

Molecular epidemiology of rabies in KwaZulu Natal, South Africa.

BY

PETER COETZEE

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
M.Sc. (MICROBIOLOGY) IN THE FACULTY OF AGRICULTURAL AND BIOLOGICAL
SCIENCES, DEPARTMENT OF MICROBIOLOGY AND PLANT PATHOLOGY,
UNIVERSITY OF PRETORIA

SUPERVISOR: PROF L.H. NEL

DECEMBER 2005

Declaration

I hereby declare that this thesis, except where indicated, is my own research, and has not been submitted in part, or as a whole, for a degree at any other university.

Peter Coetzee

Signature: _____

Date: _____

Dedications

***Dedicated to my loving parents and
brother***

Acknowledgements

I would like to thank the following people for their assistance, guidance and contributions towards the completion of this thesis.

Prof L.H. Nel for having faith in his students, and for his invaluable contributions towards the findings of this project. I really appreciate all the help that you have given me, and have learned a lot about my chosen field of study, as well as myself !

Dr. J. Randles and the staff from the Veterinary Institute at Allerton, for allowing us access their collection of rabies virus isolates from KwaZulu Natal, and for preparing these samples for transport.

Mrs. W. Markotter, for helping me with the initial viral isolations, and for providing technical assistance throughout the course of the project.

Dr. C. T. Sabeta, for allowing us the use the facilities at OVI, and for his advice and interest in this project.

Dr. F. Burt for performing the preparation of the human isolates at the NICD.

Prof G. Pietersen, for providing advice and assistance throughout the course of the project, and for allowing us to go on the field trip to the Cape, thank you !

Friends and colleagues from the department of Microbiology, especially Nobantu, Katherine, Orienka, Liz, Aletta and Jackie, thanks for providing an excellent and enjoyable lab environment.

To my parents for their love and support, and for providing me with the financial assistance which enabled me to continue my postgraduate studies.

To my Creator, who gives me my strength, and who has directed my path.

Summary

MOLECULAR EPIDEMIOLOGY OF RABIES IN KWAZULU NATAL, SOUTH AFRICA

BY

PETER COETZEE

Supervisor: Prof. L.H. Nel
Department of Microbiology and Plant Pathology
University of Pretoria

for the degree MS.c.

In South Africa, two biotypes of type species 1 of the *Lyssavirus* genus are maintained independently among the members of the *Herpestidae* and *Canidae* families, respectively. Canid rabies is a relatively new addition to the African subcontinent, having been introduced from infectious cycles, which had existed among dogs in Angola, in the early 1940s. Two epidemics, believed to have originated from dog endemic regions which had existed in the southern Maputo district of Mozambique since 1952, have occurred among domestic dogs in the KwaZulu Natal province in recent years. The first of these epidemics started in 1964, and ended by 1968, while the second epidemic which started in 1976, has proven to be intractable, despite the concerted efforts which have been implemented to bring it under control. In order to contribute to the understanding of the molecular epidemiology of rabies in the KwaZulu Natal province, and to thereby assist in future surveillance and control efforts, we conducted a molecular sequence analysis of representative panel of viral isolates which were obtained from the province during the year 2003. A 591 nt. sequence encompassing the G-L intergenic region and glycoprotein cytoplasmic domain was sequenced for 128 viral isolates, which were obtained from the different magisterial districts and affected host species of the province, and was subsequently used to characterize these viruses phylogenetically.

Summary

Characterization of the KwaZulu Natal variants, and comparison of the obtained sequence data, to sequences data which was obtained from rabies endemic regions from elsewhere in South Africa and Zimbabwe, in general supported the pattern of spread which led to the introduction of rabies into the province, as was previously suggested from the literature. The phylogeny which was established from the analyses, indicated that the viral isolates from the province were highly related to each other, and could be divided into two groups, which although belonging to the canid biotype, were in general distinguishable from canid rabies virus isolates which were obtained from elsewhere in South Africa and Zimbabwe. The observation that these subfamilies showed a low genetic divergence, as well as that they shared a unique recent common ancestor, suggested that they were introduced recently into the northern reaches of the province, probably from the same geographical region (i.e. southern Mozambique).

