Theodorus Bernardus Mollentze

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

I, Theodorus Bernardus Mollentze, declare that this thesis, which I hereby submit for the degree *Magister Scientiae* in Microbiology at the University of Pretoria, is my own work (except where indicated) and has not previously been submitted by me for a degree at this or any other tertiary institution.

_____ _____

Date Signature

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria
Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

Acknowledgements

A sincere thank you to all who have made this work possible:

- Professors Louis Nel and Wanda Markotter for input, constant guidance and support.
- Professor Dan Haydon and doctor Samuel Soubeyrand for guidance, endless patience and a lot of effort in designing these models.
- Doctors Roman Biek, Sunny Townsend and Katie Hampson for advice and critical discussions.
- Mr. Kevin le Roux and his staff at the KwaZulu Natal Department of Agriculture and Environmental Affairs for sample collection.
- My family for their endless support.
- The National Research Foundation of South Africa, the Poliomyelitis Research Foundation, the University of Pretoria and the Medical Research Council of the United Kingdom for providing funding.[1]

[1] Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to any of the funding bodies.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Summary

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

In the fight against endemic infectious diseases – which disproportionately affect the developing world – the effective use of scarce resources is of paramount importance. For vaccine preventable diseases, vaccination campaigns should be of optimal efficiency, a goal which is dependent on effective disease surveillance as well as a thorough understanding of the disease's spatial epidemiology. Several recent approaches show great promise in allowing us to understand the high resolution spatial aspects of epidemic disease spread following a single introduction, but do not account for the complexities inherent to endemic diseases. This thesis describes the development and use of novel techniques that can be applied to better understand endemic diseases and epidemics originating from multiple introductions, towards improved control and eventual elimination.

Using observation times and accurate geographic coordinates along with partial genome sequences of the virus, nearly 200 rabies cases detected in rabies endemic KwaZulu Natal province, South Africa over a 15 month period were analysed. Introductions from outside the province were found to be rare, while significant spatial-genetic clustering was shown to exist within the province. By reconstructing the causal links between cases at the individual case level, this clustering was shown to be the result of distinct transmission cycles. Although a large number of transmissions over long distances point to a significant anthropogenic influence in the transmission of rabies in the study area, the individual endemic clusters had very little contact and remained independent.

Demographic reconstructions showed that current control efforts are having a marked effect in reducing the number of cases throughout the region. Meanwhile, reconstruction of direct and indirect transmission links between cases allowed for an accurate inference of the true number of cases affecting the study area, which showed that surveillance in this region currently detects upwards of 50% of cases. Taken together, the results described here allow for the design of more directed control strategies tailored to the area under study, while also allowing for the improvement of

surveillance. These techniques can be applied to data from any directly transmitted endemic disease, allowing investigation of transmission dynamics at the individual case level.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Contents

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

January 2015

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

List of Figures

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

List of Tables

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Chapter 1.

Literature review

1.1. Introduction

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Rabies encephalitis, one of the oldest known, most feared and deadly zoonotic diseases, is caused by the globally distributed *Lyssavirus* genus of negative-stranded RNA viruses (World Health Organization, 2004). The type species of this genus, and the primary cause of rabies in most of the world, is rabies virus (RABV; World Health Organization, 2004; International Committee on Taxonomy of Viruses, 2013). RABV is primarily transmitted by biting and moves along peripheral nerves from the site of infection to the central nervous system, where it replicates and causes a serious and invariably fatal neurological disease (Warrell & Warrell, 2004). The *Lyssavirus* genus currently encompasses twelve recognized species, and three putative species – Bokeloh bat lyssavirus, Ikoma lyssavirus and Lleida bat lyssavirus – were recently described (Freuling *et al.*, 2011; Marston *et al.*, 2012; International Committee on Taxonomy of Viruses, 2013; Ceballos *et al.*, 2013). On a global scale, lyssaviruses display the highest diversity in Africa, which may indicate that this genus first evolved here: five of the recognized species as well as one putative species occur in Africa. Rabies virus has an almost global distribution, while Duvenhage virus, Lagos bat virus, Mokola virus, Shimoni bat virus and Ikoma lyssavirus are associated exclusively with Africa (although it should be noted that both Shimoni bat virus and Ikoma lyssavirus have only been isolated once; Rupprecht *et al.*, 2002; Markotter *et al.*, 2008a,b; Kuzmin *et al.*, 2010; Marston *et al.*, 2012).

Although RABV is usually transmitted by biting, with the virus inducing behavioural changes such as aggression in infected animals, the high numbers in which virus is shed in the saliva of its hosts means non-bite transmission can occasionally occur when the saliva comes into contact with the mucous membranes or broken skin of

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

susceptible animals (Rupprecht *et al.*, 2002). RABV can be transmitted by a wide range of mammals, but different variants of the virus are usually responsible for the transmission cycles in each host (Rupprecht *et al.*, 2002). Two such independently maintained variants are found in southern Africa. The canid variant circulates among members of the Canidae family, particularly domestic dogs (*Canis lupus familiaris*), jackals (*Canis mesomelas* and *Canis adustus*) and bat eared foxes (*Otocyon megalotis*), while the mongoose variant circulates among members of the Herpestidae - mainly the yellow and slender mongoose (*Cynictis penicillata* and *Galarella sanguinea*, respectively) (Von Teichman *et al.*, 1995; Sabeta *et al.*, 2003; Nel *et al.*, 2005; Sabeta *et al.*, 2007). While other wild animals, humans and livestock are susceptible to all variants of this virus, they are "dead-end" hosts that do not contribute to the maintenance or further spread of RABV (Childs, 2002).

Rabies remains endemic in much of the developing world, and it is one of 17 diseases regarded by the World Health Organisation as neglected tropical diseases (World Health Organization, 2002, 2013). Canine rabies is also endemic throughout large parts of South Africa, where it leads to between 10 and 30 human cases per year, the majority of which are in children (48.8%) and young adults (21.8%; Bishop *et al.*, 2010; Weyer *et al.*, 2011). Although these numbers are low, one should also consider the high cost of prevention, the losses incurred due to spill-over infections in livestock, and the traumatic nature of disease symptoms, leading entire communities to live in fear (Cleaveland, 1998). Rabies caused an estimated loss of 1 462 000 disability-adjusted life years globally in 2010 (95 % uncertainty interval: 852 000 to 2 659 000), down from 3 234 000 (95 % uncertainty interval: 1 866 000 to 6 509 000) disability-adjusted life years in 1990, but still a major problem (Murray *et al.*, 2013). However, quantifying the true burden of rabies with confidence is extremely difficult because rabies cases are widely believed to be vastly under-reported (Cleaveland, 1998; Nel, 2013).

Globally, more than 95% of human cases may be attributed to bites from rabid domestic dogs (World Health Organization, 2002; Cleaveland *et al.*, 2006). This is illustrated by the fact that the KwaZulu Natal province of South Africa (KZN), where rabies is most prevalent in domestic dogs, has also historically had the highest numbers of human cases (Bishop *et al.*, 2010; Weyer *et al.*, 2011). Between 1983 and 2007, 79% of the South African laboratory-confirmed human cases occurred in this province (Weyer *et al.*, 2011). For this reason, and because post-exposure treatment of humans is more expensive and does nothing to address the actual source

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

of infections, the elimination of human rabies will depend on the control of this disease in domestic dogs (Cleaveland, 1998; Nel *et al.*, 2009; Lembo *et al.*, 2010, 2011). This has been achieved in most of the developed world, through vaccination and control of dog movement (King *et al.*, 2004; Velasco-Villa *et al.*, 2008). In these countries, rabies still persists in wildlife, but human cases are extremely rare (King *et al.*, 2004; Velasco-Villa *et al.*, 2008). With this in mind, and to demonstrate the feasibility of rabies elimination in developing countries, three demonstration projects have been launched with support from the Bill and Melinda Gates Foundation and the World Health Organization – in the Visayas Archipelago of the Philippines, south-eastern Tanzania and KwaZulu Natal, South Africa (Lembo *et al.*, 2011). An overarching aim of these projects is to generate information about the challenges unique to each setting, and how they can be successfully overcome, with the view of applying the knowledge gained in future dog rabies elimination projects elsewhere (World Health Organization, 2010). A key determinant of the success of these campaigns will be an improved understanding of rabies transmission dynamics in the targeted regions. The work described here aims to contribute towards such knowledge through the use of both existing and novel techniques combining epidemiological and genetic data for the fine-scale resolution of epidemic dynamics.

1.2. Molecular epidemiology of canid RABV

1.2.1. Genomic variability

As members of the Mononegavirales, lyssaviruses have a single-stranded, negative-sense RNA genome approximately 12 000 nucleotides in length (Tordo & Poch, 1988). They share a conserved gene order, and five proteins are encoded: the phosphoprotein (P gene), RNA-dependent RNA polymerase (large, or L gene) and nucleoprotein (N gene) make up the inner capsid, while the matrix protein (M gene) and glycoprotein (G gene) respectively form part of the inner and outer layers of the envelope (Wunner *et al.*, 1988). RNA viruses evolve rapidly, a consequence of short generation times and the lack of proofreading activity in the virus-encoded RNA-dependent RNA polymerase responsible for replication of their genomes (Jenkins *et al.*, 2002). It is not surprising therefore that point mutation is one the major forces driving lyssavirus evolution (Badrane & Tordo, 2001). However, the substitution rate of RABV is relatively low compared to other RNA viruses. The synonymous

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

substitution rate in the N gene of European RABV isolates has been calculated as 5.27×10^{-4} ($\pm 0.23 \times 10^{-4}$) substitutions per site per year, while the synonymous substitution rate calculated for partial G gene sequences from the same dataset was 4.1×10^{-4} ($\pm 0.3 \times 10^{-4}$) substitutions per site per year (Holmes *et al.*, 2002).[1] By comparison, the synonymous substitution rate of of a spectrum of RNA viruses has been calculated to lie between 6.8×10^{-4} and 7.9×10^{-3} (Jenkins *et al.*, 2002).

This stochastic accumulation of point mutations is followed by strong purifying selection (Badrane & Tordo, 2001). This causes RABV to have a much lower non-synonymous substitution rate than the rate of accumulation of synonymous substitution rates, thus lowering the overall substitution rate. The non-synonymous substitution rate of European RABV isolates has been calculated as 2.85×10^{-5} ($\pm 0.265 \times 10^{-5}$) substitutions per site per year in the N gene and 5.06×10^{-5} ($\pm 0.85 \times 10^{-5}$) for partial G gene sequences (Holmes *et al.*, 2002). Meanwhile, the mean overall nucleotide substitution rates calculated for a globally representative group of RABV isolates was 2.3×10^{-4} (95% posterior interval [PI]: 1.1×10^{-4}–3.6×10^{-4}) substitutions per site per year for the N gene and 3.9×10^{-4} (95 % PI: 1.2×10^{-4}–6.5×10^{-4}) substitutions per site per year for the G gene (Bourhy *et al.*, 2008). The range of overall substitution rates for RNA viruses has been estimated as between 7.5×10^{-5} and 2.8×10^{-3} (Jenkins *et al.*, 2002). This strong purifying selection may be due to the diversity of cells that this virus infects, where an advantageous mutation in one cell type would be deleterious in other cell types or hosts, implying strong purifying selection one would not expect for many other viruses, along with the absence of immune selection which drives high mutation rates in the genes encoding viral surface proteins of viruses such as human immunodeficiency virus 1 and bovine respiratory syncitial virus (Holmes *et al.*, 2002; Johnson *et al.*, 2002). RABV spends most of its time in the central nervous system where it appears to remain hidden from view of immune cells (Murphy 1977; Charlton *et al.* 1997; Roy & Hooper 2008; section 1.3.1). A third explanation suggested by Holmes *et al.* (2002) is the influence of stochastic processes such as population bottlenecks, which would decrease the power of natural selection, including immune selection. However, Bourhy *et al.* (2008) showed that the ratio of non-synonymous to synonymous substitutions was much higher on external than internal branches of a maximum likelihood phylogenetic tree constructed from sequences representative of terrestrial (i.e. not bat-borne) RABV diversity from around the world. This indicates that

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

[1] These are the two genes most frequently used for phylogenetic analyses of lyssaviruses.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

November 2013

most of the mutations observed on external branches never become fixed (Pybus *et al.*, 2007; Bourhy *et al.*, 2008). If bottlenecks played a significant role in the evolutionary history of RABV, one would have expected some deleterious mutations to become fixed due to the stochasticity that this process introduces (Ohta, 1992; Bourhy *et al.*, 2008). Although transmission should theoretically constitute a population bottleneck for pathogens, recent studies of intra- and inter-host viral population dynamics suggest that the population bottlenecks during transmission are not necessarily narrow for all viruses (Murcia *et al.*, 2010). If this proves to be the case for RABV, it may go some way to explain why relatively few mutations become fixed, but this would also reduce the effectiveness of population bottlenecks in reducing both the genetic variation of RABV and the effect of diversifying selection.

1.2.2. Overview of the global genetic lineages of rabies virus

Although rabies has been reported in domestic dogs since antiquity (Neville, 2004; Baer, 2007), modern-day RABV appears to have originated much more recently. Phylogenetic evidence strongly supports the hypothesis that this virus first originated from other lyssaviruses infecting the Chiroptera, before spilling over to – and becoming established in – members of the Carnivora (Badrane & Tordo, 2001; Bourhy *et al.*, 2008). Through molecular clocking based on G gene sequences, Badrane & Tordo (2001) detected two separate host switches that appear to have occurred between 888 and 1459 years ago. Using a much larger dataset (almost 4 times larger), Bourhy *et al.* (2008) dated the most recent common ancestor (MRCA) of a globally representative dataset of Chiroptera- and Carnivora-associated RABV to 749 years ago (95 % PI: 363–1215 years ago) based on N gene sequences and 583 years (95 % PI: 222–1116 years ago) based on a smaller dataset of G gene sequences. However, this study included relatively few Chiropteran RABV isolates (25 sequences of bat RABV isolates versus 126 sequences of carnivore RABV isolates), and indeed the age estimate for the MRCA of all carnivore isolates included was similar to that of the whole dataset at 761 years old (95 % PI: 373–1222 years). Nevertheless, the phylogeny constructed by Bourhy *et al.* shows evidence of only one switch to the Carnivora, followed by a rapid radiation of lineages. It is possible that this host switch occured in southern India, since carnivoran isolates from the Indian subcontinent and Sri Lanka form the most divergent clade, occupying a basal position in phylogenetic reconstructions based on both the N- and G-gene (Bourhy *et al.*, 2008). Chiropteran RABV still causes fairly frequent spillovers into carnivores (Leslie *et al.*, 2006; Kuzmin *et al.*,

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. J.H. Nel & Prof. W. Markotter

27 September 2013

2012), and it is therefore likely that the lineage causing dog rabies in antiquity has since died out, followed by re-introduction from bats in more recent times, or that these cases were caused by a different lyssavirus that is now extinct (Badrane & Tordo, 2001; Bourhy *et al.*, 2008). It is however unclear why RABV is no longer found in bats outside of the Americas.

Phylogenetically, the carnivoran lineages of RABV strongly correlate to geographic location of origin, but also form distinct variants associated with specific host species (Smith *et al.*, 1992; Von Teichman *et al.*, 1995; Sabeta *et al.*, 2003; Bourhy *et al.*, 2008; Kuzmin *et al.*, 2012). On a global scale, Bourhy *et al.* (2008) found that RABV lineages from various carnivoran hosts are phylogenetically interspersed within dog lineages, which they argued reflects the role of dogs in disseminating rabies to the majority of the other carnivoran hosts. Using a 200 base pair partial sequence of the conserved N-gene, Smith *et al.* (1992) identified six distinct lineages differing by more than 10 % among 59 carnivoran RABV sequences (Table 1.1). While five of these lineages were also geographically distinct, Smith *et al.* identified a widely distributed lineage comprising of isolates from the US, Central and South America, North and Central Africa and Western Asia. This lineage also contained vaccine strains derived from historic European isolates, leading the authors to suggest that this "cosmopolitan" lineage was widely disseminated from Europe as a result of European colonialism. This hypothesis is supported by analysis of the substitution rate of *Lyssavirus* G genes, which dates the most recent common ancestor of cosmopolitan lineage RABV isolates to between 1493 and 1717 (Badrane & Tordo, 2001). Kissi *et al.* (1995) used *complete* N-gene sequences to identify eight lineages among the 69 global isolates of carnivoran RABV they examined (Table 1.1). Again, most lineages were associated with specific geographic areas, but an additional lineage (designated the Arctic lineage) was widely distributed in the northern hemisphere. In agreement with the results of Smith *et al.* (1992), some isolates from Africa and Latin America, along with all European and Middle Eastern isolates, were more closely related to each other and to the European-origin vaccine strains than to any of the other designated lineages. A subsequent study using complete sequences of both the N and G genes and a much larger sample size confirmed and expanded the wide distribution of both the Arctic-related and Cosmopolitan lineages (Bourhy *et al.*, 2008). Since the evolution of RABV is constrained by strong purifying selection (section 1.2.1), Bourhy *et al.* argued that these global phylogeographic patterns are the result of stochastic processes rather than natural selection. In this scenario,

much of the variation observed at smaller spatial and temporal scales as a result of range expansions (e.g. Biek *et al.*, 2007; Coetzee & Nel, 2007) are lost over time due to genetic drift, leading to the spatially and genetically distinct lineages observed at a global scale (Bourhy *et al.*, 2008). Of course, this also means that the picture presented in Table 1.1 may be distorted. Nevertheless, such global phylogeographic studies have value in that they allow for an accurate reconstruction of past events, although arguably the focus should move towards using full genomes if one hopes to gain an accurate picture of the origin of RABV.

Table 1.1: Global rabies virus lineages associated with the Carnivora

Smith *et al.* (1992)[a]	**Kissi *et al.* (1995)**	**Bourhy *et al.* (2008)**
Lineage I (widely distributed)	Africa 1 (North, Central and Southern Africa) / Europe/Middle East / Latin America 1[b]	Cosmopolitan
Lineage II (South Asia)	Asia	Asian
Lineage III (South-East Asia)		
Lineage IV (China)		
Lineage V (South-East Asia and China)		
	Africa 2 (West Africa)	Africa 2
	Africa 3 (Southern Africa)	Africa 3
	Arctic (widely distributed)	Arctic-related
		Indian subcontinent

[a] Adjacent cells indicate the designation used by the various authors, with many of the earlier lineages collapsed into single clades as more isolates were included.
[b] Latin America 2 is likely bat associated and therefore not included in this table.

1.2.3. African lineages of rabies virus

Surveillance for rabies is deficient throughout most of Africa, and it is difficult to get a clear view of RABV diversity on the continent, with the picture being skewed

towards southern Africa, where monitoring is more intensive (Swanepoel *et al.*, 1993). Nevertheless, four major clades of RABV belonging to three of the lineages described above have been identified in Africa and are believed to be descended from progenitor viruses introduced in different events (Kissi *et al.*, 1995; David *et al.*, 2007). The Africa 1 clade, a subset of the globally distributed Cosmopolitan lineage, is further divided into Africa 1a, found throughout North and West Africa as well as Madagascar, and Africa 1b, distributed throughout southern and East Africa, while the distinct Africa 2 clade is widely distributed throughout West Africa (Kissi *et al.* 1995; Bourhy *et al.* 2008). The Africa 3 clade represents the mongoose variant of RABV (section 1.3.2), found among the Herpestidae of the central plateau of southern Africa (Kissi *et al.*, 1995; Nel *et al.*, 2005; Davis *et al.*, 2007). In the study by Kissi *et al.* (1995), one isolate – from a human case in Egypt – did not group with any of the African lineages. After analysing RABV isolates from the Middle East and Egypt (including the isolate from Kissi *et al.*) David *et al.* (2007), designated a fourth clade (Africa 4), containing all four Egyptian isolates sequenced and a vaccine strain from Israel. Where this clade fits in the global diversity of RABV is unclear. The Africa 4 clade appears ancestral to European vaccine strains, which are in turn ancestral to the rest of the Cosmopolitan clade (David *et al.*, 2007). However, while there was strong bootstrap support for the branches leading to the Africa 4 clade (99 %), the branch leading from there to the Cosmopolitan clade received only 72 % bootstrap support. These Egyptian isolates are closely related to cases in the Cosmopolitan lineage, and from the present limited evidence it would appear that the Africa 4 clade is simply another subset of the Cosmopolitan lineage. This is also the classification used by Bourhy *et al.* (2008) for the single Egyptian RABV sequence available at the time (sequenced by Kissi *et al.* 1995).

1.2.4. The canid variant in South Africa

Despite the long history of canid-associated RABV on a global scale, this variant of the virus is relatively new in southern Africa (Swanepoel *et al.*, 1993). Phylogenetic evidence confirms that all canid RABV isolates in southern Africa stem from a common ancestor, and form part of the Africa 1b group of the Cosmopolitan RABV lineage, consistent with being introduced by European colonizers (Nel *et al.*, 1997; Davis *et al.*, 2007). The first confirmed outbreak of rabies in dogs in South Africa occurred between 1892 and 1894 in the Eastern Cape province, after importation of an infected dog from England (Swanepoel *et al.*, 1993). After eradication of this

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

epidemic, rabies was not confirmed again until 1928, when two children died following bites from a yellow mongoose (*Cynictis penicillata*; Swanepoel *et al.*, 1993). There had been occasional reports of a rabies-like disease following exposure to several members of the Herpestidae since at least the 18th century (reviewed by Swanepoel *et al.*, 1993), which was later confirmed to be caused by a genetically distinct and host-specific variant of rabies virus (Von Teichman *et al.* 1995). Despite occasional spill-over events, this variant has never become established in domestic dogs, which has limited its effect on humans (Von Teichman *et al.*, 1995; Nel *et al.*, 2005; Weyer *et al.*, 2011).

A second wave of canid rabies reportedly entered South Africa in 1950. The disease is believed to have spread from neighbouring Botswana into what is now the Limpopo province (Figure 1.1), where it first infected domestic dog populations before spreading to black-backed jackals (*Canis mesomelas*; Swanepoel *et al.*, 1993). The epidemic rapidly spread eastward, reportedly entering Mozambique in 1952, where it spread throughout the central and densely populated southern parts of the country (Swanepoel *et al.*, 1993). From here, the epidemic is believed to have spread to Swaziland in 1954 and finally to have re-entered South Africa from southern Mozambique in 1961, spreading through northern KZN into the densely populated coastal and central areas (Swanepoel *et al.*, 1993). Control efforts led to eradication of the disease in KZN by the end of 1968, but the epidemic is believed to have been re-introduced from Mozambique in 1976 when large numbers of refugees fled into South Africa from newly independent Mozambique (Swanepoel *et al.*, 1993). This second wave spread throughout KZN and into Lesotho (first detected in 1982) and the Eastern Cape province (in 1987), and has remained present and largely uncontrolled in these areas since (Swanepoel *et al.*, 1993; Bishop *et al.*, 2010). With the development of molecular phylogenetic techniques, it became possible to confirm the exact origin of new incursions. In 2002, rabies spread from Lesotho to the Free State province, an area where historically only the herpestid variant of RABV was found (Swanepoel *et al.*, 1993; Ngoepe *et al.*, 2009). This was followed by a re-introduction of rabies from Zimbabwe to dogs in the northern districts of Limpopo in 2005, which resulted in 21 confirmed human deaths and a further 9 suspected deaths (Cohen *et al.*, 2007). Although subsequent control efforts have caused a marked decline in the number of dog rabies cases in Limpopo (Figure 1.2), this newly introduced phylogenetic variant has become established in northern Limpopo (Sabeta *et al.*, 2011).

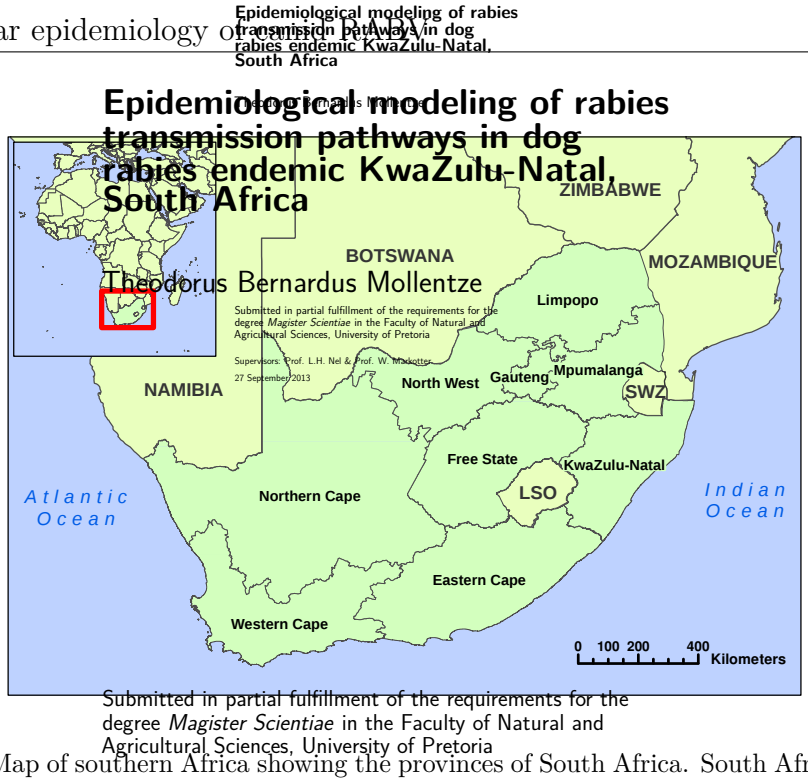

Figure 1.1: Map of southern Africa showing the provinces of South Africa. South African provinces are indicated in green, while neighbouring countries are shown in a lighter shade with names in capital letters (LSO: Lesotho, SWZ: Swaziland). Based on data from the Municipal Demarcations Board of South Africa (www.demarcation.org.za, accessed October 2012).

Since its reintroduction in 1961, RABV has spread throughout most of South Africa and is now endemic in many areas (Figure 1.2). This spread of rabies throughout South Africa and particularly eastern southern Africa has left a discernible fingerprint in the spatial-genetic distribution of RABV isolates from this region. RABV isolates from the Limpopo and North West provinces are closely related, and circulate mainly in black-backed jackals (Zulu *et al.*, 2009). Zulu *et al.* (2009) also found a close genetic relationship between isolates from Limpopo and some from Mpumalanga, but these appear to be merely the result of occasional translocation of dogs. Rabies has been under control in most of Mpumalanga for 20 years, and only recently expanded in range to cover a large part of this province (Mkhize *et al.*, 2010). In fact, more recent RABV isolates from Mpumalanga were found to be closely related and clustered with the single isolate available from Swaziland, with Limpopo isolates grouping ancestrally to isolates from Mozambique and KwaZulu Natal, which are in turn ancestral to the Mpumalanga isolates (Mkhize *et al.*, 2010). Isolates from KZN are also ancestral to those from the Eastern Cape and to the closely related isolates from Lesotho and the Free State province of South Africa (Coetzee & Nel, 2007; Ngoepe *et al.*, 2009; Weyer *et al.*, 2011). Biek *et al.* (2007) found a similar pattern of the present day spatial distribution being determined by events during

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

the initial colonisation of rabies in North America, and argued that this may be explained by the "surfing mutation" model of range expansion. This model arises from simulation studies showing that advantageous and neutral mutations (and possibly even slightly deleterious mutations) occurring at the leading edge of a range expansion are far more successful in surviving and becoming established than mutations arising later (Edmonds *et al.*, 2004; Klopfstein *et al.*, 2006). This is because the later mutations (or lineages) do not have the same opportunities of expansion, since the available niche is already filled (i.e. in the case of pathogens, less susceptible hosts are available to fuel range expansion; Biek *et al.*, 2007; Brunker *et al.*, 2012).

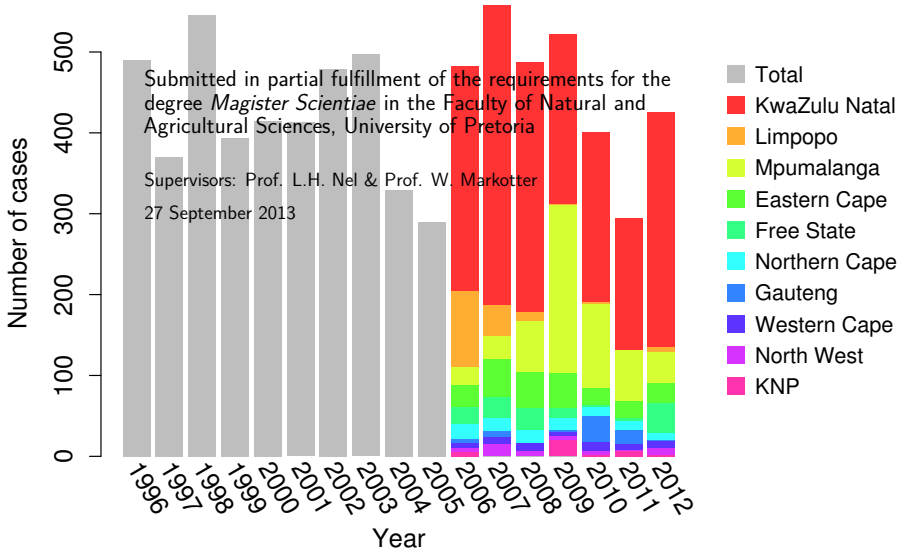

Figure 1.2: Annual incidence of animal rabies cases in South Africa as reported to the OIE (World Organisation for Animal Health) between 1996 and 2012. Data for individual provinces was only available from 2006 onwards (KNP indicates the Kruger National Park, shared between Limpopo and Mpumalanga). Based on data from OIE Handistatus 2 (1996 to 2004, www.oie.int/hs2/) and the World Animal Health Information Database (2005 to 2012, www.oie.int/wahis_2/public/wahid.php/Diseaseinformation/statusdetail).

