	2003	L10
5	03/106	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	O9
6	03/114	Canid/Dog	South Africa/KwaZulu Natal	Mount Currie	2003	D17
7	03/120	Canid/Dog	South Africa/KwaZulu Natal	Alfred	2003	H17
8	02/121	Canid/Dog	South Africa/KwaZulu Natal	Alfred	2003	G17
9	03/127	Canid/Dog	South Africa/KwaZulu Natal	Ixopo	2003	G15
10	03/129	Canid/Dog	South Africa/KwaZulu Natal	Inanda	2003	K13
11	03/133	Canid/Dog	South Africa/KwaZulu Natal	Lower Tugela	2003	L12
12	03/137	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	P7
13	03/142	Canid/Dog	South Africa/KwaZulu Natal	Durban	2003	K14
14	03/147	Canid/Dog	South Africa/KwaZulu Natal	Umzinto	2003	I17
15	03/149	Canid/Dog	South Africa/KwaZulu Natal	Umvoti	2003	I11
16	03/152	Canid/Dog	South Africa/KwaZulu Natal	Ndwedwe	2003	J13
17	03/169	Canid/Dog	South Africa/KwaZulu Natal	Eshowe	2003	L10
18	03/170	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	K6
19	03/176	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	K6
20	03/180	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	P6
21	03/192	Canid/Dog	South Africa/KwaZulu Natal	Port Shepstone	2003	H18
22	03/194	Canid/Dog	South Africa/KwaZulu Natal	Lower Tugela	2003	K12
23	03/200	Canid/Dog	South Africa/KwaZulu Natal	Lower Tugela	2003	L12
24	03/204	Canid/Dog	South Africa/KwaZulu Natal	Lower Tugela	2003	L11

Table 2.4. KwaZulu Natal rabies virus isolates analyzed during the course of this study.

Virus nr.	Allerton ref nr.	Species	Country/province	Magisterial district	Year of isolation	Grid reference number (figure 2.9.).
25	03/205	Canid/Dog	South Africa/KwaZulu Natal	Pietemaritzburg	2003	H13
26	03/209	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
27	03/213	Canid/Dog	South Africa/KwaZulu Natal	Port Shepstone	2003	H17
28	03/214	Canid/Dog	South Africa/KwaZulu Natal	Pongola	2003	L4
29	03/215	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	O7
30	03/225	Canid/Dog	South Africa/KwaZulu Natal	Umkomazi	2003	E16
31	03/230	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	J6
32	03/235	Canid/Dog	South Africa/KwaZulu Natal	Inanda	2003	J13
33	03/236	Canid/Dog	South Africa/KwaZulu Natal	Alfred	2003	F17
34	03/237	Canid/Dog	South Africa/KwaZulu Natal	Ndwedwe	2003	J13
35	03/241	Canid/Dog	South Africa/KwaZulu Natal	Mhlabatini	2003	L8
36	03/247	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	J6
37	03/251	Canid/Dog	South Africa/KwaZulu Natal	Ndwedwe	2003	J13
38	03/253	Canid/Dog	South Africa/KwaZulu Natal	Vyheid	2003	J6
39	03/254	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
40	03/256	Canid/Dog	South Africa/KwaZulu Natal	Pinetown	2003	J14
41	03/263	Canid/Dog	South Africa/KwaZulu Natal	Inanda	2003	J13
42	03/264	Canid/Dog	South Africa/KwaZulu Natal	Ngotshe	2003	L5
44	03/269	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	J5
45	03/270	Canid/Dog	South Africa/KwaZulu Natal	Eshowe	2003	M10
46	03/276	Canid/Dog	South Africa/KwaZulu Natal	Nwedwe	2003	J13
47	03/290	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	J6
48	03/292	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	O7
49	03/293	Canid/Dog	South Africa/KwaZulu Natal	Durban	2003	J14
50	03/299	Canid/Dog	South Africa/KwaZulu Natal	Durban	2003	K14
51	03/302	Canid/Dog	South Africa/KwaZulu Natal	Eshowe	2003	L10
52	03/306	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
53	03/307	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	O7
54	03/308	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
55	03/309	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
56	03/314	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	O8
57	03/315	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	J6
58	03/316	Canid/Dog	South Africa/KwaZulu Natal	Nkandhla	2003	J9
59	03/321	Canid/Dog	South Africa/KwaZulu Natal	Ngotshe	2003	L5
60	03/326	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
61	03/328	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	M4
62	03/330	Caprine/Goat	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
63	03/335	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	J5
64	03/336	Canid/Dog	South Africa/KwaZulu Natal	?	2003	?

Table 2.4. KwaZulu Natal rabies virus isolates analyzed during the course of this study.

Virus nr.	Allerton ref nr.	Species	Country/province	Magisterial district	Year of isolation	Grid reference number (figure 2.9.)
65	03/340	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
66	03/343	Canid/Dog	South Africa/KwaZulu Natal	Mhlabatini	2003	L7
67	03/358	Caprine/Goat	South Africa/KwaZulu Natal	Mhlabatini	2003	M7
68	03/359	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	O7
69	03/360	Canid/Dog	South Africa/KwaZulu Natal	Pongola	2003	M4
70	03/364	Caprine/Goat	South Africa/KwaZulu Natal	Pongola	2003	L4
71	03/366	Canid/Dog	South Africa/KwaZulu Natal	Port Shepstone	2003	H18
72	03/368	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	O9
73	03/375	Canid/Dog	South Africa/KwaZulu Natal	Eshowe	2003	J6
74	03/378	Canid/Dog	South Africa/KwaZulu Natal	Port Shepstone	2003	H17
75	03/382	Canid/Dog	South Africa/KwaZulu Natal	Ixopo	2003	G15
76	03/387	Canid/Dog	South Africa/KwaZulu Natal	Durban	2003	J14
77	03/389	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	J6
78	03/391	Canid/Dog	South Africa/KwaZulu Natal	Umzinto	2003	I16
79	03/399	Canid/Dog	South Africa/KwaZulu Natal	Umzinto	2003	I16
80	03/400	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	O7
81	03/404	Canid/Dog	South Africa/KwaZulu Natal	Ixopo	2003	G15
82	03/406	Canid/Dog	South Africa/KwaZulu Natal	Inanda	2003	K13
83	03/407	Canid/Dog	South Africa/KwaZulu Natal	Ubombo	2003	O5
84	03/409	Canid/Dog	South Africa/KwaZulu Natal	Mount Currie	2003	E17
85	03/410	Canid/Dog	South Africa/KwaZulu Natal	Eshowe	2003	L10
86	03/411	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	J6
87	03/417	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	O7
88	03/418	Canid/Dog	South Africa/KwaZulu Natal	Port Shepstone	2003	H17
89	03/425	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	L6
90	03/430	Canid/Dog	South Africa/KwaZulu Natal	Nongoma	2003	N6
91	03/431	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	O7
92	03/433	Ovine/Sheep	South Africa/KwaZulu Natal	Vryheid	2003	J6
93	03/437	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	O9
94	03/453	Canid/Dog	South Africa/KwaZulu Natal	Durban	2003	J14
95	03/454	Canid/Dog	South Africa/KwaZulu Natal	Nongoma	2003	M5
96	03/455	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	K7
97	03/461	Canid/Dog	South Africa/KwaZulu Natal	Lower Tugela	2003	K12
98	03/463	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
99	03/467	Bovine/Cow	South Africa/KwaZulu Natal	Ubombo	2003	O5
100	03/470	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
101	03/474	Canid/Dog	South Africa/KwaZulu Natal	Mount Currie	2003	D17
102	03/475	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
103	03/478	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
104	03/485	Canid/Dog	South Africa/KwaZulu Natal	Mapumulo	2003	K11

Table 2.4. KwaZulu Natal rabies virus isolates analyzed during the course of this study.

Virus nr.	Allerton ref nr.	Species	Country/province	Magisterial district	Year of isolation	Grid reference number (figure 2.9.)
105	03/491	Canid/Dog	South Africa/KwaZulu Natal	Port Shepstone	2003	H17
106	03/492	Canid/Dog	South Africa/KwaZulu Natal	Port Shepstone	2003	H17
107	03/494	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	O9
108	03/502	Canid/Dog	South Africa/KwaZulu Natal	Port Shepstone	2003	H16
109	03/503	Canid/Dog	South Africa/KwaZulu Natal	Ingwavuma	2003	Q3
110	03/507	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
111	03/509	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
112	03/510	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
113	03/513	Canid/Dog	South Africa/KwaZulu Natal	Eshowe	2003	L10
114	03/514	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
115	03/518	Canid/Dog	South Africa/KwaZulu Natal	Ubombo	2003	P6
116	03/568	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
117	03/583	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	J6
118	03/588	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	P7
119	03/589	Canid/Dog	South Africa/KwaZulu Natal	Vryheid	2003	K7
120	03/594	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	N9
121	03/620	Bovine/Cow	South Africa/KwaZulu Natal	Port Shepstone	2003	H17
122	03/621	Canid/Dog	South Africa/KwaZulu Natal	Hlabisa	2003	O8
123	03/672	Canid/Dog	South Africa/KwaZulu Natal	Lower Umfolozi	2003	O10
124	SPU/15.03	Human	South Africa/KwaZulu Natal	Lower Umfolozi	2003	*
125	SPU/48.03	Human	South Africa/Eastern Cape	Cofimvaba	2003	*
126	SPU/77.03	Human	South Africa/KwaZulu Natal	Port Shepstone	2003	*
127	SPU/272.03	Human	South Africa/KwaZulu Natal	Lower Umfolozi	2003	*
128	SPU/326.02	Human	South Africa/KwaZulu Natal	Port Shepstone	2002	*

2.4.4 Phylogenetic analysis

Phylogenetic trees were constructed as discussed in section 2.3.8. Trees constructed using the neighbourhood-joining (NJ) method largely corresponded to trees constructed by maximum parsimony (MP), with the majority of the principal groups being maintained throughout each analysis (see appendix C and D). The differences in topology which was observed for some of the reconstructed viral groups between the NJ and MP trees could possibly be ascribed to the inefficiency of the CNI heuristic search algorithm which was used to search tree space for the most parsimonious trees. This method does not examine every possible tree topology that can be reconstructed from the given dataset, and as such, the reconstructed tree may not necessarily represent the most parsimonious tree. A hierarchal system was used to name the identified monophyletic groups (OTUs sharing a common ancestor) with family, subfamily, cluster, subcluster and group

designations generally referring to successively smaller viral groupings [(* system applied differently in section 2.4.4 (A) and (B)]. The low bootstrap support values which were obtained from some of these reconstructed viral groupings during the phylogenetic analyses could possibly be ascribed to the low genetic divergence which was present between the analyzed sequence data, and probably reflects the low number of phylogenetically informative sites which were available for the bootstrap analysis from the sequence alignment.

(A) **Phylogeny of KZN rabies virus isolates, compared to canine rabies virus isolates from elsewhere in South Afica and Zimbabwe.**

In order to determine the relationship between the isolates from KZN and isolates from rabies endemic regions from elsewhere in Southern Africa and Zimbabwe, a neighbourhood-joining (NJ) tree was constructed based on an alignment of a 543 nt. sequence encompassing the glycoprotein cytoplasmic and G-L intergenic sequences for a total of 58 taxa. The analysis included selected isolates from dogs, humans and domestic livestock from KwaZulu Natal, as well as wildlife sequences which were obtained from the different rabies endemic regions of South Africa and Zimbabwe. The geographic regions (magisterial districts, towns etc.) from which the reference sequences were obtained from are presented in figure 2.5 and 2.6 respectively. Isolates from KZN were selected to represent the different viral groupings as identified in section 2.4.4 (B), while the wildlife isolates which were included in the analysis, were selected to represent the different rabies virus cycles, as identified throughout Southern Africa and Zimbabwe by Sabeta *et al.*, 2003. This study comprises one of the most extensive descriptions of canid biotype variants from throughout Southern Africa and Zimbabwe, with respect to the gene sequences utilized during the course of this study, to date. The NJ phylogenetic tree is presented in figure 2.7., while the epidemiological data of the analyzed sequences are provided in table 2.4. and 2.6., respectively. The sequence alignment used for the construction of the distance matrix is presented in appendix A. The multiple alignment of viral sequences from KwaZulu Natal and comparison of the aligned sequences to the complete genome of the PV strain, indicated that the KZN viruses contained a deletions in the first of the two TTP sites postulated for the G gene of the PV strain. This deletion has previously been demonstrated in street viruses from Europe, Zimbabwe and South Africa. (Sacramento *et al.*, 1992; Sabeta *et al.*, 2003.