Phylogenetic characterization of the KwaZulu Natal isolates further suggested that at least three enzootic fronts are currently responsible for the introduction of rabies into the northern and southern regions of the province. The first of these fronts was hypothesized to have spread directly across the southern Mozambique border (possibly via southeastern Swaziland), into the northeastern coastal regions of KwaZulu Natal, while the second front represented the south-eastwards spread of synergistic dog-jackal cycles from southeastern Mpumalanga, into the northern and northeastern regions of the province. The third front on the other hand, represented the possible spread of a remnant infectious cycle, left over from the 1964-1968 epidemic, from the northern region of the Eastern Cape, into southern KwaZulu Natal.

Phylogenetic characterization further proved useful for identifying the distribution of viral variants, and allowed us to propose a pathway by which the disease might have spread throughout the province. The proposed pathway of spread suggested that viral variants may have been translocated over long distances, and highlighted the role that major routes of human transportation may play in the dissemination of the disease. The regional characterization of viral variants from KwaZulu Natal, further demarked the location where the identified viral variants circulated in individual host populations, allowing us to place

Summary

the current epidemic into an epidemiological framework which attempts to explain the long term persistence of the disease. This provided clues as to the intractability of the second epidemic, and allowed us to develop a proposal as to how current control strategies may be altered, in order to contain the current outbreak in the province.

The initial phylogeny which was established from the study provides an epidemiological framework, which will play an important role in determining the origin of future human spillover cases, and for tracking the spread of viral variants throughout the affected regions of the province. It is further envisaged that the data which was generated during the course of the project will be utilized in future surveillance efforts, targeted to the evaluation of the efficacy of potentially implemented control campaigns.

Contents

| | PAGE |
|--|--------------|
| Declaration | 2 |
| Dedication | 3 |
| Acknowledgements | 4 |
| Summary | 5-7 |
| List of abbreviations | 14-19 |
| List of figures | 20-21 |
| List of tables | 22-23 |
| List of appendixes | 24 |
| | |
| Chapter 1 | 25-99 |
| | |
| (A) - Literature survey | 25-84 |
| | |
| 1.1 General introduction | 26-27 |
| | |
| 1.2 History of rabies | 27-31 |
| | |
| 1.3 Classification of the <i>Lyssavirus</i> genus | 31-38 |
| | |
| 1.4 Molecular biology of the rabies virus | 39-50 |
| | |
| 1.4.1 Structure of the rabies virus | 39-40 |
| 1.4.2 Organization of the rabies virus genome | 40-41 |
| | |
| 1.4.2.1. Coding regions | 42-48 |
| (a) The nucleoprotein (N) | 42-43 |
| (b) The phosphoprotein (P) | 43-44 |
| (c) The Matrix protein (M) | 44-46 |

Contents

| | | PAGE |
|---------|-----|--|
| | (d) | The Glycoprotein (G) 46-47 |
| | (d) | The RNA polymerase (L) 47-48 |
| 1.4.2.2 | | Non-coding regions 48-50 |
| | (a) | The leader RNA (Le) 48 |
| | (b) | Intergenic sequences 48-50 |
| 1.5 | | Genetic variability of the Lyssavirus genome 50-52 |
| 1.6 | | Heterogeneous nature of RNA genomes: Quasispecies concept 52-53 |
| 1.7 | | Intracellular replication cycle of the rabies virus 54-57 |
| 1.8 | | Rabies in humans: Clinical presentation and pathogenesis 58-62 |
| 1.9 | | Epidemiology of rabies 62-77 |
| 1.9.1 | | Worldwide distribution of rabies 64-68 |
| 1.9.2 | | Molecular epidemiological studies conducted on a global scale on the <i>Lyssavirus</i> genus 68-72 |
| 1.9.3 | | Distribution of rabies in South Africa 72-76 |
| 1.9.4 | | Molecular epidemiology of rabies in South Africa 76-77 |
| 1.10 | | Prevention and control of rabies 78-81 |
| 1.11 | | Vaccines 81-84 |