1.3. Epidemiology of RABV

1.3.1. Pathogenesis

Because of its use of conserved receptors to enter cells, RABV can infect all mammals (Rupprecht *et al.*, 2002). The virus is primarily transmitted by biting, although any exposure where virus-laden saliva comes into contact with the peripheral nervous

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Molentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

system (e.g. exposure involving mucous membranes, cuts or grazes) can result in successful transmission in some cases (Murphy, 1977). RABV is known to replicate to some extent in striated muscle cells at the inoculation site before it enters the peripheral nervous system (Murphy & Bauer, 1974). However, it is thought that the virus may also enter the peripheral nervous system immediately if the inoculum is large (Fishbein & Robinson, 1993). From the edges of the peripheral nervous system, RABV particles are passively transported to the central nervous system (CNS), where the majority of replication takes place (Murphy, 1977). Clinically, rabies typically presents in either a furious form, involving behavioural changes and particularly aggression which accommodates the bite transmission of RABV, or a paralytic form, characterised by progressive paralysis. There is evidence to suggest that these different manifestations may be caused by the exact regions of the CNS in which the virus replicates, although it is unclear what causes such differential infection (Smart & Charlton, 1992; Mitrabhakdi *et al.*, 2005). From the CNS, virus particles spread passively along the remainder of the peripheral nerves to reach most organs, including the salivary glands, where the virus replicates further and is shed in high numbers into saliva (Yamamoto *et al.*, 1965; Murphy, 1977). Although this process is believed to be broadly conserved amongst all mammals and RABV variants, some variations have been noted. For example, there is little evidence of bite-transmission among greater kudu (*Tragelaphus strepsiceros*), the only mammal outside the Carnivora and Chiroptera known to maintain rabies independently (Barnard *et al.*, 1982; Mansfield *et al.*, 2006).

1.3.2. The reservoir host

An important concept in understanding the epidemiology of RABV and any other pathogen is that of the host or maintenance population. Traditionally, the reservoir host of an infectious disease has been defined as one that is capable of independent maintenance of the disease (Cleaveland & Dye, 1995). However, Haydon *et al.* (2002) argued that the concept of a reservoir host must be defined with reference to a target population, the population that one seeks to protect. They noted that multiple sub-populations (or species) might constitute a maintenance population even if some or indeed all of these sub-populations are incapable of maintaining the pathogen independently of the others. Thus, they defined a reservoir as "one or more epidemiologically connected populations or environments in which the pathogen can be permanently maintained and from which infection is transmitted to the

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Marker

27 September 2013

target population" (Haydon *et al.*, 2009). Despite the large number of animals that can be infected, only a few species will successfully transmit RABV and act as potential reservoirs. Domestic dogs are the primary reservoir of rabies throughout the developing world, but several wildlife hosts also exist, sometimes apparently acting as a part of the reservoir for infection of humans (*sensu* Haydon *et al.*, 2002). Lembo *et al.* (2008) found that although there is no evidence for independent maintenance of RABV in wildlife in the Serengeti, there is strong evidence to suggest frequent transmission between species. Thus, the wildlife hosts still form part of the maintenance population, even though they are not necessarily an indispensable component of it (Lembo *et al.*, 2008).

Numerous phylogenetic studies have reported that rabies appears to spread in several geographically distinct cycles in black-backed jackals (*Canis mesomelas*), side-striped jackals (*Canis adustus*) and bat-eared foxes (*Otocyon megalotis*) in southern Africa, with much more intra-species transmission than transmission to or from domestic dogs (Bingham *et al.*, 1999; Sabeta *et al.*, 2003, 2007; Zulu *et al.*, 2009; Sabeta *et al.*, 2011). That these species have sufficiently high densities to be able to maintain RABV independently of dogs has been disputed (Rhodes *et al.*, 1998), and is in contrast with results from elsewhere in Africa, where the virus was found to co-circulate in domestic dogs and wild canids (Cleaveland & Dye, 1995; Lembo *et al.*, 2007). Evidence for such co-circulation between dogs, jackals and foxes has also been found in South Africa (Sabeta *et al.*, 2007; Zulu *et al.*, 2009). However, in some areas with low human population densities (to which dog densities are closely correlated) RABV is isolated almost exclusively from bat-eared foxes (Sabeta *et al.*, 2007). Similarly, the phylogenetic clusters associated exclusively with jackals are found in western Limpopo, an area dominated by large game ranches and lower human densities (Zulu *et al.*, 2009; Sabeta *et al.*, 2011). Rhodes *et al.* (1998) concluded that jackal populations would need an average population density of 1.4 individuals/km^2, while the actual density of side-striped jackals observed was only 1 individual/km^2 in the 150 km^2 of commercial farmland they studied. From this they concluded that jackals in Zimbabwe would be unable to maintain RABV indefinitely in the absence of re-introductions from domestic dogs. Although it is likely that the demographic parameters and contact rates observed in this population (and therefore their estimate of the threshold density for RABV maintenance) will be broadly similar in both side-striped and black-backed jackals across southern Africa, it is entirely possible that different habitats could sustain higher densities of jackals.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Indeed, large populations of black-backed jackals have been reported in the game ranches and commercial farmland of Limpopo (Zulu *et al.*, 2009).

In addition to reports of independent cycles in these hosts, RABV seems to circulate in shared cycles in other areas. This situation lends further support to the idea of a metapopulation reservoir. Zulu *et al.* (2009) found that in the commercial farming areas of Limpopo and North West, viruses isolated from dogs and jackals clustered together, indicative of co-circulation, while clusters from the communal areas of Limpopo and Mpumalanga included only one jackal-associated isolate. It would appear therefore that in areas where both dogs and jackals are present, they have sufficient contact to maintain a shared epidemiological cycle, an observation with significant implications for control efforts. This is in agreement with the observations of Rhodes *et al.* (1998), which showed that the incidence of rabies in dogs and jackals were closely correlated in commercial farming areas in Zimbabwe. In areas where the virus co-circulates among hosts, controlling RABV in dogs should also lead to a reduction in the numbers of cases in jackals (Cleaveland & Dye, 1995; Rhodes *et al.*, 1998). However, data for the whole of Limpopo do not show a decrease in jackal cases despite a marked decrease in the number of cases in dogs in response to recent vaccination campaigns (Sabeta *et al.*, 2011). This again indicates that at least in some parts of Limpopo, rabies in jackals is independent from the disease in dogs. In the arid and sparsely populated Northern Cape where bat-eared foxes are prevalent, lack of water and food would seem to preclude high population densities. However, these same factors could increase the contact rate between individuals as they are forced to roam over wide areas. To date, no study has combined detailed epidemiological observations and modeling with phylogenetic data (an approach which would undoubtedly clarify this situation), but the phylogenetic observations are compelling. In fact, it is difficult to imagine an alternative explanation for the phylogenetic observations, unless very large numbers of dog cases (and thereby genetic diversity of RABV) are being missed. However, such an explanation might still not explain the persistence of RABV in large parts of the sparsely populated Northern Cape. Bingham (2005) argued that jackals in Zimbabwe lack the spatial separation required for them to show a metapopulation structure, and that this explains why RABV is not independently maintained in these animals. In dogs, such structure is thought to be critical in the long-term maintenance of RABV in the face of local extinction caused by its high virulence (section 1.3.4), with dogs having distinct sub-populations wherever there are human populations (Bingham, 2005).

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Although this argument could help explain the lack of RABV persistence in jackal populations throughout most of Africa, it does not explain why RABV appears to be maintained independently by jackals in parts of South Africa, or why South African jackal populations would be structured differently from those in the rest of southern Africa. In reality, it is likely a combination of host density and population structure that determine whether independent maintenance of this deadly virus is possible.

In theory, such co-circulation means the (meta)population that must be vaccinated to achieve rabies control is much larger than it would be otherwise (and less accessible if part of the reservoir is wildlife). Bingham (2005) argued that this means rabies control will require ever increasing resources and that current control strategies are unlikely to work in future. This may be true if the goal is to achieve total elimination of RABV. However, to protect a specific population (e.g. humans or endangered wildlife) it should be sufficient to control rabies in the sub-population that transmits the virus to this "target population". In the case of humans, this source of infection is domestic dogs in more than 99 % of cases (World Health Organization, 2002; Cleaveland *et al.*, 2006). Indeed, although rabies persisted in foxes throughout most of Europe until recently, cases of spill-over events to humans were rare (Malerczyk *et al.*, 2011; Freuling *et al.*, 2013). Although it could be argued that the free-roaming nature of dogs in Africa makes it much more difficult to prevent re-introduction to dog populations in Africa than it was in Europe, recent experience suggests otherwise. Despite the presence of seemingly independent wildlife reservoirs in the Limpopo province of South Africa, rabies was successfully brought under control (Sabeta *et al.*, 2011). In this province, the eventual source of re-introduction was not black-backed jackals, amongst which rabies remained endemic in some parts of the province, but domestic dogs in neighbouring Zimbabwe (Cohen *et al.*, 2007; Sabeta *et al.*, 2011).

1.3.3. Epidemiology of rabies in dogs

Although much more is known about the epidemiology of wildlife rabies, which affects the developed world, several recent studies have focussed on understanding the transmission of dog rabies in natural settings in Africa in more detail. Understanding the transmission dynamics of RABV has important applications in the design of control measures, where these parameters inform mathematical models. In Tanzania, where only the Africa 1b variant of canid RABV is present (which is also the only canid-associated lineage in South Africa, section 1.2.3 and section 1.2.4), the mean incubation period in dogs has been found to be 22.3 days (95% confidence interval

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W.T. Markotter

27 September 2013

[CI]: 20.0–25.0 days)(Hampson *et al.*, 2009). Although the incubation period is variable and extremely long incubation periods have been reported (reviewed by Fekadu, 1993), epidemiological models show that this has limited impact on the population-level dynamics of rabies (Cleaveland & Dye, 1995). This may be because of the short life-span of dogs in the developing world (2.4 to 3.4 years, Kitala *et al.* 2001), which means very few dogs would survive long enough to become infectious in the rare situations where a longer incubation period is required for some reason (Cleaveland & Dye, 1995). Following the incubation period, the salivary glands become infected (section 1.3.1), allowing the virus to be spread when other animals or humans are bitten. However, the time-window for transmission is quite small – dogs are infectious for only 3.1 days on average (95 % CI: 2.9–3.4 days) before succumbing to rabies, causing the number of secondary cases to be low in most cases (Hampson *et al.*, 2009).

Another consequence of this small infection window and the fatality of RABV infection is that rabies epidemics are cyclical (Childs *et al.*, 2000; Hampson *et al.*, 2007). The number of potential hosts rapidly decreases (Bingham, 2005), thus decreasing the number of new infections (a factor which is density-dependent; Beyer *et al.*, 2011). At this time, RABV may become extinct in the local population, only to be reintroduced from a neighbouring population once the local dog population has grown beyond a miminum threshold (section 1.3.4), or could persist with low levels of infection in larger populations (Beran & Frith, 1988; Bingham, 2005). In southern and eastern African dog populations, these cycles have been found to occur over a period of between 3 and 6 years (Hampson *et al.*, 2007). Hampson *et al.* (2007) argued that responsive vaccination campaigns actually sustain this periodicity and decrease the period between epidemics. This is because fewer dogs in the population are killed by the virus, which means the population recovers more rapidly. Thus, if vaccination is not sustained, the majority of these animals will be susceptible to infection in the next cycle of the epidemic (Hampson *et al.*, 2007). Carroll *et al.* (2010) showed that higher population growth rates or carrying capacity also lead to rabies epidemics occurring at increasing intervals.

Synchrony between rabies epidemic cycles in southern and eastern Africa has been found over distances of up to 1000 km, which Hampson *et al.* (2007) ascribed to possible repeated long-distance transportation of latently infected dogs or to the involvement of wide-ranging wildlife such as hyenas. Indeed, road-distance has been found to be a better predictor of rabies spread than geographic distance in North

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Africa, suggesting the involvement of humans – either by transporting infected, RABV incubating dogs or simply by providing accessible corridors for dog movement (perhaps linked to the availability of food along roads; Talbi *et al.*, 2010). In contrast however, contract tracing in Tanzania, which is more-or-less in the middle of the region studied by Hampson *et al.* (2007), showed that the mean distance of RABV transmission among dogs in this area was 0.88 km, and although the distribution of observed distances follows a long-tailed gamma distribution with transmission of up to 16 km, these were rare and all values within the 95% confidence interval was below 1 km (0.83–0.92 km; Hampson *et al.*, 2009). The involvement of humans in transporting rabid animals has frequently been cited as a major contributor to the long-distance dispersal of RABV (Denduangboripant *et al.*, 2005; Coetzee & Nel, 2007; Lardon *et al.*, 2010; Talbi *et al.*, 2010). Bourhy *et al.* (2008) noted that the strong subdivisions between globally distributed RABV sub-populations indicates that historically, humans have not been the major driving force behind RABV dispersal, although they undoubtedly play a role in the transcontinental transport of RABV, such as must have occurred to explain the distribution of the Cosmopolitan lineage of RABV (section 1.2.2). However, these global subdivisions could simply be a reflection of historical or current limitations of human movement by for example geopolitical boundaries, and very little is known about the transmission of rabies and the involvement of humans at the local, within-country level. Indeed, Talbi *et al.* (2010) found that transmission rarely crossed geopolitical borders in North Africa, which one would not expect from natural animal movement. Thus, a key unanswered question that remains is whether human-mediated movement may blur metapopulation boundaries to the extent of creating large, continuous endemic cycles, and at what scale independent disease control programmes will remain effective if this is the case.

A large amount of information regarding the factors that influence RABV transmission among sub-populations of dogs was generated by a recent study using bite-histories as an indirect way of measuring disease occurrence in rural Tanzania (Beyer *et al.*, 2011). This study evaluated several competing spatial models of rabies spread, each incorporating different combinations of epidemiological factors which might explain the observed patterns of spread. From the four best supported models (which were more-or-less equally well supported), it could be inferred that rabies transmission between villages is strongly influenced by the distance between the villages (negatively, i.e. decreasing with increasing distance) and the size of the

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magistar Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter
September 2013

dog population in the village receiving infection (positively). The size of the dog population in the village transmitting the infection seems to be less important – this parameter only mattered in the case of very small populations of less than 150 dogs, which have a lower probability of transmitting rabies. Crucially, these metapopulation models showed that vaccination targeting only the modelled district lead to a predicted 50% reduction in rabies occurrence, while eliminating external sources of infection through a broader vaccination campaign would lead to a 81.7% reduction in rabies occurrence (Beyer *et al.*, 2011). Clearly, it is important to investigate at what spatial scale vaccination should be applied and to understand the source of infections in a specific area before effective control strategies can be designed.

1.3.4. Mechanisms of persistence

Submitted in partial fulfillment of the requirements for the degree *Magistar Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter
September 2013

For a pathogen to persist, the rate of contact between susceptible hosts must be high enough to ensure continued transmission. Equally important however, particularly in the case of a pathogen such as RABV which rapidly kills its hosts after only a short transmission window, is a constant supply of susceptible hosts through new births or immigration (Cleaveland & Dye, 1995). These factors imply a minimum population size, often termed the critical community size, below which the epidemic will die out (Lembo *et al.*, 2008). These two factors also form the basis for control efforts against rabies and other diseases of domestic animals – restricting the movement of susceptible and infected hosts and more importantly reducing the number of susceptible hosts through vaccination (World Organisation for Animal Health & Food and Agriculture Organization, 2012; World Health Organization, 2004). Given the fatality of RABV, it may be expected that either the host birth rate or the critical community size would have to be extremely large to maintain a constant supply of hosts. At a small scale, Beran & Frith (1988) found that rabies was endemic and continuously present in parts of the city of Guayaquil, Ecuador that had dog densities higher than 680 unvaccinated dogs/km^2, but occurred only sporadically in areas with lower dog densities. However, RABV has been found to persist in populations as sparse as 5.3 dogs/km^2 in Tanzania (Cleaveland & Dye, 1995). Although several explanations involving atypical disease manifestations such as an infectious carrier state or recovery from illness can be suggested for persistence within host populations with such low densities (Fekadu, 1993; Cleaveland & Dye, 1995), there is very little evidence to support the existence of these phenomena (Rupprecht

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

et al., 2002). Bingham (2005) argued that the persistence takes place only at the metapopulation level, with frequent extinction and reintroduction in various local populations. This argument is in agreement with Haydon *et al.*'s definition of a reservoir host (section 1.3.2). This scenario would explain the conflicting results from Ecuador and Tanzania – in the latter setting examined by Beran & Frith (1988), high dog densities may lead to high contact rates, removing the metapopulation structure that allows rabies to persist in much smaller sub-populations in the rural regions of Tanzania examined by Cleaveland & Dye (1995).

In reality, persistence of RABV in many regions is likely to be the result of a combination of factors, including constant maintenance in densely populated areas with introductions to outlying areas as well as transmission between these more rural areas. The key point is that sub-populations smaller than the critical community size can still be integral to the maintenance of a pathogen, provided there is sufficient contact between it and other sub-populations of susceptible hosts, a fact which provides both challenges and opportunities for control. While long-distance transmission may lead to self-perpetuating epidemics over large areas, targeting control efforts at key areas or transmission links between small sub-populations has the potential to disrupt maintenance without having to reach the entire population (Figure 1.3).

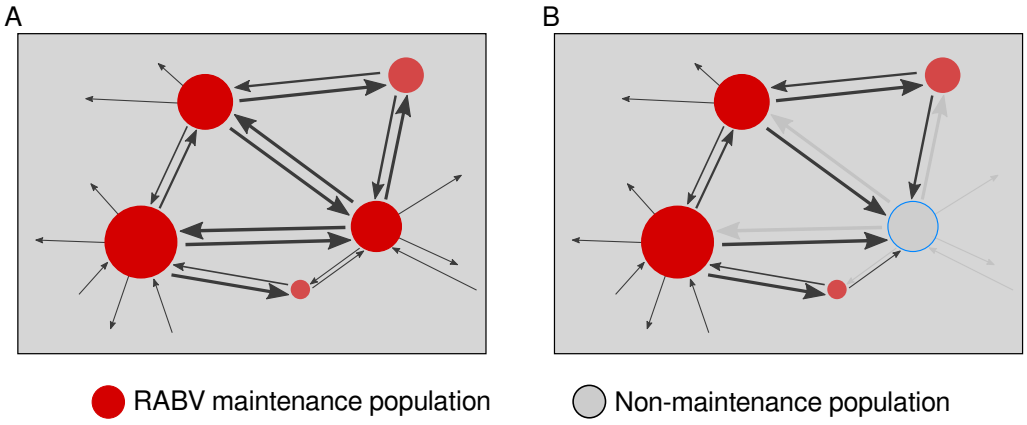

Figure 1.3: Maintenance of rabies in a metapopulation and potential targets for control. **A:** In a continuous, heterogeneous landscape, different populations of hosts may be involved in maintenance of RABV to varying degrees because of various factors such as size of the host population, population turn-over, etc. (indicated by colour between red [capable of independent maintenance] and gray [incapable of indefinite maintenance]). These populations seed cases to other populations to varying degrees (indicated by arrow size), which is again determined by various intrinsic and extrinsic factors. **B:** If the key connections between populations can be identified, these can be prioritised in control programmes to disrupt the continuous cycle of re-introduction, before targeting specific maintenance populations.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

1.3.5. Understanding transmission at the individual case level

A key component of effective control of infectious diseases is an understanding of the various factors that influence transmission through the heterogeneous (meta)populations and across the heterogeneous landscape in which it occurs (Ferguson *et al.*, 2003; Keeling *et al.*, 2003; Beyer *et al.*, 2011). Towards this goal, the traditional approach in epidemiology has been to study contact patterns between hosts. This has yielded great success in the study of infectious diseases of humans and some infectious diseases of agricultural crops and livestock where records of human contacts or of plant or animal movements are often available (e.g. Ramstedt *et al.*, 1990; Kiss *et al.*, 2006), or where the memorable nature of disease symptoms allow for detailed contact tracing after disease spread (e.g. Lembo *et al.*, 2008; Hampson *et al.*, 2009). However, not all contacts result in disease transmission, and establishing the exact causal relationships between cases from such data can be subjective, particularly when many cases occur over a short time period in close proximity to each other or when cases may be introduced from further afield.

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

A more rigorous method of establishing causal relationships is to consider the genetic relationships of the infecting agents. Early attempts at understanding spatial transmission patterns by integrating spatio-temporal data with genetics involved inferring some sort of phylogeny or parsimony network from the genetic data (in some cases also including temporal data), before considering spatial and other epidemiological information related to cases in a second step (e.g. Biek *et al.*, 2007; Cottam *et al.*, 2008). However, this ignores phylogenetic uncertainty and does not adequately accommodate the covariance between the genetic, spatial and temporal data (Lemey *et al.*, 2009). This is problematic, because not only are these three data types influenced by the same transmission process, but they also influence each other both directly and indirectly through their own influence on transmission, and it is this very system which we are trying to understand (Figure 1.4).

More recently, two distinct approaches to solving this problem have been developed in parallel, based on advances in population genetic and epidemiological modeling respectively. Both these methods analyse all three data types in an integrated, model-based setting, thus allowing all sources of uncertainty and correlation to be accommodated. In addition, both are Bayesian and investigate the relationships between cases using Monte Carlo Markov chain (MCMC) sampling techniques, thus allowing the incorporation of prior knowledge and the simultaneous estimation of various unknown model parameters (Drummond *et al.*, 2002; Ypma *et al.*, 2012).

Because they approach the problem from entirely different angles (and in fact from entirely different fields), these methods have quite different strengths and weaknesses, and arguably complement each other. While the exact choice of method will of course be driven by the question at hand, care should also be taken to account for the assumptions of each model.

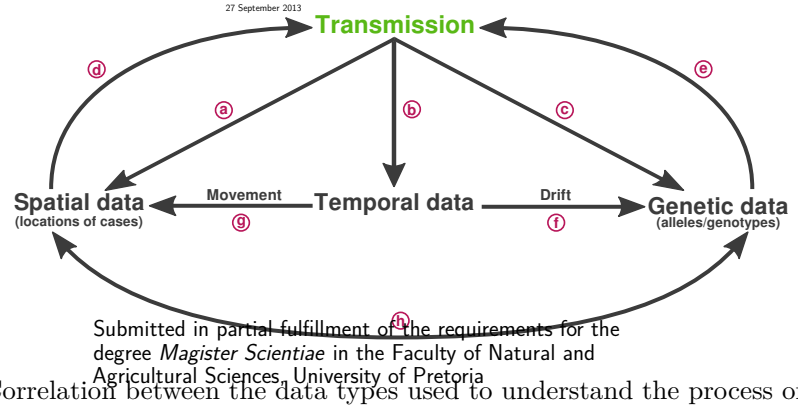

Figure 1.4: Correlation between the data types used to understand the process of transmission. Spatial, temporal and genetic data can be used to understand the unseen process of transmission because each transmission involves movement of the pathogen (a; section 1.3.3), is sequential in time, and the timing between one transmission and the next is determined by, and can be predicted from knowledge of, the pathogen's epidemiology (b; section 1.3.3), and each transmission event is a population bottleneck for the transmitted pathogen, potentially leading to the fixation of observable genetic changes (c; section 1.2.1). However, because these processes occur on the same time-scale (Grenfell *et al.*, 2004) and are all influenced by the same process of transmission (Pybus *et al.*, 2012) and also in turn influence this process, they indirectly affect each other – circulation of the pathogen in a given area decreases the number of available hosts, leading to a decrease in the rate of transmission (d; section 1.3.4), while accumulation of mutations over time may eventually lead to subtle changes in the epidemiology of the pathogen, thus also affecting the rate of transmission (e). Concomitantly, the time between transmissions determines the amount of mutations that can accumulate between observed cases (f; section 1.2.1), and also how far infected hosts (or virus particles) can move, thus determining the spatial influence of every infection (g; section 1.3.3). In addition, the relatively small spatial influence of every infection compared to the size of the area under study means transmission is a local process, causing spatial clustering of genetically related cases (h).

1.3.5.1. Approaches based on population genetic modeling

The first of these approaches uses standard nucleotide or amino acid substitution models in conjunction with a coalescent model that relates the timing and genetic divergence between cases to population genetic processes, with a diffusion model to account for movement of the pathogen across geographic space (Lemey *et al.*, 2009, 2010). These models can simultaneously be fitted to the observed data while co-estimating a phylogeny, which is used to record the genetic relationships between cases (Lemey *et al.*, 2010).

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

The coalescent model was designed to accommodate correlation between data-types caused by the genetic and spatial relatedness of descendants from a common ancestor (Kingman, 1982; h in Figure 1.4). It has since been expanded in a variety of ways to accommodate different population genetic processes and to study different aspects of evolution (Kingman, 2000). The core concept of all of these variants remains the same however, and involves working backwards from the genomic or sub-genomic sequences of n individuals sampled from a large population to find common ancestors for pairs of sequences, thus building a genealogy for the sampled individuals which expresses when in the past individuals shared common ancestors (Kingman, 1982). This is done by using a model of genetic drift to calculate the probability that the sequences from two individuals coalesce (i.e. that the sequences are identical) in the preceding generations (Kingman, 1982, 2000). Although it is possible to use other models of genetic drift, the original coalescent model used the neutral Wright-Fisher model, a model which assumes a neutrally evolving locus, non-overlapping generations and which Kingman (1982) noted essentially means that each individual chooses it ancestor independently and at random from all possible sequences in the previous generation. This means the probability (and hence the rate) of coalescence is determined by the total size of the population[2] (N) from which the n individuals were sampled (Kingman, 2000). Variants of the original coalescent model accommodating a wide variety of demographic processes have been developed by simply allowing the value of N to vary in different steps according to the assumed demographic process (e.g. exponential growth following an introduction or logistic growth typical of adaptation before rapid population growth and finally depletion of hosts; Kingman, 1982; Pybus & Rambaut, 2002). More recent advances also allow the demographic process in question to be inferred directly from sequence data while estimating the genealogy (Pybus & Rambaut, 2002; Minin *et al.*, 2008; Drummond *et al.*, 2005). Drummond *et al.* (2002) developed a method of rescaling the number of generations between coalescent events to calendar time by taking into account known dates at internal nodes (e.g. dates estimated from known fossil ancestors of a set of individuals) or at the tips of the tree (e.g. pathogen sequences sampled at different times during an epidemic) or by taking into account the rate of sequence evolution under a particular model of nucleotide or amino acid substitution.

[2]More accurately this should be the *effective* population size, N_e, which is the number of individuals in an idealised population (i.e. a population conforming to the simplifying assumptions made by the specific model of genetic drift) with the same distribution of alleles as the sampled population.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

The framework designed by Drummond *et al.* (2002) forms the basis of the now widely used BEAST (Bayesian evolutionary analysis by sampling trees) package, a suite of programs for the testing of evolutionary hypotheses and estimation of evolution-related parameters with a focus on the integration of genetic and temporal data (Drummond & Rambaut, 2007).

This assemblage of evolutionary models has been coupled with models of spatial diffusion, the first of which considered diffusion between discrete locations (e.g. towns or countries). This model acts in a manner similar to the techniques used for ancestral state reconstruction, modeling rates of changes in location states through a continuous time Markov process[3] (Lemey *et al.*, 2009). From these pairwise instantaneous transition rates, it is relatively straightforward to calculate the probability of a descendant having transitioned to a different state from its ancestor during the time between them (Schluter *et al.*, 1997). This timing between the occurrence of a descendent and its ancestor can be calculated from their genetic divergence using the framework of Drummond *et al.* (2002) described above. The most parsimonious migration process will invoke the smallest number of transitions between different locations, which means most of these transition rates (or 'migration rates') will have a high probability of being zero (Lemey *et al.*, 2009). Lemey *et al.* adapted a sampling technique called Bayesian stochastic search variable selection in order to estimate which rates should actually remain zero while separately estimating rates of migration that explain the phylogeographic patterns observed, thus discarding inferred migration scenarios invoking large numbers of small migration rates in favour of the most parsimonious scenario. Following the Markov chain Monte Carlo inference, a Bayes factor test can be used to determine the statistical support for invoking each type of migration (e.g. from location A to B) in explaining the observed spatial-genetic patterns. Thus, this model allows the most parsimonious migration history to be inferred for the pathogen, whilst allowing for a measure of the confidence in this inference, in contrast to explicitly parsimony-based analyses. This technique has been used to test different hypotheses regarding the spread of rabies in North Africa by using different combinations of prior distributions for the probability that particular transition rates are zero (i.e. allowing migration between some locations to be more likely to be invoked to explain the observed data than others; Talbi *et al.*, 2010). Various parameters can be used to inform the choice of these priors, such as

[3] A Markov process is a process in which the probability of the next state or value is only conditional on the present state, and not on any earlier states.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

simple geographic distance between the different locations (Lemey *et al.*, 2009) or measures of accessibility (Tatem *et al.*, 2010).

The discrete diffusion model can only account for migration between sampled locations, and hence requires very thorough sampling, and also does not allow migration rates to vary over time (Lemey *et al.*, 2009). These problems are overcome by continuous diffusion models where movement from the sampled locations is modelled by a stochastic process which has been termed a relaxed random walk (Lemey *et al.*, 2010). By treating the landscape as a continuous entity rather than focussing on discrete locations, unsampled locations can be invoked as sources. As with the discrete reconstructions, the reconstruction of ancestral locations is analogous to ancestral state reconstruction, where random walks have been used to reconstruct continuous traits for some time (Schluter *et al.*, 1997). In this case, the random walk is used to directly model the diffusion process and thus to infer probable locations for the ancestors of the observed cases (the internal nodes) by constructing a Markov chain of 2 dimensional locations sampled from a mutivariate normal distribution for each migration along a branch to the previous node (Lemey *et al.*, 2010). Although the directionality of such a diffusion process may seem arbitrary, constraints are imposed by the phylogeny, because the diffusion processes along multiple branches must converge on a single location at the root of the tree. The multivariate normal distribution describes the probability of any given location being sampled at every step of the migration process, which means the rate of migration can be modelled by changing the shape of the distribution so that more distant locations become less or more likely to be chosen. However, in the case of location change, a simple random walk drawing the incremental location changes from a single probability distribution would only model Brownian diffusion, where the rate of migration from any one location to another is fixed in time and over the entire phylogenetic history (Lemey *et al.*, 2010). To allow variation in the migration rates among branches, Lemey *et al.* used the same strategy as that used in the relaxed clock model (Drummond *et al.*, 2006). In this approach, independent scalars alter the migration rates of different branches of the phylogeny by rescaling the probability distribution from which new locations are drawn. These rate scalars are themselves drawn from a second probability distribution, thus allowing variation in rates whilst still maintaining control over the type and amount of variation[4]. Assuming a gamma

[4]Note that the rate scalars are drawn *independently and identically* (Lemey *et al.*, 2010), i.e. drawing a specific rate scalar does not affect the probability of drawing any other rate scalars subsequently.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

prior distribution for the rate scalers means new locations are drawn from a Student's t-distribution, which has heavier tails than the normal distribution (Lemey *et al.*, 2010). This means locations further from the mean become more likely, and results in a range of dispersal processes termed Lévy flights (Lemey *et al.*, 2010), which have been shown to accurately approximate the movement of foraging animals for example (reviewed by Reynolds & Rhodes, 2009).