From the tree it was evident that viral isolates from South African canid hosts (canid biotype) were distinguishable from the mongoose biotype (m669.90), as well as from PV for the 591 nt. region analysed. The separation between m669.90, the PV sequence, and isolates of the South African canid biotype was supported by a bootstrap value of a 100%. A further division was evident between the viruses of the canid biotype from Southern Africa which was supported by a bootstrap value of 99% (cluster 1 and 2). Cluster 1 could further be divided into two subclusters 1-1 and 1-2. The bootstrap value for the node segregating these subclusters was not significant (<50%), however they were considered to represent phylogenetically separable groups, since the tree topology separating their division was reconstructed when using the MP algorithm. Subcluster 1-1 could be further subdivided into two groups (1-1-1, 1-1-2), based on a bootstrap value of 69%. The first group (1-1-1) contained the sequences which were obtained from human, dogs, and domestic livestock isolates from KZN, as well as two jackal isolates, both of which were isolated from the Limpopo province of South Africa. Group 1-1-2 on the other hand consisted of dog isolates which were obtained from the northern and northeastern regions of KZN, as well as dog and jackal isolates which were obtained from the eastern Mpumalanga region bordering western Swaziland. Subcluster 1-2 could be divided into three groups (1-2-1, 1-2-2, 1-2-3) based on bootstrap values above 90%. Group 1-2-1 was isolated from dogs from the northeastern districts of Zimbabwe, while group 1-2-2 represented cycles circulating among dog and jackal in the central, northern and eastern districts of Zimbabwe. The remaining group (1-2-3) consisted of dog isolates which were made from southern Zimbabwe.

Cluster 2 could be divided into three subclusters (2-1, 2-2, 2-3), based on moderate, to high bootstrap values (58-86%). Subcluster 2-1 consisted of isolates which were made from dogs and jackal in the central region of Zimbabwe as well as from the Limpopo province of South Africa, while subcluster 2-2 consisted of isolates which were made from jackal from the northern regions of South Africa and south western Zimbabwe. Subcluster 2-3 on the other hand consisted of isolates which were made from bat eared foxes from the western regions of South Africa.

The intrinsic divergence between viral isolates belonging to subcluster 1-1, subcluster 1-2 and cluster 2 was equal to 1.6%, 3.1% and 3% percent, respectively. The divergence between isolates from subcluster 1-1 and 1-2, and subcluster 1-1 and cluster 2 on the other hand was equal to 4.3% and 3.9%. The higher percentage intrinsic divergence

observed for viral isolates belonging to subcluster 1-2 and 2, when compared to isolates from subcluster 1-1, could possibly reflect the fact that viral isolates belonging to these groups were isolated from different host species, over a broad geographical region, and over different time periods. The low variation present among the KZN isolates could in comparison however, be indicative of the relatively recent emergence of these cycles in the province. The principal groups which could be identified from the phylogeny, could be distinguished from each other by specific nucleotide substitutions in the sequence alignment as presented in table 2.5. (also see appendix A).

Table 2.5. Nucleotide substitutions responsible for the clustering of the major groups identified from the phylogeny in section 2.4.4 (A)

Nucleotide position	Viral groups	Substitutions
13	Cluster 1 and cluster 2	G - A
328	Subcluster 1-1, subcluster 1-2 and cluster 2	G - A - A
418	Subcluster 1-1, subcluster 1-2 and cluster 2	G - C - C

Figure 2.7. (page 126) Neighbourhood joining tree constructed from selected isolates from KZN, as well as isolates from other canine endemic regions in South Africa and Zimbabwe. Sequences from KwaZulu Natal are preceded by a prefix KZN, while host species of isolation are indicated by the following prefixes (dg, dog, cw, cow, gt, goat, shp, sheep, hm, human, o, *O. megalotis*, j, *C. mesomelas*, m, *C. penicillata*). Horizontal lines are proportional to the evolutionary distances between sequences, as given by the scale bar in nucleotide substitutions per site, while vertical lines have no meaning and are provided for purposes of clarity only. A mongoose isolate m669.90 was used as an outgroup to root the tree.

Table 2.6. Reference sequences used to conduct the analysis in 2.4.4 (A) (Tordo *et al.*,1986, Von Teichman, 1995; Sabeta *et al.*, 2003).

Nr.	Lab ref. nr.	Species	Country	Year of isolation	Locality	Coordinates (long-lat)	Genbank Acc. Nr.
1	dgA95.755	Dog/ *Canis familiaris*	South Africa	1995	Amanzintoti	30°54'-30°04'	AF303081
2	dg90.57	Dog/ *Canis familiaris*	South Africa	1990	Durban	30°53'-30°03'	AF177101
3	j101.90	Black backed jackal/ *C. mesomelas*	South Africa	1990	Soutpansberg	28°55'-22°49'	AF079902
4	j378.90	Black backed jackal/ *C. mesomelas*	South Africa	1990	Potgietersrus	28°31'-23°10'	AF177106
5	dg373.97	Dog/ *Canis familiaris*	South Africa	1997	Baberton	31°48'-25°42'	AF303069
6	dg224.98	Dog/ *Canis familiaris*	South Africa	1998	Ermelo	29°59'-26°31'	AF177098
7	j596.99	Black backed jackal/ *C. mesomelas*	South Africa	1999	Piet Retief	31°12'-29°33'	AF303063
8	d21057	Dog/ *Canis familiaris*	Zimbabwe	1992	Muzarabani	31°12'-16°19'	AF177064
9	d21869	Dog/ *Canis familiaris*	Zimbabwe	1993	Nyakasoro, Pfungwe	32°15'-16°49'	AF177069
10	d22547	Dog/ *Canis familiaris*	Zimbabwe	1994	Kumutsenzere, Masoso	31°47'-16°22'	AF177070
11	d19385	Dog/ *Canis familiaris*	Zimbabwe	1991	Zaka	31°34'-20°12'	AF303080
12	j19273	Black backed jackal/ *C. mesomelas*	Zimbabwe	1991	Chiredzi	31°58'-20°54'	AF303076
13	d24505	Dog/ *Canis familiaris*	Zimbabwe	1996	Gutu	31°10'-19°37'	AF177075
14	d16387	Dog/ *Canis familiaris*	Zimbabwe	1986	Zhombe	29°22'-18°41'	AF177057
15	d21428	Dog/ *Canis familiaris*	Zimbabwe	1993	Chipinge	32°43'-20°22'	AF177065
16	j23275	Side striped jackal/ *C. adustus*	Zimbabwe	1995	Bindura	31°21'-16°56'	AF177089
17	d24465	Dog/ *Canis familiaris*	Zimbabwe	1996	Wendza	31°43'-18°52'	AF177074

Table 2.6.Reference sequences used to conduct the analysis in 2.4.4 (A) (Tordo *et al.*,1986, Von Teichman, 1995; Sab eta *et al.*, 2003).

Nr.	Lab ref. nr.	Species	Country	Year of isolation	Locality	Coordinates (long- lat)	Genbank Acc. Nr.
18	j22642	Side striped jackal/ *C. adustus*	Zimbabwe	1994	Glendale	31°06'-17°19'	AF177088
19	j19301	Side striped jackal/ *C. adustus*	Zimbabwe	1991	Chegutu	30°14'-18°09'	AF177081
20	j19319	Side striped jackal/ *C. adustus*	Zimbabwe	1991	Selous	30°28'-18°09'	AF177082
21	j17711	Black backed jackal/ *C. mesomelas*	Zimbabwe	1988	Turk Mine	30°49'-19°27'	AF177087
22	j306.98	Black backed jackal/ *C. mesomelas*	South Africa	1998	Warmbaths	28°07'-24°51'	AF177105
23	j673.99	Black backed jackal/ *C. mesomelas*	South Africa	1999	Potgietersrus	28°36'-22°43'	AF303061
24	j717.99	Black backed jackal/ *C. mesomelas*	South Africa	1999	Pietersburg	29°29'-23°42'	AF303064
25	j23374	Black backed jackal/ *C. mesomelas*	Zimbabwe	1995	Bulawayo	28°47'-20°15'	AF177091
26	j631	Black backed jackal/ *C. mesomelas*	South Africa	1999	Soutpansberg	30°05'-22°03'	AF303060
27	o377.99	Bat eared fox/ *O. megalotis*	South Africa	1999	Gordonia	30°28'-18°09'	AF177119
28	o414.96	Bat eared fox/ *O. megalotis*	South Africa	1996	Beaufort West	22°47'-32°22'	AF177112
29	o469.99	Bat eared fox/ *O. megalotis*	South Africa	1999	Kimberley	24°27'-28°47'	AF303073
30	o491.98	Bat eared fox/ *O. megalotis*	South Africa	1998	Petrusburg	25°29'-29°23'	AF303059
31	o578.95	Bat eared fox/ *O. megalotis*	South Africa	1995	Strydenburg/ Hopetown	23°46'-29°55'	AF177113
32	o774.95	Bat eared fox/ *O. megalotis*	South Africa	1995	Carnarvon	21°56'-31°13'	AF177115
33	m669.99	Mongoose/ *C. penicillata*		1999	Ermelo/ Mpumalanga	29°52'-26º42'	*
34	PV	*	*	*	*	*	*

100

Figure 2.5. Approximate geographic region of isolation for the reference sequences which were obtained from South Africa (Sabeta et al., 2003)
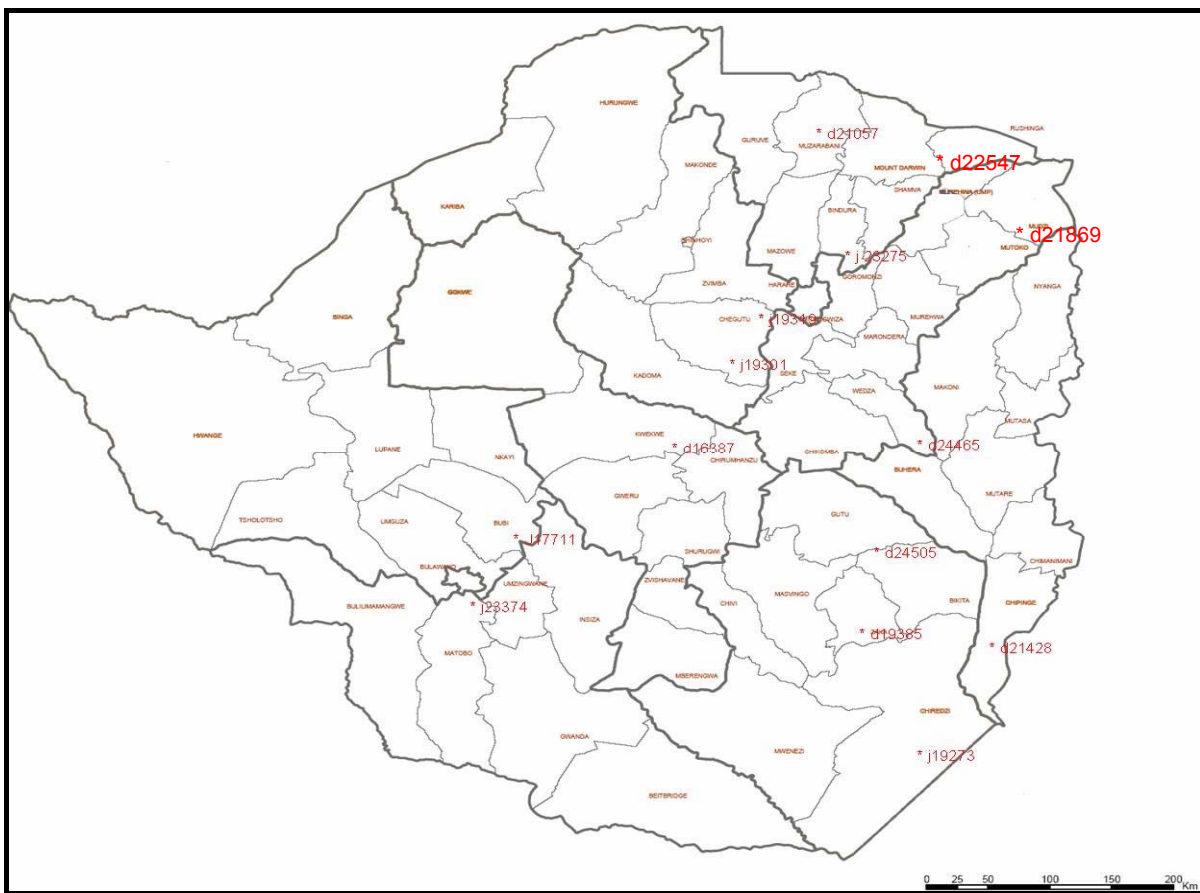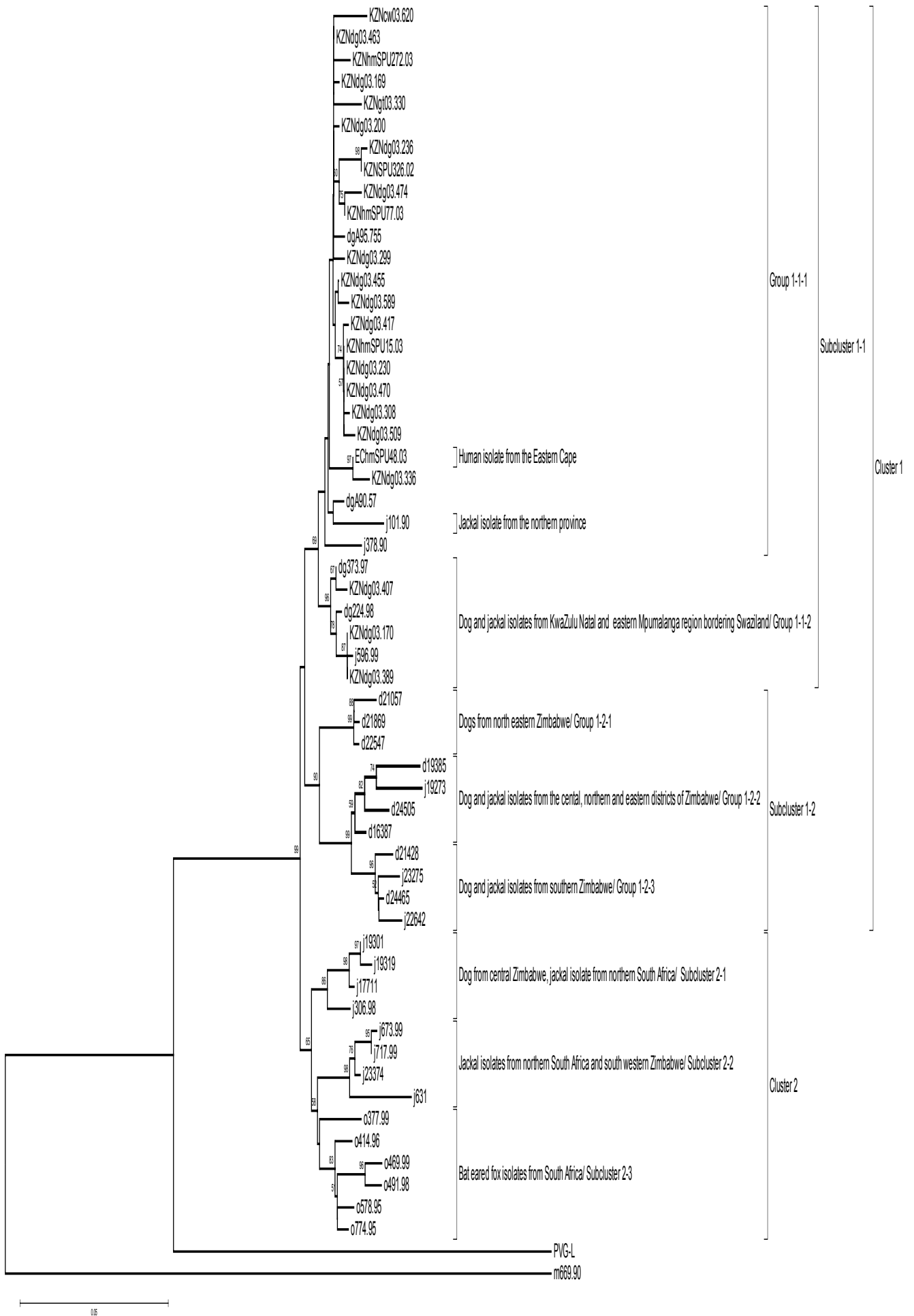