Contents

| | | PAGE |
|----------|--|--------------|
| | (B) – Themes in molecular epidemiology | 84-97 |
| 1.12 | What is molecular epidemiology? | 84-85 |
| 1.13 | Methods to generate nucleotide sequence data | 85-87 |
| 1.13.1 | Polymerase chain reaction | 85-86 |
| 1.13.2 | Automated fluorescent nucleotide sequencing | 86-87 |
| 1.14 | Methods used to infer phylogenetic relationships | 87-95 |
| 1.14.1 | Multiple sequence alignment | 89-90 |
| 1.14.2 | Methods used in tree construction | 90-95 |
| 1.14.2.1 | Distance methods | 93-94 |
| 1.14.2.2 | Discrete character methods | 94-95 |
| 1.15 | Bootstrap analysis | 95-96 |
| 1.16 | Molecular clock analysis | 96 |
| 1.17 | Advantages and disadvantages of different tree construction methods. | 97 |
| | (C) - Aims of this study | 98-99 |

Contents

| | | PAGE |
|--|--|-----------------|
| Chapter 2 – Sequence analysis of rabies virus isolates from KwaZulu Natal (KZN). | | 100- 136 |
| 2.1 | Summary | 101 |
| 2.2. | Introduction | 101-108 |
| 2.2 | Materials and methods | 109-115 |
| 2.2.1 | Rabies virus isolates | 109 |
| 2.2.2 | RNA extraction | 109-110 |
| 2.2.3 | Primer selection | 110 |
| 2.2.4 | Reverse transcription | 110 |
| 2.2.5 | Polymerase chain reaction (PCR) | 111-112 |
| 2.2.6 | Purification of PCR products | 112 |
| 2.2.7 | Nucleotide sequencing | 112-113 |
| 2.2.8 | Phylogenetic analysis | 113-115 |
| 2.3 | Results | 115-135 |
| 2.3.1 | Rabies virus isolates | 115 |
| 2.3.2 | RNA extraction, cDNA synthesis, and PCR of rabies virus isolates | 115 |
| 2.3.3 | Purification and nucleotide sequence determination | 115-119 |
| 2.3.4 | Phylogenetic analysis | 119-136 |
| (A) Phylogeny of KwaZulu Natal rabies virus isolates, compared to canine rabies virus isolates obtained from elsewhere in South Africa and Zimbabwe. | | 120-126 |
| (B) Phylogeny of rabies virus isolates from KwaZulu Natal. | | 126-136 |

Contents

| | PAGE |
|--|-----------------|
| Chapter 3 - Discussion | 137- 159 |
| 3.1 Summary | 137 |
| 3.2. Phylogeny of KwaZulu Natal rabies virus isolates, compared to canine rabies virus isolates obtained from elsewhere in South Africa and Zimbabwe. | 138-143 |
| 3.2.1 Phylogenetic and geographical divisions present for viral isolates belonging to subcluster 1-1 (KZN family). | 139-140 |
| 3.2.2. Jackal rabies in the northern and north-eastern regions of KwaZulu Natal. | 140-141 |
| 3.2.3 Reconstruction of the events which led to the introduction of rabies into KwaZulu Natal. | 141-142 |
| 3.2.4 Introduction of KwaZulu Natal variants into other regions of South Africa. | 143 |
| 3.2.5 Future studies required to clarify the molecular epidemiology of rabies in Mozambique and the northern Limpopo province of South Africa | 143 |
| 3.3 Phylogeny of rabies virus isolates from KwaZulu Natal | 143-159 |
| 3.3.1 Pattern of spread and its associated factors, which led to the introduction of viral variants from subfamily B into the affected regions of KZN. | 145-146 |
| 3.3.2 Pattern of spread and its associated factors, which led to the introduction of viral variants from subfamily A into the affected regions of KZN. | 146-149 |
| 3.3.3. Correlation between the variants which were isolated from human infections, and the geographically discrete variants identified from KZN during the course of this study. | 150-152 |