This combination of evolutionary and geographic diffusion models have primarily been used to investigate various aspects of the past invasion dynamics of epidemics (Lemey *et al.*, 2009, 2010; Raghwani *et al.*, 2011). However, Pybus *et al.* (2012) showed that it can also be used to calculate spatial epidemiological parameters. These authors argued that a phylogeny can be used to correct for the spatial autocorrelation between genetically related cases because its branching pattern represents a record of the history of transmission. While this is true, Jombart *et al.* (2011) noted that the phylogeny is actually only an approximation of the transmission tree, which would represent the true branching structure of transmission. This is because phylogenies assume a structure where sampled cases only form tips of the tree, while in reality a cross-sectional sample from an epidemic would contain both ancestors (i.e. internal nodes) and descendants (Jombart *et al.*, 2011). In addition, the fact that the structure of the phylogeny takes only genetic data and the date of sampling into account may limit its resolution. Resolution of epidemic dynamics at the individual case level would require at least one mutation per transmission, which is unlikely for rabies virus (section 1.2.1). Furthermore, although it is straightforward to calculate spatial epidemiological parameters from the inferred ancestral locations and branching patterns (Pybus *et al.*, 2012), the lack of an explicit epidemiological formulation makes the calculation of other parameters describing the transmission process more difficult. However, recent progress has been made in linking epidemiological and coalescent models (Volz *et al.*, 2009; Frost & Volz, 2010), which, if implemented into the framework described above may remedy this last problem. As an alternative to coalescent models in the above framework, it may also be possible to use the birth-death models developed by Stadler *et al.* (2012). These models do have an epidemiological formulation and have been used within the BEAST framework to estimate the basic reproductive number of HIV from sequentially sampled genetic data and their date of sampling.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

1.3.5.2. Approaches based on epidemiological modeling

A more direct approach to understanding transmission is to directly reconstruct the series of transmission events linking cases or areas (thus using the actual transmission tree to account for spatial-temporal-genetic autocorrelation). If the aim is to understand transmission (rather than evolution), this approach has the advantage of having a direct link to the epidemiological processes that generated the data, making the investigation of epidemiological parameters much simpler. Directly modeling the epidemiological processes also means that this technique can be tailored to specific pathogens simply by incorporating existing knowledge of epidemiological parameters such as the incubation and infectious periods as prior information, which should increase its power to discriminate between closely related cases.

Numerous statistical approaches have been developed for the reconstruction of pathogen transmission trees (e.g. Haydon *et al.*, 2003; Cottam *et al.*, 2008; Heijne *et al.*, 2009; Cauchemez *et al.*, 2011; Ypma *et al.*, 2012), but very few make use of genetic data. Cottam *et al.* (2008) showed that using genetic data in addition to epidemiological data significantly reduces the number of possible transmission trees and allows for a better assessment of the likelihood of different trees. This is even better illustrated in an integrated framework, where Ypma *et al.* (2012) showed that a combination of temporal and genetic data leads to significantly better resolution when compared to transmission trees constructed from just geographic and temporal data, while the trees are even better resolved when all three data types are combined.

As discussed above, early approaches to estimating transmission trees from a combination of epidemiological and genetic data used a sequential approach, which has several disadvantages. The first attempt at combining the different data types in a single analysis was by Ypma *et al.* (2012), who designed a single, modular likelihood function to calculate the probability of transmission trees sampled from the space of all possible trees using an MCMC sampling technique. This likelihood function took into account the date of infection, date of culling, geographic location and a genome sequence of each farm infected with avian influenza A virus (H7N7) during a recent epidemic in The Netherlands. The time component of their likelihood function calculated the probability of potential source farms being infectious at the date that a given farm was infected, with the probability of a farm acting as a source declining exponentially after all poultry on it has been culled (this is necessary because influenza virus can be transmitted via fomites remaining on the farm). The geographic component of the likelihood function used a distance kernel that

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

describes the relationship between transmission and geographic distance such that closely located farms are more likely to act as sources. The genetic component consisted of a likelihood function that assigns different probabilities to transition, transversion and deletion mutations occurring between the source and the sink farm (the exact probabilities were co-estimated with the transmission tree and other unknown parameters). Using a single likelihood function allowed Ypma *et al.* to take into account the sources of uncertainty arising in the different components that determine the transmission tree. However, these authors made a number of simplifying assumptions, many of which may influence the resolution and accuracy of this method when applied to diseases with more complex epidemiological features. For example, they assumed that the dates of infection of each farm was known without error, and also did not account for variability in the incubation period (Ypma *et al.*, 2012). This last assumption would be particularly problematic for RABV, which is known to have variable and occasionally extremely long incubation periods (section 1.3.3). Another key assumption acknowledged by Ypma *et al.* is that the different components were assumed to be independent. Thus, while the reconstructed transmission tree can be used to account for the spatial-genetic autocorrelation between observations (h in Fig. Figure 1.4) when estimating spatial/epidemiological parameters, it does not account for many other sources of correlation, such as the accumulation of mutations over time (Ypma *et al.*, 2012).

Morelli *et al.* (2012) addressed these problems by constructing a joint likelihood function taking into account the major sources of correlation in their data of a foot-and-mouth disease virus epidemic in the United Kingdom. This likelihood function assesses the probability of a potential source having infected a given farm by using the date on which infection was first reported, the suspected date of infection, the date on which the host-population was culled, the location and a genome sequence for each farm. Like Ypma *et al.*, the causal relationships between farms were assessed in a Bayesian setting, allowing efficient sampling from the space of all possible transmission links using MCMC simulation. In this approach, the probability of a given farm acting as a source of infection to another farm was determined by its infectious period compared to the date of infection of the sink farm, its geographic distance from the sink farm, and the probability of the sequence observed at the sink farm arising from that observed at the source farm in the time that elapsed between the observations (Morelli *et al.*, 2012). The probability of a farm being infectious at any given time and the dates of farms becoming infectious was calculated using

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

an epidemiological model. In this model, the paths of symptoms first appearing are drawn from a probability distribution centred around estimates from experts based on disease symptoms, thus taking into account uncertainty in these estimates. Farms first become infectious on the same day that symptoms appear and remain infectious until the date of culling, which was known with certainty. The probability of specific dates of infection was calculated by drawing the unknown duration of the incubation period between infection and the appearance of symptoms from another probability distribution (whose shape was co-estimated in the analysis). Once the range and probability of possible infection dates were known, Morelli *et al.* calculated the probability of the observed genetic sequence at the sink farm arising in the time between potential source farms and every subsequently infected farm using a simplified model of mutation which affords all types of mutations the same probability (equivalent to the Jukes-Cantor model of nucleotide substitution; Jukes & Cantor, 1969). The geographic distance between farms was accommodated in the joint likelihood equation using an exponentially declining transmission kernel, dependent only on the great-circle distance between farms.

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Although both these approaches were developed to reconstruct the transmission tree linking infected farms, this was merely due the limited granularity of the data available. Both approaches should theoretically be able to reconstruct causal relationships down to the individual case level if sufficiently resolved data are available (Ypma *et al.*, 2012; Morelli *et al.*, 2012). However, exactly how much resolution is required in each of the data types (spatial, temporal and genetic), and how feasible it is to generate large volumes of such data remains to be seen. These factors can be expected to vary with the epidemiology of different pathogens, since the genetic data of infections with shorter incubation and infectious periods will necessarily contain less differences between every infection. Morelli *et al.* showed that their technique was accurate using simulated data, but it failed to resolve some clusters of cases when applied to observed data, with some competing transmission links having fairly high posterior probabilities. This was despite the fact that the dates of symptoms first appearing was known with high confidence and despite a mean substitution rate of 2×10^{-5} substitutions per site per day during one of the epidemics analysed (Morelli *et al.*, 2012). However, the amount of resolution gained from genetic data by Morelli *et al.*'s technique could be improved considerably using a more realistic model of nucleotide substitution, preferably one tested for its fit to the data in question using techniques such as those described by Posada & Crandall (1998). The model

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

of Morelli *et al.* also assumed that every farm contains just one sequence, and that all cases are observed. Increasing the resolution of sampling and accounting for the contribution of unsampled farms should increase resolution even further.

What all of the approaches described above have in common is their focus on relatively simple epidemic scenarios. All assume that the epidemic arises from a single index case and proceeds from there with no further introductions and with all or nearly all cases being observed. Thus, significant improvements will be needed if we hope to apply such techniques to endemic diseases.

1.4. Aims of this study

Although rabies has been successfully eliminated from domestic dogs in North America and most of Europe, it remains firmly established throughout much of the developing world. If we hope to control the disease and save human lives in these areas, sustainable, epidemiologically informed strategies that make the best possible use of the scarce resources available for this task will be needed. In recent years, an international collaboration funded by the Bill and Melinda Gates Foundation and coordinated by the World Health Organisation was established to show that canine rabies elimination is both feasible and sustainable in the developing world (Lembo *et al.*, 2011). The work described here forms part of a scientific support programme funded by the UK Medical Research Council aiming to help make this possible. A key requirement of all strategies developed for the three demonstration sites is that they should be broadly applicable and transferable to other sites, towards eventual elimination of rabies in domestic dogs.

The principle aim of this study was to develop a coherent strategy for better understanding the spatial epidemiology of endemic rabies, which would allow for the design of more effective and efficient vaccination campaigns. Specifically:

1. To investigate the current rabies situation in the KwaZulu Natal province of South Africa

   Using existing molecular epidemiology tools, the goal was to compare the present situation after several years of active rabies control programmes to the picture given by historic data. In addition to providing important data that could be used to assess the efficacy of current control strategies in the face of incomplete surveillance, this analysis was required to better understand

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

the available data prior to the following steps, and to place results from more detailed analyses in context.

2. To develop an approach for reconstructing the transmission trees linking cases of endemic rabies

   Although recent advances in the field of spatial infectious disease epidemiology allow detailed investigation of the transmission dynamics of epidemic diseases at the single case-level, these methods are not compatible with endemic infectious diseases. Thus, a key aim was to expand the scope of such methods to allow investigation of rabies transmission in unprecedented detail. At the same time, it was important to keep the approach general enough to allow its application to other endemic diseases.

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

3. To apply this newly developed technique to data from the canine rabies elimination project in KwaZulu Natal

   This would allow for a proof of concept using real-world data, and would also lead to a better understanding of the evolving endemic situation in this area.

4. To make specific recommendations that will allow improvement of the rabies elimination programme in this area.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Chapter 2.

Molecular epidemiology of rabies in KwaZulu Natal, South Africa

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

2.1. Introduction

Despite the availability of safe and effective vaccines against rabies virus (RABV), rabies remains endemic throughout most of the developing world (World Health Organization, 2002). Two genetically distinct and host-specific variants (previously termed 'biotypes') of rabies virus are known to circulate within South Africa, one affecting small mammals of the Herpestidae family, particularly mongooses, and the other affecting members of the Canidae (Von Teichman *et al.*, 1995). The Canidae-associated variant appears to have several wildlife hosts in southern Africa (Sabeta *et al.*, 2003, 2007; Zulu *et al.*, 2009), but the source of most human infections is domestic dogs, the primary host of this variant, and it is in these animals that the disease must be controlled if we hope to eliminate the disease in humans (Sabeta *et al.*, 2003, 2007; Zulu *et al.*, 2009; Weyer *et al.*, 2011; Lembo *et al.*, 2011).

In South Africa, the rabies problem is particularly acute in the KwaZulu Natal province (KZN). Apart from during recent epidemic outbreaks in other provinces – and despite the presence of active control programmes – the overwhelming majority of South African cases in both domestic dogs and humans occur in KZN (Figure 1.2; Bishop *et al.*, 2010; Weyer *et al.*, 2011; Lembo *et al.*, 2011). The success of RABV in this province has been ascribed to the poverty affecting this area (Brown, 2010), but the true causes are likely to be much more complex. KZN is the second most populous province in South Africa, and is home to approximately 10 267 300 people (19.8 % of the South African population; Statistics South Africa, 2012). As is common throughout South Africa and many other sub-Saharan African countries, a significant

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfilment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter
September 2013

proportion of the population can be classified as migratory labourers, with frequent circular migration both within KZN and between different provinces of South Africa (Posel & Casale, 2003; Posel & Marx, 2013). KZN shares international borders with three countries and provincial borders with two separately administrated provinces (Figure 2.1), and rabies is endemic throughout the region (Swanepoel *et al.*, 1993; Coetzee & Nel, 2007; Ngoepe *et al.*, 2009; Mkhize *et al.*, 2010). These factors, together with the wide distribution of dwellings in rural areas and the hilly terrain in most areas, creates unique problems for rabies control and the establishment of a rabies free area.

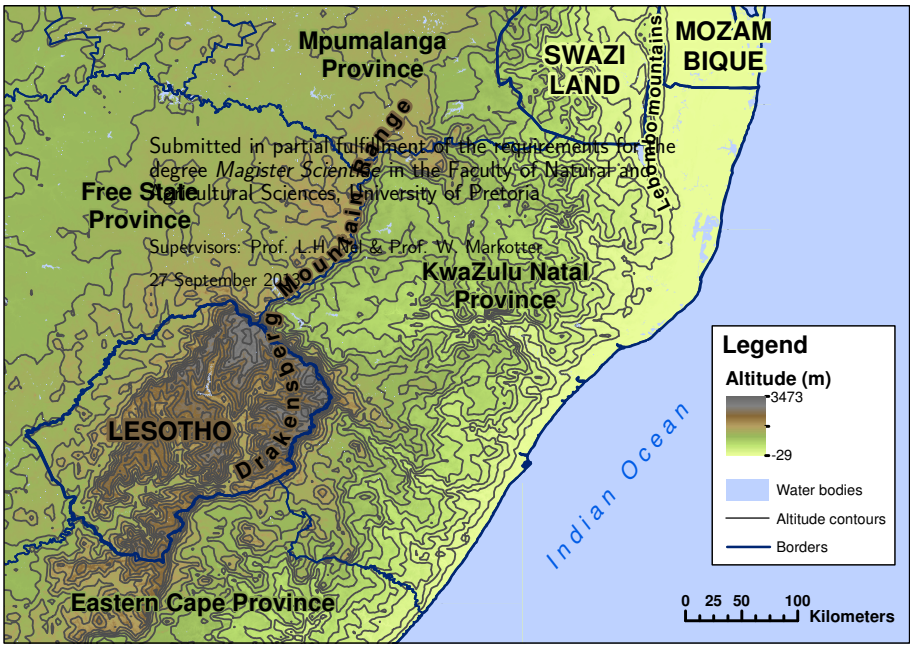

Figure 2.1: Topographic map of the KwaZulu Natal province of South Africa showing international and provincial borders as well as major mountain ranges. Background colour indicates altitude based on version 2.1 of data from the Shuttle Radar Topography Mission (Farr *et al.*, 2007), at a resolution of 3″ (approximately 90 m), while contour lines indicate altitude in steps of 150 m. For clarity, contour lines were smoothed to a resolution of 50″ (approximately 1500 m).

Terrestrial RABV isolates typically show discernible genetic clustering by location of origin even at relatively small scales (section 1.2; Kissi *et al.*, 1995; Coetzee & Nel, 2007; Talbi *et al.*, 2009). RABV isolates from South African domestic dogs fall into two broad phylogenetic lineages, with isolates from the Limpopo and North West provinces grouping entirely separate from isolates from eastern southern Africa (the Mpumalanga, Free State and KwaZulu Natal provinces of South Africa as well as Mozambique and the Kingdoms of Swaziland and Lesotho; Ngoepe *et al.*, 2009; Mkhize *et al.*, 2010). Further clustering by province or country can be seen

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

within these lineages (although the separation of isolates from KZN and Mozambique receive only moderate bootstrap support; Mkrtize *et al.*, 2010). At an even finer scale, Coetzee & Nel (2007) showed that although the KZN isolates they sequenced share a 98.9% sequence identity based in the variable G-L intergenic region, they form two major RABV subfamilies with an average sequence divergence of 1.9% between the families. Even more specific spatial-genetic clustering was discernible within these subfamilies, despite the relatively short region sequenced (592 nucleotides) and the fact that all isolates represented cases from a single year. Such clustering is typical of RNA viruses, with their short generation times and rapid mutation rate, and finding ways to relate observed spatial-genetic patterns to the epidemiological and population genetic processes that caused them is an active area of research (reviewed by Brunker *et al.*, 2012).

Coetzee & Nel argued that the two KZN subfamilies they identified represent independent epidemic fronts. The largest subfamily, designated subfamily A, consisted mainly of viruses isolated in the coastal districts, stretching from the Mozambique border downwards to southern KZN (Figure 2.1). The majority of cases detected in KZN belonged to this subfamily, and were more closely related to isolates from the neighbouring Eastern Cape province than the viruses in subfamily B (Coetzee & Nel, 2007). The viruses in subfamily B were more closely related to each other, and were isolated from northern KZN, bordering the Mpumalanga Province of South Africa as well as Swaziland. In fact, isolates from this subfamily group closely with isolates from the Mpumalanga province (Coetzee & Nel, 2007; Weyer *et al.*, 2011). Coetzee & Nel also identified a small cluster of cases grouping with viral sequences from the Eastern Cape province. These cases seem to point to fairly frequent introductions across provincial borders, which would provide challenges for rabies control programmes. A similar situation exists at the border between KZN and the Free State province of South Africa – all sequenced RABV isolates from past human cases in the Free State have been found to group closely with KZN isolates from dogs (Weyer *et al.*, 2011). It is unclear whether this is the result of several unsuccessful and unrecorded incursions of RABV into the Free State prior to the reported emergence of RABV in this province in 2005, or simply due to exposures that occurred in KZN. The border between KZN and the Kingdom of Lesotho can be expected to be less amenable for such interactions, with the Drakensberg mountain range forming a significant barrier (Figure 2.1). Nevertheless, KZN isolates group in an ancestral position to isolates from a shared epidemiological cycle affecting Lesotho

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Wollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

and the Free State Province of South Africa (Ngoepe *et al.*, 2009). Although the phylogenetic support for this assessment is poor, such a relationship would be in agreement with historic reports describing the spread of RABV from KZN to Lesotho in 1982 (Swanepoel *et al.*, 1993), from where the virus eventually spread to the Free State province of South Africa (Ngoepe *et al.*, 2009), illustrating the potentially long lasting impact of such cross-border transmissions.

The independently administered provinces of South Africa have had varying levels of success in controlling rabies, and very little is known regarding the epidemiological situation in the countries bordering KZN. Given the important implications of cross-border introductions in disease control programmes and the wealth of partial genome sequence information available to elucidate the patterns of RABV spread in eastern southern Africa, surprisingly little is known about the dynamics of RABV transmission in the region. In particular, while recent assessments have been made of the origin of new introductions in other parts of South Africa, no comprehensive, statistically rigorous analyses of cross-border transmissions have been attempted. This chapter describes the analysis of recent genetic and epidemiological data from KZN, as well as existing data from the surrounding region, towards an improved understanding of both the fine-scale and regional dynamics of RABV transmission.

2.2. Materials and Methods

2.2.1. Data collection

All cases testing positive for rabies virus by fluorescent antibody test (Dean *et al.*, 1996) at the Allerton provincial veterinary laboratory of the KZN Department of Agriculture and Environmental Affairs during routine passive surveillance in KZN from 1 March 2010 to 8 June 2011 were selected for analysis (n=195; Appendix A, Table A.1). Five cases were negative by PCR (see below) after multiple attempts and were excluded from further analysis. The sequence generated for one isolate, from an unrecorded wildlife species, matched the herpestid variant of rabies virus by BLAST (Altschul *et al.*, 1990) and was also excluded. Brain samples were stored at −80 °C before and after processing.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

2.2.2. Primer design

To amplify as large a possible region of the G-L intergenic region for subsequent analysis (chapter 3), new primers were designed based on an alignment of full genomes for terrestrial RABV, which included a recently sequenced herpestid RABV variant genome (unpublished). These primers were designated GL4614F (5'-GATTTTGTAGAGGTTCACC-3') and GL5632R (5'-GACCTGGAGCAATTGTCTG-3'), and amplify a 1019 nucleotide region from position 4614 to position 5632 on the Pasteur rabies virus genome (Tordo *et al.*, 1988; GenBank accession number NC_001542).

2.2.3. RNA extraction

RNA extraction was performed using TRIzol reagent (Invitrogen), following the manufacturer's instructions. Approximately 30 mg to 100 mg of original brain material was homogenized by repeated pipetting in 700 µl TRIzol reagent. Following incubation at room temperature for 5 minutes, 200 µl chloroform (Merck) was added before gently mixing the solution. This solution was then incubated at room temperature for a further 3 minutes, before centrifugation at 12 000 × g for 15 minutes. The aqueous phase was transferred to a new 1.5 ml Eppendorf tube and mixed with 500 µl isopropyl alcohol (Merck). This solution was then incubated at room temperature for 10 minutes, before centrifugation at 12 000 × g for 10 minutes. After removing the supernatant, the RNA pellet was washed with 1 ml 75 % ethanol (Merck) by gentle mixing followed by centrifugation at 7500 × g for 5 minutes. The supernatant was removed and the RNA pellet allowed to air-dry for 10 minutes. The RNA pellet was then redissolved in 50 µl nuclease-free water (Promega) by incubating it at 55 °C for 10 minutes, and stored at −20 °C until use.

2.2.4. Reverse-transcription polymerase chain reaction

For reverse transcription, 2 pmol GL4614F was incubated with 5 µl of a 1:4 or 1:9 dilution of RNA at 70 °C for 5 minutes. After a further 5 minutes in an ice bath, 4.1 µl nuclease-free water, 4 µl Improm-II reaction buffer (Promega), 3 mM $MgCl_2$ (Promega), 0.5 mM of each deoxyribonucleotide (dATP, dTTP, dGTP and dCTP; Roche), 20 units Protector RNase inhibitor (Roche) and 1 unit Improm-II reverse transcriptase (Promega) was added. This was followed by incubation at 25 °C for 5 minutes, 42 °C for 60 minutes and finally 70 °C for 15 minutes, using an ABI 2720 thermal cycler (Applied Biosystems). The entire reaction mixture was used

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

for PCR by adding 10 pmol GL4614F, 12.5 pmol GL5632R, 10 µl DreamTaq buffer (Fermentas), 1.25 units DreamTaq polymerase (Fermentas) and 67.5 µl nuclease free water. This reaction was incubated at 94 °C for 1 minute, followed by 40 cycles of 94 °C for 30 seconds, 56 °C for 30 seconds and 72 °C for 90 seconds, before a final incubation step of 72 °C for 7 minutes, using the same thermal cycler as above.

2.2.5. Visualisation and purification of PCR products

To allow visual inspection of amplicons, 10 µl of each PCR product was added to 5 µl sucrose loading buffer (0.25 % bromophenol blue, 40 % sucrose; both from Merck). This was electrophoresed at 100 V for 40 minutes in an agarose gel containing 1 % (w/v) agarose (Lonza) in 40 ml TAE buffer (40 mM tris-acetate, 1 mM EDTA, pH = 8.5; Promega) and 0.0001 mg ml^{-1} ethidium bromide (Sigma-Aldrich).

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

PCR amplicons were purified using the Wizard SV gel and PCR cleanup system (Promega), following the manufacturer's instructions. In short, the remainder of the PCR reaction mixture was electrophoresed as before, but with 10 µl sucrose loading buffer. Excised gel slices containing DNA bands of the appropriate size were melted in 1 µl mg^{-1} membrane binding solution (Promega) at 60 °C, before centrifugation through a silica membrane minicolumn assembly at 12 700 × g for 1 minute. The membrane-bound DNA was then washed twice with membrane wash solution (Promega) before elution with 50 µl nuclease-free water (Promega). Purified PCR products were stored at −20 °C.

2.2.6. Sequencing

The purified PCR amplicons were sequenced with both the forward (GL4614F) and reverse primer (GL5763) in separate reactions using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Reactions contained 2 µl 2.5× BigDye reaction mix, 1 µl 5× BigDye sequencing buffer, 3.2 pmol primer, 3.9 µl template and 0.9 µl nuclease-free water (Promega). Cycle sequencing was performed by using the same thermal cycler as above to incubate reactions at 94 °C for 1 minute, followed by 25 cycles of incubation at 94 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes.

Next, the sequencing reactions were precipitated by adding 1 µl of a 125 mM EDTA solution (ethylenediaminetetraacetic acid; Promega), 1 µl of a 3 M sodium acetate solution (CH_3COONa; Sigma-Aldrich) and 25 µl 100 % molecular-grade ethanol

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

(Merck) to each reaction. The reactions were mixed by vortexing before incubation at room temperature for 15 minutes, followed by centrifugation at 12 700 × g for 22 minutes. The pellet was washed twice with 100 µl 70 % molecular-grade ethanol (Merck) by centrifugation at 12 700 × g for 12 minutes. The supernatant was removed and the pellet dried at 94 °C for 3 minutes. The reactions were then frozen at −20 °C, before being submitted to the sequencing facility of the Faculty of Natural and Agricultural Sciences, University of Pretoria, where the pellets were reconstituted in 20 µl formamide and analysed on an ABI PRISM® 3100 or 3500xL automated DNA sequencer (Applied Biosystems).

Sequence reads were assembled using CLC Main Workbench version 6.6.2 (CLC Bio) and manually checked for conflicting base-calls, after which a BLAST search (Altschul *et al.*, 1990) was performed to ensure sequences belonged to the canid genetic variant of RABV.

2.2.7. Phylogenetic analysis

All canid variant consensus sequences were aligned using the FFT-NS-i algorithm of MAFFT version 6 (Katoh & Toh, 2008). Sequences were trimmed to equal length (760 nucleotides, encompassing the last 224 nucleotides of the glycoprotein gene, the G-L intergenic region, and 118 nucleotides of the polymerase gene, based on the genome of the Pasteur rabies virus strain). The overall mean distance between sequences in the alignment was calculated using MEGA version 5 (Tamura *et al.*, 2011). Next, the best fitting DNA substitution model was selected using Akaike's information criterion in jModelTest version 2.1 (Akaike, 1974; Darriba *et al.*, 2012). The best fitting model was identified as the 3-parameter model described by Kimura (1981). This model was used to construct a Bayesian phylogeny in BEAST version 1.7.3 (Drummond *et al.*, 2012), using a strict molecular clock with a broad substitution rate prior consisting of a uniform distribution between 1×10^{-5} and 1×10^{-2}. Relaxed clock models were tested but showed no evidence of rate-variation among branches, supporting the assumption of a strict molecular clock. The best-fitting demographic model was determined to be one describing exponential growth using both path sampling and stepping stone sampling (which were in agreement) of individual Markov chains of 50 million iterations each, saving every 5000th step (Baele *et al.*, 2012). This model was used to construct two Markov chains of 50 million steps each, saving every 10 000th step. The resulting estimates were checked for convergence and the posterior estimates of trees were combined after a burn-in of 10 % of each chain, and

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

then summarised as a majority-consensus phylogeny using Dendroscope version 3.2.2 (Huson & Scornavacca, 2012).

2.2.8. Analysis of spatial-genetic structure

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

The alignment described above was re-aligned using the FFT-NS-i algorithm of MAFFT after the exclusion of 13 cases lacking coordinates (Appendix A, Table A.1). Next, a Mantel test was performed to test for evidence of spatial-genetic correlation using version 1.5-2 of the "ade4" package for the R statistical programming language (version 2.15.1), with 9999 permutations (Mantel, 1967; Dray & Dufour, 2007; R Core Team, 2012). This was based on a genetic distance matrix calculated under the model of Kimura (1980) using version 3.0-8 of the "APE" R package, and a matrix of great-circle geographic distances calculated using the haversine formula described by Sinnott (1984) with the mean radius of earth approximated to 6371 km (Moritz, 1980; Paradis *et al.*, 2004).

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

September 2013

Single nucleotide polymorphisms were identified from the alignment of all sequences with coordinates using the "DNAbin2genind" function of version 1.3-7 of the "adegenet" R package (Paradis *et al.*, 2004; Jombart, 2008). In this calculation, all polymorphism was included by setting the minimum frequency of mutations considered to $\frac{1}{176}$ (176 is the number of sequences in the alignment). This was used in a spatial principal components analysis (sPCA) based on inverse geographic distances with a minimum distance of 1×10^{-5} before points are considered as separate, again using the "adegenet" package (Jombart, 2008; Jombart *et al.*, 2008). The statistical support for this spatial clustering was assessed using the global and local Monte Carlo tests described by Jombart *et al.* (2008) with 9999 permutations.

2.2.9. Bayesian discrete-state phylogeography and demographic reconstruction

In addition to the 189 sequences generated above, all G-L intergenic region sequences for the canid genetic variant of RABV publicly available from the countries and South African provinces surrounding KZN were included in a discrete-state phylogeographic analysis (Appendix A, Table A.2). These sequences were aligned as before and trimmed to equal length (610 nucleotides).

The fit of various DNA substitution models to this data were compared with MODELTEST via the PALM parallel computing pipeline, using Akaike's information

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. LTH. Nel & Prof. W. Markotter

27 September 2013

criterion to compensate for varying numbers of free parameters between models (Akaike, 1974; Posada & Crandall, 1998; Chen *et al.*, 2009). The best-fitting substitution model (a transversion model with equal base frequencies and gamma-distributed rate variation among sites) was used in a phylogeographic analysis in BEAST version 1.7.5 (Drummond *et al.*, 2012), with a constant size coalescent model, a uniform distributed molecular clock rate prior between 1×10^{-5} and 1×10^{-2} substitutions per site per year and calibrated using the sampling years of all sequences. Locations for internal nodes were inferred using an asymmetric continuous-time Markov model with Bayesian stochastic search variable selection to identify the most parsimonious diffusion process (Lemey *et al.*, 2009). One Markov chain of 50 million steps was constructed for each of three molecular clock models (a strict clock model and a relaxed clock model with either log-normal or exponentially distributed rate variation among branches), sampling every 5000th step. There was no evidence to reject the assumption of rate variation among branches. Of the two specifications for the relaxed clock model, the log-normal version showed less autocorrelation and was chosen for subsequent analyses.