Figure 2.6. Approximate geographic region of isolation for the reference sequences which were obtained from Zimbabwe (Sabeta et al., 2003)

101

(B) Phylogeny of rabies virus isolates from KZN

In order to describe the regional genetic variation of rabies viruses from KwaZulu Natal, a NJ tree was constructed based on an alignment of the 128 isolates sequenced during the course of this project (see figure 2.8 and 2.9). The tree was based on the alignment of a 591 nt. sequence encompassing the cytoplasmic domain of the glycoprotein as well as a major part of the G-L intergenic region. The details of the reference and KZN isolate sequences are supplied in table 2.4. and 2.6. respectively, while the alignment used in the construction of the distance matrix is presented in appendix B. In order to increase the resolution of the NJ phylogenetic tree, it was rooted with a bat eared fox isolate (o03.179), which has previously been shown to be distinguishable from the dog isolates from KZN and the cosmopolitan PV, but which nevertheless falls within the radius defined by the South African canid biotype cluster.

Phylogenetic analysis indicated that the KZN family could be sub-divided into two subfamilies (subfamily A and B) based on a bootstrap support value of a 100%. Subfamily A was composed of 109 isolates and was by far the biggest, appearing to be the principal group responsible for the rabies epidemic in the province. Isolates belonging to this subfamily were primarily isolated from the coastal regions of the province, from as far north as the region bordering Mozambique, to as far south as the Eastern Cape border. A number of clusters could be identified within subfamily A, many of which contained viruses that were found to be a 100% homologous with respect to their G-L intergenic sequences.

Subfamily B was composed of 21 isolates, which were separable into two subclusters (bootstrap support value of 88%), both of which were isolated from the northern and northeastern regions of the province bordering southern Swaziland and Mozambique. A low genetic divergence was also evident within this subfamily, with a large number of isolates belonging to each of the subclusters showing identical sequences with respects to the nucleotide sequences analysed. Subfamily A showed a mean divergence of 3.6% from the outgroup o03.179, while the divergence between subfamily B and the outgroup was equal to 3.3%. Subfamily A and B in turn, showed a mean divergence of 1.9% from each other (Calculated using Mega3). The principal groups which could be identified from the phylogeny, could be distinguished from each other by specific nucleotide substitutions in the sequence alignment as presented in table 2.7. (also see appendix B).

Table 2.7. Nucleotide substitutions responsible for the clustering of the major groupings identified from the phylogeny in section 2.4.4 (B).

Nucleotide position	Viral groups	Substitutions
98	Subfamily A and B	T – A
369	Subfamily A and B	A – G
374	Subfamily A and B	T – C
439	Subfamily A and B	A – G
444	Subfamily A and B	T – C
459	Subfamily A and B	T,G – A
516	Subfamily A and B	T – C

Phylogenetic and geographical divisions present between viral isolates of subfamily A

Subfamily A could be divided into three groups (A, B, C) based on bootstrap support values of >50%. These groups consisted of cluster 1A-5A, 6A–7A and cluster 8A, respectively. The division between group A and B was supported by a low bootstrap value of 55%, while the division between group A and B, and group C was supported by a high bootstrap value of 77%. Group A consisted of 80 taxa isolated from the region surrounding Durban, northwards, up till the border region between the KZN province and Mozambique (North Coast). A few isolations from this group were also made from the internal northern and northeastern regions of the KZN province, south of the Swaziland border. Group B contained 22 taxa which were obtained from the regions south of Durban up to the southern border between KZN and the Eastern Cape (South Coast). Group C on the other hand consisted of 3 isolates which were obtained from the far southern regions of the KZN province, bordering the Eastern Cape. This cluster also contained a single human isolate which was obtained from the Eastern Cape district of Cofimvaba. Group A and B were more closely related to each other, than to isolates of group C, as is evident from the tree topology, as well as the bootstrap values for the nodes segregating these groups from each other. The percentage divergence between group A and B was equal to 1.3%, while the divergence between groups A and C, and ,B and C was equal to 1.3% and 2%, respectively.

Phylogenetic and geographical divisions present between the viral isolates of group A in subfamily A.

As previously mentioned group A could be divided into 5 clusters (1A-5A). Cluster 1A consisted of 37 isolates, of which 36 were obtained from domestic dogs, and of which one, was obtained from a human patient (KZNhmSPU15.03) who had become infected in the Lower Umfolozi district. Virus isolates belonging to this cluster were isolated from magisterial districts including Hlabisa, Ubomobo, Mtonlaneni, Mhlabatini, Nongoma and Nongotshe.

Cluster 2A was closely related to viruses of cluster 1A and consisted of 4 isolates which were obtained from 3 dogs and 1 sheep from the magisterial district of Vryheid. The bootstrap support value separating cluster 1A and 2A was below 50%, however these clusters were considered to represent phylogenetically separable groups due to localized distribution of these groups in their respective regions of isolation. The topology illustrating the division between these clusters could however not be reconstructed when using the MP method (see appendix D).

Cluster 1A and 2A were closely related to a third cluster of viruses (3A) consisting of 35 isolates which were obtained from the coastal regions surrounding Richards Bay southwards up till the magisterial district of Durban. Isolates (32) in this cluster were obtained from magisterial district including the Lower Umfolozi, Eshowe, Mtunzini, Lower Tugela, Ndwedwe, Inanda, Pinetown, Durban, Msinga and Chatsworth. A further three isolations were made from the internal magisterial districts of Pietermaritzburg and Umvoti, as well as the southern district of Port Shepstone. A total of thirty-four isolates belonging to this cluster were made from domestic dogs, while a single isolate was obtained from a human patient (KZNhmSPU272.03) who had become infected with this variant in the magisterial district of the Lower Umfolozi.

Cluster 4A was represented by a single isolate, made from a cow from the Port Shepstone region, and may be a representative of a group of viruses that may be circulating in dogs in this area. Cluster 5A on the other hand, consisted of 3 isolates which were made from the Pongola district, just south of the Swaziland border as well as a single isolate which was obtained from the Lower Umfolozi. Two of the isolates belonging to this cluster were made from domestic dogs, while the remaining two isolates were made from goats. Since

isolations from cluster 5A were made in a region of the KZN province bordering southern Swaziland, it was thought that these isolates might be representative of a dog rabies cycle that is currently circulating in Swaziland itself, and that had spread southwards across the border into northern KZN. The bootstrap values for the nodes segregating cluster 4A and 5A from each other, and from viral isolates in cluster 1A-3A were not significant (<50%) and cluster 4A-5A were thus considered to fall together with cluster 1A-3A to form Group A.

Phylogenetic and geographical divisions present between viral isolates in group B in subfamily A.

Group B consisted of two clusters (6A-7A) whose division was supported by a low bootstrap value of 54%. Although the bootstrap value supporting the separation between these clusters was low, they were considered to represent phylogenetically distinguishable groups because this division could also be demonstrated when reconstructing the tree with the use of the MP method (appendix D). Cluster 6A consisted of 7 isolates which were obtained from the magisterial districts of Port Shepstone, Alfred and Kokstad, while cluster 7A was a bit more widely distributed, and consisted of 15 isolates which were obtained from the magisterial districts of Durban, Umzinto, Port Shepstone, Ixopo, Alfred and Umkomazi. Five of the isolates in cluster 6A were obtained from dogs, while the remaining two isolates were obtained from a cow, and a human patient (KZNhmSPU77.03), who had become infected in the magisterial district of Port Shepstone. Of the total number of isolates (15) in cluster 7A, 14 isolations were made from domestic dogs, while a single isolate was obtained from a human patient (KZNhm326.02), who had become infected in the magisterial district of Port Shepstone.

Phylogenetic and geographical divisions present between viral isolates in group C in subfamily A.

Group C consisted of a single viral cluster (8A). This cluster contained four isolates, three of which were obtained from dogs in the magisterial districts of Port Shepstone and Mount Currie, and a single isolate (KZNhmSPU48.03) which was obtained from a human, who had become infected in 2003 in the Eastern Cape district of Cofimvaba. Since the three dog isolates from cluster 8A was obtained from a region of KZN bordering the Eastern Cape, and since the human exposure apparently occurred in the Eastern Cape itself, it

was thought that the 4 isolates in cluster 8A were representative of a cycle circulating in dogs in the northern regions of the Eastern Cape, which had recently spread northwards across the southern border of KZN.

Phylogenetic and geographical divisions present between viral isolates belonging to subfamily B

Subfamily B consisted of two viral clusters, cluster 1B and 2B, that could be separated from each other based on a bootstrap value of 88%. Cluster 1B had a wide distribution throughout the northern districts of KZN, and consisted of 9 isolates which were obtained from the magisterial districts of Pongola, Ngotshe, Ubombo and Ingwavuma. Eight of the isolates were made from domestic dogs, while a single isolate was made from a cow. Cluster 2B on the other hand, consisted of 12 isolates, that had a narrower distribution when compared to cluster 1B, being confined to the districts of Vryheid and Ngotshe. All of the isolates from cluster 2B were made from domestic dogs and shared a 100% nucleotide sequence homology with respect to the gene regions analyzed.

Phylogenetic and geographical subdivisions present between the identified viral groups from KZN.

Within a large number of the cluster divisions present on the phylogenetic tree, a number of subdivisions or subclusters could be identified. Typically these subclusters consisted of isolates that were highly related, and which were obtained from geographically proximate loci. These subclusters were thus thought to correspond to highly localised outbreaks of the disease. The epidemiological and phylogenetical data of the identified viral groups, as well as the symbols and colours used to denote them on the phylogenetic tree and map in figure 2.8. and 2.9., are presented in table 2.8.