Contents

| | PAGE |
|---|----------------|
| 3.3.4 Factors that influence the maintenance and persistence of the rabies virus in the KZN province. | 153 |
| 3.3.4.1 Maintenance and persistence of rabies in local dog populations in KwaZulu Natal. | 153-155 |
| 3.3.4.2 Maintenance and persistence of rabies among jackal in KwaZulu Natal. | 155-156 |
| 3.3.5 Suggestions of how current strategies for rabies control in the KZN province may be altered in order to bring the epidemic under control. | 156-158 |
| 3.3.6 Implications of this study for the future surveillance of the epidemic in KwaZulu Natal | 159 |
| Chapter 4 – Conclusion | 160-167 |
| 4.1 Introduction | 161-162 |
| 4.2 Principal findings of this study | 162-167 |
| 4.3 Recommendations for the future | 167 |
| References | 168-187 |
| Communications | 188-189 |
| Appendix | 190-233 |

List of abbreviations

| | |
|---------|--------------------------------------|
| µm | micrometer |
| A | Adenosine |
| Ab | antibody |
| aa | amino acid |
| A.D. | Anno Domino |
| ABI | Applied Biotechnologies Incorporated |
| ABLV | Australian Bat Lyssavirus |
| AChR | acetylcholine receptor |
| AIDS | Acquired Immune Deficiency Syndrome |
| approx. | approximately |
| Arg | arginine |
| ATP | adenosine triphosphate |
| B.C. | before Christ |
| BCG | bacille Calmette-Guerin |
| bp | base pair |
| C | cytosine |
| ca. | calculated |
| cDNA | complementary DNA |
| CNI | close neighbor interchange |
| CNS | central nervous system |
| CRU | cellular receptor unit |
| CVS | challenge virus standard |
| cw | cow |
| Da. | Dalton |
| ddNTP | dideoxyribonucleotide triphosphate |

List of abbreviations

| | |
|---------|---------------------------------------|
| DFA | direct fluorescent antibody test |
| dg | dog |
| DI | defective interfering particle |
| DNA | deoxyribonucleic acid |
| dNTP | deoxyribonucleotide triphosphate |
| DTT | dithiothreitol |
| DUVV | Duvenhage Virus |
| e.g. | for example |
| EBLV-1 | European Bat Lyssavirus 1 |
| EBLV-2 | European Bat Lyssavirus 2 |
| EC | Eastern Cape |
| Eds. | edition |
| EM | electron microscopy |
| ER | endoplasmic reticulum |
| ERA | Evelyn Rokitnicki Abelseth |
| et. al. | and others |
| g | acceleration of gravity |
| Gu | guanosine |
| G | glycoprotein |
| GDN | glycine, aspartic acid and asparagine |
| Gln | glutamine |
| Gly | glycine |
| GPS | Global positioning system |
| GT | genotype |
| HCL | hydrochloric acid |

List of abbreviations

| | |
|------|---|
| HDCV | human diploid cell vaccine |
| HEP | Flury high egg passage |
| hm | human |
| HRIG | human rabies immunoglobulin |
| i.e. | in other words |
| ICTV | International Committee for the Taxonomy of Viruses |
| IDT | Integrated DNA Technologies |
| IFA | indirect fluorescent antibody test |
| IL | interleukin |
| Ile | isoleucine |
| Inc. | incorporated |
| IHNV | Infectious Hematopoietic Necrosis Virus |
| j | <i>Canis mesomelas</i> |
| kb | kilobase pairs |
| KCl | potassium chloride |
| km | kilometer |
| KZN | KwaZulu Natal |
| L | polymerase (L) protein |
| LBV | Lagos Bat Virus |
| Le | leader RNA |
| Lys | lysine |
| m | <i>Cynictis penicillata</i> |
| M | matrix protein |
| MAb | monoclonal antibody |

List of abbreviations

| | |
|-------------------|--|
| Met | methionine |
| MgCl ₂ | magnesium chloride |
| ml | millilitre |
| ML | maximum likelihood |
| mM | millimolar |
| M-MLV | Moloney Murine Leukemia Virus |
| MOKV | Mokola Virus |
| MP | maximum parsimony |
| mRNA | messenger ribonucleic acid |
| N | nucleoprotein |
| NV | non-virion protein |
| NaOH | sodium hydroxide |
| ng | nanogram |
| NICD | National Institute for Communicable Diseases |
| NJ | neighbourhood joining |
| nm | nanometres |
| NT | neurotrophin |
| nt. | nucleotide |
| o | <i>Otocyon megalotis</i> |
| ORF | open reading frame |
| OVI | Onderstepoort Veterinary Institute |
| °C | degree Celsius |
| OTU | operational taxonomic unit |
| P | phosphoprotein |