Next, the constant size coalescent model described above was replaced with a Bayesian skyride model with a time-aware Guassian Markov random field smoothing prior, to co-estimate the demographic patterns in the region under study (Minin *et al.*, 2008). Under this specification, four Markov chains were sampled for 50 million steps each, retaining every 5000th step. The chains were inspected visually for convergence, after which the associated posterior distributions of trees were combined, keeping every 10000th sample after removing a burn-in of 20% of the samples from each chain. The phylogenies in this sample were summarised as a majority consensus phylogeny using Dendroscope version 3.2.2, while a Bayesian skyride plot was generated by summarising the sampled demographic parameters into 100 bins using Tracer version 1.5.0 (Minin *et al.*, 2008; Rambaut *et al.*, 2009; Huson & Scornavacca, 2012). A Bayes factor test was performed to identify supported migration rates between the analysed regions using SPREAD version 1.0.4 (Lemey *et al.*, 2009; Bielejec *et al.*, 2011).

2.3. Results and Discussion

To gain an up-to-date view of the current epidemiological situation in KZN, all RABV-positive cases detected between 1 March 2010 and 8 June 2011 were selected

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

for further analysis (195 cases; Figure 2.2). Cases occurred throughout most of the province, but primarily clustered around the informal settlements surrounding cities and major towns (Figure 2.2 B). The majority of cases – both in terms of the actual number of cases and in terms of their density – were reported in the densely populated southern coastal region, particularly in the Ugu district municipality and the eThekwini metropolitan municipality (Figure 2.3). However, when viewing this data in terms of the human population in the affected areas (which is generally closely correlated to dog numbers), eThekwini appears less affected, with the Ugu and uThukela district municipalities carrying most of the burden (Figure 2.3 D). It should be noted however that this is only a "snapshot" of the current rabies situation – analyses of a much longer period than the 15 months presented here would be needed to critically assess which areas need to be prioritised for control measures.

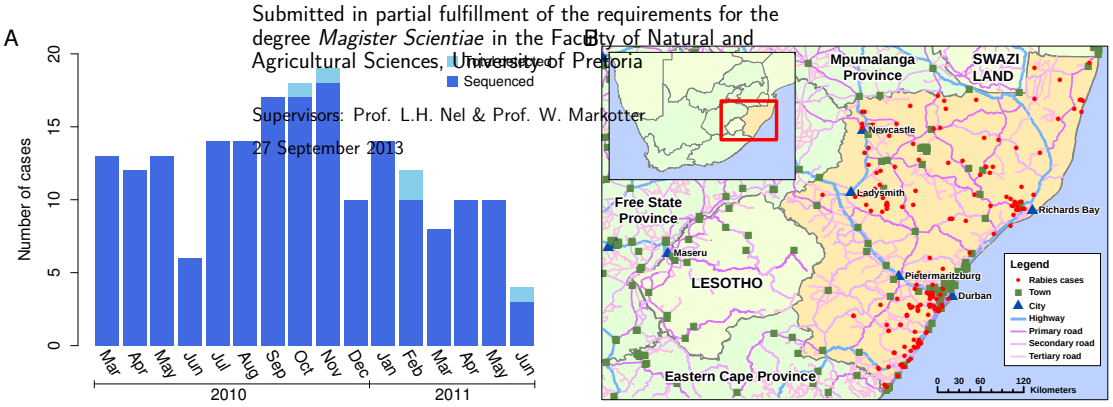

Figure 2.2: Dates of detection and location of origin of all cases included in this study. **A:** Time-series of all cases detected between 1 March 2010 and 8 June 2011 in KwaZulu Natal. **B:** Map of KwaZulu Natal showing the cases detected in the context of major roads, towns and cities. Note that in addition to the 180 cases shown on this map, a further 15 cases were detected for which coordinates were not recorded (road and town data copyright OpenStreetMap contributors).

The G-L intergenic region of 190 (97.44 %) of the detected cases was sequenced (Appendix A, Table A.1). This genome region was chosen due to its high variability and the large numbers of partial sequences available from previous studies of RABV in southern Africa (e.g. Sabeta *et al.*, 2003; Coetzee & Nel, 2007; Mkhize *et al.*, 2010; Ngoepe *et al.*, 2009; Zulu *et al.*, 2009; Weyer *et al.*, 2011). Despite the high variability of this genomic region and of RNA viruses in general, many sequences were identical, a reflection of the dense sampling undertaken in this study. Nevertheless, clear phylogenetic clustering of cases can be distinguished, which correspond to their geographic region of origin (Figure 2.4). This pattern was confirmed by a Mantel test (Mantel, 1967), which shows clear evidence of correlation between geographic

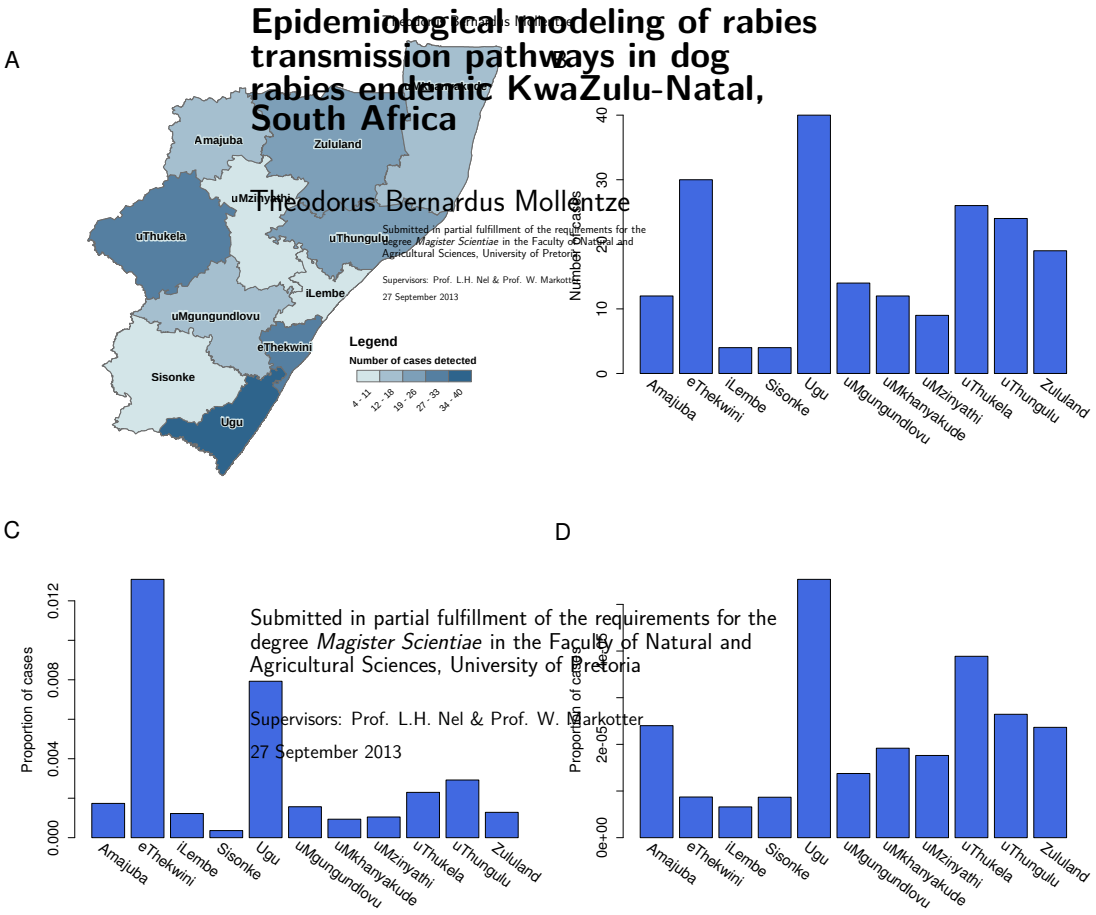

Figure 2.3: Rabies cases detected in animals by fluorescent antibody test between 1 March 2010 and 8 June 2011 in each of the administrative districts of KwaZulu Natal (KZN). **A:** Map of the administrative districts of KZN with districts shaded according to the relative number of cases detected there during the study period. The same data are displayed in **B**. **C:** The proportion of cases detected in each district as a function of district size, with the y-axis in cases/km). **D:** The proportion of cases detected in each district as a function of the human population in the affected area (based on data from Statistics South Africa, 2012).

and genetic distance with a correlation coefficient (r) of 0.5068 and a simulated p-value of 0.0001. As discussed in section 1.3.5, such spatial-genetic correlation can be expected from local transmission processes, and it is thus likely that these clusters reflect either separate, diverging branches of the same "global transmission tree", independent transmission cycles (i.e. unlinked transmission trees), or a combination of both. Based on the large distances between several of the phylogenetic clusters, the latter option seems most likely.

To gain a more quantitative view of the spatial clustering of genetically related cases, a spatial principal component analysis (sPCA) was performed. This technique operates in a similar manner to principal component analysis, which summarises

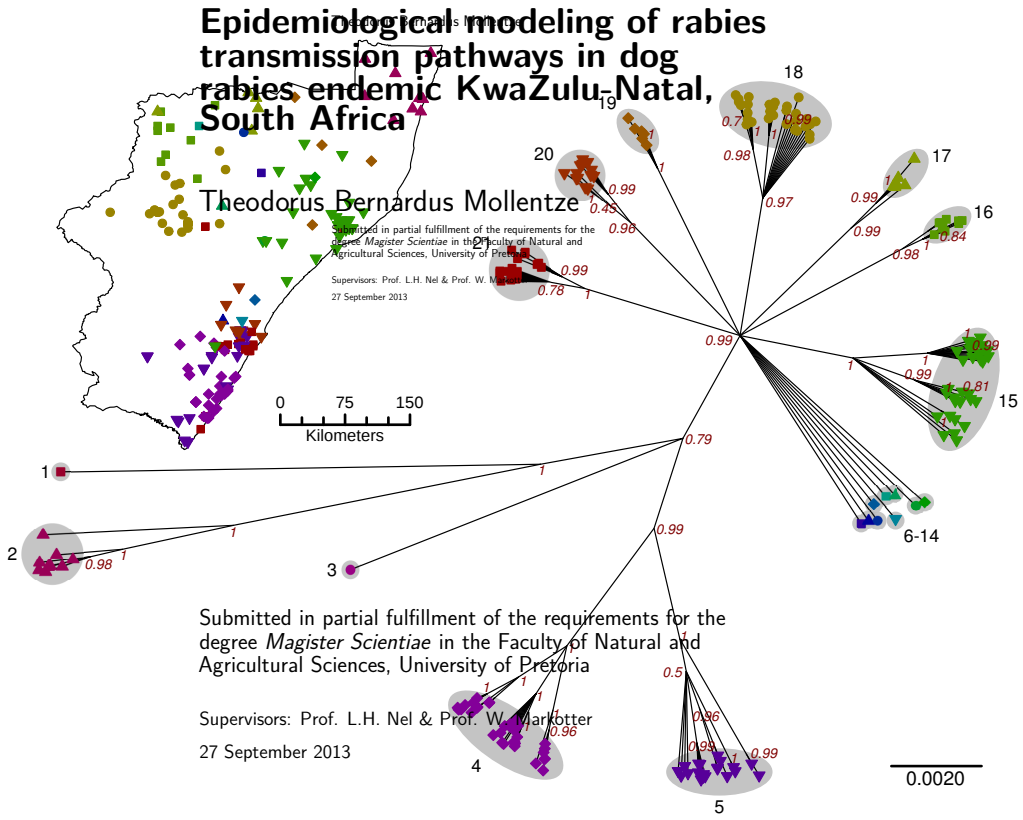

Figure 2.4: Phylogenetic clustering of the cases sequenced in this study. The main figure shows a majority consensus phylogeny constructed using BEAST version 1.7.3 with a strict molecular clock and an exponential growth demographic model, while the inset shows the sampling locations of the cases from each cluster for which coordinates were available. Phylogenetic clusters are indicated by differing symbols and colours and labelled with numbers as used in the text. Branch lengths are in substitutions per site per year, as indicated by the bottom scale bar, and numbers at internal nodes indicate the posterior probabilities of the inferred branching patterns.

multivariate data as a number of uncorrelated "components" by maximizing the amount of variance in the data reflected by each component (Jombart *et al.*, 2008). In the case of sPCA, it is the product of variance and Moran's spatial autocorrelation index (Moran's I) that is maximised when defining each component (which encompasses groups of cases here), thus allowing complex spatial-genetic patterns to be discerned (Moran, 1948, 1950; Jombart *et al.*, 2008). This analysis also revealed statistically significant spatial structure, with a simulated p-value of 0.0001, leading to rejection of the null hypothesis of no spatial structure in favour of the alternative hypothesis of positive spatial autocorrelation, or "global structures". By contrast, there was no significant evidence to support an alternative hypothesis of negative spatial autocorrelation (i.e. highly dissimilar virus variants in close quarters), with a simulated p-value of 0.3326 when the null hypothesis was that

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W.J. Markotter

27 September 2013

there is no spatial structure in the data. The majority of genetic variance can be explained by four sPCA components (Figure 2.5 A), each revealing different global spatial structures which may be related to different causes of genetic separation (Jombart *et al.*, 2008). The first of these components encompasses a significant proportion of both the genetic variance and spatial autocorrelation, and clearly separates cases from the majority of KZN from those observed along the southern coast (Figure 2.5 B). The second component shows differences between cases in the extreme north of KZN from most other cases, with some cases in the southern coastal region showing the same, but weaker spatial-genetic pattern. These northern-most cases could be expected to be different from those in the rest of KZN due to their separation by the Lebombo mountain range and Lake St. Lucia (Figure 2.1), but it is unclear why the geographically distant cases in the densely populated south would follow the same pattern (note that these results do not necessarily reflect genetic similarity between cases in the extreme north and those in the south – it merely indicates that they share the same spatial-genetic pattern). The third component distinguishes between cases in eastern and western KZN, and again also separates a number of cases in the south from their close neighbours (though with fairly weak spatial-genetic structure), while the fourth component further delineates clusters of cases in the west and south (Figure 2.5 B). Taken together, the results from this analysis delineate at least 9 spatial-genetic clusters (Figure 2.6). These clusters broadly correspond to the phylogenetic clusters, although the sPCA analysis provided less resolution in some cases, with the spatially intermingled phylogenetic clusters 20 and 21 and the more distant phylogenetic cluster 16 showing very little dissimilarity from each other. This analysis did however resolve the single, unresolved nodes making up phylogenetic clusters 6 to 14 to various clusters, but support for this is fairly weak (Figure 2.6).

The results from both the phylogenetic analysis (which takes genetic data and observation dates into account) and the sPCA analysis (which takes genetic data and observation locations into account) show a number of diverging groups of cases. Since genetic separation of pathogen sequences can only occur through time or independent evolution and the cases analysed here were sampled over a relatively short period, these clusters likely reflect independent epidemiological cycles or transmission trees. Such independence may be a sampling artefact, with many dogs being infected already by the time sampling was started as well as a potentially large number of undetected cases. However, it would be surprising if either analysis contained

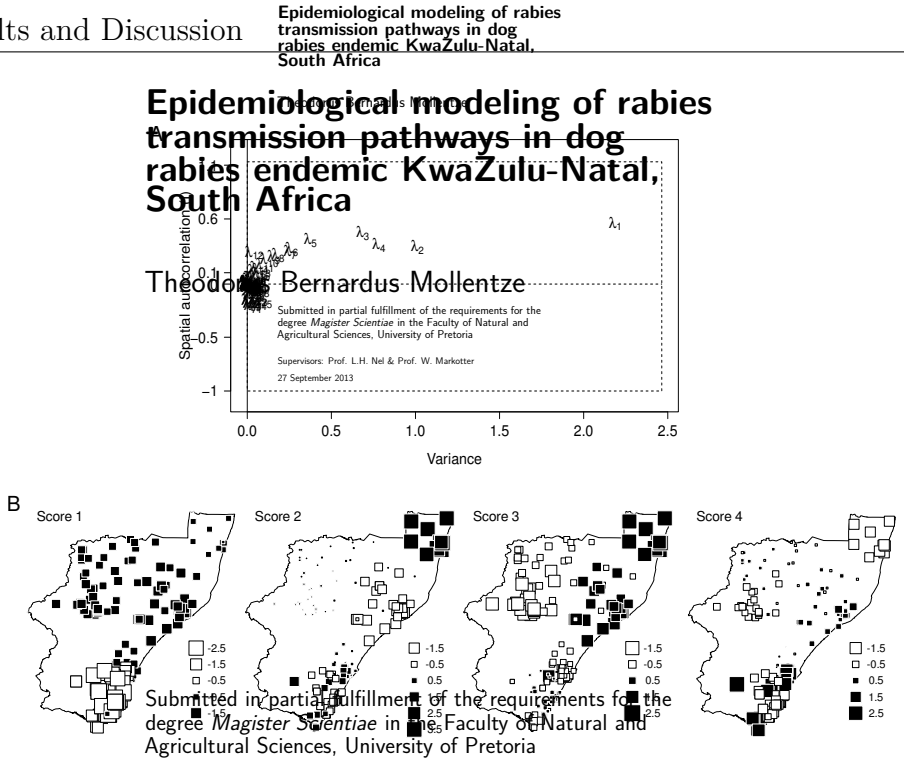

Figure 2.5: Spatial-genetic clustering among cases. **A:** A decomposition of the eigenvalues or "scores" for identified spatial principal components (λ_i represents the eigenvalue for component i, with i being sorted from the highest to the lowest score). The x-axis shows the amount of genetic variance encompassed by each principal component, while the y-axis shows the amount of spatial autocorrelation explained by each component. Dashed horizontal lines show the total range of variation of Moran's I for different groupings of the analysed cases, while the dashed vertical line indicates the amount of genetic variance when considering all alleles as a single component/group. Since the first four scores explain the majority of genetic variance, only these were interpreted further. **B:** Scores from component 1–4 plotted by sampling location, showing the spatial-genetic structuring between clusters of cases. The sizes of squares indicate the values of the scores, with large black squares (representing highly positive scores) being well distinguished from large white squares (representing highly negative scores; Jombart, 2012).

sufficient resolution to discriminate between branches of the same transmission tree separated by only a few unsampled cases. For some of these clusters, geographic causes of separation are obvious (e.g. the northern-most cases separated by the Lebombo mountains or the separation of western and eastern clusters in central KZN due to distance), but the observation of a number of spatially overlapping yet independent clusters is harder to explain.

On a broader scale, both sets of results point to the existence of four isolated spatial clusters of cases (Figure 2.4 and Figure 2.6), which appear to be largely independent. Indeed, the better resolved phylogenetic analysis shows only one overlap between these broad spatial clusters, with one case from phylogenetic cluster 20, which occurs primarily in southern KZN, appearing in eastern KZN. The northern-most cluster is homogenous and distinct from all other cases, and appears to be part of a cross-border

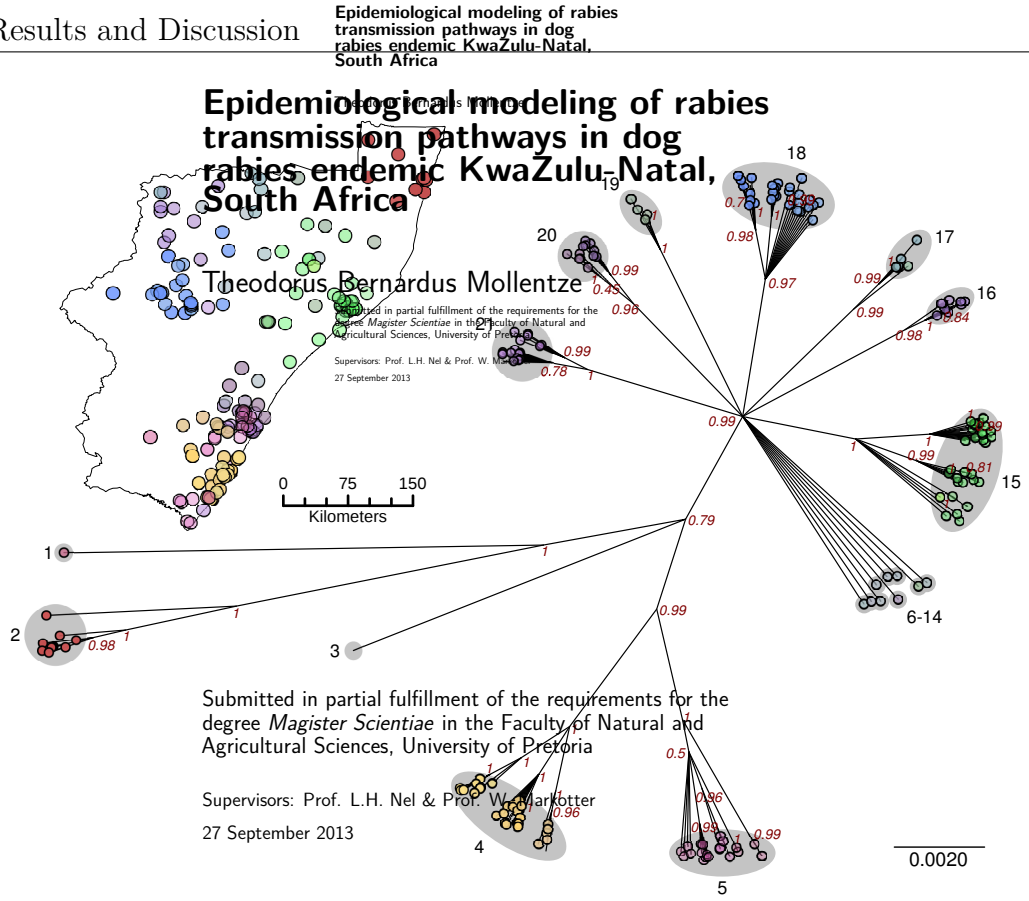

Figure 2.6: Combined scores from the first four spatial principal components indicated on the majority-rule consensus phylogeny from Figure 2.4, showing close agreement with the clusters identified from genetic data and date of sampling. Numbers indicate the phylogenetic clusters designated in Figure 2.4. The value of scores from each component determine the relative amount of colour contributed, with scores from component one contributing red, scores from component two contributing green and scores from component three contributing blue. Scores from component four determine the intensity of the colour produced. The inset shows the spatial distribution of these clusters.

endemic cycle shared with neighbouring Mpumalanga or Mozambique (see below). The eastern and western clusters are also fairly homogenous, with little mixture between different phylogenetic clusters. However, the southern cluster is much more diverse, consisting of five freely mixing phylogenetic clusters. This region of KZN is densely populated and urbanised, and it can be expected that movement of both humans and dogs occurs with greater frequency and ease there, although there is clearly some barrier to prevent complete mixing.

Since rabies is endemic throughout eastern South Africa and in the countries bordering KZN, it can be expected that at least some of the more distantly related phylogenetic clusters observed arose as a result of independent introductions from outside the province. This possibility was further investigated by reconstructing the locations of origin of the ancestors of all cases from KZN and surrounding

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae (Biochemistry) in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

provinces and countries for which G-L intergenic region sequences were available. This dataset spanned a period of 31 years (1980–2011) and consisted of 637 sequences (Appendix A, Table A.2). The most recent common ancestor (MRCA) of the analysed isolates was estimated to have occurred in 1978 (95 % PI: 1977–1979) in KZN (Table 2.1). Historical reports date the re-introduction of RABV from Mozambique into KZN to 1976, after first entering Mozambique in 1952 from what is now the Mpumalanga province of South Africa (Swanepoel *et al.*, 1993). Only 6 sequences were available from Mozambique and no historic sequences exist from either Mpumalanga or northern KZN. It is thus possible that the later date given by the analysis here is because the older lineages have died out completely or simply were not sampled. Indeed, while rabies would have been established in Mozambique by 1978, the demographic reconstruction shows a pattern typical of a newly emerged epidemic in the first decade after this date, before reaching an endemic stage in 1988 (Figure 2.8). It should be noted however that the date estimate for the MRCA is based on a slightly low effective sample size of 147.72, while >200 is typically recommended (the effective sample sizes of all parameter estimates not relating to the molecular clock were well over 200).

Table 2.1: Posterior probabilities of the estimated origin of the most recent common ancestor of all available, sequenced isolates from eastern southern Africa

Location of origin	Posterior probability
KwaZulu Natal	0.72
Mpumalanga	0.18
Mozambique	0.05
Swaziland	0.03
Eastern Cape	0.02
Lesotho	0.01
Free State	0.00

The analysis detected several introductions from KZN to other locations included in the analysis soon after the estimated date for the MRCA, although the location estimates for most nodes occurring close to the root – including that of the root itself – were fairly ambiguous, with low posterior probabilities (Figure 2.7; Table 2.1). A close relationship was inferred between cases from Mpumalanga and the few available from Mozambique, but the direction of spread remained ambiguous. Indeed, this relationship is so ambiguous that other scenarios of spread involving Mozambique,

many of which make little geographically sustained significant levels of support (Table 2.2). Two major introductions of rabies from KZN to the Eastern Cape (in 1983 [95 % PI: 1982–1984] and 1984 [95 % PI: 1981–1986]) were much better supported, with a posterior probability that KZN was the source of 0.85 and 0.95 respectively (the posterior probabilities for the phylogenetic groupings at these node were 0.99 and 1 respectively). These dates are earlier than the first reports of rabies in the Eastern Cape from 1987 (Swanepoel *et al.*, 1993).

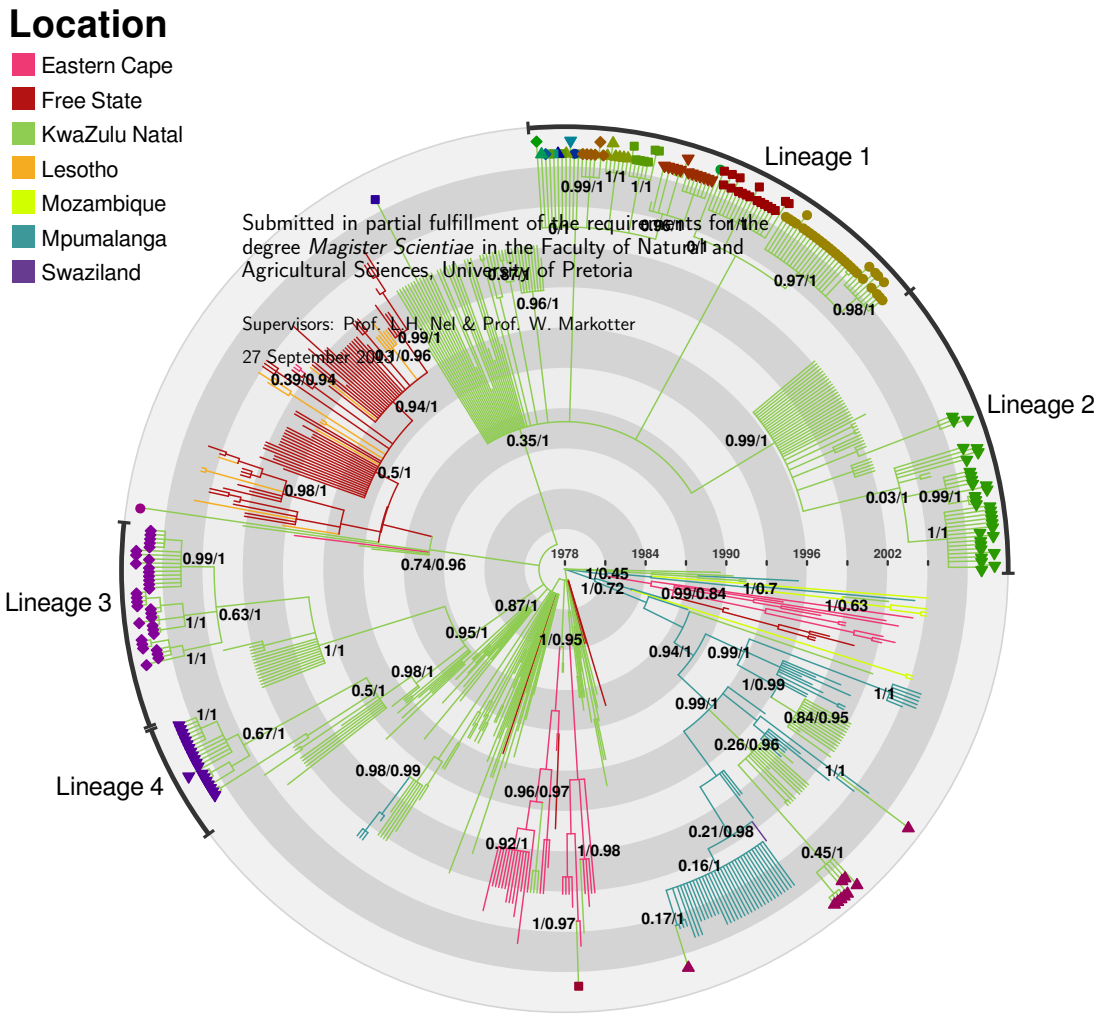

Figure 2.7: Majority consensus phylogeny showing the most probable locations of hypothetical ancestors (internal nodes) reconstructed using the method described by (Lemey *et al.*, 2009). Branch lengths indicate time, estimated under a relaxed clock with rates for different branches varying according to a log-normal distribution. The colour of branches indicates their inferred geographic location, while symbols at the tips of the tree indicate the phylogenetic clusters shown in Figure 2.4. Numbers at nodes are posterior probabilities, with the first number representing the posterior support for that specific phylogenetic clustering and the second representing the support for the inferred geographic location of that node.