Table 2.8. Viral groups identified from KZN, their epidemiological information, and the symbols used to denote them on the phylogenetic tree and map in figure 2.8. and 2.9. respectively.

| Subfamily	Group	Cluster	Subcluster	Bootstrap support value (%)	Percentage divergence between viral isolates in group	Region/District of isolation	Host species of isolation	Map symbol
A	A	1A	-	76	0.2%	North coast, from Richards bay northwards till the Ubombo district. Few isolates also made from the internal northeastern regions of the province (details discussed in text)	Dog (35), goat (1), human (1)	Green
A	A	1A	1	<50	0%	Lower Umfolozi, Eshowe	Dog (4)	□
A	A	1A	2	<50	0.3%	Lower Umfolozi	Dog (2)	▽
A	A	1A	3	53	0.3%	Hlabisa, Lower Umfolozi	Dog (3)	■
A	A	1A	4	<50	0.3%	Hlabisa	Dog (2)	▲
A	A	1A	5	64	0%	Vryheid, Ngotshe	Dog (2)	△
A	A	1A	6	64	0%	Lower Umfolozi	Dog (2)	○
A	A	2A	-	52	0.2%	Vryheid	Dog (3), sheep (1)	Orange
A	A	3A	-	63	0.3%	North Coast, from Richards bay southwards up till Durban (details discussed in text).	Dog (34), human (1)	Purple
A	A	3A	1	64	0%	Eshowe	Dog (2)	■
A	A	3A	2	95	0.1%	Mtunzini, Umvoti, Pietermaritzburg	Dog (3), human (1)	□
A	A	3A	3	65	0%	Durban and surrounding districts, also isolated from Ndwedwe and Port Shepstone	Dog (8)	▲
A	A	4A	-	<50	-	Port Shepstone	Cow (1)	Dark yellow

108

Table 2.8. Viral groups identified from KZN, their epidemiological information, and the symbols used to denote them on the phylogenetic tree and map in figure 2.8. and 2.9., respectively.

Subfamily	Group	Cluster	Subcluster	Bootstrap support value (%)	Percentage divergence between viral isolates in group	Region/District of isolation	Host species of isolation	Map symbol
A	A	5A	-	97	0.2%	Pongola, Lower Umfolozi	Dog (2), goat (1)	Light purple
A	B	6A	-	55	0.4%	Port Shepstone, Alfred, Mount Currie	Dog (5), cow (1), human (1)	Light blue
A	B	7A	-	97	0.2%	Southwards from Durban, till the southern border of the Eastern Cape (details discussed in text)	Dog (14), human (1)	Pink
A	B	7A	1	66	0%	Ixopo, Umkomazi, Alfred	Dog (3)	▲
A	C	8A	-	89	0.3%	Southern most districts of KZN, bordering the Eastern Cape (details supplied in text)	Dog (3), human (1)	Brown
A	C	8A	1	-	-	Eastern Cape, Cofimvaba	Human (1)	●
A	C	8A	2	89	0.3%	Port Shepstone, Mount Currie	Dog (2)	■
B	-	1B	-	89	0.1%	Northeastern districts of KZN northwards, up till the border between Mozambique, Swaziland and KZN	Dog (8), cow (1)	Dark blue
B	-	1B	1	64	0%	Ubombo, Ingwavuma	Dog (1), goat (1)	■
B	-	2B	-	96	0%	Vryheid	Dog (12)	Red

Figure 2.8 (page 135). A neighbourhood-joining tree constructed from an alignment of 128 nucleotide sequences, encompassing the cytoplasmic domain of the glycoprotein and the G-L intergenic region of canid rabies viruses from KZN only (see table 2.4). Principal bootstrap values, as discussed in the text, are indicated on the relevant branches. Horizontal branch lengths are proportional to the similarity of the sequences within and between clusters, while the scale bar at the bottom of the figure indicates the evolutionary distances between taxa, as calibrated according to the number of nucleotide substitutions per site. Vertical lines on the tree have no meaning and are presented for purposes of clarity of presentation only. A sequence from a bat eared fox (03.179) was used to root the tree. Colours and symbols were used to denote the identified clusters and subclusters as presented in table 2.8. and figure 2.9., respectively. Sequences are preceded by the prefix KZN or EC to indicate whether the isolations were made in either the KwaZulu Natal or the Eastern Cape province. The area of isolation prefix is followed by a second prefix, which was used to denote the host species of origin. These prefixes are as follows, dg, dog, cw, cow, gt, goat, shp, sheep, hm, human and j, *C. mesomelas.*

Figure 2.9. (page 136) The approximate geographic location of the isolates which were analyzed during the course of this study, as well as the identified subfamilies, clusters and subcluster divisions as described in the text in section 2.4.4 (B), the phylogenetic tree in figure 2.8., and table 2.8. respectively. The locations where human exposures occurred, are indicated by arrows on the map of the province, whereas the large block arrows indicate the proposed pathway for the three enzootic fronts through which it was thought that rabies is currently being introduced into the province. . The circular region highlighted on the map refers to viral groups which may have been directly introduced into southern KwaZulu Natal from southern Mozambique, probably through the translocation of infected dogs via the N2 highway.Solid black lines correspond to the major routes of human transportation in the province.

University of Pretoria etd, Coetzee P (2006)

111
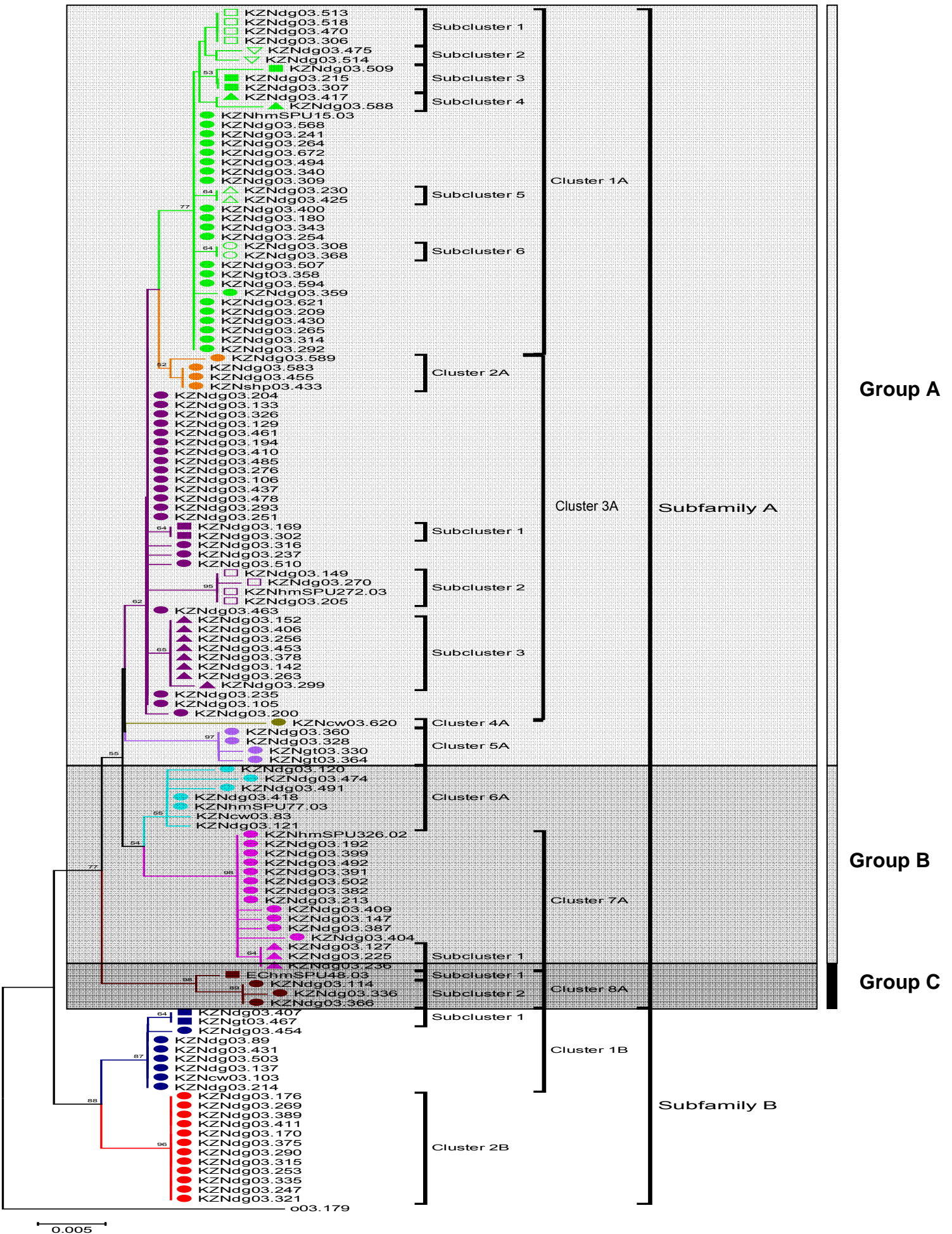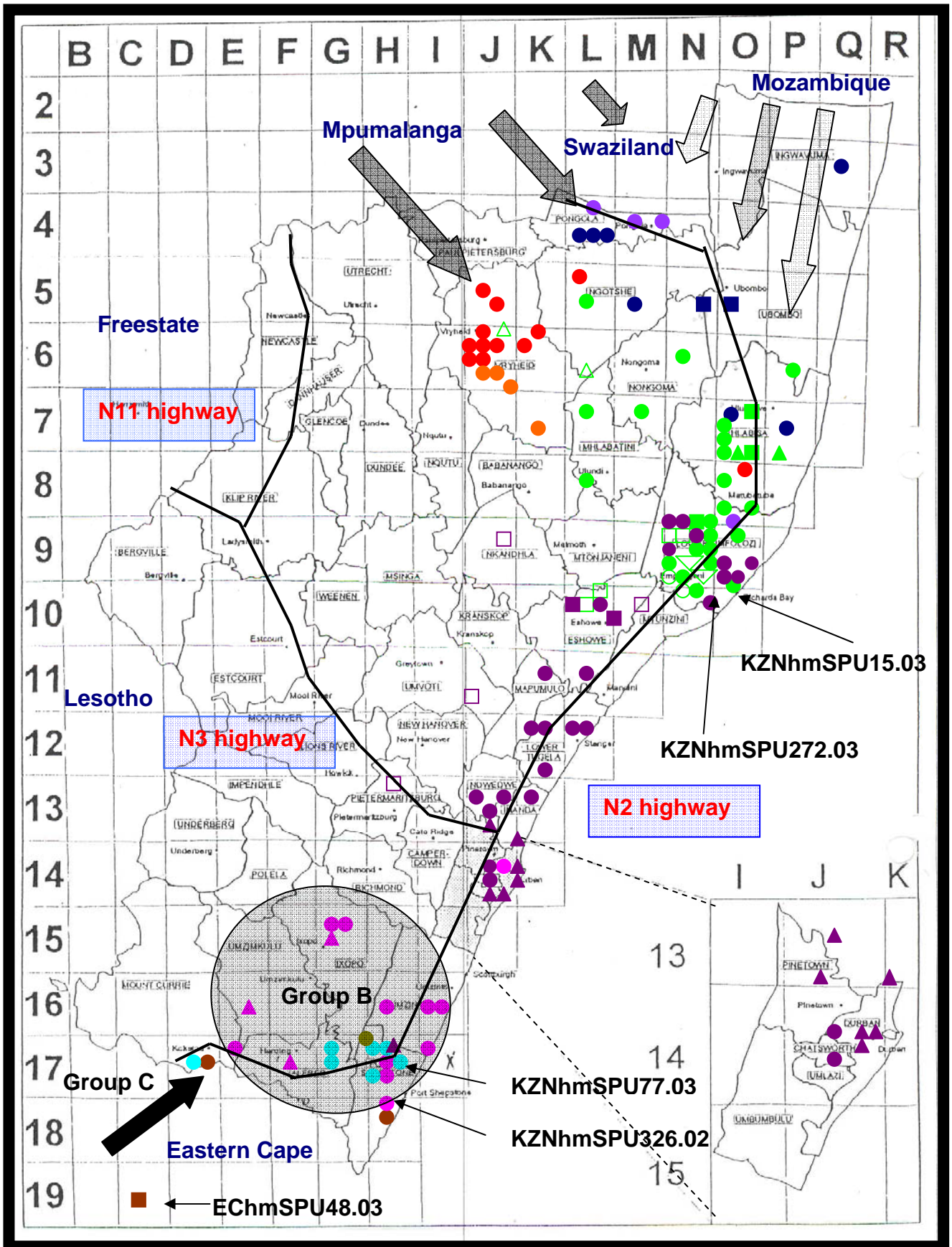
Chapter 3

Discussion

3.1 Summary

In this chapter the results which were obtained from the phylogenetic analyses conducted in chapter 2 will be discussed in detail. Particular emphasis was placed on correlating the reconstructed phylogenies to historical surveillance data which is available from the literature, allowing conjectures to be made as to the events which led to the emergence of rabies within the province. We were also able to demonstrate the possible role jackal may play in the epidemiology of the disease in the northern and northeastern regions of the province, and were able to use the phylogenetic data to reconstruct human case histories from the province. Finally we also placed the phylogenetic data which was generated during the course of this project, into a framework, that attempts to explain the long term persistence of the current epidemic, and also illustrated the potential role that long range translocation may play in the reintroduction and spread of viral variants throughout the affected regions oft the province. Finally, based on the epidemiological data which was generated during the course of this project, we made suggestions as to how control strategies may be updated, in order to bring the current epidemic under control.