List of abbreviations

| | |
|-----------|---|
| PAGE | polyacrylamide gel electrophoresis |
| PAHO/AMRO | Panamerican Health Organization/Americas |
| PBS | phosphate buffered saline |
| PCEC | purified chick embryo cell vaccine |
| PCR | polymerase chain reaction |
| PEP | post exposure prophylaxis |
| PM | Pitman Moore |
| pMol | pico molar |
| pp. | page |
| Pu | purine |
| PV | Pasteur virus |
| Py | pyrimidine |
| RABV | rabies virus |
| RFLP | restriction fragment length polymorphism |
| RI | replicative intermediate |
| RIG | rabies immunoglobulin |
| RNA | ribonucleic acid |
| RNP | ribonucleoprotein |
| RT | reverse transcriptase |
| RT-PCR | reverse transcriptase polymerase chain reaction |
| SAD | Street Alabama Dufferin |
| SEARG | South and East African Rabies Group |
| Ser | serine |
| SDS | sodium dodecyl sulphate |

List of abbreviations

| | |
|------------------|---|
| Shp | sheep |
| SP | signal peptide |
| SPU | Special Pathogens Unit |
| T | thiamine |
| <i>Taq</i> | <i>Thermus aquaticus</i> |
| TBSV | Tomato Bushy Stunt Virus |
| TMV | Tobacco Mosaic Virus |
| TRAIL | Tumor Necrosis Factor Related Apoptosis Inducing Ligand |
| TTP | Transcription termination and polyadenylation signal |
| U | unit |
| U | uracil |
| UPGMA | unweighted pair group method with arithmetic mean |
| USA | United States of America |
| UV | ultraviolet |
| V | volt |
| VNA _b | virus neutralizing antibody |
| VRG | vaccinia recombinant glycoprotein |
| VSV | Vesicular Stomatitis Virus |
| WHO | World Health Organization |
| x | times |
| mg | milligram |
| μl | microliter |
| Ψ | G-L intergenic region, or pseudogene |

List of figures

| Figure number | Page |
|---|-------------|
| 1.1. UPGMA phylogenetic tree illustrating the genotype and phylogroup divisions within the <i>Lyssavirus</i> genus (Smith <i>et al.</i> , 2002). | 37 |
| 1.2. Schematic presentation of the rabies virus particle (Tordo and Poch, 1988). | 40 |
| 1.3. Schematic presentation of the Lyssavirus genome (Levy <i>et al.</i> , 1994). | 42 |
| 1.4. Similarity profile generated by comparing the divergent rabies and Mokola virus genomes (Bourhy <i>et al.</i> , 1993; Bourhy <i>et al.</i> , 1995; Nadin-Davis, 2000). | 52 |
| 1.5. Infection and replication cycle of the rabies virus. | 57 |
| 1.6.. Route of spread of the rabies virus, from a peripheral inoculation site, to the central nervous system, and from there to other sites throughout the body (Internet reference). | 59 |
| 1.7. Global distribution of the primary host species which are responsible for the maintenance of the rabies and rabies related viruses (Rupprecht <i>et al.</i> , 2002). | 63 |
| 1.8. The distribution of the primary host species responsible for rabies endemicity within South Africa (Bishop <i>et al.</i> , 2003). | 72 |
| 2.1 Provinces of South Africa. | 102 |
| 2.2. Rabies virus trends for KwaZulu Natal between 1992-2002 (Randles, 2003). | 105 |
| 2.3. Total number of dogs vaccinated in KwaZulu Natal between 1992-2002 (Randles, 2003). | 105 |