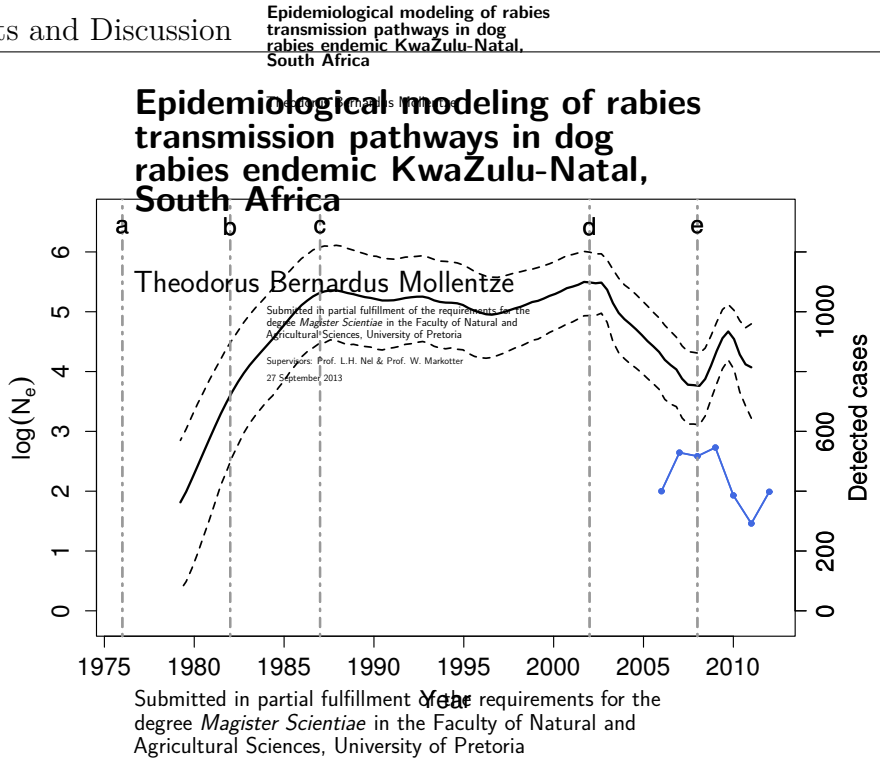

Figure 2.8: Time-smoothed Bayesian skyride plot estimated for eastern southern Africa, showing changes in the effective population size (N_e) over time. The solid black curve shows the mean effective population size, while the dashed curves indicate the upper and lower bounds of the 95 % posterior interval for this estimate. The blue line shows the detected cases in this region (where data are available), based on data from the OIE World Animal Health Information Database (http://www.oie.int/wahis_2/public/wahid.php/Diseaseinformation/statusdetail). Vertical lines indicate the reported dates of historic events relating to RABV emergence in this region (reviewed by Swanepoel *et al.*, 1993): *a* – the re-emergence of RABV in KZN, *b* – the spread of RABV from KZN to Lesotho, *c* – the spread of RABV to the Eastern Cape province of South Africa, *d* – the spread of canid-associated RABV from Lesotho to the Free State province of South Africa (Ngoepe *et al.*, 2009), and *e* – the start of a large epidemic outbreak in the Mpumalanga province of South Africa (Mkhize *et al.*, 2010).

Coetzee & Nel (2007) suggested that RABV may have entered the Eastern Cape during the first KZN rabies epidemic between 1964 and 1968, where it could have remained undetected in the former apartheid homeland of Transkei before being re-introduced to KZN after the virus was eliminated there. This argument was used to explain the existence of two distinct subfamilies of RABV in KZN, one of which was closely related to rabies sequences from the Eastern Cape (Coetzee & Nel, 2007). Although the phylogeograhic analysis did detect an earlier introduction to the Eastern Cape with an unresolved origin, this was estimated to have occurred in 1981 (95 % PI: 1979–1984) at the earliest, and the resulting cases were not related to cases from KZN (Figure 2.7). With the inclusion of many much older sequences from KZN than those analysed by Coetzee & Nel, the present analysis shows that the relationship between the majority of cases from the Eastern Cape and KZN is not as close as previously thought. In addition, the direction of spread for the major

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria
Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

Table 2.2: Reconstructed migrations between locations in eastern southern Africa occurring at rates supported to be different from zero by a Bayes factor of more than three, the generally recommended cut-off level for rates to be considered significantly supported (Lemey *et al.*, 2009)

Migration	Bayes factor
Eastern Cape province → Mpumalanga province	84 401.20
Mozambique → Free State province	282.80
KwaZulu Natal province → Free State province	279.88
Free State province → Mozambique	102.94
Mozambique → Swaziland	60.87
KwaZulu Natal province → Eastern Cape province	32.61
Free State province → KwaZulu Natal province	**31.73**
Free State province → Eastern Cape province	29.63
Eastern Cape province → KwaZulu Natal province	**16.37**
Mpumalanga province → Free State province	15.50
Mozambique → Lesotho	13.05
Lesotho → Free State province	9.20
Swaziland → Lesotho	4.04
Mpumalanga province → KwaZulu Natal province	**3.16**

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria
Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

introductions is inferred to be *from* KZN *to* the Eastern Cape with a high level of support.

Although rabies seems to have remained in a stable endemic state in this region for more than a decade, this is no longer the case (Figure 2.8). The demographic reconstruction showed a marked decrease in the effective population size from 2003 onwards, which continued until the 2008 outbreak in Mpumalanga (e in Figure 2.8). This epidemic appears to have peaked during 2009, after which the effective population size resumed the decline of previous years (this peak in incidence in Mpumalanga is visible in both the demographic reconstruction and the surveillance data in Figure 1.2). Worryingly, surveillance data from 2012 shows a marked increase in rabies incidence in KZN (Figure 1.2). This may be due to an increase in surveillance activity, as appears to have happened during 2006–2008 when the strong decrease in effective population was not reflected in the surveillance data. However, the upper estimate of the 95 % posterior interval of the demographic curve also shows some evidence of an upward trend in 2011, and recent reports suggest that RABV has in fact spread to areas of KZN where the virus had previously been eliminated (K. le Roux, personal comunication).

In addition to introductions of RABV from KZN to the surrounding regions, the

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

phylogeographic analysis detected some instances of re-introductions to KZN from these regions. Three neighbouring areas received significant support as sources of introduction to KZN required to explain the observed spatial distribution of genetic variation (Table 2.2). Phylogenetic cluster 2 from the previous analyses appears to represent the modern-day descendants of Coetzee & Nel's "subfamily B" – which they suggested may have been introduced from Mpumalanga (Coetzee & Nel, 2007; Figure 2.7) – although there is insufficient support to resolve this relationship beyond a common ancestor which was estimated to have occurred in Mpumalanga with a posterior probability of 1. This "subfamily" was more widespread in KZN in 2003 (Coetzee & Nel, 2007), while canid rabies was not endemic in most of Mpumalanga, including the areas neighbouring KZN, before 2008 (Mkhize *et al.*, 2010). The ancestral relationship of KZN and Mozambique sequences to these isolates was also observed by Mkhize *et al.* (2010), who also found a close relationship to the single sequenced case available from Swaziland (although this relationship was poorly supported in the phylogeographic analysis). Thus, the present evidence suggests that the cases of subfamily B represent a cross-border endemic cycle present in parts of Mozambique, Mpumalanga and KZN and possibly spread throughout Swaziland situated between them.

Apart from the series of introductions from the Eastern Cape represented by phylogenetic cluster 1 discussed above, all other phylogenetic clusters of cases sequenced here represented diversity previously observed in KZN (Figure 2.7). With the inclusion of many older sequences from KZN, it becomes clear that the very distant phylogenetic clusters 4 and 5 from southern KZN represent distinct lineages which have been endemic in KZN for some time (lineage 3 and 4 in Figure 2.7). Phylogenetic cluster 15 – which represents almost all of the diversity observed in eastern KZN (Figure 2.4 B) – is also resolved as a distinct, but younger lineage (designated lineage 2 in the figure). The ancestry of the remaining clusters (indicated as lineage 1 in the figure) could not be resolved, although much of the diversity *within* these clusters appears to have arisen after 2007 (95 % PI: 2005–2008). This last observation is at odds with the marked decline in effective population size observed at a regional level during this time (Figure 2.8), although it should be noted that the density of sampling was much higher for these cases than for most of the dataset analysed, with previous authors sequencing only limited numbers of cases from any given region and year.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfilment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

2.4. Conclusion

From the "snapshot" of current RABV diversity in KZN presented here, it is clear that rabies remains widespread throughout KZN. The problem is particularly acute along the densely populated southern coast, where the majority of cases and four independent lineages occur. These lineages have been distinct for several decades, with the split between the ancestor of lineages 1 and 2 and the ancestor of lineages 3 and 4 occurring shortly after the introduction of rabies to KZN in 1976. In some cases, this separation may simply be due to spatially distinct endemic cycles (e.g. the separation of lineage 2 from the unresolved clusters grouped into lineage 1), but the observation of co-occurring lineages in southern KZN is more difficult to explain. A likely explanation would be that these lineages arose elsewhere and were subsequently introduced by migrants to the metropolitan areas of southern KZN. The re-introduction of rabies to KZN has historically been ascribed to the movement of dogs incubating RABV by migrants from Mozambique (Swanepoel *et al.*, 1993), and there is no reason to suppose that this introduction involved just one lineage. However, the phylogeographic analysis including all available G-L region sequences from surrounding areas (which is by far the most commonly sequenced genomic region in this area) showed no evidence of introductions from elsewhere. Although historic introductions from further afield may a possibility, no previous studies using many of the older sequences included here have found any evidence for this. Thus, the only remaining solution is that the extremely limited number of samples from Mozambique skewed the analysis, leading to historic introductions being missed. Indeed, the inferred location for the root of the tree is fairly uncertain, and Mpumalanga – where the closest relatives of the few RABV isolates from Mozambique occur (Mkhize *et al.*, 2010) – has a small level of support as the location of the root (Table 2.1). Such an explanation would also explain the inference of the Free State province as the origin of the epidemiological cycle shared with Lesotho, in contrast to the fact that rabies was not detected in this province until fairly recently (Ngoepe *et al.*, 2009). Thus, one can propose a scenario in which at least the oldest lineages, representing the ancestor of lineages 1 and 2 and that of lineages 3 and 4, may have arisen from independent introductions from Mozambique during the initial epidemic from 1976 onwards. In the case of lineages 1 and 2, whose grouping was poorly supported, this may even hold true for the individual lineages themselves. However, in the absence of more data from Mozambique this is mere speculation.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. J. H. Nel & Prof. W. Markotter

27 September 2013

The sparse distribution of cases throughout most of KZN and particularly the north-east seems to suggest that many cases are not being detected by surveillance (Figure 2.2 B). Although there is no direct way of linking effective population size to the exact number of cases, the demographic reconstruction presents a possible solution to allow examination of epidemiological trends despite poor surveillance. From the currently available genetic data it can be inferred that control measures applied prior to 2003 had very little effect at a regional level. Thus, the changes in control strategies from late 2002 onwards are worth investigating, and may yield clues relevant to designing new control strategies elsewhere. It is also worth noting that the extent of this decline is not reflected in surveillance data (although provincial level data are not available prior to 2006), which points to an increase in surveillance efficacy at the same time as a reduced number of infections.

At a smaller scale, several distinct phylogenetic clusters are discernible within KZN. This is supported by both phylogenetic (Figure 2.4) and spatial-genetic data (Figure 2.6). The decreased resolution of the latter can be ascribed to the fact that the sPCA takes only the presence/absence of single nucleotide polymorphisms into account, while in reality the type of nucleotide change also carries some information, since some changes are more likely than others (e.g. transitions are more likely than transversions, substitutions are more likely than insertions/deletions, etc.). None of the analyses employed here take into account epidemiological factors, which have the potential of increasing resolution even further. Thus it is not possible to determine the exact causes of the fine scale genetic clustering observed – although it seems likely that they may reflect individual branches of the network of transmission, determining the order of transmissions and the causality between individual cases would be required to understand transmission dynamics at this scale.

As an encouraging sign for potential rabies elimination, the majority of these phylogenetic clusters do not reflect recent introductions to the province, and indeed very few historic introductions to KZN were detected. Thus, introductions are either rare or they do not become established in the province and are therefore not detected. This means that elimination of canid RABV in KZN, as well as maintenance of this rabies-free status, should be feasible in theory. This goal will however require an increasingly cost-effective approach in the face of declining political will with declining numbers of human deaths from rabies. In particular, the finding of seemingly independent clusters of cases suggests that independent, smaller-scale control programmes may be feasible. However, there is some mixing between clusters,

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

and it is difficult to establish what the effect or extent of such mixing will be given the possibility of large numbers of undetected cases. There is currently no quantitative data on the efficacy of surveillance in this region, although the general agreement between surveillance data and the demographic trends detected from available genetic data provides some level of assurance that at least a sizeable proportion of cases is being detected.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Chapter 3.

A novel approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data[1]

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

3.1. Introduction

Understanding the spatial aspects of disease transmission is increasingly recognised as an essential component of successful control strategies (Ferguson *et al.*, 2003; Keeling *et al.*, 2003; Beyer *et al.*, 2011). However, disease transmission is usually a highly elusive event and reconstructing the causal relationships between cases (i.e. 'who-infected-who') in outbreaks of infectious disease remains a challenging problem. In this regard, the availability of high throughput, affordable pathogen genome sequencing to complement conventional space-time incidence data promises a step-change in our ability to understand pathogen transmission at the level of individual cases. However, if we are to make full use of these advances in sequencing technology, they must be matched by advances in statistical methodology.

Two different but complementary approaches that use spatial, temporal and pathogen genetic information to reconstruct the dynamics of epidemics have been

[1]The models described in this chapter were designed by Dr. Samuel Soubeyrand (Institut National de la Recherche Agronomique) with input from me and Prof. Daniel Haydon (University of Glasgow). Quality control and all additional analyses of the output were performed by me. This work will soon be published as: Mollentze N., Nel L.H., Haydon D., & Soubeyrand S. (2013). A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data (submitted).

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. DH. Nel & Prof. JW. Markotter

27 September 2013

developed in recent years. The first approach is based on coalescent models that assume some form of population dynamic model to relate the demography of the pathogen to its evolution, while implementing a diffusion model to account for the movement of the pathogen over geographic space (Lemey *et al.*, 2010). This approach has the advantage of being relatively robust to the intensity of epidemiological sampling, but because such models do not have an explicit epidemiological formulation, the inferences cannot easily be related to real epidemiological processes. Although coalescent models with an explicit epidemiological focus have recently been developed (Volz *et al.*, 2009; Frost & Volz, 2010), they have yet to be integrated into a framework that accounts for the spatial aspects of disease transmission. The second approach is more forensic in nature, explicitly recognizing the host population structure and the epidemiological processes that govern the interaction of host and pathogen. This second approach is based on epidemiological models of transmission and reconstructs the transmission tree reflecting 'who-infected-whom', thus allowing direct investigation of epidemiological processes (Ypma *et al.*, 2012; Morelli *et al.*, 2012). However, current methods cannot handle large numbers of missing infections, and therefore require a high proportion of infected hosts from the outbreak to be present within the sample.

Both these techniques have been applied within epidemic contexts and to data that are assumed to be monophyletic (i.e. arising from a single introduction). When pathogens are sampled from infected hosts in an endemic context, the epidemiological situation is potentially more complex. In this context, the connection between cases applies at two scales (Figure 3.1 A). At the scale of the entire endemic region, all cases may be related in some way (through the global transmission tree), leading to genetic relatedness and spatial autocorrelation between sampled cases. However, in a given study region (even one that has been exhaustively sampled), only some cases will be directly related through chains of transmission, and many chains of transmission may exist that are only indirectly related to each other by virtue of sharing a common ancestor outside the sampled area. The sample of pathogens within the study area is therefore polyphyletic. The picture is further complicated because surveillance is unlikely to be exhaustive, and therefore the sampling will be incomplete. Undetected or unsampled cases will reduce the detectable correlation between cases that are nevertheless causally related. Thus, if we hope to use genetic data to understand the detailed transmission biology of endemic pathogens the challenge will be to develop algorithms that can accommodate the polyphyletic nature of pathogen population

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

structure while accounting for and allowing inferences to be made regarding the unobserved and unsampled infections.

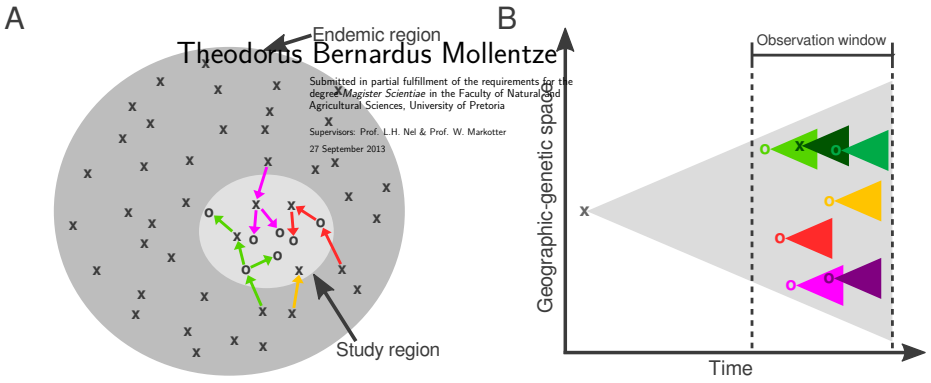

Figure 3.1: Modeling the transmission of endemic diseases. **A:** All cases in the study region are in some way related both genetically and spatially because they form part of a larger epidemic that originated from a single progenitor. This, along with the fact that some cases go undetected, makes determining dependence among transmission chains difficult. O's represent sampled cases, while X's represent unsampled cases. **B:** The pathogen radiates both in terms of genetic diversity and in terms of the geographic space it has invaded. In the relatively short observation window (compared to the time for which the pathogen has been established in the area), three types of relationships are apparent. Cases directly infected by another sampled case are easily linked into a chain of transmission (pink/purple). Cases caused by a case outside the study area will be more closely related to the common ancestor of all sampled cases than to any other cases (red and yellow), while some cases can be linked to sampled cases through unsampled intermediary cases (green).

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

This chapter describes the development of a novel technique for the reconstruction of transmission trees of endemic diseases and polyphyletic epidemics. This technique represents an extension of previous work in this field (Morelli *et al.*, 2012) to accommodate the complexities inherent to polyphyletic and partially sampled outbreak data containing space, time and genetic information. By accounting for the potentially high number of unsampled cases of infection, this technique also allows inference of the infected host population size over the study period and region, thus providing upper and lower estimates of the number of undetected or unsampled cases. This approach can be expected to be particularly useful for investigation of endemic RNA viruses, because their mutation rates are high enough for population genetic and epidemiological processes to occur on similar timescales, and spatial expansion and epidemiology leaves a discernible fingerprint on the genetic structure of these viruses (Grenfell *et al.*, 2004; Davis *et al.*, 2007; Biek *et al.*, 2007).

The technique was applied to endemic rabies virus (RABV) in the KwaZulu Natal province of South Africa (KZN) to demonstrate how it allows for a better understanding of the spatial epidemiology of endemic viruses. Such knowledge is

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

crucial for advancing the effectiveness of large scale vaccination campaigns – some of which have been in place for decades, but have failed to eliminate the disease in question. RABV has become endemic throughout the developing world (World Health Organization, 2002). It is typically transmitted through direct contact of saliva from an infected animal with broken or perforated skin of a susceptible host (Rupprecht *et al.*, 2002), but the epidemiological dynamics of rabies are complicated by two factors. First, rabies has a highly variable incubation period (Charlton *et al.*, 1997; Hampson *et al.*, 2009) and second, rabies has a very large host range that includes all mammals, many of which would play no part in the onward transmission of the virus (i.e. many hosts are 'dead-end' hosts; Rupprecht *et al.* 2002). Nevertheless, the majority of infections in humans are associated with rabid domestic dogs (*Canis lupus familiaris*; World Health Organization 2002; Cleaveland *et al.* 2006), and it is in dogs that the disease must be controlled if the burden on humans is to be reduced (Lembo *et al.*, 2011).

3.2. Materials and Methods

3.2.1. Data preparation

Thirteen of the 189 cases sequenced in chapter 2 lacked GPS coordinates and were therefore excluded from the transmission tree reconstruction (Appendix A, Table A.1). The remaining 176 trimmed sequences (section 2.2.7) were realigned using the FFT-NS-i algorithm of MAFFT version 6 (Katoh & Toh, 2008).

The hypothetical ancestral sequence of all cases was reconstructed from this alignment using the FastML server under the generalised time reversible model (Rodriguez *et al.*, 1990; Ashkenazy *et al.*, 2012).

3.2.2. Transmission tree construction[2]

We then designed an inference algorithm and a post-processing analysis in a Bayesian framework providing joint and marginal posterior distributions for the unknown parameters (see below). Inference was performed using an interacting Markov chain Monte Carlo (MCMC) algorithm based on Metropolis-Hastings updates. This

[2]The following model descriptions (section 3.2.2 and section 3.2.3) are given in their intuitive form. A more formal, mathematical description can be found in Mollentze N., Nel L.H., Haydon D., & Soubeyrand S. (2013). A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data (submitted).

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

algorithm simultaneously performs the estimation of model parameters, hidden variables and transmission trees.

3.2.2.1. Generalisation of the basic model

The core algorithm used here is a generalisation of the algorithm of Morelli *et al.* (2012; described in section 1.3.5.2) to allow its application to any directly transmitted disease. Morelli *et al.* calculated the date of infection of farms from a probability distribution centred around assessments by experts of the age of foot-and-mouth disease virus lesions on each farm. However, the ability to determine the age of infection is rare amongst viral diseases, and even when this is possible, such data may not be recorded. We therefore employed the epidemiological model already used to estimate the probability of cases being infectious at any given moment to also estimate the probability of all possible dates of infection. As a simplification, we assumed that cases are detected shortly after death, thus reducing the epidemiological model to one akin to the widely used Susceptible-Exposed-Infectious-Removed (SEIR) compartmental model (Figure 3.2). In this model, a susceptible host i becomes infectious at time T_i^{inf}, thus entering the 'exposed' compartment. It remains in this compartment for time-period $\mathbf{L}_i$, drawn from a probability distribution of incubation periods with a strict prior distribution (based on the results of Hampson *et al.* (2009) in the case of rabies; section 1.3.3), before becoming infectious. The host remains infectious for time-period $\mathbf{D}_i$, drawn from a probability distribution of infectious periods (again based on the results of Hampson *et al.* for rabies) and is then either immune against re-infection for life or, as is the case for rabies, dead. Because we assume that cases are observed shortly after the end of the infectious period and the observation date is known, we can estimate both the probable infectious period and the probable date of infection for all cases. From this data it is possible calculate the probability of a transmission from any host j to any host i based on the probability of j being infectious at the time of i's infection (Figure 3.2), with various values for T_i^{inf} being drawn from its calculated probability distribution during the MCMC estimation procedure (Morelli *et al.*, 2012).

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

However, as described in section 1.3.5.2, this forms only part of the probability of transmission between hosts. The spatial component of the likelihood equation was modified to accommodate a wide variety of spatial transmission patterns by replacing the exponential transmission kernel with the exponential-power spatial transmission kernel described by Austerlitz *et al.* (2004). This kernel is often used in

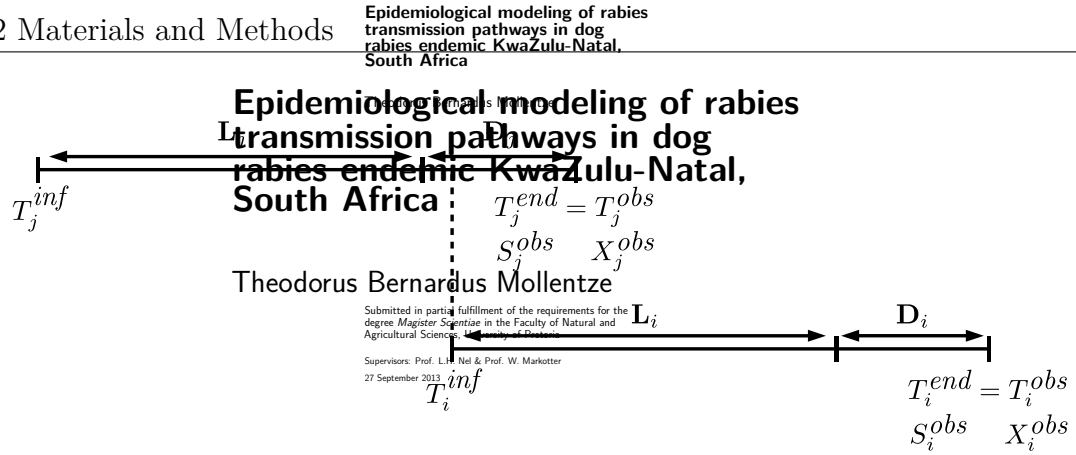

Figure 3.2: Calculating the probability of possible transmission events. Disease progression is modelled in two stages: susceptible host i becomes infected at time T_i^{inf}, and incubates the virus for duration $\mathbf{L}_i$. At the end of the incubation period, the host becomes infectious for duration $\mathbf{D}_i$ before being removed from the population (either because of death or life-long immunity). At this point, the case is observed, with the date (T_i^{obs}) and location (X_i^{obs}) being recorded and a partial pathogen genome sequence (S_i^{obs}) being observed. Thus, the probability of a transmission from any host j to host i can be calculated from the temporal overlap of the probability distributions of i's date of infection and j's infectious period, across with the probability of S_j^{obs} evolving to S_i^{obs} in the time between observations and of the pathogen dispersing from location X_j^{obs} to location X_i^{obs} in the time between observations.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa
Theodorus Bernardus Mollentze
Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria
Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

dispersal studies and can take a variety of shapes, making it well suited to a range of endemic situations where often very little is known regarding spatial transmission patterns. We also replaced the simplified substitution model of Morelli *et al.* with the Kimura 3-parameter model, which was the best fitting DNA substitution model for the dataset according to Akaike's information criterion (determined using jModelTest version 2.1; Kimura, 1981; Akaike, 1974; Darriba *et al.*, 2012). This model takes into account the different probabilities of transitions (mutations from U $\leftrightarrow$ C and A $\leftrightarrow$ G), and two types of transversions (U $\leftrightarrow$ A and C $\leftrightarrow$ G versus U $\leftrightarrow$ G and A $\leftrightarrow$ C) and can thus be expected to increase the discriminatory power of the approach.

3.2.2.2. Extension to polyphyletic transmission trees

In a partially sampled outbreak any given infected host which was sampled might have been infected by: (1) another sampled host (through direct transmission), (2) an unsampled host which has been infected directly or indirectly by a sampled infected host (termed "indirect transmission" here) or (3) an unsampled host which has no ancestors within the sample, i.e. transmission from an exogenous source (Figure 3.1). The model of Morelli *et al.* (2012), allows for only a single virus introduction (i.e. a single "exogenous" transmission) followed by direct transmissions for the rest of the outbreak. Thus, the model is only appropriate for exhaustively

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* (Veterinary Epidemiology) in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

sampled, monophyletic outbreaks. We extended this model by allowing multiple unobserved cases to arise anywhere in both space and time within the set of inferred transmissions.

The likelihood equation of Morelli *et al.* (2012) models the spatial radiation and genetic evolution of cases over time to determine the likelihood of various parameters at any point in time and thus calculate the probability of different transmissions. In our model, this is equivalent to the approach taken for direct transmisions, where each sampled infected host species able to transmit the virus can be a source of infection (in the case of rabies, typically carnivores but not humans or livestock). These are modelled by the probability distribution $\mathcal{P}_{direct}$, defined over the geographic-genetic space and evolving in time (represented by coloured cones in Figure 3.1 B). $\mathcal{P}_{direct}$ is dependent on the infection time of the host (estimated as described above), its incubation duration (estimated), its removal or observation time (observed), a spatial dispersal kernel (estimated) and substitution rates for the sequence evolution (estimated).

Each sampled infected host which can spread the disease can also be an indirect source of observed infections even after its removal, as a consequence of unsampled intermediate hosts: Case A (sampled) infects B (unsampled) which infects C (unsampled) which infects D (sampled). Since these unsampled cases extend the influence of Case A in both geographic and genetic space, their effect can be modelled by allowing cases to continue moving and evolving after their death (represented by the green and red cones in Figure 3.3). This is represented by probability distribution $\mathcal{P}_{indirect}$, again defined over the geographic-genetic space and evolving in time. As is the case for $\mathcal{P}_{direct}$, $\mathcal{P}_{indirect}$ depends on the infection time of the host, its incubation duration, a dispersal kernel and the substitution rates. The spatial influence contributed by sampled cases is harder to determine. We considered two different specifications for the dispersal kernel governing indirect transmissions ($\mathcal{K}_{indirect}$). In the first specification, we conservatively assume that $\mathcal{K}_{indirect}$ is the same as the dispersal kernel used for the direct transmissions, thus allowing only movement over transmission distances observed for (single) direct transmissions. However, this does not adequately accommodate a scenario encompassing multiple unsampled intermediate cases, where greater distances between the indirectly connected cases would be possible. We therefore also considered a more liberal specification, where $\mathcal{K}_{indirect}$ is a uniform distribution over the whole study region, thus allowing unsampled intermediate hosts to carry the virus to any location within the sampled region.

These two scenarios form extremes between which the true process can reasonably be expected to occur.

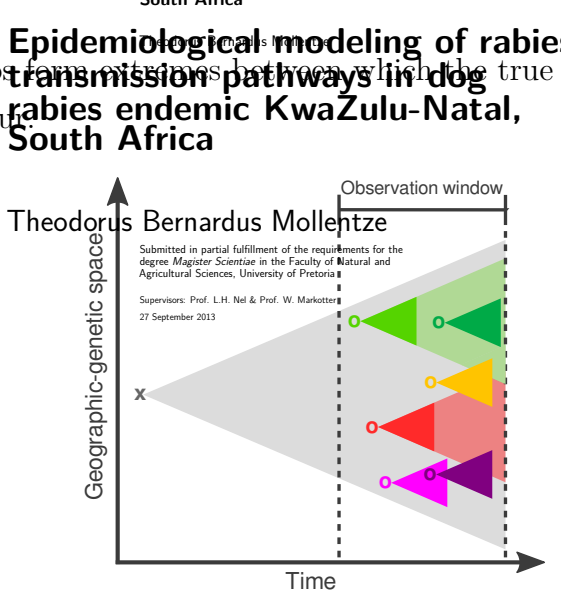

Figure 3.3: Inferring indirect transmissions. The unsampled index case is represented by an x, while o's represent observed cases. The influence of cases in both the genetic and geographic space will be extended by cases further down the transmission chain. Thus, the effect of unsampled cases can be modelled by allowing cases to continue moving and evolving after their death, as illustrated by the green and red cones. In this way, we can detect the indirect causal connection between the light and dark green cases in the figure caused by an unsampled intermediate case whose host was infected by the light green case and which then went on to infect the host of the dark green case.

In a similar vein, the source of exogenous transmissions can be modelled as a probability distribution $\mathcal{P}_{exo}$ defined over the geographic-genetic space and evolving in time (represented by the grey cone in Figure 3.1 B). $\mathcal{P}_{exo}$ can be completely specified based on an ancestral virus sequence (determined *a priori* through ancestral state reconstruction, in our case using FastML, although any algorithm can be used), a time for the ancestral sequence, and the same substitution rates as above (both of which are co-estimated with the transmission tree). The ancestral sequence and the sampled infected hosts generate a mixture $\mathcal{M}$ of spatio-temporal-genetic distributions ($\mathcal{P}_{exo}$, $\mathcal{P}_{direct}$ and $\mathcal{P}_{indirect}$) from which the infection events are drawn. Estimating the source that infected a given host involves assessing in which component of the mixture model $\mathcal{M}$ the infection of the host arose.