3.2. Phylogeny of KwaZulu Natal rabies virus isolates, compared to canine rabies virus isolates from elsewhere in South Africa and Zimbabwe.

From the NJ phylogenetic tree which is presented in figure 2.7., it was apparent that viruses obtained from dogs, jackal and bat-eared foxes in South Africa and Zimbabwe belong to a single group (canid biotype), that is separable from the PV strain, and a virus of the mongoose biotype (m669.90). The high bootstrap value for the clustering of canid isolates from KZN, with wildlife and dog isolates from other regions of southern Africa (99%), combined with the calculated percentage nucleotide sequence homology values between the KZN viruses, the PV virus and the mongoose virus (m669.90 – also see section 2.4.3), provided clear evidence that the viral isolates from the province belonged to the South African canid biotype. The demonstration that viruses from the different canid hosts in South Africa clustered together in a single group, further provided evidence for a hypothesis that presumes that viruses of the canid biotype were all derived from a single viral progenitor. This progenitor virus was according to historical information, introduced into Namibia from Angola in the 1940s, from where it subsequently spread throughout the entire subcontinent (Swanepoel, 1994). These canid biotype viruses have previously been

shown to be highly adaptable, being able to establish infectious cycles and spread rapidly throughout whatever canid host species are capable of sustaining the virus' transmission (Sabeta *et al.*, 2003). However, it could also be demonstrated that the viral isolates obtained from KZN all shared a more recent common ancestor, and these isolates could thus on the basis of phylogeny also be distinguished from the canid rabies virus isolates which were obtained from elsewhere in South Africa and Zimbabwe. This finding lends credence to a hypothesis which presumes that the KZN rabies viruses were all introduced from a single geographical location, possibly from southern Mozambique. It was also thus evident that we were dealing with a unique family of KZN rabies viruses.

3.2.1 Phylogenetic and geographical divisions present for viral isolates belonging to subcluster 1-1 (KZN family).

Viruses isolated from KZN (subcluster 1-1) could be divided into two groups (group 1-1-1 and 1-1-2) based on moderate bootstrap support value of 69%. Group 1-1-1 viruses were isolated from throughout KZN, especially from the coastal regions, and were obtained from dogs, humans and domestic livestock. Group 1-1-2 on the other hand had a more localized distribution and was made from dogs in the northern and northeastern districts of the province. Dog and jackal isolates obtained from the eastern region of Mpumalanga, bordering western Swaziland, also clustered within this group, suggesting that group 1-1-2 isolates that circulate among dogs in KZN, may have been derived from a focus of infection that is maintained simultaneously by jackal and dogs in this region.

Group 1-1-1 and 1-1-2 may be representative of two separate fronts of the disease that had spread from the dog rabies endemic regions which are present in the southern district of Maputo in Mozambique. The first front (group 1-1-1) could possibly have spread directly southwards across the southern border of Mozambique, or the south-eastern border of Swaziland with South Africa, from where it entered into the highly populated regions of the North Coast, and from where it subsequently spread southwards to affect the entire eastern coastal region of KZN. The second front (represented by group 1-1-2) on the other hand may have spread with dogs into northeastern Swaziland from southern Mozambique. From the north-eastern regions of Swaziland this epidemic front may have spread westwards throughout that country, to establish infectious cycles among whatever dog or jackal populations had a sufficient population density to sustain the transmission of the disease. Eventually the front reached the western border of Swaziland, which it

115

crossed to establish the identified dog-jackal rabies cycles in the southeastern districts of Mpumalanga. It can be hypothesized that these viruses then spread southwards and south-eastwards from these rabies endemic regions, across the northern border regions of KZN, leading to the establishment of rabies cycles among dogs, and possibly jackal, in the northern and north eastern districts of the province. Alternatively group 1-1-2 viruses may be representative of cycles of the disease among jackal or dogs that are currently circulating further north in the Mpumalanga province, and that had spread southwards through Swaziland into northern regions of KwaZulu Natal. These observations are further discussed in detail in the following section (section 3.3).

Evidence that cross border movement of infected jackal and dogs may have been responsible for the emergence of rabies in the northern regions of KZN, is provided by the fact that the cross border movement of these animals from bordering countries and regions, have frequently been observed to coincide with outbreaks of the disease among these vector species in South Africa. For example the cross border movement of dog and jackal have been responsible for outbreaks of rabies in the eastern regions of South Africa, following the introduction of the disease from across the border regions lying between Mozambique, Swaziland, and the region of the northern Limpopo province, lying between Swaziland and the southern border of the Kruger National park. This suggested that a similar cross border incursions was responsible for the establishment of the viruses of group 1-1-1 and 1-1-2 in KZN (Swanepoel, 1994).

3.2.2. Jackal rabies in the northern and north-eastern regions of KwaZulu Natal.

It can further be hypothesized that rabies virus isolates belonging to group 1-1-2 which circulate among dogs in the northern and northeastern regions of KZN may be maintained simultaneously among dog and jackal, even though no jackal isolates from the province were available from the year 2003, in order to substantiate this hypothesis. Evidence in support of this hypothesis, is however provided by the observation that group 1-1-2 viruses have a limited distribution in the northern regions of KZN, which coincides with the known distribution of the blacked back jackal (Internet reference – biology of the black backed jackal). This region is also the location of agricultural activities such as game ranching and cattle farming, that have previously been shown to provide ideal ecological conditions for jackal, allowing them to reach a sufficient population density at which they are able to sustain the transmission of the disease (Bingham and Foggin, 1993;

116

Swanepoel, 1994). Although the precise reason for the propensity of jackal rabies to predominate in regions where such activities are practiced is not known, it appears to be related to abundant resources, increased demographic turnover, and the absence of competitors such as dogs and wild scavengers (Bingham, 2005).

It has been demonstrated that rabies cycles that initially occur among black backed jackal appear to closely correlate with outbreaks of the disease among dogs. This indicates that spillover events from dogs were initially responsible for the establishment of the disease among these animals. Once the disease had becomes established among jackal however, it appears to be maintained independently from cycles in dogs (Swanepoel, 1994). Some controversy however still exists with respect to this observation, since it is thought that outbreaks among black backed jackal soon reduces the population density below a threshold that is required to sustain the virus' transmission (Bingham, 2005). This issue and its implications in terms of potential control strategies in the KZN province will be further discussed in section 3.3.4.2.

3.2.3 Reconstruction of the events which led to the introduction of rabies into KwaZulu Natal.

From the NJ tree constructed in section 2.4.4. (A) (figure 2.7) it was apparent that isolates from KZN and the Mpumlanga region bordering western Swaziland (subcluster 1-1) were more closely related to viruses of subcluster 1-2, which were isolated from jackal and dogs in the northern, central, eastern and southern districts of Zimbabwe, than to isolates of subcluster 2, which were isolated from dogs, jackal and bat eared foxes from south-western and central Zimbabwe, as well as the northern and western districts of South Africa.

It is believed that rabies was introduced into the regions of southern and south western Zimbabwe, from Botswana and the northern Limpopo province of South Africa in the early 1950s, following the movement of people (and their dogs) who crossed these borders illegally to buy grain (Swanepoel, 1994). From here the disease spread widely, to encompass all the districts of Zimbabwe (Swanepoel, 1994). Control measures were implemented in the 1960s, and managed to, except for resistant foci in the eastern and western regions of Zimbabwe bordering Mozambique and Botswana, reduce the incidence of rabies in the affected districts of the country (Swanepoel, 1994). These control

measures broke down in 1965 due to a civil war which had erupted in that year, with the result that rabies cases rose to reach record levels in 1981 (Swanepoel, 1994). The war ended in 1980, but control of rabies in the eastern and western districts of the country has continued to be hampered by continued political unrest in the districts of Matabeleland in the southwest, and the influx of refugees from the civil war raging in Mozambique in the east (Swanepoel, 1994).

From the phylogeny presented for the isolates in figure 2.7, as well as the historical surveillance information described above, it can be postulated that at least two separate fronts of the epidemic led to the establishment of rabies in jackal and dog populations in two separate regions within Zimbabwe. The separation between these outbreaks or fronts represented by cluster 1 and 2 on the phylogenetic tree, was supported by a high bootstrap value of 99%. Cluster 2 could represent the introduction of rabies into the dog and jackal populations of southern and south-western Zimbabwe, from Botswana and the Northern Limpopo province, from where it subsequently spread into central Zimbabwe. The second front on the other hand as represented by isolates from subcluster 1-2, could represent the northeastward spread of the epidemic from the initial focus of rabies in southern and soutwestern Zimbabwe, into the dog and jackal populations of the southern, central, eastern and northern districts of the country.

The cross border movement of people and their dogs from eastern districts of Zimbabwe into Mozambique, and vice versa, probably could have led to the establishment of the cluster 1 lineage in regions throughout Mozambique. It is likely that this lineage may have been introduced into Swaziland and the northern districts of KZN from the dog rabies endemic region which were established in the southern Mozambique district of Maputo, possibly through the translocation of domestic dogs with refugees fleeing across the northern border of KZN, as was documented to have occurred in 1976. The demonstration that viral isolates from subcluster 1 (KZN isolates), were more closely related to isolates from subcluster 1-2 than to isolates belonging to cluster 2, support the above hypothesis for the spread of rabies from the northern regions of South Africa into Zimbabwe, and from there on south-eastwards into Mozambique and the Mpumalanga province, from where it was subsequently introduced into northern KZN. The proposed regional path of spread which led to the introduction of rabies into KZN, is presented on a map of southern Africa, in appendix E.

118

3.2.4 Introduction of KwaZulu Natal variants into other regions of South Africa.

Finally, it was also of interest to note that two jackal isolates, j101.90 and j378.90, which were obtained from the northern Limpopo province of South Africa, clustered together with isolates from group 1-1-1 which have thus far have only been isolated from the eastern coastal districts of KZN. A possible explanation for this observation could be given by considering that the movement of animals from the eastern coastal regions of KZN, north-eastwards into the northern regions of South Africa, may have occurred due to human intervention. Transport of a group 1-1-1 infected dog from KZN, into the northern Limpopo province of South Africa, and spillover of the virus from infected dogs, back into jackal, could represent a possible explanation for the emergence of the KZN variant among wildlife in the northern regions of South Africa.

3.2.5 Future studies required to clarify the molecular epidemiology of rabies in Mozambique and the northern Limpopo province of South Africa.

Since viral isolates from Mozambique were not available for inclusion in this analysis, it was not possible to demonstrate the presence of the viral lineage which was presumed to be the most likely to cycle throughout Mozambique, viz. cluster 1 (figure 2.7). Future studies of the epidemiology of rabies in the eastern regions of South Africa, should thus include a phylogenetic comparison of viral isolates which were obtained from Mozambique in order to complete the currently available molecular epidemiological picture in this region. Furthermore, future studies should also focus on the characterization of additional rabies virus isolates from the northern regions of South Africa, in order to validate the presence of, and delineate the possible distribution of the group 1-1-1 outbreak among jackal and dogs in the northern Limpopo province of South Africa

3.3. Phylogeny of rabies virus isolates from KwaZulu Natal

Phylogenetic analysis was used in order to determine the distribution of rabies viral variants throughout the magisterial districts of the province. The regional characterization of these variants, combined with their respective topologies on the NJ phylogenetic tree as presented in section 2.4.4 (B), allowed us to make inferences about the pattern of spread of rabies into, and throughout the KZN province. The regional characterization of these variants, further allowed us to correlate known human case histories to the regions where

these exposures are likely to have occurred in, and in this way we could demonstrate the utility of the phylogenetic analysis of the sequence data which was generated during the course of this project, towards the reconstruction of human case histories from the province. The identification of the regional variation of the RABV variants in KZN furthermore allowed us to investigate the factors which are likely to play a role in the long term persistence of the disease, allowing us to devise updated vaccination strategies with which the current epidemic might be brought under control.