List of figures

| Figure number | Page |
|---|-------------|
| 2.4. Schematic presentation of the annealing positions and orientations of primers, used to amplify the glycoprotein cytoplasmic domain and G-L intergenic regions (Tordo <i>et al.</i> , 1986; Sacramento <i>et al.</i> , 1991). | 111 |
| 2.5. Approximate geographic region of isolation for reference sequences which were obtained from South Africa (Sabeta <i>et al.</i> , 2003) | 125 |
| 2.6. Approximate geographic region of isolation for reference sequences which were obtained from Zimbabwe (Sabeta <i>et al.</i> , 2003). | 125 |
| 2.7. Neighbourhood joining tree constructed from selected isolates from KZN as well as isolates from other canine rabies endemic regions in South Africa and Zimbabwe. | 126 |
| 2.8. Neighbourhood-joining tree, constructed from an alignment of 128 nucleotide sequences, encompassing the cytoplasmic domain of the glycoprotein, and the G-L intergenic region of canine rabies viruses from KZN. | 135 |
| 2.9. The approximate geographic location of isolation for rabies virus isolates which were analyzed during the course of this study. | 136 |

List of tables

| Table number | Page |
|--|-------------|
| 1.1. Abridged classification of the Mononegavirales order (ICTV, 2005). | 38 |
| 1.2 . Some common terms used in the field of phylogenetics (Li and Graur, 2000). | 88 |
| 1.3. Some methods used in phylogenetic tree construction. and the strategies they are based on (Vandamme, 2003). | 91 |
| 1.4. Advantages and disadvantages of different tree construction methods (Bourhy <i>et al.</i> , 1995). | 97 |
| 2.1. Total number of infections in domestic dogs, livestock, humans and wildlife from KwaZulu Natal for the year 2003 (Randles, 2003). | 106 |
| 2.2. Summary of the number of cases per magisterial district for KwaZulu Natal from the calendar year 2003 (Randles, 2003). | 107 |
| 2.3. Annealing positions and nucleotide sequences of the primers used to amplify the glycoprotein cytoplasmic domain and G-L intergenic regions (Tordo <i>et al.</i> , 1986; Sacramento <i>et al.</i> , 1991). | 112 |
| 2.4. Rabies virus isolates analyzed during the course of this study. | 116-119 |
| 2.5 Nucleotide substitutions responsible for the clustering of the major groupings identified from the phylogeny in section 2.4.4 (A) | 122 |
| 2.6. Reference sequences used to conduct the analysis in 2.4.4 (A) (Tordo <i>et al.</i> , 1986, Von Teichman, 1005; Sabeta <i>et al.</i> , 2003). | 123-124 |

List of tables

| Table number | Page |
|---|-------------|
| 2.7. Nucleotide substitutions responsible for the clustering of the major groupings identified from the phylogeny in section 2.4.4 (B) | 128 |
| 2.8. Viral groups which were identified from KwaZulu Natal, their epidemiological information, and the symbols and colours used to denote them on the phylogenetic tree and map in figure. 2.8. and 2.9., respectively. | 132-133 |
| 3.1. Case histories from human infections, which had occurred in the KwaZulu Natal and Eastern Cape provinces from 2002-2003. | 152 |

Appendix

| | Page |
|--|-------------|
| Appendix A. Multiple sequence alignment of a 591 nt. sequence encompassing the cytoplasmic domain of the glycoprotein and the G-L intergenic regions for 58 isolates, obtained from KwaZulu Natal and the other rabies endemic regions of southern Africa and Zimbabwe. | 191-199 |
| Appendix B. Multiple sequence alignment of a 591 nt. sequence encompassing the cytoplasmic domain of the glycoprotein and the G-L intergenic region, for 128 isolates obtained from KwaZulu Natal. | 200-218 |
| Appendix C. Maximum parsimony tree constructed from a 591 nt. sequence encompassing the glycoprotein cytoplasmic domain and the G-L intergenic region for 58 isolates, obtained from KwaZulu Natal and the other rabies endemic regions of southern Africa and Zimbabwe. | 219-220 |
| Appendix D. Maximum parsimony tree constructed from a 591 nt. sequence encompassing the glycoprotein cytoplasmic domain and the G-L intergenic region, for 128 isolates obtained from KwaZulu Natal. | 221-222 |
| Appendix E. The proposed regional path of spread of the epidemic throughout the African subcontinent, which led to the introduction and the establishment of the different rabies virus variants, which were identified from neighboring regions and countries during the course of this study. | 223-224 |
| Appendix F. The proposed path of spread of the epidemic, which led to the introduction and establishment of the different rabies virus variants which were identified from KwaZulu Natal during the course of this study. | 225-229 |
| Appendix G. Magisterial districts of KZN. | 230-233 |