Conceptually however, the source of both types of transmissions involving unobserved ancestors (indirect and exogenous) can be modelled in the same way – as being external to the sampled dataset, meaning the transmissions arise in $\mathcal{P}_{exo}$. Thus, to reduce complexity and computation time we distinguished only between direct and "unsampled" sources in the first instance (only $\mathcal{P}_{direct}$ and $\mathcal{P}_{exo}$ are used to define $\mathcal{M}$), with a post-processing algorithm to distinguish between indirect and true exogenous transmissions. In the previously described monophyletic model (Morelli

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Wollentze

Submitted in partial fulfillment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

et al., 2012), the posterior distributions of the incubation and infectious period durations can be determined by indirect links between cases. We used narrow priors for the parameters governing these distributions, essentially forcing a decision between direct transmission or linkage to an exogenous source in the first step. To distinguish between exogenous and indirect transmissions, the post-processing analysis applies a Metropolis-Hastings update to the unsampled" transmission links determined by the MCMC algorithm, which involves comparing the probability that the transmission was really from an exogenous source (based on $\mathcal{P}_{exo}$) with the probability that it was merely indirect (based on $\mathcal{P}_{indirect}$). This post-processing was applied under both the conservative and liberal specifications of $\mathcal{K}_{indirect}$ described above.

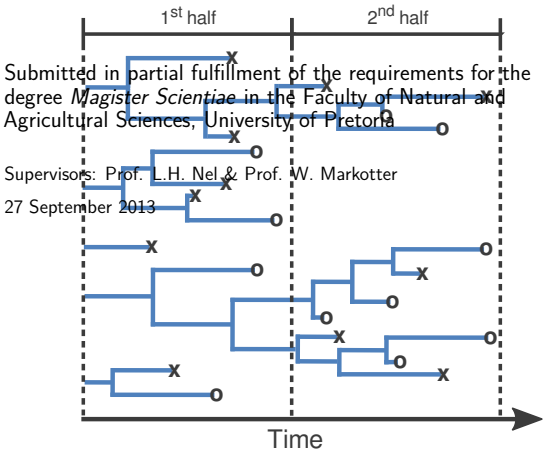

Figure 3.4: The inferred causal relationships among detected cases can be used to estimate the true number of cases despite incomplete surveillance using a mark-recapture strategy. Here, lineages in the second half of the observation period are considered 'recaptured' if one of its ancestors was sampled in the first half of the study period.

3.2.3. Population size estimation

To determine the true number of cases represented by indirect transmission links, a mark-recapture technique was applied to the virus lineages identified in the previous analysis. In this approach, the cases are divided into two categories based on their observation dates. Any host sampled in the second time-period is considered as recaptured if it was directly or indirectly infected by a host sampled in the first part of the data set (Figure 3.4). Thus, if the exact transmission tree were known, the true size of the population of infected hosts could be determined using standard mark-recapture statistics taking into account uncertainty regarding changes in the

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Barnardus Wollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

population size from the first to the second period. In our case, the size of the dataset and the large number of unknown parameters means we only have access to the posterior distribution of probable transmission trees. However, for each element of this sample, the number of recaptured virus lineages can be calculated and therefore the posterior distribution of the number of recaptured virus lineages can be assessed.

3.3. Results and Discussion

A novel method for reconstructing transmission trees using space-time and genetic data from partially sampled endemic diseases was developed and applied on 176 canid-associated rabies cases for which detailed epidemiological data were available (Appendix A, Table A.1). This represents all but 13 of the canid-associated cases detected in KZN over a 15 month period (90.7% of cases). The dataset contained 153 rabies cases detected in domestic dogs, 1 case detected in a jackal (which species was not recorded but given its location of origin it was most likely *Canis mesomelas*) and 22 cases detected in domestic livestock (Appendix A, Table A.1). Livestock typically do not transmit rabies, and these cases were explicitly treated as dead-ends for transmission in the model. When considering only direct transmissions, there were several independent chains of transmission and many transmissions over long distances. (Figure 3.5 and Figure 3.6 A and B). The mean distance between the most probable directly connected cases was 14.9 km (0.025- and 0.975-quantiles: 0.0 km and 56.1 km; Figure 3.6 A and B), in stark contrast to the mean transmission distance of 0.88 km observed by contact tracing in rural Tanzania (Hampson *et al.* 2009; section 1.3.3). Occasional long-distance transmissions in this area, particularly along the major highways that follow the KZN coast (Figure 2.2 B), have been identified before based on phylogenetic patterns and have been ascribed to motorised transportation of dogs (Coetzee & Nel, 2007). Road distances have also been shown to be a better predictor of rabies dissemination than absolute distances in northern Africa, again implying that humans are responsible for long distance transmission of rabies (Talbi *et al.*, 2010). The long distances and short time-periods between cases in the transmission tree (Figure 3.6) provide further evidence for motorised transportation of infected dogs, but such transmissions were not restricted to any one area and instead appear to be a common feature of the epidemiology of rabies in this area. This might be due to the high prevalence of circular human migration and migrant labour in many parts of KZN, with urban-dwelling migrants visiting their

secondary rural households (and it would seem taking their dogs with them) on a regular basis (Posel & Marx, 2013).

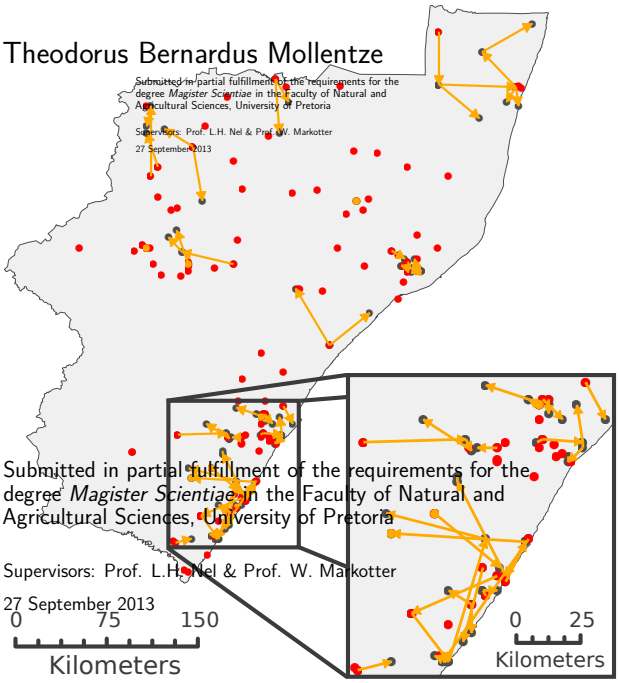

Figure 3.5: Transmission tree showing the direct pairwise transmissions with highest posterior probabilities. Transmission links between cases are represented by orange arrows (and dots when a transmission links cases at the same location), while red dots represent cases for which no direct ancestor was detected. The inset shows an enlarged view of connections in the densely populated southern coast of KZN, were the majority of cases where detected.

The majority of cases could not be linked through direct transmissions – 69 (95 % posterior interval [PI]: 60-79) direct transmissions were identified, while unsampled sources were the most likely link for the remaining 107 (95 % PI: 97–117) cases. The conservative specification of our post-processing algorithm identified a further 37 (95 % PI: 27–47) indirect transmission links over the 15-month study period, while the liberal version of the algorithm identified 67 (95 % PI: 57–78) indirect transmissions (Figure 3.7). When considering both direct and indirect connections, there are many separate, unjoined transmission trees (Figure 3.8). For the most probable connections under both the conservative and liberal specifications of the algorithm, these transmission trees can be grouped into 8 distinct spatial clusters (Figure 3.8; although the clusters do not represent exactly the same cases between the two specifications). Transmission between different spatial clusters was rare – only one such transmission was detected with the conservative specification of the algorithm, and ten such transmissions with the liberal specification. In addition,

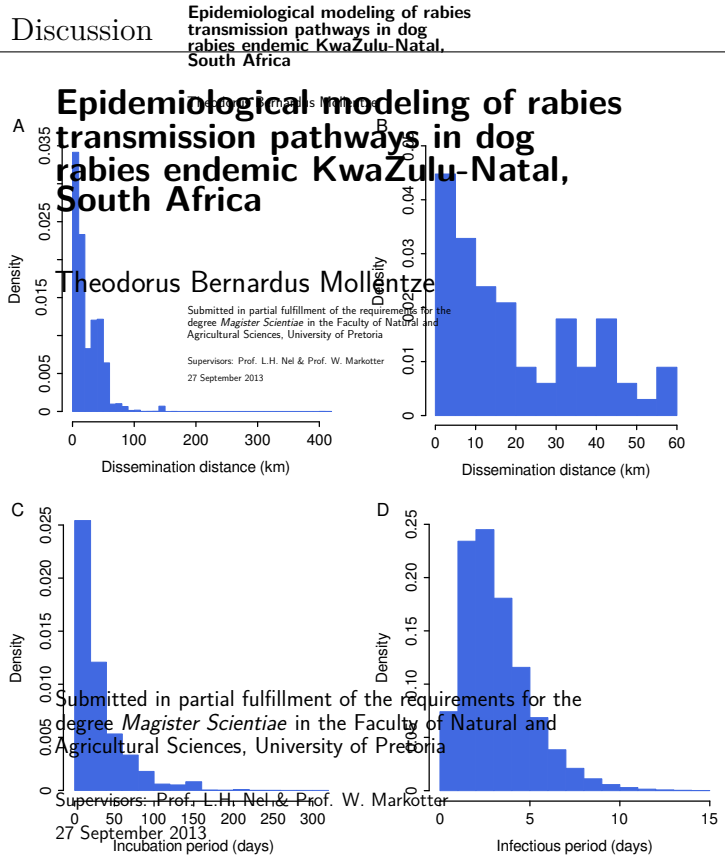

Figure 3.6: Posterior distributions of the transmission distances, incubation period and infectious period of directly connected cases. **A:** Transmission distances between all *a posteriori* directly connected cases. **B:** Distances between connected cases corresponding to the direct transmission links with the highest posterior probabilities (i.e. only the most probable links). **C & D:** Posterior distributions of the incubation period and infectious period of directly connected cases. The distributions in C and D were obtained by aggregating the respective posterior distributions of all cases responsible for direct onward transmission in the transmission tree.

such transmissions do not appear to seed substantial additional numbers of cases, as only one instance of onward spread in the new cluster was detected under either specification, causing just one additional case in both instances. Interestingly, four of the inter-cluster transmissions identified under the liberal specification involved transmission from one cluster to another and then back to more-or-less the same location, before onward transmission in the original cluster, further supporting the hypothesis of migrants moving dogs back-and-forth between their urban and rural homes.

Among the most probable connections, between 44 % and 24 % of all cases had no detectable origin in the study area (conservative and liberal specifications, respectively), while between 29 % and 12 % of cases were also not linked to any subsequent cases (grey squares in Figure 3.8). Both versions of the post-processing algorithm detected no indirect transmissions in the first quarter of the sampling

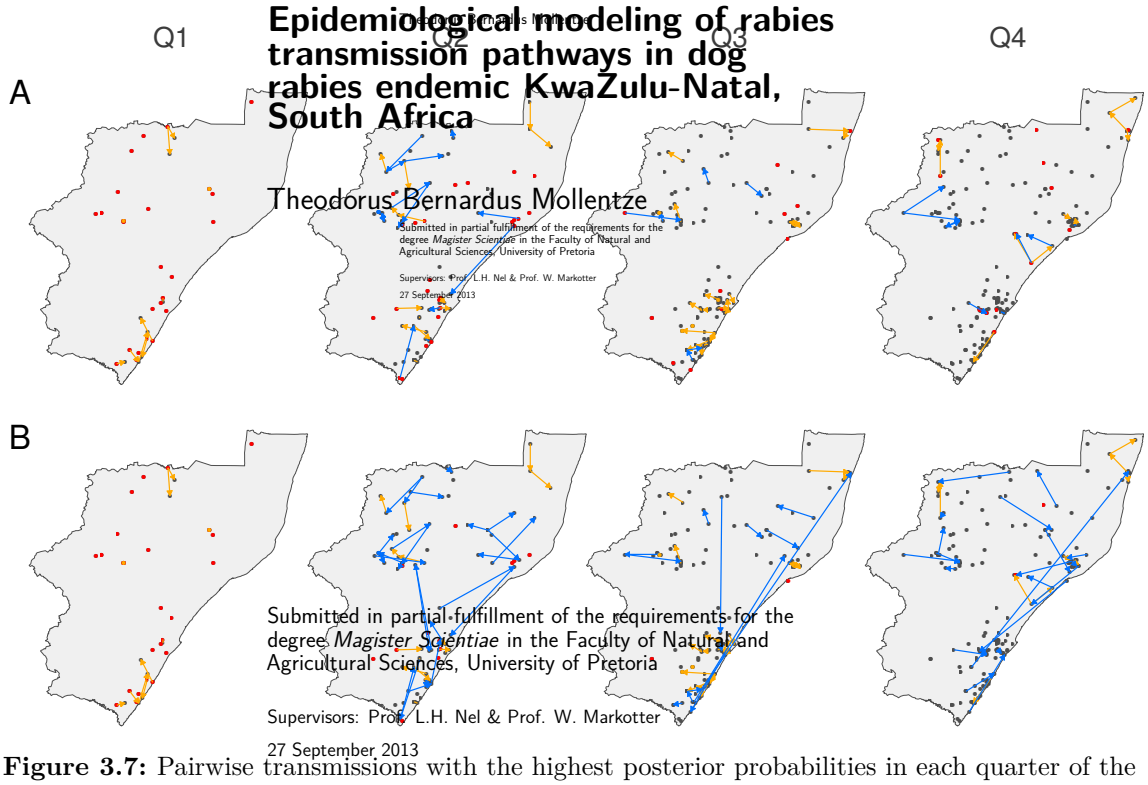

Figure 3.7: Pairwise transmissions with the highest posterior probabilities in each quarter of the sampled period, including indirect transmissions. Grey dots represent all cases since the start of the sampling period, while red dots represent cases appearing in that quarter that have an exogenous source. Orange arrows represent direct transmission events, with orange dots representing direct transmission between cases at the same location. Blue arrows and dots represent indirect transmissions inferred using the conservative (**A**) and liberal specification (**B**) of the post-processing algorithm. Note that detected cases (grey dots) are displayed cumulatively. Q1-Q4: First to fourth quarter of the sampling period.

period (Figure 3.7) and it is likely that the majority of cases missing ancestors during this period represent dogs already infected by the start of the sampling period. This interpretation is further supported by the observation that the number of unconnected cases (cases inferred as infected by an external source) dwindles in the subsequent quarters, before stabilising late in the second quarter (Figure 3.7 and Figure 3.9). Thus it would appear that the variable and occasionally long incubation period, along with incomplete surveillance, necessitates sampling for at least 6 months before the dataset contains the majority of lineages present in the area. The number of unconnected cases stabilises over the last period of the data series at a posterior median of 2.27 (95 % PI: 0.6–5.1) and 0.2 (95 % PI: 0.0–1.3) per month for the conservative and liberal post-processing algorithms respectively (Figure 3.9). As described above, the reality is likely between these two extremes. Nevertheless, the relatively high numbers of unconnected cases means that these values likely still represent a combination of lineages that have not been previously

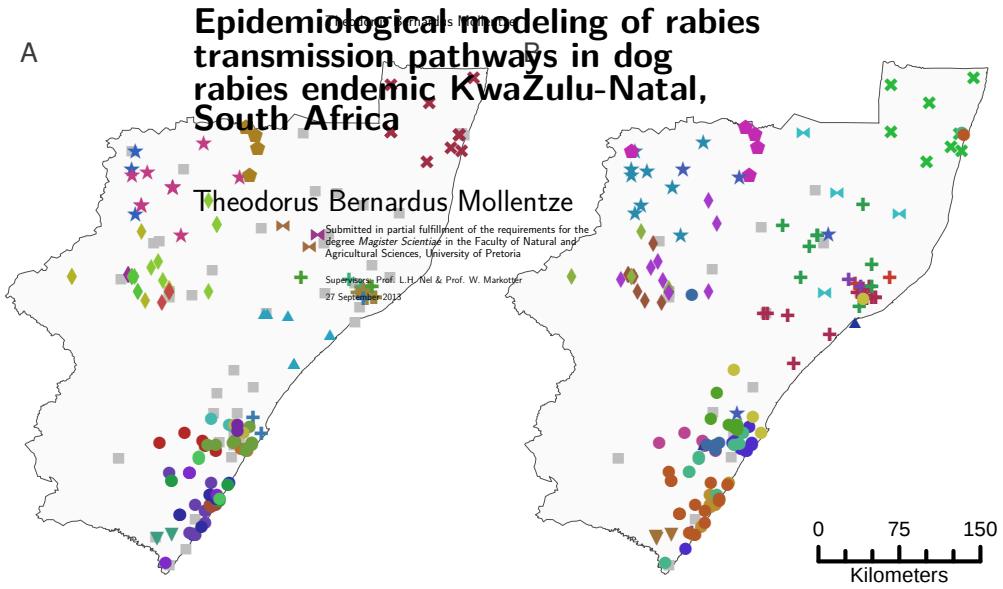

Figure 3.8: Cases belonging to independent transmission trees (indicated by colour) and spatial clusters (indicated by symbols) and completely unconnected cases (indicated by grey squares) when considering the most probable direct and indirect connections between cases. Indirect connections were determined by a conservative (**A**) and liberal (**B**) specification of the post-processing algorithm. When multiple cases of two or more independent transmission trees occur within 14.9 km of each other, these transmission trees are considered part of the same spatial cluster (14.9 km is the mean transmission distance observed amongst the most probable direct transmission links).

detected in addition to introductions from outside the study area. Indeed, the phylogeographic analyses described in chapter 2 detected very few introductions, and comparison with these results shows that the majority of these unconnected cases represent lineages previously sampled in KZN (Appendix A, Figure A.1). In the phylogeny, the relationships between many of these cases could not be resolved beyond common ancestors existing several decades ago. However, some unconnected cases in lineage 1 show no more genetic diversity than other cases which are connected by the transmission tree reconstruction algorithm, particularly under the conservative specification (Figure A.1). It may be that there is an upper limit to the number of undetected cases between any two cases beyond which the relationship is no longer detectable (at least with the method described here) although the genetic separation between these cases and other, unconnected cases show no evidence for this. Alternatively, the choice of strict prior distributions conditioned on data from Tanzania for the incubation and infectious periods of infections, necessitated by the need to detect and account for unsampled cases, may lead to some connections involving outliers being missed if the true distribution of RABV incubation and infectious periods is slightly different between KZN and Tanzania.

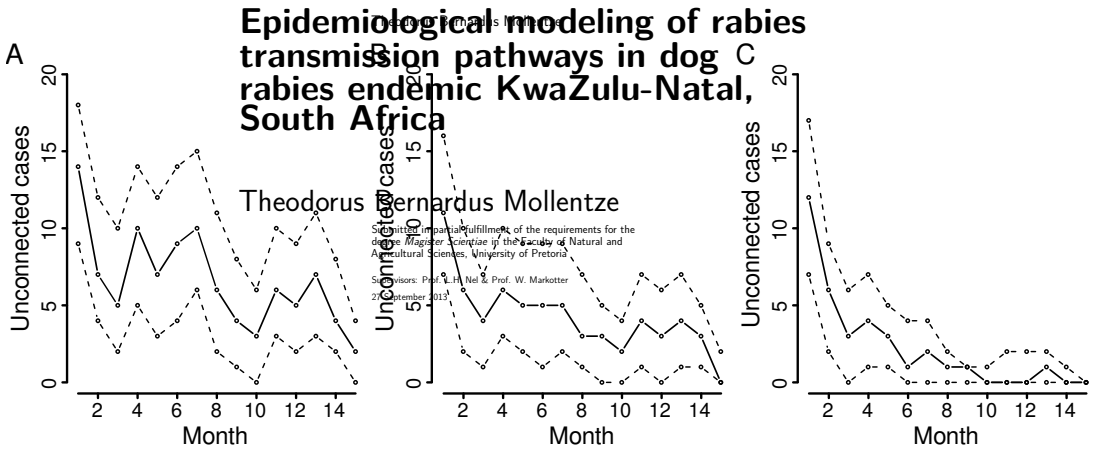

Figure 3.9: Number of cases with no detectable source in the dataset per month of the sampling period. **A:** Number of cases infected by an exogenous source under the main inference algorithm allowing only direct or exogenous connections. **B:** Cases still infected by an exogenous source after applying the post-processing algorithm using the conservative specification, and **C**, after applying the post-processing algorithm using the liberal specification. Solid lines indicate the median of the posterior distributions, while dashed lines show the 95 % posterior intervals.

To gain a better understanding of the surveillance failures leading to the high number of indirect connections detected, the true number of cases occurring in the study area was estimated by using a mark-recapture-based approach. This yielded a posterior median estimate of 389 cases (95 % PI: 260–881) using the conservative specification of the post-processing algorithm, and 195 cases (95 % PI: 182–298) using the liberal specification, over the 15 month study period (Figure 3.10). Again, these values can be interpreted as a lower and upper bound of the true estimate. Although the apparent undetected connections described above may lead to a slight over-estimation of the true number of cases, this is balanced by the undetected lineages evident from the phylogeographic analysis (Appendix A, Figure A.1). The herpestid-associated genetic variant of rabies virus is rare in KZN, which means the five cases which could not be sequenced were most likely representatives of the canid-associated variant. Thus, surveillance detected 194 cases of the canid-associated variant, or between 49.9 % and 99.5 % of all canid-associated cases (based on the posterior medians of the conservative and liberal specifications, respectively). The areas where surveillance is poor can be deduced from the locations of indirect links (Figure 3.7), providing a powerful tool for improving detection rates.

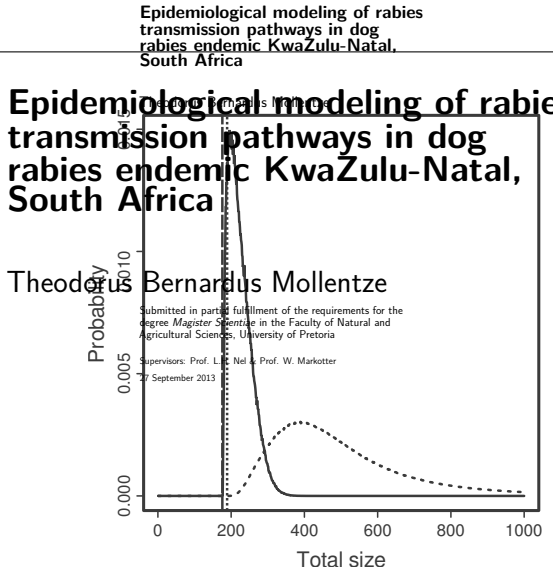

Figure 3.10: Estimated total number of animals infected by the canid-associated variant of rabies virus in KZN over the sampled period. Curves show the posterior distribution of cases under the conservative (dotted curve) and liberal (solid curve) specification of the transmission kernel ($\mathcal{K}_{indirect}$). As discussed in the text, the true number lies somewhere between these two extremes. The dashed vertical line shows the number of cases included in the analysis, while the dotted vertical line shows the number of cases detected by surveillance (including 5 which could not be confirmed as belonging to the canid-associated variant – see text for details).

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria
Supervisors: Prof. L. H. Nel & Prof. W. Markotter
27 September 2013

3.4. Conclusions

To successfully control rabies and other endemic diseases in a changing landscape, a detailed understanding of their spatial epidemiology is required. The method described here allows for the detailed reconstruction of the transmission events of endemic infectious diseases, providing information that can be used both in designing more efficient control strategies and to measure and improve the quality of surveillance programmes. Although the reconstructed connections are in agreement with phylogenetic data, the resolution is significantly improved. In addition, the reconstructed transmission tree allows the direction of transmission to be discerned, an important factor to consider in the design of control programs, particularly when disparate areas are connected by only few transmissions as the case is here.

The long distances characterising many internal transmissions point to a significant anthropogenic influence on the epidemiology of rabies in KZN. The causes of this phenomenon require further study, but it is clear that education campaigns should form an integral part of the rabies control programmes in this area. Such long transmission distances complicate the detection of unsampled cases – cases linked by unsampled intermediary cases would be indistinguishable from a scenario where a case with a long incubation period was transported over a long distance. Despite these long-distance transmissions, clear spatial groupings could be discerned (Figure 3.8).

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

In addition, the frequent, long-distance transmissions cause most of these spatial clusters to consist of a small core area and numerous surrounding cases (Figure 3.7). Identifying the connections of surrounding cases to specific clusters enables more directed vaccination, where targeting the much smaller core areas would allow control of rabies over large areas. In addition, identifying the spatial scale at which independent control strategies can be applied means it is possible to replace the current thin spread of limited resources across the province with intense, focused campaigns that move across the province on an annual basis.

By applying the methods described here to data from multiple years, important information will be revealed about how to iteratively improve surveillance and adapt rabies control strategies by identifying areas to be prioritised during annual vaccination campaigns. In addition, these methods can easily be adapted to other endemic diseases, and the high mutation rate of RNA viruses make them ideal candidates for this approach. Particularly encouraging is the fact that the small genome region sequenced here provided sufficient resolution for this analysis, making the generation of adequate data for large numbers of cases feasible even in resource-poor areas.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Chapter 4.

Concluding remarks

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Despite marked decreases in the number of cases evident in both surveillance data and demographic reconstructions, rabies remains widespread throughout most of KZN. Increasing success in the fight against rabies also generates new challenges, and subsequent campaigns will likely need to be increasingly efficient given that a disease with an ever decreasing public-health impact is unlikely to remain a public health and political priority. Thus, sustainable and efficient rabies control strategies are needed, particularly because the control strategies designed and tested in KZN are intended for eventual application to other areas, where resources may be much more limited. The design of these strategies are crucially dependent on detailed knowledge of rabies transmission and epidemiology in the settings where they are to be applied, which in turn depends on increasing surveillance efficacy and improved statistical analyses.

The results presented here point to several important recommendations for KZN. Given the relatively independent nature of spatial clusters of cases, as well as the low number of introductions from outside the province, it may be pertinent to examine alternative rabies control strategies consisting of compartmentalised campaigns targeting key areas. This can easily be achieved using mathematical modeling, which should consider the potential impact of the small numbers of inter-cluster transmissions and introductions that do occur, and how their effect can be counteracted using various configurations of temporally overlapping campaigns targeting the different areas. Redesigning control strategies would also require a better understanding of the potential effect of introductions to areas where RABV is near elimination. For example, the transmission trees in KZN show that the rare inter-cluster transmissions which could be detected do not become established, which may also explain the low number of introductions from outside KZN that

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

were detected. It may be that this phenomenon is simply because of resource limitation (see the discussion of the "surfing mutation" model of range expansion in section 1.2.4), in which case eliminating rabies locally would require constant vigilance as well as alternative control strategies such as dog-movement restrictions to prevent re-emergence. Thus, further investigation to establish the cause of this phenomenon will be an important next step. The spatial barriers preventing or slowing the spread of rabies also require investigation, as these can be reinforced with less resources than other strategies aimed at stopping transmission between specific areas (Russell *et al.*, 2006). In this regard, the ability to reconstruct transmission trees for endemic diseases using only partial genome sequencing will allow relatively simple generation of large datasets of transmissions in various areas spanning several years.

A further challenge to achieving sustainable dog rabies elimination is the lack of adequate surveillance throughout much of the developing world. When very few cases are detected, it will be difficult to confirm the efficacy of control programmes (or indeed to determine where to apply control programmes) and there will be no clear indication of when elimination is achieved. Townsend *et al.* (2012) developed a mathematical model that could be used to calculate the time delay between RABV no longer being detected by inadequate surveillance and the point at which the virus is truly eliminated. This model showed that vaccination could be halted after six months of no longer detecting cases, provided that surveillance is able to detect at least 10 % of cases (Townsend *et al.*, 2012). The methods to estimate the true proportion of cases being detected demonstrated here can be used to improve surveillance and could be combined with the model of Townsend *et al.* to calculate the required length of control programmes more precisely. In the case of KZN, where upwards of 50 % of cases are being detected, the time period before vaccination can be halted is likely to be much shorter. However, in rabies endemic areas such as southern Africa, there remains a constant threat of re-introduction. Although introductions appear not to become established in clusters of existing cases at present, the exact causes of this remain to be established. Thus, while rabies remains endemic in the surrounding area it is likely that buffer zones and constant surveillance to allow rapid response to new outbreaks will be needed indefinitely. This means that the only way to achieve sustainable elimination is by using a concerted, regional approach. A first step, which would directly benefit the KZN rabies control programme, would be to improve the quality of surveillance in neighbouring regions. By detecting and

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

characterising more cases from neighbouring countries and provinces, it would be possible to examine the source of new outbreaks to allow a rapid response. To make these data truly useful, and to allow the use of analyses such as those presented here to determine the source of introduced cases, accurate geographic coordinates should be collected for all cases. From the results presented here, it is clear that the surveillance system in KZN is effective, and this expertise should be shared at a regional level.

There is no technical barrier to prevent effective control of canid-associated rabies across the region. Most of Mpumalanga managed to remain rabies free until recently despite the presence of endemic canine rabies in two neighbouring countries as well as two neighbouring provinces, while dog rabies had largely been under control in Limpopo for several years prior to re-emergence, despite the presence of Canid-associated rabies in wildlife reservoirs (Mkhize *et al.*, 2010; Sabeta *et al.*, 2011). There is no evidence of independent wildlife reservoirs in the eastern parts of southern Africa, making elimination of canine rabies even more feasible. These examples also highlight the difficulties of maintaining a rabies free status however – rabies spread from a remaining endemic area to once again affect most of Mpumalanga, while the spread of rabies from neighbouring Zimbabwe into Limpopo led to numerous human deaths and now remains out of control in many areas (Cohen *et al.*, 2007; Mkhize *et al.*, 2010; Sabeta *et al.*, 2011).

The novel transmission tree reconstruction algorithm demonstrated here can be used to investigate the epidemiology of various other endemic pathogens for which sufficient data are available. In particular, the posterior distribution of possible trees can be tested for statistically significant patterns correlating with specific landscape or sociological features, to test various hypotheses about the spread of specific pathogens. It may also be of interest to study the evolution of pathogens at the single case level, where the increased resolution of this method, together with its explicit incorporation of epidemiological knowledge, may lead to novel insights. In this regard, some improvements remain to be added, such as a variety of models of nucleotide evolution as well as more complex epidemiological models to accommodate for example vector-borne pathogens or other types of data that may be available. For many datasets of endemic infectious disease, the ability to accommodate missing data may also be needed. Although the results of Ypma *et al.* (2012) show that transmission tree reconstruction with missing sequence data is highly uncertain, it should be relatively straightforward to overcome the unavailability of geographic

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Molentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

coordinates for limited number of cases. The accuracy of the inferred transmission links will undoubtedly be increased by incorporating uncertainty in the dates of detection, since many cases may not be detected immediately. Meanwhile, the resolution of this method would be much greater if full genome sequences are used, the generation of which is increasingly becoming more accessible. Although it would require much greater computational power, it may also be possible to increase the resolution even further by modeling the actual sequences that are *transmitted*, rather than relying on the differences between observed sequences.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Appendix A.