A strong regional variation was present between the isolates from the different identified viral groups, with the groups increasing in nucleotide similarity based on their geographic proximity to each other. This was suggestive of a strong regional pattern of virus evolution, similar to that described for fox rabies virus isolates (genotype 1) from France (Sacramento *et al.*, 1991). Viruses from the mongoose biotype were not found to be present in the analyzed sample set. This result was not unexpected as the Drakensberg mountains which lies along the western border regions of KZN, provides an effective barrier to the movement of small animals such as yellow mongoose from the mongoose biotype endemic regions of the central plateau.

Viruses from KZN could be divided into two subfamilies (A and B) based on a bootstrap support value of a 100%. Subfamily B (corresponds to group 1-1-2 in the analysis in section 2.4.4. (A)) consisted of two clusters 1B and 2B that had a separate but localized distribution in the northern and north eastern regions of the province. These two clusters were thought to represent dog-jackal rabies cycles that had spread from a focus of infection in Swaziland or southeastern Mpumalanga, southwards and southeastwards, into the northern and northeastern districts of KZN (see section 3.2). Subfamily A on the other hand comprised the largest number of isolates, and were isolated primarily from along the coastal regions of the KZN province. It was thought that these viruses were directly introduced into the northeastern coastal regions of KZN, from the southern Maputo region of Mozambique.

3.3.1 Pattern of spread and its associated factors, which led to the introduction of viral variants from subfamily B into the affected regions of KZN.

Isolates of cluster 1B had a wider distribution in comparison to that of cluster 2B. Indeed the cluster 1B isolates were made from four magisterial districts - including Pongola, Ngotshe, Ubombo and Ingwavuma in the northern and northeastern regions of the province. Two isolations from this cluster were also made from the eastern coastal district of Hlabisa. Cluster 1B on the other hand had a more localized distribution, being isolated only from the northern magisterial districts of Vryheid and Ngotshe. A single isolate from this cluster was however also made from the eastern coastal district of Hlabisa. The translocation of infected dogs across the Golela border station, present between the southern border of Swaziland and the northern KZN district of Pongola may have been responsible for the introduction of viral isolates from these clusters from southeastern Swaziland into the northern regions of KZN. Alternatively these clusters may have been introduced through the cross border movement of jackal from southern Swaziland into the northern reaches of the province.

The precise reason for the phylogenetic division between isolates of cluster 1B and 2B is not known, although it could be speculated that the genetic bottleneck of the ancestral population that led to the independent regional evolution of these viral groups may have occurred either somewhere within the northern or eastern regions of South Africa, or alternatively in Swaziland itself. Separate introductions of cluster 1B and 2B viruses into the regions of the KZN province these viral groups circulate in, could then have led to the observed distribution. These clusters may alternatively have evolved in KZN itself, following the introduction of the common ancestor (possibly in jackal) from Swaziland or the northern Limpopo province, into the discrete regions of the KZN province from which these variants have been isolated from. The evolution of these viral groups may thus have occurred in these geographically separate regions, possibly in association with parallel cycles in jackal.

Finally it is of interest to note that a large number of isolates within, cluster 1B and 2B, shared a 100% sequence homology in respects of the nucleotide sequence regions analyzed. This suggests that the cycles represented by these clusters are highly active, with the respective variants having spread rapidly through their susceptible host populations with little to no genetic variation. The isolation of a single isolate from cluster

2B, which primarily circulates in the districts of Vryheid and Ngotshe from the district of Hlabisa on the eastern coast, suggested that the translocation of infected dogs through human intervention, may have played a role in the dissemination of this variant. The N2 highway runs through the district of Pongola southeastward along the eastern coastal district into Hlabisa, and it is likely that the translocation of infected animals, specifically via this highway may have led to the outbreak of this variant in the Hlabisa magisterial district. A similar translocation event from Pongola via the N2 highway, could also further have been responsible for the isolation of cluster 1B viruses from this district. In fact it appears that the distribution of cluster 1B fits the route of the N2 highway exactly as it runs southeastwards through the districts of Pongola, Ngotshe, Ubombo and Hlabisa. The single isolate of cluster 1B made from the northeastern magisterial districts of Ingwavuma on the other hand, could represent the additional eastward spread of this cluster from southeastern Swaziland, across the region lying between that country and the northeastern regions of KZN.

We suggest that future studies focused on determining the regional spread of rabies throughout the eastern regions of southern Africa, should include a panel of viral isolates from the Mpumlanga province (South Africa), as well as isolates from Swaziland. The characterization of these viruses would clarify the events leading to the emergence and regional spread of cluster 1B and 2B variants in their respective geographical regions of isolation.

3.3.2 Pattern of spread and its associated factors, which led to the introduction of viral variants from subfamily A into the affected regions of KZN.

Subfamily A viruses from KZN (represented by group 1-1-1, figure 2.7) could be divided into three groups (A, B and C) that all shared a recent common ancestor (figure 2.8). The divisions between these groups were supported by bootstrap values between 58-77%. Group A was by far the largest of the three and was composed of viruses that were primarily isolates from the North Coast (Durban district northwards up till the border region between South Africa and Mozambique) as well as the internal northern and northeastern districts of the province. This group is representative of the present-day core of the rabies epidemic in the province. A few of the members of this group (cluster 5A) were also isolated from the northern district of Pongola south of the Swaziland border. Group B viruses were primarily isolated from the South Coast, while Group C viruses had a

localized distribution in the southern border region between the Eastern Cape and KwaZulu Natal. A single isolate made from a human patient who had become infected in the Eastern Cape in 2002, also belonged to group C. This case led us to surmise that these viruses may represent a cycle that is being maintained in dogs in the Eastern Cape, and which had spread northwards in recent times into the southern regions of KwaZulu Natal.

The order in which viral clusters within group A, B and C branched from the reconstructed NJ tree (figure 2.8.), concomitant with historical information, permitted some degree of conjecture on the events leading to the introduction and spread of rabies throughout the magisterial districts of the KZN province. The hypothesis which presumed that subfamily A viruses (represented by group 1-1-1) were originally introduced into KZN from the southern Mozambique district of Maputo has already been outlined in section 3.2. Group C viruses, isolated from the southernmost regions of the KZN province were the first to diverge from the subfamily A collection of KZN viruses. This result would be highly unexpected if it was presumed that viruses isolated from along the continuous stretch of coastal region of the province all belonged to a single outbreak of the disease (commonly introduced). Indeed, the branching order of the groups suggest that Group C is not representative of an epidemic that has spread sequentially down the coastal routes of the province, contrary to the common opinion expressed in the public domain (Swanepoel, 1994; Bishop *et al.*, 2003). A possible explanation for this observation could be found by considering that two epidemics had occurred among dogs in KZN in recent times. The first outbreak started in 1964, was brought under control in 1968, and apparently did not enter the Eastern Cape province. The second epidemic which started in 1976, and which still rages among dogs in KZN today, is however known to have spread as far southwards as the Eastern Cape, from where it was reported in 1987. We suggest that the 1964-1968 epidemic in KZN did in fact spread as far southwards as the northern regions of the Eastern Cape Province, which it entered to initiate local outbreaks among domestic dogs. Although the first epidemic in KZN was brought under control, the focus in the Eastern Cape has managed to persist, and has in recent times spread northwards to re-establish cycles among dogs in the southern regions of the KZN province.

Groups A and B thus represents viruses that had spread southwards during the course of the second epidemic, while viruses of group C on the other hand represented a remnant focus of infection left over from the 1964-1968 epidemic. The order of events as described

above are supported by the phylogeny, and seems to account why group C viruses are distinguishable from virus isolates from group A and B, as well as why group C viruses were the first group to diverge from the collection of subfamily A viruses on the phylogenetic tree (figure 2.8).

The node separating cluster 4A and 5A from cluster 1A to 3A was not significant, and these clusters were thus considered to fall under the group A collection of viruses within subfamily A (tree interpreted as if the nodes under question were collapsed to the next significant node). Group A viruses were isolated from the "North Coast", while Group B consisted of two viral clusters (6A and 7A), which were isolated from the southern coastal region of the province. Group A and B viruses were separated from each other by a node that had a low bootstrap support value (55%). Group A was however considered to be separable from Group B because the topology illustrating their distinction could also be reconstructed when using the MP method. Again the observation that group B viruses were distinguishable from group A viruses, as well as that group B branched from the phylogenetic tree at the same time period than group A, was unexpected. This discrepancy could be explained by the possibility that the group A and B common ancestor was introduced into southern KZN directly from Maputo in southern Mozambique, independently from the front (represented by group A) that had entered into the north eastern coastal districts of the province. It is presumed that at the same time as this common ancestor entered into the northern regions of KZN, that this common ancestor was also introduced by a separate translocation event into the southern districts of the KZN province. The independent spread of this common ancestor into discrete dog populations present in this region is then presumed to have subsequently led to independent regional evolution of cluster 6A and 7A viruses.

The following events leading to the establishment of group A, B and C viruses in their respective regions throughout the province could thus be reconstructed. During the start of the second epidemic, a rabies cycle spread from Maputo southwards (possibly through southeastern Swaziland) leading to the introduction of the common ancestor of group A and B in the northern regions of KZN, from where it spread sequentially southwards to affect the "North Coast" of the province. The independent maintenance of this virus in local dog populations in this region of KZN led to the regional evolution of cluster 1A-3A. The common ancestor of group A and B also spread southwards from Swaziland into the Pongola district, via the Golela border crossing, leading to the regional evolution of cluster

5A among susceptible dog populations in this region of the province. At the same time the transport of the common ancestor of group A and B from the Maputo region of Mozambique, via the N2 highway, into the South Coast region of the province, led to the establishment of the group B lineage in the southern regions of KwaZulu Natal. The independent spread of this common ancestor into different discrete dog populations in southern KwaZulu Natal, would then subsequently have led to the regional evolution of cluster 4A viruses, and the common ancestor of clusters 6A and 7A. The common ancestor of cluster 6A and 7A would then be presumed to have spread into different individual dog populations present in the South Coast, leading to the regional evolution and observed distribution of these clusters. Unfortunately no more isolates from cluster 4A were available from this analysis, and its regional distribution throughout southern KZN, could thus not be determined.

The independent maintenance of cluster 3A viruses in the North Coast region of the province furthermore also led to the evolution of the common ancestor of cluster 1A and 2A. This common ancestor spread westwards, possibly through the transportation of infected dogs via the N2 highway, into the northern reaches of the province, leading to the establishment of, and evolution, of cluster 1A and 2A viruses in their respective regions of isolation.

Translocation via the major N2 highway also appears to have played a role in the spread of virus isolates belonging to cluster 3A, 5A, 6A and 7A. This could explain the isolation of a single viral isolate belonging to cluster 3A (subcluster 3), and cluster 7A, which circulate in the magisterial districts of Durban, and the southernmost districts of KZN, in Durban and Port Shepstone, respectively. Furthermore, the isolates from cluster 6A and 7A appear to have a distribution in the southern regions of the province fitting the route of the N2 national highway, as it moves westwards from Port Shepstone to Kokstad. This further illustrates how the transportation of infected animals through this route, most likely could have led to the observed spread of virus isolates belonging to these clusters, further inland, westwards from Port Shepstone.

The demonstration that a viral isolate from a human patient, belonging to cluster 8A, was isolated from the Eastern Cape Province, suggests that the cross border movement of viruses from the Eastern Cape region northwards may play a role in the epidemiology of rabies in KZN. We have previously illustrated that this group of viruses may represent a

remnant infectious cycle left over from the 1964-1968 epidemic, which was brought under control to the north in KZN. Nothing is however known about the current regional variation and distribution of these viruses in the Eastern Cape, and we would thus propose that additional studies be performed to characterize these variants with respect to the relevant gene regions utilized during this analysis. This would provide not only information on the molecular epidemiology of the disease in the Eastern Cape province, but would also provide information that may allow for the clarification of the role that this viral group plays in the epidemiology of rabies in the KZN province itself. In addition, it is recommended that rabies in the bordering country of Lesotho should also be characterized, in order to complete the epidemiological picture in this region of southern Africa. The proposed path of spread of rabies virus variants into and throughout the KZN province is illustrated on 5 sequential maps, presented in appendix F.

3.3.3. Correlation between the variants which were isolated from human infections, and the geographically discrete variants identified from KZN during the course of this study.

Eight human infections were diagnosed from KZN for the year 2003 (NICD, 2003) the majority of which occurred in children. Cases such as these among young children, can complicate the task of reconstructing case histories from these infections, an essential step which is required to identify possible contact routes, and for determining where appropriate measures such as vaccination of domestic dogs should be undertaken in order to prevent the transmission of the disease to humans. In order to illustrate the utility of the phylogenetic analysis of the sequence data generated during the course of this project, in assisting health officials in obtaining this information, we decided to compare the sequence data which was obtained from five human infections that had occurred during the year 2002-2003 in KZN and the EC, to the sequence data which were obtained from the domestic dog isolates from KZN during the course of this study. In so doing, we were able to correlate the known human case histories to the regions of the province these infections had presumably occurred in, therefore allowing for the identification of the main regional variants which were responsible for these infections. A summary of the case histories for the human infections are supplied in table 3.1 (NICD, 2003), while the geographic regions of KZN and the EC these exposures occurred in, are presented in figure 2.9. The NJ phylogenetic tree comparing the phylogeny of the sequence data obtained from the human isolates, to the sequence data obtained from isolates from KZN,

is presented in figure 2.8. By correlating the human case histories, to the reconstructed phylogeny for the sequence data obtained from human and animal isolates the following conclusions could be made.