Additional figures and tables

Table A.1: Rabies cases detected in KwaZulu Natal between 1 March 2010 and 8 June 2011

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Case number	Date	Host species	Latitude (degrees)	Longitude (degrees)	Accession number
10/129	2010/03/03	Unspecified caprine species	-30.62	30.45	KC660293
10/128	2010/03/05	*Canis lupus familiaris*	-30.78	30.13	KC660323
10/146	2010/03/12	*Canis lupus familiaris*	-27.35	30.87	KC660234
10/155	2010/03/19	*Canis lupus familiaris*	-27.75	30.90	KC660179
10/157	2010/03/20	*Bos taurus*	-28.60	29.92	KC660255
10/153	2010/03/21	*Canis lupus familiaris*	-28.60	29.92	KC660207
10/154	2010/03/21	*Canis lupus familiaris*	-29.98	30.65	KC660233
10/164	2010/03/22	*Bos taurus*	-28.30	30.15	KC660254
10/167	2010/03/22	*Canis lupus familiaris*	-28.25	31.47	KC660224
10/173	2010/03/23	*Bos taurus*	-28.73	30.23	KC660183
10/168	2010/03/25	*Canis lupus familiaris*	-29.38	30.77	KC660240
10/175	2010/03/26	*Bos taurus*	-29.52	30.93	KC660227
10/188	2010/03/29	*Canis lupus familiaris*	-28.25	31.47	KC660196
10/183	2010/04/02	*Canis lupus familiaris*	-27.00	32.08	KC660341
10/184	2010/04/02	*Canis lupus familiaris*	-30.02	30.85	KC660282
10/212	2010/04/10	*Canis lupus familiaris*	-28.72	30.23	KC660230
10/202	2010/04/11	*Canis lupus familiaris*	-27.52	30.97	KC660201
10/203	2010/04/12	*Canis lupus familiaris*	-27.70	30.35	KC660167
10/195	2010/04/13	*Canis lupus familiaris*	-30.45	30.65	KC660329
10/205	2010/04/16	*Canis lupus familiaris*	-29.83	30.80	KC660244
10/209	2010/04/16	*Canis lupus familiaris*	-30.58	30.32	KC660285
10/207	2010/04/19	*Canis lupus familiaris*	-28.72	30.23	KC660229
10/220	2010/04/22	*Canis lupus familiaris*	-30.12	30.48	KC660327
10/229	2010/04/24	*Canis lupus familiaris*	-30.73	30.42	KC660351
10/233	2010/04/27	*Canis lupus familiaris*	-28.32	31.52	KC660222
10/236[2]	2010/05/02	*Canis lupus familiaris*	N.R.	N.R.	KC660235
10/237	2010/05/03	*Canis lupus familiaris*	-29.90	30.77	KC660328

N.R: not recorded; N.S: not sequenced (RT-PCR unsuccessful)

[1] Herpestid-associated variant of RABV (excluded from analysis)

[2] Excluded from analysis (RT-PCR unsuccessful or coordinates not recorded)

Table A.1 – continued from previous page

Case number	Date	Species	Latitude	Longitude	Accession number
10/245[2]	2010/05/04	*Canis lupus familiaris*	N.R.	N.R.	KC660166
10/252	2010/05/12	*Canis lupus familiaris*	-27.48	30.52	KC660163
10/261	2010/05/16	*Canis lupus familiaris*	-28.62	29.83	KC660228
10/277	2010/05/17	*Canis lupus familiaris*	-30.75	30.25	KC660322
10/268	2010/05/18	*Canis lupus familiaris*	-30.42	30.57	KC660305
10/267	2010/05/19	*Canis lupus familiaris*	-29.88	30.80	KC660243
10/269[2]	2010/05/19	*Canis lupus familiaris*	N.R.	N.R.	KC660265
10/272	2010/05/19	*Canis lupus familiaris*	-28.72	30.23	KC660275
10/274	2010/05/23	*Canis lupus familiaris*	-30.68	30.50	KC660299
10/271	2010/05/24	*Canis lupus familiaris*	-28.73	31.53	KC660209
10/278	2010/05/26	*Canis lupus familiaris*	-30.75	30.45	KC660303
10/286[1]	2010/06/08	Unspecified wildlife species	-28.03	29.98	KC660352
10/294	2010/06/22	*Canis lupus familiaris*	-28.54	30.59	KC660185
10/295	2010/06/22	*Canis lupus familiaris*	-30.32	30.58	KC660297
10/305	2010/06/29	*Canis lupus familiaris*	-30.23	30.23	KC660315
10/307	2010/06/29	*Canis lupus familiaris*	-28.58	29.89	KC660211
10/308	2010/06/30	*Canis lupus familiaris*	-28.58	29.89	KC660226
10/309	2010/06/30	*Canis lupus familiaris*	-28.72	30.56	KC660216
10/310	2010/07/01	*Canis lupus familiaris*	-29.85	30.77	KC660246
10/314	2010/07/05	*Canis lupus familiaris*	-28.63	30.18	KC660184
10/317	2010/07/06	*Canis lupus familiaris*	-28.07	32.15	KC660238
10/318	2010/07/06	*Canis lupus familiaris*	-28.73	31.82	KC660248
10/321	2010/07/06	*Canis lupus familiaris*	-27.41	30.95	KC660221
10/324	2010/07/07	*Canis lupus familiaris*	-29.83	30.78	KC660326
10/325	2010/07/07	*Bos taurus*	-28.19	31.00	KC660218
10/327	2010/07/09	*Canis lupus familiaris*	-29.98	30.15	KC660311
10/331	2010/07/16	*Canis lupus familiaris*	-28.68	31.92	KC660194
10/342	2010/07/23	*Bos taurus*	-30.10	30.75	KC660325
10/347	2010/07/28	*Canis lupus familiaris*	-30.73	30.40	KC660298
10/351	2010/07/30	*Canis lupus familiaris*	-29.90	30.85	KC660318
10/352	2010/07/30	*Canis lupus familiaris*	-29.73	30.60	KC660200
10/353	2010/07/30	*Bos taurus*	-30.10	30.48	KC660320
10/355	2010/08/02	*Canis lupus familiaris*	-31.00	30.23	KC660324
10/366	2010/08/06	*Canis lupus familiaris*	-28.81	30.17	KC660161
10/367	2010/08/06	*Canis lupus familiaris*	-28.70	31.85	KC660273
10/369	2010/08/11	*Canis lupus familiaris*	-28.60	31.33	KC660217
10/370	2010/08/11	*Canis lupus familiaris*	-28.52	30.08	KC660165
10/376	2010/08/16	*Canis lupus familiaris*	-28.78	31.85	KC660219
10/379[2]	2010/08/17	*Canis lupus familiaris*	N.R.	N.R.	KC660269
10/385	2010/08/23	*Canis lupus familiaris*	-28.32	30.10	KC660262
10/387	2010/08/24	*Canis lupus familiaris*	-30.45	30.62	KC660287
10/392	2010/08/25	*Canis lupus familiaris*	-29.77	30.93	KC660350
10/393	2010/08/25	*Canis lupus familiaris*	-30.47	30.65	KC660291
10/395	2010/08/27	*Canis lupus familiaris*	-27.39	32.08	KC660340

N.R: not recorded; N.S: not sequenced (RT-PCR unsuccessful)

[1] Herpestid-associated variant of RABV (excluded from analysis)

[2] Excluded from analysis (RT-PCR unsuccessful or coordinates not recorded)

Table A.1 – continued from previous page

Case number	Date	Species	Latitude	Longitude	Accession number
10/397	2010/08/30	*Canis lupus familiaris*	-30.98	30.20	KC660331
10/400	2010/08/31	*Canis lupus familiaris*	-29.85	30.80	KC660317
10/401	2010/09/01	*Canis lupus familiaris*	-28.60	32.07	KC660241
10/410	2010/09/06	*Canis lupus familiaris*	-29.97	30.51	KC660295
10/420	2010/09/13	*Canis lupus familiaris*	-28.00	30.00	KC660195
10/425	2010/09/14	*Canis lupus familiaris*	-27.85	30.26	KC660225
10/440	2010/09/22	*Canis lupus familiaris*	-28.17	31.18	KC660223
10/441	2010/09/23	*Canis lupus familiaris*	-28.72	29.97	KC660278
10/445	2010/09/27	*Canis lupus familiaris*	-28.77	30.23	KC660172
10/447	2010/09/27	*Bos taurus*	-28.60	29.94	KC660181
10/452	2010/09/29	*Canis lupus familiaris*	-28.25	30.33	KC660162
10/453[2]	2010/09/29	*Canis lupus familiaris*	N.R.	N.R.	KC660236
10/456	2010/09/29	*Canis lupus familiaris*	-30.23	30.40	KC660330
10/458	2010/09/30	Unspecified caprine species	-30.11	29.81	KC660334
10/459	2010/09/30	*Canis lupus familiaris*	-30.42	30.63	KC660319
10/460	2010/09/30	*Canis lupus familiaris*	-27.77	30.82	KC660212
10/461	2010/09/30	*Canis lupus familiaris*	-27.75	29.92	KC660164
10/462	2010/09/30	*Canis lupus familiaris*	-28.16	30.63	KC660176
10/463	2010/09/30	*Canis lupus familiaris*	-29.57	30.63	KC660245
10/472	2010/10/05	*Canis lupus familiaris*	-29.97	30.78	KC660231
10/479	2010/10/07	*Canis lupus familiaris*	-27.64	32.38	KC660348
10/481	2010/10/08	*Canis lupus familiaris*	-28.00	31.85	KC660247
10/485	2010/10/12	*Canis lupus familiaris*	-29.85	30.77	KC660250
10/488	2010/10/12	*Canis lupus familiaris*	-28.75	30.42	KC660215
10/492	2010/10/14	*Canis lupus familiaris*	-28.73	30.23	KC660252
10/495	2010/10/15	*Canis lupus familiaris*	-30.52	30.58	KC660313
10/496[2]	2010/10/18	*Canis lupus familiaris*	-28.73	30.23	N.S.
10/498	2010/10/19	*Canis lupus familiaris*	-30.00	30.62	KC660281
10/504	2010/10/21	*Canis lupus familiaris*	-29.98	30.76	KC660177
10/506	2010/10/21	*Canis lupus familiaris*	-29.98	30.92	KC660204
10/508	2010/10/22	*Canis lupus familiaris*	-27.90	31.63	KC660173
10/509	2010/10/25	*Canis lupus familiaris*	-30.01	30.53	KC660294
10/515	2010/10/26	*Canis lupus familiaris*	-28.75	31.87	KC660272
10/517	2010/10/27	*Canis lupus familiaris*	-28.60	29.92	KC660180
10/518	2010/10/27	*Canis lupus familiaris*	-28.98	31.78	KC660263
10/524	2010/10/28	*Canis lupus familiaris*	-30.50	30.57	KC660314
10/525	2010/10/28	*Bos taurus*	-30.23	30.40	KC660321
10/527[2]	2010/11/01	*Canis lupus familiaris*	-29.92	31.00	N.S.
10/528	2010/11/01	*Canis lupus familiaris*	-29.85	30.90	KC660214
10/530	2010/11/01	*Canis lupus familiaris*	-29.85	30.77	KC660178
10/531	2010/11/03	*Canis lupus familiaris*	-30.33	30.72	KC660304
10/532	2010/11/03	*Canis lupus familiaris*	-30.58	30.32	KC660300
10/533	2010/11/03	*Canis lupus familiaris*	-30.58	30.32	KC660283
10/536	2010/11/04	*Canis lupus familiaris*	-28.47	30.14	KC660169

N.R: not recorded; N.S: not sequenced (RT-PCR unsuccessful)

[1] Herpestid-associated variant of RABV (excluded from analysis)

[2] Excluded from analysis (RT-PCR unsuccessful or coordinates not recorded)

Table A.1 – continued from previous page

Case number	Date	Species	Latitude	Longitude	Accession number
10/541	2010/11/09	*Canis lupus familiaris*	-29.83	30.73	KC660168
10/552	2010/11/12	*Canis lupus familiaris*	-30.00	30.92	KC660199
10/557	2010/11/13	*Canis lupus familiaris*	-30.45	30.65	KC660296
10/560[2]	2010/11/16	*Canis lupus familiaris*	N.R.	N.R.	KC660242
10/562	2010/11/17	*Canis lupus familiaris*	-30.30	30.25	KC660308
10/564	2010/11/17	Unspecified jackal species	-28.50	31.92	KC660259
10/572	2010/11/17	*Canis lupus familiaris*	-28.60	29.92	KC660174
10/579[2]	2010/11/18	*Canis lupus familiaris*	N.R.	N.R.	KC660258
10/583	2010/11/19	*Canis lupus familiaris*	-29.99	30.82	KC660316
10/593	2010/11/24	*Bos taurus*	-30.11	29.81	KC660335
10/600	2010/11/26	*Canis lupus familiaris*	-28.60	29.42	KC660276
10/608	2010/11/30	*Canis lupus familiaris*	-28.73	31.80	KC660274
10/613	2010/12/02	*Canis lupus familiaris*	-28.77	31.95	KC660271
10/614	2010/12/02	*Canis lupus familiaris*	-28.77	31.95	KC660191
10/624	2010/12/06	Unspecified caprine species	-27.96	30.56	KC660253
10/633	2010/12/08	*Canis lupus familiaris*	-30.00	30.55	KC660280
10/635	2010/12/09	*Bos taurus*	-30.77	30.12	KC660332
10/641	2010/12/13	*Equus ferus caballus*	-29.78	30.58	KC660251
10/644	2010/12/14	*Canis lupus familiaris*	-28.35	31.40	KC660190
10/655	2010/12/20	*Canis lupus familiaris*	-30.30	30.25	KC660292
10/656	2010/12/21	*Canis lupus familiaris*	-30.05	30.87	KC660267
10/659	2010/12/21	*Canis lupus familiaris*	-27.72	30.05	KC660171
11/02	2011/01/04	*Canis lupus familiaris*	-30.03	30.87	KC660232
11/15[2]	2011/01/07	*Canis lupus familiaris*	N.R.	N.R.	KC660312
11/16[2]	2011/01/10	*Canis lupus familiaris*	N.R.	N.R.	KC660188
11/28	2011/01/14	*Canis lupus familiaris*	-29.90	30.36	KC660309
11/29	2011/01/14	*Canis lupus familiaris*	-27.41	32.65	KC660342
11/51	2011/01/24	*Canis lupus familiaris*	-28.62	31.73	KC660337
11/54	2011/01/24	*Canis lupus familiaris*	-30.50	30.62	KC660302
11/55	2011/01/24	*Canis lupus familiaris*	-30.65	30.53	KC660307
11/56	2011/01/24	*Canis lupus familiaris*	-30.55	30.53	KC660286
11/61	2011/01/26	*Canis lupus familiaris*	-29.90	31.00	KC660239
11/62	2011/01/26	*Canis lupus familiaris*	-27.52	32.58	KC660347
11/63	2011/01/27	*Canis lupus familiaris*	-28.78	31.88	KC660197
11/66	2011/01/28	*Canis lupus familiaris*	-30.50	30.45	KC660284
11/67[2]	2011/01/28	*Canis lupus familiaris*	N.R.	N.R.	KC660339
11/84	2011/02/02	*Canis lupus familiaris*	-29.73	30.80	KC660249
11/96	2011/02/08	*Canis lupus familiaris*	-27.42	32.69	KC660345
11/99	2011/02/09	*Canis lupus familiaris*	-30.87	30.37	KC660213
11/100[2]	2011/02/11	*Canis lupus familiaris*	N.R.	N.R.	KC660288
11/101	2011/02/14	*Canis lupus familiaris*	-28.70	30.23	KC660277
11/108	2011/02/15	*Canis lupus familiaris*	-28.85	31.82	KC660260
11/115	2011/02/16	*Canis lupus familiaris*	-29.32	31.27	KC660256

N.R: not recorded; N.S: not sequenced (RT-PCR unsuccessful)

[1] Herpestid-associated variant of RABV (excluded from analysis)

[2] Excluded from analysis (RT-PCR unsuccessful or coordinates not recorded)

Table A.1 – continued from previous page

Case number	Date	Species	Latitude	Longitude	Accession number
11/120	2011/02/17	*Bos taurus*	-27.40	32.67	KC660349
11/121	2011/02/17	*Canis lupus familiaris*	-28.80	30.03	KC660210
11/124[2]	2011/02/21	*Canis lupus familiaris*	N.R.	N.R.	N.S.
11/127	2011/02/23	*Canis lupus familiaris*	-29.08	31.57	KC660192
11/129[2]	2011/02/24	*Canis lupus familiaris*	-28.75	29.87	N.S.
11/144	2011/03/07	*Bos taurus*	-28.90	31.02	KC660270
11/178	2011/03/22	*Canis lupus familiaris*	-28.90	31.04	KC660338
11/181	2011/03/23	*Canis lupus familiaris*	-27.40	31.35	KC660279
11/185	2011/03/24	*Canis lupus familiaris*	-30.04	30.62	KC660187
11/186	2011/03/25	*Canis lupus familiaris*	-27.15	32.40	KC660343
11/188	2011/03/28	*Canis lupus familiaris*	-30.00	30.52	KC660264
11/191	2011/03/29	*Canis lupus familiaris*	-28.77	31.92	KC660261
11/195	2011/03/29	*Canis lupus familiaris*	-30.05	30.62	KC660310
11/240	2011/04/05	*Canis lupus familiaris*	-28.68	31.90	KC660198
11/241	2011/04/05	*Canis lupus familiaris*	-28.68	31.83	KC660189
11/203	2011/04/07	*Canis lupus familiaris*	-28.65	31.78	KC660336
11/208	2011/04/08	*Canis lupus familiaris*	-28.07	29.95	KC660170
11/209	2011/04/11	*Bos taurus*	-28.22	30.00	KC660206
11/212	2011/04/14	*Bos taurus*	-28.72	30.23	KC660175
11/217	2011/04/18	*Canis lupus familiaris*	-30.32	30.73	KC660301
11/221[2]	2011/04/19	*Canis lupus familiaris*	N.R.	N.R.	KC660205
11/224	2011/04/20	*Canis lupus familiaris*	-27.55	29.95	KC660160
11/232	2011/04/28	*Bos taurus*	-27.88	31.45	KC660266
11/251	2011/05/06	*Canis lupus familiaris*	-28.24	31.56	KC660182
11/252	2011/05/09	*Bos taurus*	-27.55	29.92	KC660220
11/257	2011/05/09	*Canis lupus familiaris*	-30.00	30.77	KC660333
11/262	2011/05/11	*Canis lupus familiaris*	-26.94	32.77	KC660344
11/268	2011/05/16	*Canis lupus familiaris*	-30.03	30.83	KC660186
11/270	2011/05/17	*Canis lupus familiaris*	-30.05	30.88	KC660202
11/272	2011/05/17	*Canis lupus familiaris*	-30.75	30.43	KC660290
11/273	2011/05/17	*Canis lupus familiaris*	-30.75	30.44	KC660289
11/284	2011/05/27	*Canis lupus familiaris*	-28.90	31.05	KC660193
11/296	2011/05/31	*Canis lupus familiaris*	-27.70	29.92	KC660268
11/300[2]	2011/06/01	*Canis lupus familiaris*	N.R.	N.R.	KC660306
11/305	2011/06/06	*Canis lupus familiaris*	-27.55	32.67	KC660346
11/306[2]	2011/06/07	*Canis lupus familiaris*	N.R.	N.R.	N.S.
11/308	2011/06/08	Unspecified caprine species	-28.92	31.22	KC660257

N.R: not recorded; N.S: not sequenced (RT-PCR unsuccessful)

[1] Herpestid-associated variant of RABV (excluded from analysis)

[2] Excluded from analysis (RT-PCR unsuccessful or coordinates not recorded)

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Table A.2: Published rabies G-L intergenic region sequences from eastern southern Africa

Location	Host	Count	Date	Accession numbers
Lesotho	Unknown	17	2000–2007	EU163385 – EU163401
Mozambique	*Canis lupus familiaris*	5	2006	EU123929 – EU123934
	Unknown	1	2006	EU123930
Swaziland	*Canis lupus familiaris*	1	2004	FJ842763
Eastern Cape	*Bos taurus*	1	2005	DQ431338
	Canis lupus familiaris	33	1988 – 2007	AF303072, DQ431253, DQ431254, DQ431257, DQ431262, DQ786033, DQ841404 – DQ841422, GQ918293, GQ918295, GQ918321, GQ983417, GQ983478 – GQ983490, GQ983525, GQ983539
	Otocyon megalotis	7	1998 – 2004	AF177097, DQ431276 – DQ431278, DQ431282, DQ431283, DQ431305
Free state	*Canis lupus familiaris*	72	1984 – 2007	EU163289 – EU163291, EU163294, EU163297, EU163298, EU163308, EU163310, EU163312 – EU163317, EU163319, EU163322 – EU163325, EU163327 – EU163333, EU163335 – EU163341, EU163343 – EU163350, EU163352 – EU163361, EU163363 – EU163368, EU163372 – EU163380, EU163382 – EU163384, GQ918282, GQ983491, GQ983538
	Caracal caracal	1	2005	GQ918329
	Otocyon megalotis	5	1991–2001	AF303059, DQ431289, DQ431291, DQ431294, DQ431295
	Unknown	2	1984	GQ983482, GQ983510
Mpumalanga	*Canis lupus familiaris*	67	2008 – 1986	AF177098, AF303069, EF686057, EF686058, EF686068, EF686071, EF686072, EF686077 – EF686080, EF686086, EF686096, EF686097, EF686101 – EF686104, EF686106, EF686113, EF686120, EF686125, EF686131, EF686140, EF686146, EF686151, FJ842721 – FJ842736, FJ842739, FJ842740, FJ842742 – FJ842762, GQ918318, GQ983495
	Canis mesomelas	1	1999	AF303063

Table A.2 – continued from previous page

Location	Host	Count	Date range	Accession numbers
KwaZulu Natal	*Bos taurus*	21	2003–2011	DQ841424 – DQ841426, DQ841545, KC660175, KC660181, KC660183, KC660206, KC660218, KC660220, KC660227, KC660254, KC660255, KC660266, KC660270, KC660320, KC660321, KC660325, KC660332, KC660335, KC660349
	Canis lupus familiaris	378	1980 – 2011	AF079904, AF177100 – AF177102, AF303081, AY605042, DQ841427 – DQ841541, GQ918283, GQ918284, GQ918286, GQ918287, GQ918289, GQ918290, GQ918296, GQ918298, GQ918301 – GQ918303, GQ918307, GQ918309 – GQ918311, GQ918316, GQ918317, GQ918320, GQ918322, GQ918323, GQ918325 – GQ918328, GQ918330, GQ983385, GQ983387, GQ983390, GQ983393, GQ983394, GQ983397, GQ983400 – GQ983403, GQ983410, GQ983413, GQ983415, GQ983420, GQ983422, GQ983425, GQ983427 – GQ983430, GQ983434, GQ983436 – GQ983438, GQ983440 – GQ983442, GQ983444, GQ983447, GQ983450, GQ983452, GQ983454, GQ983457, GQ983458, GQ983461, GQ983465, GQ983468, GQ983469, GQ983472 – GQ983474, GQ983477, GQ983480, GQ983486, GQ983489, GQ983492, GQ983496, GQ983501 – GQ983503, GQ983505, GQ983508, GQ983512, GQ983519, GQ983520, GQ983522, GQ983526, GQ983531 – GQ983533, GQ983540 – GQ983542, KC660160 – KC660174, KC660176 – KC660180, KC660182, KC660184 – KC660205, KC660207 – KC660217, KC660219, KC660221 – KC660226, KC660228 – KC660250, KC660252, KC660256, KC660258, KC660260 – KC660265, KC660267 – KC660269, KC660271 – KC660319, KC660322 – KC660324, KC660326 – KC660331, KC660333, KC660336 – KC660348, KC660350, KC660351
	Felis catus	1	1999	GQ983527

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Table A.2 – continued from previous page

Location	Host	Count	Date range	Accession numbers
KwaZulu Natal	*Equus ferus caballus*	2	2003–2010	AY605005, KC660251
	Ovis aries	1	2003	DQ841403
	Unspecified caprine species	7	2003–2011	DQ841542 – DQ841544, KC660253, KC660257, KC660293, KC660334
	Unspecified jackal species	1	2010	KC660259
	Unknown	11	1984–2005	GQ918281, GQ918299, GQ918306, GQ918331, GQ983414, GQ983464, GQ983470, GQ983493, GQ983518, GQ983534, GQ983535

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

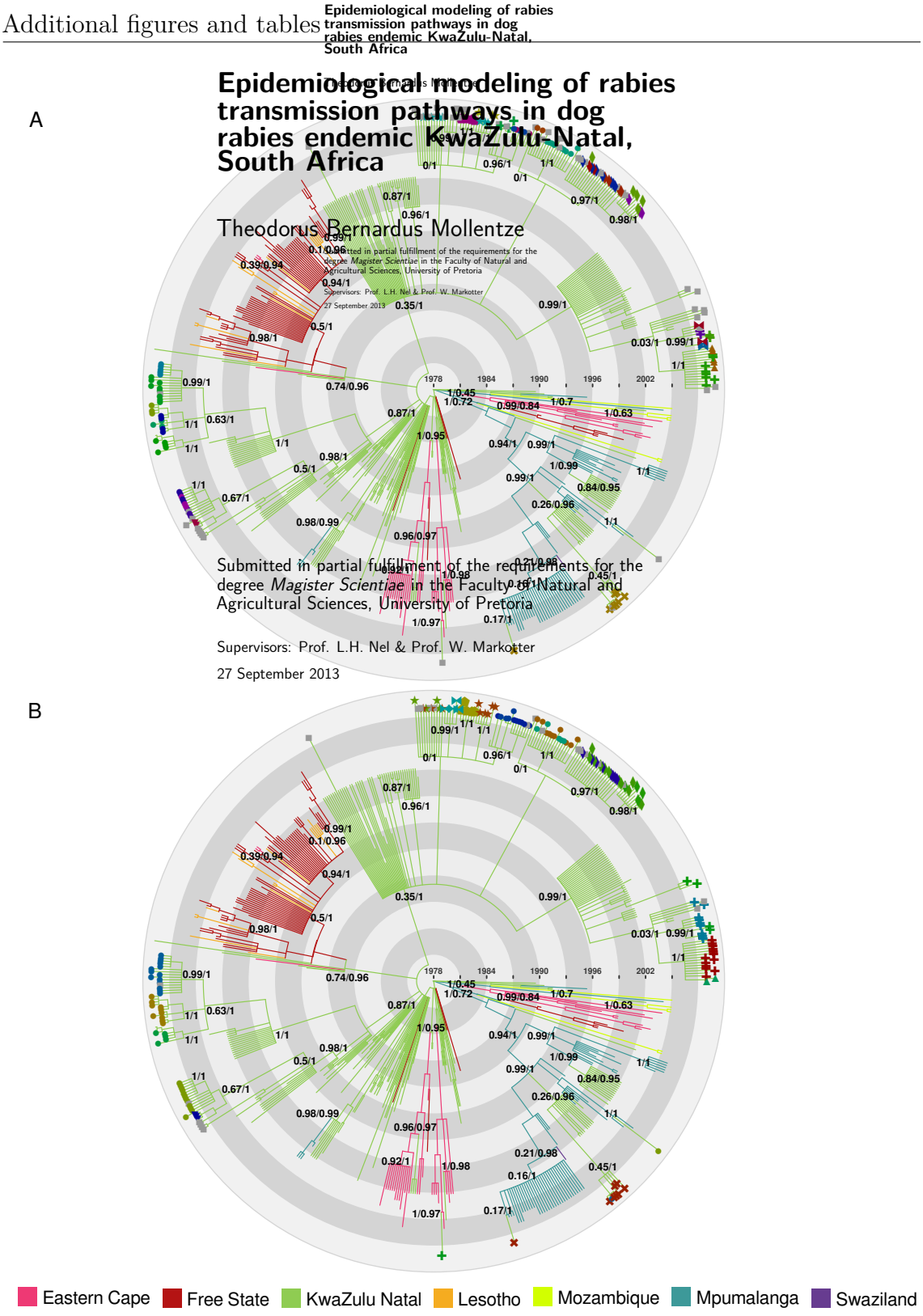

Figure A.1: Cases belonging to independent transmission trees (indicated by colour) and spatial clusters (indicated by symbols) and completely unconnected cases (indicated by grey squares) mapped onto a phylogeny of all G-L intergenic region sequences available for eastern southern Africa. The colours of branches represent their most likely location of origin reconstructed using a discrete diffusion model, as described in Chapter 2. Symbols and colours at the tips of the phylogeny match those in Figure 3.8, with figure **A** representing results obtained under the conservative specification of the transmission tree reconstruction algorithm and figure **B** representing results obtained under the liberal specification.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Bibliography

Akaike H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6): 716–723.

Altschul S., Gish W., Miller W., Myers E., & Lipman D. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215**(3): 403–410.

Ashkenazy H., Penn O., Doron-Faigenboim A., Cohen O., Cannarozzi G., Zomer O., & Pupko T. (2012). FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Research*, **40**: W580–W584.

Austerlitz F., Dick C.W., Dutech C., Klein E.K., Oddou-Muratorio S., Smouse P.E., & Sork V.L. (2004). Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**(4): 937–954.

Badrane H. & Tordo N. (2001). Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders. *Journal of Virology*, **75**(17): 8096–8104.

Baele G., Lemey P., Bedford T., Rambaut A., Suchard M.A., & Alekseyenko A.V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, **29**(9): 2157–2167.

Baer G. (2007). The history of rabies. In: *Rabies: Scientific basis of the disease and its management* (A. Jackson & W. Wunner, eds.), 2nd edition, chapter 1 (pp. 1–22). London: Academic Press.

Barnard B., Hassel R., Geyer H., & De Koker W. (1982). Non-bite transmission of rabies in kudu (Tragelaphus strepsiceros). *The Onderstepoort Journal of Veterinary Research*, **49**(4): 191.

Beran G.W. & Frith M. (1988). Domestic Animal Rabies Control : An Overview. *Reviews of Infectious Diseases*, **10**(S4): S672–S677.

Beyer H.L., Hampson K., Lembo T., Cleaveland S., Kaare M., & Haydon D.T. (2011). Metapopulation dynamics of rabies and the efficacy of vaccination. *Proceedings of the Royal Society B: Biological Sciences*, **278**(1715): 2182–2190.

Biek R., Henderson J.C., Waller L.A., Rupprecht C.E., & Real L.A. (2007). A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(19): 7993–7998.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Bielejec F., Rambaut A., Suchard M.A., & Lemey P. (2011). SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics*, **27**(20): 2910–2912.

Bingham J. (2005). Canine rabies ecology in southern Africa. *Emerging Infectious Diseases*, **11**(9): 1337–1342.

Bingham J., Foggin C.M., Wandeler A.I., & Hill F.W.G. (1999). The epidemiology of rabies in Zimbabwe. 2. Rabies in jackals (Canis adustus and Canis mesomelas). *Onderstepoort Journal of Veterinary Research*, **66**: 11–23.

Bishop G.C., Durrheim D.N., Kloeck P.E., Godlonton J.D., Bingham J., Speare R., & 2010 Rabies Advisory Group (2010). *Rabies guide for the medical, veterinary and allied professions*. Second edition. Pretoria: Government Printer.

Bourhy H., Reynes J., Dunham E.J., Dacheux L., Larrous F., Thi V., Huong Q., Xu G., Yan J., Miranda M.E.G., & Holmes E.C. (2008). The origin and phylogeography of dog rabies virus. *Journal of General Virology*, **89**: 2673–2681.

Brown K. (2010). Rabid epidemiologies: The emergence and resurgence of rabies in twentieth century South Africa. *Journal of the history of biology*, **44**(1): 81–101.

Brunker K., Hampson K., Horton D., & Biek R. (2012). Integrating the landscape epidemiology and genetics of RNA viruses: rabies in domestic dogs as a model. *Parasitology*, **139**(S14): 1899–1913.

Carroll M., Singer A., Smith G., Cowan D., & Massei G. (2010). The use of immunocontraception to improve rabies eradication in urban dog populations. *Wildlife Research*, **37**(8): 676–687.

Cauchemez S., Bhattarai A., Marchbanks T.L., Fagan R.P., Ostroff S., Ferguson N.M., & Swerdlow D. (2011). Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(7): 2825–2830.

Ceballos N., Morón S., Berciano J., Nicolás N., López C., Juste J., Nevado C., Setién A., & Echevarría J. (2013). Novel Lyssavirus in bat, Spain. *Emerging Infectious Diseases*, **19**(5): 793–795.

Charlton K.M., Nadin-Davis S., Casey G.A., & Wandeler A.I. (1997). The long incubation period in rabies: delayed progression of infection in muscle at the site of exposure. *Acta Neuropathologica*, **94**(1): 73–77.

Chen S.H., Su S.Y., Lo C.Z., Chen K.H., Huang T.J., Kuo B.H., & Lin C.Y. (2009). PALM: A paralleled and integrated framework for phylogenetic inference with automatic likelihood model selectors. *PloS One*, **4**(12): e8116.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfilment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Childs J.E. (2002). Epidemiology. In: *Rabies* (A.C. Jackson & W.H. Wunner, eds.), (p. 113). San Diego: Academic Press.

Childs J.E., Curns A.T., Dey M.E., Real L.A., Feinstein L., Bjø rnstad O.N., & Krebs J.W. (2000). Predicting the local dynamics of epizootic rabies among raccoons in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, **97**(25): 13 666–13 671.

Cleaveland S. (1998). Epidemiology and control of rabies: the growing problem of rabies in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **92**: 131–134.

Cleaveland S. & Dye C. (1995). Maintenance of a microparasite infecting several host species: rabies in the Serengeti. *Parasitology*, **111**(S1): S33–S47.

Cleaveland S., Kaare M., Knobel D., & Laurenson M.K. (2006). Canine vaccination – providing broader benefits for disease control. *Veterinary Microbiology*, **117**(1): 43–50.

Coetzee P. & Nel L.H. (2007). Emerging epidemic dog rabies in coastal South Africa: a molecular epidemiological analysis. *Virus Research*, **126**(1-2): 186–195.

Cohen C., Sartorius B., Sabeta C.T., Zulu G., Paweska J., Mogoswane M., Sutton C., Nel L.H., Swanepoel R., Leman P.A., Grobbelaar A.A., Dyason E., & Blumberg L. (2007). Epidemiology and molecular virus characterization of reemerging rabies, South Africa. *Emerging Infectious Diseases*, **13**(12): 1879–1886.

Cottam E.M., Thébaud G., Wadsworth J., Gloster J., Mansley L., Paton D.J., King D.P., & Haydon D.T. (2008). Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proceedings of the Royal Society B: Biological Sciences*, **275**(1637): 887–895.

Darriba D., Taboada G.L., Doallo R., & Posada D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**(8): 772.

David D., Hughes G.J., Yakobson B.A., Davidson I., Un H., Aylan O., Kuzmin I.V., & Rupprecht C.E. (2007). Identification of novel canine rabies virus clades in the Middle East and North Africa. *The Journal of General Virology*, **88**(3): 967–980.

Davis P.L., Rambaut A., Bourhy H., & Holmes E.C. (2007). The evolutionary dynamics of canid and mongoose rabies virus in Southern Africa. *Archives of Virology*, **152**(7): 1251–1258.

Dean D.J., Abelseth M.K., & Atanasiu P. (1996). The fluorescent antibody test. In: *Laboratory techniques in rabies* (F.X. Meslin, M.M. Kaplan, & H. Koprowski, eds.), 4 edition, chapter 7 (pp. 88–95). Geneva: World Health Organization.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfilment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Marwarth

27 September 2013

Denduangboripant J., Wacharapluesadee S., Lumlertdacha B., Ruankaew N., Hoonsuwan W., Hemachudha T., Kuntawunginn W. & Tepsumethanon V. (2005). Transmission dynamics of rabies virus in Thailand: implications for disease control. *BMC Infectious Diseases*, **5**(52).

Dray S. & Dufour A. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**(4).

Drummond A.J., Ho S.Y.W., Phillips M.J., & Rambaut A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**(5): e88.

Drummond A.J., Nicholls G.K., Rodrigo A.G., & Solomon W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, **161**: 1307–1320.

Drummond A.J. & Rambaut A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**(214).

Drummond A.J., Rambaut A., Shapiro B., & Pybus O.G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**(5): 1185–1192.

Drummond A.J., Suchard M.A., Xie D., & Rambaut A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**(8): 1969–1973.

Edmonds C.A., Lillie A.S., & Cavalli-Sforza L.L. (2004). Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(4): 975–979.

Farr T.G., Rosen P.A., Caro E., Crippen R., Duren R., Hensley S., Kobrick M., Paller M., Rodriguez E., Roth L., Seal D., Shaffer S., Shimada J., Umland J., Werner M., Oskin M., Burbank D., & Alsdorf D. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, **45**(RG2004): 2005RG000 183.

Fekadu M. (1993). Canine rabies. *Onderstepoort Journal of Veterinary Research*, **60**: 421–427.

Ferguson N.M., Keeling M.J., Edmunds W.J., Gani R., Grenfell B.T., Anderson R.M., & Leach S. (2003). Planning for smallpox outbreaks. *Nature*, **425**(6959): 681–685.

Fishbein D.B. & Robinson L.E. (1993). Rabies. *The New England Journal of Medicine*, **329**(22): 1632–1638.

Freuling C., Beer M., Conraths F., Finke S., Hoffman B., Keller B., Kliemt J., Mettenleiter T., Mühlbach E., Teifke J., Wohlsein P., & Müller T. (2011). Novel lyssavirus in Natterer's bat, Germany. *Emerging Infectious Diseases*, **17**(8): 10–13.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Freuling C.M., Hampson K., Selhorst T., Schröder R., Meslin F.X., Mettenleiter T.C., & Müller T. (2013). The elimination of canine rabies from Europe: determinants of success and lessons for the future. *Philosophical Transactions of the Royal Society B*, **368**(1623): 20120 142.

Frost S.D.W. & Volz E.M. (2010). Viral phylodynamics and the search for an 'effective number of infections'. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**(1548): 1879–1890.

Grenfell B.T., Pybus O.G., Gog J.R., Wood J.L.N., Daly J.M., Mumford J.A., & Holmes E.C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, **303**(5656): 327–332.

Hampson K., Dushoff J., Bingham J., Brückner G., Ali Y.H., & Dobson A. (2007). Synchronous cycles of domestic dog rabies in sub-Saharan Africa and the impact of control efforts. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(18): 7717–7722.

Hampson K., Dushoff J., Cleaveland S., Haydon D.T., Kaare M., Packer C., & Dobson A. (2009). Transmission dynamics and prospects for the elimination of canine rabies. *PLoS Biology*, **7**(3): e1000 053.

Haydon D.T., Chase-Topping M., Shaw D.J., Matthews L., Friar J.K., Wilesmith J., & Woolhouse M.E.J. (2003). The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proceedings of the Royal Society B*, **270**(1511): 121–127.

Haydon D.T., Cleaveland S., Taylor L.H., & Laurenson M.K. (2002). Identifying reservoirs of infection: a conceptual and practical challenge. *Emerging Infectious Diseases*, **8**(12): 1468–1473.

Heijne J.C.M., Teunis P., Morroy G., Wijkmans C., Oostveen S., Duizer E., Kretzschmar M., & Wallinga J. (2009). Enhanced hygiene measures and norovirus transmission during an outbreak. *Emerging Infectious Diseases*, **15**(1): 24–30.

Holmes E.C., Woelk C.H., Kassis R., & Bourhy H. (2002). Genetic constraints and the adaptive evolution of rabies virus in nature. *Virology*, **292**(2): 247–257.

Huson D.H. & Scornavacca C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology*, **61**(6): 1061–1067.

International Committee on Taxonomy of Viruses (2013). *Virus Taxonomy: 2012 Release (EC 44, Leuven, July 2012; Email ratification 2013 [MSL #27])*. Technical report, International Union of Microbiological Societies, Leuven.

Jenkins G.M., Rambaut A., Pybus O.G., & Holmes E.C. (2002). Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *Journal of Molecular Evolution*, **54**(2): 156–165.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfilment of the requirements for the degree Magister Scientiae in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Johnson N., McElhinney L.M., Smith J., Lewis P., & Fooks A.R. (2002). Phylogenetic comparison of rabies virus isolates using distal coding sequences of the glycoprotein and nucleoprotein genes. *Archives of Virology*, **147**(11): 2111–2123.

Jombart T. (2008). Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**(11): 1403–1405.

Jombart T. (2012). *A tutorial for the spatial Analysis of Principal Components (sPCA) using adegenet 1.3-4*. Vignette for the adegenet R package. Available from: http://adegenet.r-forge.r-project.org/files/adegenet_1.3-4.tar.gz.

Jombart T., Devillard S., Dufour A.B., & Pontier D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**: 92–103.

Jombart T., Eggo R.M., Dodd P.J., & Balloux F. (2011). Reconstructing disease outbreaks from genetic data: a graph approach. *Heredity*, **106**(2): 383–390.

Jukes T. & Cantor C. (1969). Evolution of protein molecules. In: *Mammalian protein metabolism* (H.N. Munro, ed.), volume III, (pp. 21 – 132). New York: Academic Press.

Katoh K. & Toh H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**(4): 286–298.

Keeling M.J., Woolhouse M.E.J., May R.M., Davies G., & Grenfell B.T. (2003). Modelling vaccination strategies against foot-and-mouth disease. *Nature*, **421**(6919): 136–142.

Kimura M. (1980). A simple method for estimating evolutionary rates op base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**: 111–120.

Kimura M. (1981). Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences of the United States of America*, **78**(1): 454–458.

King A.A., Fooks A.R., Aubert M., & Wandeler A.I. (Eds.) (2004). *Historical Perspective of Rabies in Europe and the Mediterranean Basin: A testament to rabies by Dr Arthur A King*. Paris: World Organisation for Animal Health (OIE).

Kingman J. (1982). On the genealogy of large populations. *Journal of Applied Probability*, **19**(1982): 27–43.

Kingman J. (2000). Origins of the coalescent: 1974-1982. *Genetics*, **168**(3): 1097–1104.

Kiss I.Z., Green D.M., & Kao R.R. (2006). The network of sheep movements within Great Britain: Network properties and their implications for infectious disease spread. *Journal of the Royal Society Interface*, **3**(10): 669–677.

Kissi B., Tordo N., & Bourhy H. (1995). Genetic polymorphism in the rabies virus nucleoprotein gene. *Virology*, **209**(2): 526–537.

Kitala P., McDermott J., Kyule M., Gathuma J., Perry B., & Wandeler A. (2001). Dog ecology and demography information to support the planning of rabies control in Machakos District, Kenya. *Acta Tropica*, **78**(3): 217–230.

Klopfstein S., Currat M., & Excoffier L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, **23**(3): 482–490.

Kuzmin I.V., Mayer A.E., Niezgoda M., Markotter W., Agwanda B., Breiman R.F., & Rupprecht C.E. (2010). Shimoni bat virus, a new representative of the Lyssavirus genus. *Virus Research*, **149**(2): 197–210.

Kuzmin I.V., Shi M., Orciari L.A., Yager P.A., Velasco-Villa A., Kuzmina N.A., Streicker D.G., Bergman D.L., & Rupprecht C.E. (2012). Molecular inferences suggest multiple host shifts of rabies viruses from bats to mesocarnivores in Arizona during 2001–2009. *PLoS Pathogens*, **8**(6): e1002 786.

Lardon Z., Watier L., Brunet A., Bernède C., Goudal M., Dacheux L., Rotivel Y., Guillemot D., & Bourhy H. (2010). Imported episodic rabies increases patient demand for and physician delivery of antirabies prophylaxis. *PLoS Neglected Tropical Diseases*, **4**(6): e723.

Lembo T., Attlan M., Bourhy H., Cleaveland S., Costa P., de Balogh K., Dodet B., Fooks A.R., Hiby E., Leanes F., Meslin F.X., Miranda M.E., Müller T., Nel L.H., Rupprecht C.E., Tordo N., Tumpey A., Wandeler A., & Briggs D.J. (2011). Renewed global partnerships and redesigned roadmaps for rabies prevention and control. *Veterinary Medicine International*, **2011**: 923 149.

Lembo T., Hampson K., Haydon D.T., Craft M., Dobson A., Dushoff J., Ernest E., Hoare R., Kaare M., Mlengeya T., Mentzel C., & Cleaveland S. (2008). Exploring reservoir dynamics: a case study of rabies in the Serengeti ecosystem. *Journal of Applied Ecology*, **45**(4): 1246–1257.

Lembo T., Hampson K., Kaare M.T., Ernest E., Knobel D., Kazwala R.R., Haydon D.T., & Cleaveland S. (2010). The feasibility of canine rabies elimination in Africa: dispelling doubts with data. *PLoS Neglected Tropical Diseases*, **4**(2): e626.

Lembo T., Haydon D.T., Velasco-Villa A., Rupprecht C.E., Packer C., Brandão P.E., Kuzmin I.V., Fooks A.R., Barrat J., & Cleaveland S. (2007). Molecular epidemiology identifies only a single rabies virus variant circulating in complex carnivore communities of the Serengeti. *Proceedings of the Royal Society B*, **274**(1622): 2123–2130.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa
Theodorus Bernardus Mollentze
Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria
Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter
November 2013

Lemey P., Rambaut A., Drummond A.J., & Suchard M.A. (2009). Bayesian phylogeography finds its roots. *PLoS Computational Biology*, **5**(9): e1000 520.

Lemey P., Rambaut A., Welch J.J., & Suchard M.A. (2010). Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, **27**(8): 1877–1885.

Leslie M.J., Messenger S., Rohde R.E., Smith J., Cheshier R., Hanlon C., & Rupprecht C.E. (2006). Bat-associated rabies virus in skunks. *Emerging Infectious Diseases*, **12**(8): 1274–1277.

Malerczyk C., Detora L., & Gniel D. (2011). Imported human rabies cases in Europe, the United States, and Japan, 1990 to 2010. *Journal of Travel Medicine*, **18**(6): 402–407.

Mansfield K., McElhinney L., Hübschle O., Mettler F., Sabeta C., Nel L.H., & Fooks A.R. (2006). A molecular epidemiological study of rabies epizootics in kudu (Tragelaphus strepsiceros) in Namibia. *BMC Veterinary Research*, **2**(2).

Mantel N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**(2): 209–220.

Markotter W., Kuzmin I.V., Rupprecht C.E., & Nel L.H. (2008a). Phylogeny of Lagos bat virus: challenges for lyssavirus taxonomy. *Virus Research*, **135**(1): 10–21.

Markotter W., Van Eeden C., Kuzmin I.V., Rupprecht C.E., Paweska J.T., Swanepoel R., Fooks A.R., Sabeta C.T., Cliquet F., & Nel L.H. (2008b). Epidemiology and pathogenicity of African bat lyssaviruses. *Developments in Biologicals*, **131**: 317–325.

Marston D.A., Horton D.L., Ngeleja C., Hampson K., McElhinney L.M., Banyard A.C., Haydon D., Cleaveland S., Rupprecht C.E., Bigambo M., Fooks A.R., & Lembo T. (2012). Ikoma lyssavirus, highly divergent novel lyssavirus in an African civet. *Emerging Infectious Diseases*, **18**(4): 664–667.

Minin V.N., Bloomquist E.W., & Suchard M.A. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, **25**(7): 1459–1471.

Mitrabhakdi E., Shuangshoti S., Wannakrairot P., Lewis R.A., Susuki K., Laothamatas J., & Hemachudha T. (2005). Difference in neuropathogenetic mechanisms in human furious and paralytic rabies. *Journal of the Neurological Sciences*, **238**(1-2): 3–10.

Mkhize G.C., Ngoepe E.C., Du Plessis B.J.A., Reininghaus B., & Sabeta C.T. (2010). Re-emergence of dog rabies in Mpumalanga province, South Africa. *Vector Borne and Zoonotic Diseases*, **10**(9): 921–926.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria.
Supervisors: Prof. L.H. Nel & Prof. W. Markotter
September 2013

Moran P.A.P. (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society*, **10**(2): 243–251.

Moran P.A.P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, **37**(1/2): 17–23.

Morelli M.J., Thébaud G., Chadœuf J., King D.P., Haydon D.T., & Soubeyrand S. (2012). A Bayesian Inference Framework to Reconstruct Transmission Trees Using Epidemiological and Genetic Data. *PLoS Computational Biology*, **8**(11): e1002 768.

Moritz H. (1980). Geodetic reference system 1980. *Journal of Geodesy*, **54**(3): 395–405.

Murcia P.R., Baillie G.J., Daly J., Elton D., Jervis C., Mumford J.A., Newton R., Parrish C.R., Hoelzer K., Dougan G., Parkhill J., Lennard N., Ormond D., Moule S., Whitwham A., McCarthy J.W., McKinley T.J., Holmes E.C., Grenfell B.T., & Wood J.L.N. (2010). Intra- and interhost evolutionary dynamics of equine influenza virus. *Journal of Virology*, **84**(14): 6943–6954.

Submitted in partial fulfilment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria
Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

Murphy F.A. (1977). Rabies Pathogenesis. *Archives of Virology*, **54**(4): 279–297.

Murphy F.A. & Bauer S.P. (1974). Early street rabies virus infection in striated muscle and later progression to the central nervous system. *Intervirology*, **3**(4): 256–268.

Murray C., Vos T., Lozano R., Naghavi M., Flaxman A., Michaud C., Ezatti M., Shibuya K., Salomon J., Abdalla S., Aboyans V., Abraham J., Ackerman I., Aggarwal R., Ahn S., Ali M., Alvarado M., Anderson H., Anderson L., Andrews K., *et al.* (2013). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, **380**(9859): 2197–2223.

Nel L.H. (2013). Discrepancies in data reporting for rabies, Africa. *Emerging Infectious Diseases*, **19**(4): 529–533.

Nel L.H., Jacobs J., Jaftha J.B., & Meredith C. (1997). Natural spillover of a distinctly Canidae-associated biotype of rabies virus into an expanded wildlife host range in southern Africa. *Virus Genes*, **15**(1): 79–82.

Nel L.H., Le Roux K., & Atlas R. (2009). Meeting the rabies control challenge in South Africa. *Microbe*, **4**(2): 61–65.

Nel L.H., Sabeta C.T., von Teichman B., Jaftha J.B., Rupprecht C.E., & Bingham J. (2005). Mongoose rabies in southern Africa: a re-evaluation based on molecular epidemiology. *Virus research*, **109**(2): 165–173.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Neville J. (2004). Rabies in the ancient world. In: *Historical Perspective of Rabies in Europe and the Mediterranean Basin* (M. Tignani, A.R. Fooks, M. Aubert, & A.I. Wandeler, eds.), chapter 1 (pp. 1–13). Paris: World Organisation for Animal Health (OIE).

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria
Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

Ngoepe C.E., Sabeta C.T., & Nel L.H. (2009). The spread of canine rabies into Free State province of South Africa: A molecular epidemiological characterization. *Virus research*, **142**(1-2): 175–180.

Ohta T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, **23**: 263–286.

Paradis E., Claude J., & Strimmer K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**(2): 289–290.

Posada D. & Crandall K. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**(9): 817–818.

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria
27 September 2013

Posel D. & Casale D. (2003). What has been happening to internal labour migration in South Africa, 1993–1999? *South African Journal of Economics*, **71**(3): 455–479.

Posel D. & Marx C. (2013). Circular Migration: A View from Destination Households in Two Urban Informal Settlements in South Africa. *Journal of Development Studies*, **In Press**.

Pybus O.G. & Rambaut A. (2002). GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics*, **18**(10): 1404–1405.

Pybus O.G., Rambaut A., Belshaw R., Freckleton R.P., Drummond A.J., & Holmes E.C. (2007). Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Molecular Biology and Evolution*, **24**(3): 845–852.

Pybus O.G., Suchard M.A., Lemey P., Bernardin F.J., Rambaut A., Crawford F.W., Gray R.R., Arinaminpathy N., Stramer S.L., Busch M.P., & Delwart E.L. (2012). Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(37): 15 066–15 071.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raghwani J., Rambaut A., Holmes E.C., Hang V.T., Hien T.T., Farrar J., Wills B., Lennon N.J., Birren B.W., Henn M.R., & Simmons C.P. (2011). Endemic dengue associated with the co-circulation of multiple viral lineages and localized density-dependent transmission. *PLoS Pathogens*, **7**(6): e1002 064.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

Rambaut A., Ho S.Y.W., Drummond A.J., & Shapiro B. (2009). Accommodating the effect of ancient DNA damage on inferences of demographic histories. *Molecular biology and evolution*, **26**(2): 245–8.

Ramstedt K., Hallhagen C., Lundin B., Häkansson C., Johannisson G., Löwhagen G., G N., & Giesecke J. (1990). Contact tracing for human immunodeficiency virus (HIV) infection. *Sexually Transmitted Diseases*, **17**(1): 37–41.

Reynolds A. & Rhodes C. (2009). The Lévy flight paradigm: random search patterns and mechanisms. *Ecology*, **90**(4): 877–887.

Rhodes C.J., Atkinson R.P., Anderson R.M., & Macdonald D.W. (1998). Rabies in Zimbabwe: reservoir dogs and the implications for disease control. *Philosophical Transactions of the Royal Society B*, **353**(1371): 999–1010.

Rodriguez F., Oliver J.L., Marin A., & Medina J.R. (1990). The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology*, **142**: 485–501.

Roy A. & Hooper D.C. (2008). Immune evasion by rabies viruses through the maintenance of blood brain barrier integrity. *Journal of Neurovirology*, **14**(5): 401–411.

Rupprecht C.E., Hanlon C.A., & Hemachudha T. (2002). Rabies re-examined. *The Lancet Infectious Diseases*, **2**(6): 327–343.

Russell C.A., Real L.A., & Smith D.L. (2006). Spatial control of rabies on heterogeneous landscapes. *PloS One*, **1**(1): e27.

Sabeta C.T., Bingham J., & Nel L.H. (2003). Molecular epidemiology of canid rabies in Zimbabwe and South Africa. *Virus Research*, **91**(2): 203–211.

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

Sabeta C.T., Mkhize G.C., & Ngoepe E.C. (2011). An evaluation of dog rabies control in Limpopo province (South Africa). *Epidemiology and Infection*, **139**(10): 1470–1475.

Schluter D., Price T., Mooers A.O., & Ludwig D. (1997). Likelihood of ancestor states in adaptive radiation. *Evolution*, **51**(6): 1699–1711.

Sinnott R. (1984). Virtues of the Haversine. *Sky and Telescope*, **68**(2): 159.

Smart N.L. & Charlton K.M. (1992). The distribution of challenge virus standard rabies virus versus skunk street rabies virus in the brains of experimentally infected rabid skunks. *Acta Neuropathologica*, **84**: 501–508.

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria
Supervisors: Prof. L.H. Nel & Prof. W. Markotter
27 September 2013

Smith J.S., Orciari L.A., Yager P.A., Seidel H.D., & Warner C.K. (1992). Epidemiologic and historical correlations among 87 rabies virus isolates as determined by limited sequence analysis. *The Journal of Infectious Diseases*, **166**(2): 296–307.

Stadler T., Kouyos R., von Wyl V., Yerly S., Böni J., Bürgisser P., Klimkait T., Joos B., Rieder P., Xie D., Günthard H.F., Drummond A.J., & Bonhoeffer S. (2012). Estimating the basic reproductive number from viral sequence data. *Molecular Biology and Evolution*, **29**(1): 347–357.

Statistics South Africa (2012). *Census 2011: Statistical release – P0301.4*. October. Pretoria: Statistics South Africa.

Swanepoel R., Barnard B.J., Meredith C.D., Bishop G.C., Brückner G.K., Foggin C.M., & Hübschle O.J. (1993). Rabies in southern Africa. *Onderstepoort Journal of Veterinary Research*, **60**(4): 325–346.

Talbi C., Holmes E.C., de Benedictis P., Faye O., Nakouné E., Gamatié D., Diarra A., Elmamy B.O., Sow A., Adjogoua E.V., Sangare O., Dundon W.G., Capua I., Sall A.A., & Bourhy H. (2009). Evolutionary history and dynamics of dog rabies virus in western and central Africa. *Journal of General Virology*, **90**(4): 783–791.

Talbi C., Lemey P., Suchard M.A., Abdelatif E., Elharrak M., Jalal N., Faouzi A., Echevarría J.E., Vazquez Morón S., Rambaut A., Campiz N., Tatem A.J., Holmes E.C., & Bourhy H. (2010). Phylodynamics and Human-Mediated Dispersal of a Zoonotic Virus. *PLoS Pathogens*, **6**(10): e1001 166.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., & Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**(10): 2731–2739.

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

Tordo N., Poch O., Ermine A., Keith G., & Rougeon F. (1988). Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative-strand RNA viruses. *Virology*, **165**(2): 565–576.

Townsend S.E., Lembo T., Cleaveland S., Meslin F.X., Miranda M.E., Putra A.A.G., Haydon D.T., & Hampson K. (2012). Surveillance guidelines for disease elimination: A case study of canine rabies. *Comparative immunology, microbiology and infectious diseases*, **36**(3): 249–261.

Velasco-Villa A., Reeder S.A., Orciari L.A., Yager P.A., Franka R., Blanton J.D., Zuckero L., Hunt P., Oertli E.H., Robinson L.E., & Rupprecht C.E. (2008). Enzootic

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Epidemiological modeling of rabies transmission pathways in dog rabies endemic KwaZulu-Natal, South Africa

Theodorus Bernardus Mollentze

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Supervisors: Prof. L.H. Nel & Prof. W. Markotter

27 September 2013

rabies elimination from dogs and reemergence in wild terrestrial carnivores, United States. *Emerging Infectious Diseases*, **14**(12): 1949–1954.

Volz E.M., Kosakovsky Pond S.L., Ward M.J., Leigh Brown A., & Frost S.D.W. (2009). Phylodynamics of infectious disease epidemics. *Genetics*, **183**: 1421–1430.

Von Teichman B.F., Thomson G.R., Meredith C.D., & Nel L.H. (1995). Molecular epidemiology of rabies virus in South Africa: evidence for two distinct virus groups. *Journal of General Virology*, **76**: 73–82.

Warrell M.J. & Warrell D.A. (2004). Rabies and other lyssavirus diseases. *The Lancet*, **363**(9413): 959–969.

Weyer J., Szmyd-Potapczuk A.V., Blumberg L.H., Leman P.A., Markotter W., Swanepoel R., Paweska J.T., & Nel L.H. (2011). Epidemiology of human rabies in South Africa, 1983-2007. *Virus Research*, **155**(1): 283–290.

World Health Organization (2002). *World Survey of Rabies number 35 for the Year 1999*. Geneva: World Health Organization.

World Health Organization (2004). *WHO Expert Consultation on Rabies: First Report*. Geneva: World Health Organization.

World Health Organization (2010). *Bill & Melinda Gates Foundation fund WHO-coordinated project to control and eventually eliminate rabies in low-income countries*. http://www.who.int/rabies/bmgf_who_project/en/index.html (accessed 2011/03/22).

World Health Organization (2013). *The 17 neglected tropical diseases*. http://www.who.int/neglected_diseases/diseases/en/ (accessed 2013/03/15).

World Organisation for Animal Health & Food and Agriculture Organization (2012). *The global foot and mouth disease control strategy*. Technical report.

Wunner W., Larson J., Dietzschold B., & Smith C. (1988). The molecular biology of rabies viruses. *Reviews of Infectious Diseases*, **10**: 771–784.

Yamamoto T., Otani S., & Shiraki H. (1965). A study of the evolution of viral infection in experimental herpes simplex encephalitis and rabies by means of fluorescent antibody. *Acta Neuropathologica*, **5**(2): 288–306.

Ypma R.J.F., Bataille A.M.A., Stegeman A., Koch G., Wallinga J., & Van Ballegooijen W.M. (2012). Unravelling transmission trees of infectious diseases by combining genetic and epidemiological data. *Proceedings of the Royal Society B*, **279**(1728): 444–450.

Zulu G.C., Sabeta C.T., & Nel L.H. (2009). Molecular epidemiology of rabies: focus on domestic dogs (Canis familiaris) and black-backed jackals (Canis mesomelas) from northern South Africa. *Virus Research*, **140**(1-2): 71–78.