The exposure of the human case SPU15/2003 occurred in the Lower Umfolozi district. Sequence data from this isolate clustered together with isolates from Subfamily A, group A, cluster 1A,which has a high prevalence among domestic dogs in the magisterial district of the Lower Umfolozi among others.

The human case SPU272/2003 occurred due to the bite from an infected dog in the town of Tugela Ferry in the magisterial districts of Msinga. This virus isolate clustered together with isolates from Subfamily A, group A, cluster 3A, subcluster 2, which had also been isolated from dogs from the Mtunzini, Umvoti and the Pietermaritzburg districts.

The human case SPU77/2003 occurred due to a bite from an infected dog in the southern district of Port Shepstone. The viral sequence obtained from this isolate, clustered with dog isolates of Subfamily A, Group B, cluster 6A, which had a localized distribution in the magisterial districts of Port Shepstone, Alfred, and Mount Currie.

For the human case SPU326/2002 the source of infection was unknown, although it was thought that the exposure had occurred in Port Shepstone. The neighbors' dog had apparently died of unknown causes two days before the patient had succumbed to the disease, but the animal was however not tested for rabies. According to the NJ phylogenetic tree presented in figure 2.8. this patient became infected with a variant belonging to Subfamily A, Group B, cluster 7A, which has a localized distribution in the magisterial district of Port Shepstone as well as its surrounding districts. It is thus possible, even likely, that a lick or a scratch from the neighbors dog on the face of the child may have led to the cryptic transmission of this variant to a human.

SPU48/2003 was isolated from a 13 year old boy in the Eastern Cape district of Cofimvaba. The phylogenetic data presented in figure 2.8. indicated that the sequence from this isolate clustered together with the Subfamily A, group C, cluster 8A variants, which have only been isolated from along the southern border region of KZN with the Eastern Cape. It is thus likely that this person either became exposed to the virus in the southern regions of KZN, or that the exposure to this variant had occurred somewhere in

the Eastern Cape province itself. The latter assumption seems most probable, since the human isolate from the EC was phylogenetically separable from the dog isolates from this cluster, which were isolated from KZN (98% bootstrap support value).

Table 3.1. Case histories of human infections that had occurred in the KwaZulu Natal and Eastern Cape province from 2002-2003 (NICD, 2005).

Patient name	SPU nr	Nature of exposure	Age/Sex	District exposure occurred in
NM	15/2003	Bitten by dog 17 December 2003, died on the 8 January 2004	5f	KwaZulu Natal, Lower Umfolozi
ZS	272/2003	Bitten by dog , died 2 September 2003	12m	Tugela Ferry, Msinga
NK	77/2003	Bitten by dog, February 2003, died 27 March 2003	3f	Port Shepstone
NB	326/2002	Source of infection not confirmed, died 9 November 2002, neighbors dog died on the 12 November 2002 but was not tested for rabies	6m	Port Shepstone
MSS	48/2003	Bitten by dog in November 2003, died on the 15 February 2004	13m	Eastern Cape, Cofimvaba

3.3.4 Factors that influence the maintenance and persistence of the rabies virus in the KwaZulu Natal province.

Bingham (2005) suggested that the ecology of rabies in African canid hosts should be understood within a new terminological framework. This framework is underpinned by two new concepts; the maintenance and persistence of rabies within, and between geographically localized populations (individuals that live in the same habitat patch and can therefore interact) of susceptible hosts. Maintenance encompasses the idea of an open cycle of transmission within a local population, provided that there is an endless supply of susceptible hosts, and provided that the transmission ratio is equal to, or larger than one (Bingham, 2005). Maintenance of the virus within a single population is often unstable and leads to the extinction of the virus, due for example to the decimation and renewal of host population numbers as the virus moves through and then recedes within a specific local population (Bingham, 2005). Following the extinction of an outbreak, the disease is then usually reintroduced from other locally affected populations (Bingham, 2005). Persistence of an infection, or the notion of the long term continuity of an infection within an affected geographic region, depends on the maintenance of rabies in at least one or more of many localized heterogeneous local populations of susceptible hosts (metapopulation), and is further also dependent on the requirement for a degree of movement of infected animals being possible between these local populations (Bingham, 2005). Other factors which could possibly affect the maintenance and persistence of rabies includes the presence of a carrier state, as well as long-term incubators of the disease. The presence of these phenomena have however been questioned, and they have not as of yet been shown to play a role in the maintenance or long-term persistence of the disease in Africa (Bingham, 2005). The maintenance and persistence of rabies is however affected by population immunity, which is in turn almost exclusively determined by vaccination coverage in terrestrial carnivores (Bingham, 2005).

3.3.4.1 Maintenance and persistence of rabies in local dog populations in KwaZulu Natal.

The framework as outlined above could be extended to explain why the second epidemic, which started in 1976 in the northern districts of the province of KZN, has proven to be intractable. With respect to the first epidemic, which started in 1964, it can be hypothesized that rabies outbreaks were unable to persist, since too few local dog

populations were available for maintaining a large enough number of independent rabies cycles, as would be required for the persistence mechanism of the disease to function. The decimation of local host populations by rabies in the affected districts, aided by the targeted vaccination of local "hotspots", proved sufficient for bringing this tentatively established epidemic under control by 1968. However during the period of the second epidemic (1976-2005), a number of social and political changes had occurred within South Africa, which had led to the increased movement of people to areas surrounding the agricultural, industrial and recreational centres of KZN. This led to the development of numerous informal settlements, with large associated independent local dog populations, in which low vaccination coverage exists, due to either logistical or financial difficulties on the part of the veterinary authorities. Movement of rabies into and between the increased numbers of local dog populations subsequently led to a situation in which there was an increase in the ability of the disease to persist, thereby complicating subsequent efforts to control the epidemic.

The information garnered from this study demonstrated that the conditions for the continued persistence of dog rabies within KZN were present during the year 2003. Phylogenetic analysis indicated the presence of numerous related outbreaks, in different geographically proximate foci throughout the province. Furthermore it was demonstrated how the disease may spread from one local dog population to another, either through the natural movement of infected dogs between adjacent dog populations, or through the introduction of variants from distant geographic regions through human translocation. In other words, it could be observed that there are a large number of susceptible local dog populations in which the disease is independently being maintained, coupled with the observation of the movement of variants from one maintenance cycle to another. This provides the ideal conditions for the long term persistence of the disease among the dog populations of the province, as illustrated by the framework presented by Bingham, 2005.

The data to demonstrate whether the resurgence of outbreaks in any particular local dog populations in KZN is due to either the re-emergence of a local variant which had been maintained in that area at an undetectable level, or a new variant that had been introduced from an adjoining or a distant outbreak, has only become possible following the genetic characterization of KZN viruses during the course of this study. From present data it appears that movement of variant from adjoining, as well as distant endemic foci, plays a role in the reintroduction of the disease into the local dog populations of the province.

For example, the demonstration of genetically related clusters of viruses throughout the "North Coast" of KZN seems to support the spread of the virus between adjoining dog populations, while the long distance translocation of rabies variants from Mozambique into the southern regions of KZN via the major N2 highway, as previously demonstrated, shows that re-introductions of the disease into the province may also occur from geographically distant foci of infection.

Future surveillance efforts using the data generated during the course of this study may be applied to identify additional routes by which the disease may be reintroduced, and may thus allow for the implementation of control strategies which can prevent the regional reintroductions of rabies from occuring. In this way a mechanism could be provided by which the persistence of the virus may be undermined, allowing for the natural extinction of the outbreaks among the individual local dog populations over time. Coupled with targeted vaccination strategies for the most severely affected regions, this approach could provide a framework for updated control strategies in the province.

3.3.4.2 Maintenance and persistence of rabies among jackal in KwaZulu Natal.

For jackal rabies a different set of host specific circumstances may influence the maintenance and persistence of the disease. As previously mentioned controversy exists as to whether jackal can maintain rabies cycles independently of the cycles in dogs. Two opposing branches of evidence support these competing hypotheses. On the one hand, it has been shown by mathematical modelling of epidemics within jackal populations that these animals cannot maintain the disease independently from dogs (Bingham, 2005), while on the other hand surveillance data indicates that jackal outbreaks initially appear to be associated with outbreaks of the disease among dogs, after which they become independently maintained (Swanepoel, 1994). These observations can also be explained in the framework that describes the maintenance and persistence of the disease, and may further also be applied to the evaluation of the factors that influence the maintenance and persistence of the disease among jackal in the northern regions of KZN. According to Bingham, 2005, jackal populations should not be considered to represent metapopulations, because they do not form the spatially separate host populations required for the persistence of the disease (Bingham, 2005) A failure of the virus to be reintroduced from dogs into a jackal population, could thus be postulated to eventually lead to the extinction of the maintenance cycles among these animals, over time. This

observation may have important implications in terms of future rabies control strategies in jackal.

The two putative jackal cycles identified from the northern and northeastern regions of KZN are maintained independently in their respective geographic locales, and we could not demonstrate that the movement of variants occurred between these geographically separate outbreaks. If we consider the framework for explaining the persistence of rabies suggested by Bingham (2005), these cycles would thus not be expected to remain over the long term. Nevertheless, rabies among jackal in this region of the province will continue to represent a potential problem, due to the presence of rabies in overlapping dog reservoirs. These dog populations represent a source from which the disease may be re-established among jackal. Also it would be expected that the possible transmission of these viruses from jackal back into dogs would continue to hamper the control of the disease among domestic dogs in this region of KZN. Different control strategies would thus have to be applied simultaneously to domestic dogs and jackal in order to bring these cycles under control. Control may however be simplified by the observation that the jackal front is not likely to spread further southwards, as it would terminate at the geographic limit of the distribution of the black backed jackal in the northern regions of the province.

3.3.5 Suggestions of how current strategies for rabies control in the KwaZulu Natal province may be updated in order to bring the epidemic under control.

During the course of this investigation we were able to demonstrate the following important facets for the epidemic in KZN, which might influence the implementation, evaluation, and effectiveness of possible future control strategies.

- Rabies appears to be introduced into the province from three different geographic locations. Two fronts, one associated with domestic dogs, and the other associated with domestic dog and jackal, appear to be entering the province from the north. A third front associated with domestic dogs appears to be entering the province from the south.
- We were able to demonstrate that a large number of variants are distributed throughout the province, and demarked the location where these variants circulate in individual host populations.

- We were able to demonstrate the regional movement of these variants into and throughout the local magisterial districts of the province.

- We were able to demonstrate how the translocation of variants from distant, or proximate geographic foci of infection, either naturally, or by major routes of human transportation, can lead to the reintroduction of rabies into susceptible local dog populations in the province.

- We were able to place the phylogenetic data generated during the course of this study into a framework that attempts to explain the long-term persistence of the epidemic in KZN.

These findings allowed us to propose new vaccination strategies that may be applied to the future control of rabies in the province. The demonstration that rabies enters the province from three different regions, suggested that appropriate control measures should be implemented to prevent the cross border incursion of these viruses from neighboring provinces and countries. Cordon vaccination has proven useful in the past for creating immune barriers that can halt the spread of the disease. We would thus suggest that if economically feasible, a similar strategy be implemented to prevent the introduction of rabies from occurring from the regions bordering northern and southern KZN. For this purpose it is recommended that dogs should be vaccinated in a 20 to 30 km cordon along the border regions between Swaziland, Mozambique, Mpumalanga and the northern KZN border, and between the northern border of Eastern Cape province and the southern border of KZN. House to house visits combined with mass vaccination should prove apt for this task, and should allow for the achievement of the 70% vaccination target required to halt the spread of the disease.

These strategies may further need to be augmented by the oral vaccination of jackal using edible baits, containing for example the SAG2 attenuated vaccine in the northern regions of the province, in order to prevent the spread of rabies from jackal into domestic dogs or vice versa. The use of oral vaccination would further also be recommended for controlling the disease among feral dogs. Oral vaccination has been applied with some success in a few limited studies among jackal in Zimbabwe, and in dogs in South Africa, and it is thus likely that this strategy will also be applicable to the situation which currently exists within the province (Bishop, 2003). Alternatively the control of rabies among only dogs in the northern region of the province, would prevent the reintroduction of the disease from occurring back into jackal, leading to the eventual natural extinction of the maintenance

cycles which are present among these animals. This strategy is based on the subversion of the mechanism responsible for the persistence of rabies among geographically proximate but discrete host populations, by preventing the reintroduction of the disease into any local maintenance population, in which it has become extinct (Bingham, 2005).

The phylogenetic analysis conducted on the isolates from KZN, allowed us to identify the distribution of unique rabies virus variants that circulate among different local dog populations from throughout the province. The description of these variants allowed us to demark the regional distribution of rabies maintance cycles, which may be individually targeted by appropriate control strategies. Vaccination campaigns should initially focus on red zone districts (>10 cases/annum), present at the time period during which these possible vaccination strategies will be implemented in, and should where possible coincide with the drop in cases observed at the end and start of the three to four year movement average of the epidemic.

The phylogenetic data generated during the course of this project further proved useful by allowing the delineation of the possible routes by which translocation of infected animals may occur into and throughout the KZN province. The finding that viral variants had been introduced from Mozambique, Swaziland, the Mpumalanga and the EC province into KZN, suggest that cooperation between national as well as international veterinary authorities, from the affected regions and countries will be required to bring the epidemic within the province under control. In light of this proposal the dataset constructed during the course of this study should be expanded upon, to include isolates from these bordering regions and countries. This will allow for the correlation of the molecular epidemiology of the disease between these geographic regions, and could allow for the design and implementation of control strategies, targeted to bringing the disease under control throughout the larger geographic region in question.

3.3.6 Implications of this study for the future surveillance of the epidemic in KwaZulu Natal.

Future surveillance efforts using molecular phylogeny will play an essential role in determining the regional spread of variants throughout the magisterial districts of KZN, and will also play a role in the evaluation of the effectiveness of possible control strategies. It is probable that the genetic characteristics of the variants which were identified from KZN during the course of this study, will continue to be preserved over the long term, allowing surveillance studies to remain feasible for the foreseeable future. This assumption is based on the observation that viral isolates from KZN have been shown to accumulate mutations very slowly, showing only 1.5% divergence over a 15 year period (Jacobs, 1997).

Chapter 4

Conclusion

4.1 Introduction

Canine rabies is a relatively recent addition to the African subcontinent, having been introduced from dog endemic regions which had existed in Angola since the 1940s. It has since spread widely throughout Southern Africa and its bordering countries, and has established infectious cycles among numerous canid host species including the bat eared fox, black backed jackal and domestic dog. Canid rabies was introduced into the southern regions of Zimbabwe from Botswana and the northern Limpopo province of South Africa in the early 1950s. From here it is hypothesized that the disease moved eastwards into Mozambique, from where it subsequently spread southwards into Swaziland and the northern regions of KZN. Two epidemics believed to have been introduced from the southern Mozambique district of Maputo, has since then broken out among dogs in this province. The first of these epidemics started in 1964 and ended by 1968, while the second epidemic which started in 1976 still manages to persist, despite efforts to bring it under control (Swanepoel, 1993; Bishop *et al.*, 2003) . Reasons for this are varied and include a number of social and political changes which have occurred in South Africa in recent years, and which has resulted in the increased urbanization of human populations, the development of informal settlements and an associated parallel growth in domestic dog populations (Randles, 2003). This situation appears to be compounded by the AIDS pandemic, which has resulted in the development of feral dog packs that contribute to the spread of the disease (Bateman, 2005). Rabies continues to pose a significant public health threat in KZN, with the majority of human rabies cases diagnosed each year from South Africa being the result of bite exposures from infected dogs from this province (Randles, 2003; Bishop *et al.*, 2003). Most of these cases occur in children, complicating the reconstruction of human case histories which are required to establish possible contact routes, and for identifying the regions where appropriate control measures among dogs should be implemented in (NICD, 2003).

In this study we contributed to the understanding of the molecular epidemiology of rabies in the KZN province with the express aim of assisting in future surveillance and control efforts. Molecular sequence analysis allowed us to the describe the regional genetic variation that exists between the viral isolates which were obtained from the different magisterial districts of KZN, and allowed for the reconstruction of the possible events leading to the introduction and regional spread of the first and second epidemic into the affected regions of the province. Characterization of these viral isolates furthermore

allowed us to show that a relationship exists between viral isolates which were obtained from dogs from the northern and northeastern regions of KwaZulu Natal, and jackal isolates which were obtained from the southeastern Mpumalanga region bordering western Swaziland, possibly implicating the role that these animals may play in the maintenance and persistence of the disease in the province. We were also able to demonstrate the possible role that the translocation of infected animals, probably via major highways, may have played in the long range transportation of the disease. The characterization of the variation that exists within the KZN family of viruses further also allowed us to compare the obtained sequence data, to the sequence data which were obtained from human infections, which had been diagnosed in the province in the period between 2002-2003. The major viral variants which were responsible for these infections, and the geographic regions they circulate in were consequently pinpointed .Finally, the data described here, enabled a clear elucidation of the factors that govern the maintenance and persistence of the epidemic, allowing us to suggest how currently used vaccination strategies may be improved in order to bring the current epidemic in the province under control This study comprises the first description of the molecular epidemiology of rabies in the South African province of KZN. The regional characterization of viral isolates from the province, provides an epidemiological framework, which will play an important role in surveillance efforts used to track the spread of the disease in the future. The data which was generated from the project should in addition also be applicable to the evaluation of the efficacy of potentially implemented future control strategies, in this region of South Africa.

4.2 Principal findings of this study

With referral to the spread of the rabies epidemic into and throughout the KwaZulu Natal province, this study provided evidence in support of some of the documented hypothesis and popular conjectures, which have been made from the literature. However our data provided evidence of some digression in some of the currently expressed opinions, and allowed for a revision of the underlying dynamics of the epidemiology of the disease in this part of South Africa. The most important findings which was made from the data which was generated during the course of this project, could be summarized as follows:

1. Comparison of the KwaZulu Natal isolates to isolates which were obtained from Zimbabwe and the northern regions of South Africa, indicated that the KZN isolates were more closely related to a viral lineage which circulated in the eastern regions of Zimbabwe, than to a second lineage which was obtained from canid host species from western Zimbabwe and northern South Africa. This finding provided evidence of a closer relationship between the isolates from KZN and isolates which were obtained from western Zimbabwe, thereby providing supportive evidence for the pattern of spread throughout the northern and eastern regions of South Africa which was presumed to have led to the introduction of rabies into the KZN province, as was suggested from historical surveilance data from the literature.

2. Viral isolates from KwaZulu Natal could be divided into two groups which shared a common ancestry and which were distinguishable from rabies virus isolates which were obtained from elsewhere in South Africa and Zimbabwe. The first group (Subfamily A) was isolated primarily from the eastern coastal regions while the second group (subfamily B) had a localized distribution in the northern regions of the province bordering Swaziland and southeastern Mpumalanga. The observation that these groups shared a common ancestry suggested that they were both introduced from the same geographic region, possibly from Maputo in southern Mozambique, as has been suggested previously in the literature. Furthermore the distribution of these viral groups in KwaZulu Natal suggested that at least two epidemic fronts spread from Mozambique into the KwaZulu Natal province, resulting in the observed distribution of these viral groups in the province. Subfamily B was thought to represent an epidemic front which had spread westwards from Maputo into northern Swaziland and southeastern Mpumalanga, from where it was introduced into northern KZN, while subfamily A was thought to represent a direct introduction of the virus into the northeastern coastal districts of the province.

3. The illustration that dog isolates from the northern and northeastern regions of KZN (subfamily B) clustered together with viral isolates which were obtained from jackal from the southeastern Mpumalanga region bordering western Swaziland, implicated the possible role that these animals might play in the epidemiology of the disease within KZN itself, even though sequence data from the province itself was not available for the confirmation of this hypothesis

4. The analysis indicated that the epidemic within KZN is being introduced through three different enzootic fronts. The first two of these fronts have been described in point 2 and corresponded to two sepperate introductions in the northern regions of the province, which led to the establishment of subfamily A and B viruses in their respective geographic regions of isolation. The third introduction was represented by viral group C of subfamily A, which formed the earliest divergent lineage of the subfamily A lineage of viruses. As previously mentioned the reconstructed topology for this group on the phylogenetic tree, concomitant with the groups observed geographic distribution was unexpected, as it was thought that earlier divergent lineages would occur in the northern regions of the province near the origin of the epidemic. We suggested that this group might represent a remnant infectious cycle left over from the 1964-1968 epidemic, which has managed to persists in the Eastern Cape, and which is currently being reintroduced across the southern border of the province.

5. Our analysis suggested that a single translocation event of the group A and B common ancestor occurred into southern KZN, at the same time as this common ancestor was being introduced into the northeastern coastal regions of the province (1976- 2003). This suggestion was supported by the observed distribution of viral group A and B in northern and southern KZN, as well as by the observation that group A and B (subfamily A) diverged from each other similtaneously, at a later time period than groups C. The reconstructed topology on the phylogenetic tree would be highly unexpected if it was assumed that the epidemic spread sequentially down the coast, as was previously suggested from the literature.

6. Characterization of the regional variation of rabies virus isolates from throughout the province indicated how the translocation of variants may have led to the outbreak of a KZN variant among jackal in the northern regions of Southern Africa, as well as why individual isolates belonging to particular viral groups were isolated from districts within the province which were geographically distant from the regions these variants generally circulate in. The regions where these isolates were obtained from, as well as the distribution of these viral groupings on a map of KZN, suggested that transportation of infected dogs, primarily via the N2 national

highway, may be responsible for the dissemination of the disease throughout the magisterial districts of the province.

7. The characterization of the genetic variation which is present between the viral variants which were obtained from the different magisterial districts of the province, allowed us to reconstruct a number of human case histories. These human cases had occurred in the KZN and EC province in the time period between 2002 to 2003.

8. The regional characterization of rabies virus isolates from KZN allowed us to make conjectures as to the factors which play a role in the maintenance and persistence of the disease in the province, and allowed us to explain why the second epidemic has managed to persist, even though vaccination campaigns were successful in bringing the first epidemic under control. We concluded that during the initial 1964-1968 epidemic, fewer maintenance cycles were present in the province. Fewer individual maintenance populations created a situation in which the disease was less able to persist over the long term. This combined with the targeted vaccination strategies which were implemented during the 4 year time span of the first epidemic, proved sufficient to bring it under control. This situation changed however in the time period preceding and following the outbreak of the second epidemic. Social and political changes, especially those which were associated with the fall of apartheid in 1994, gave new freedom of movement to people, and sparked a migration in the human population from the rural to the urban areas of the province. This led to the development of numerous informal settlements with large associated dog populations. This created a situation in which a larger number of individual maintenance populations existed in the larger geographic region as whole, thereby increasing the overall ability of the disease to persist over the long term.

9. The results which were obtained from our analysis allowed us to make suggestions as to how the current epidemic in the province may be brought under control. The illustration that rabies may be introduced into the province through three different enzootic fronts in the northern and southern regions of KZN, suggested that cordon vaccination be implemented in order to prevent the introduction of the disease. Phylogenetic characterization of the viral isolates which circulate in the distinct geographic regions throughout the province, further allowed us to demark the locations where the disease circulates among individual dog populations, which

may be targeted individually by control measures. Control of the disease among individual maintenance populations, coupled with more stringent legislation to prevent the movement and reintroduction of the disease from occuring, should provide a means by which the long term persistence mechanism of rabies in the province may be undermined.

4.3 Recommendations for the future

The study conducted on the molecular epidemiology of rabies in KwaZulu Natal identified a number of areas and countries in southern Africa in which molecular epidemiological data is either non-existent or incomplete. In order to complete the epidemiological picture of rabies in southern Africa towards the optimal efficacy of regional control campaigns, we advise that additional viral isolates from these regions should be characterized with respect to the relevant gene regions which were utilized during the course of this study. We recommend that molecular epidemiological studies be conducted in the countries of Mozambique, Lesotho and Swaziland, as well as the Eastern Cape province, in order to allow for the completion of the events which were hypothesized to have led to the introduction and spread of rabies into these regions of southern Africa. We also advise that additional jackal isolates be obtained from the northern and northeastern regions of KZN for inclusion in this dataset, to allow for the confirmation of our hypothesis which suggested that rabies is currently being maintained simultaneously in parallel dog and jackal cycles in the northern regions of the province. We also recommend that future studies on the molecular epidemiology of rabies in the African subcontinent should include a molecular clock analysis of the currently available sequence data, in order to provide additional confirmatory evidence as to the timing of the events which was hypothesized to have led to establishment of the disease withinin the province, as well as in southern Africa as a whole. In addition we suggest that sequence data from South Africa and its neigbouring countries (i.e. SEARG member states), should be submitted to a central sequence repository, which may be accessed over the Internet, or which may be distributed in a standalone electronic format such as CD-ROM or DVD. Such a database would allow for easy access to regional sequence data where internet access is not available, and may be designed to incorporate additional features such as real time phylogenetics, homology search options, and cartographic and GPS software. Such a database may represent a show case for Africa, and could be expanded upon in the future

to include sequence data from regions and host species from throughout the African continent.