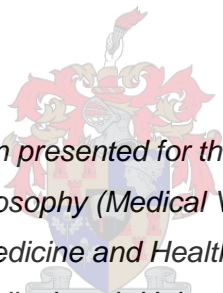# THE DIVERSITY OF CORONAVIRUSES
IN SOUTHERN AFRICAN
BAT POPULATIONS

by Nadine Cronjé

Dissertation presented for the degree of
Doctor of Philosophy (Medical Virology) in the
Faculty of Medicine and Health Sciences at
Stellenbosch University

Supervisor: Prof. Wolfgang Preiser

Co-supervisors: Prof. Corrie Schoeman and Dr Ndapewa Ithete

December 2017

i

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2017

Summary

Coronaviruses are RNA viruses encompassing four genera. The alpha- and betacoronaviruses have commonly been associated with mild disease in humans. However, outbreaks of severe respiratory disease in 2002 and 2012 led to the identification of novel highly pathogenic human coronaviruses, SARS- and MERS-CoV, respectively. Bats, order Chiroptera, are believed to be the reservoir host from which all mammalian coronaviruses have emerged.

To date, few studies have been published on coronaviruses in South African bats. With little known about the diversity and prevalence of bat coronaviruses in this region; this study aimed to describe the existing coronavirus diversity within South African bat populations as well as factors that might influence bat-coronavirus ecology. It detected nine different coronavirus species, eight alphacoronaviruses and one betacoronavirus, from ten different bat species. The study not only demonstrated that diverse coronaviruses can be found in different bat species of Southern Africa but lends additional support to an ongoing circulation of MERS-related betacoronaviruses in South African bats, with divergent variants detected in two different vespertilionid bat species.

A species-specific surveillance of *Neoromicia capensis (*Cape serotine) bats detected three different bat coronavirus species and revealed genetic diversity across different geographic regions. Several instances of coinfection with two different coronaviruses were detected, demonstrating the potential for recombination that could lead to the emergence of a new coronavirus that might have zoonotic potential. This study demonstrated that both host and environmental factors may influence CoV ecology. Female *Neoromicia capensis* bats trapped at low altitude sites within the Forest biome had the highest likelihood of being coronavirus positive. Discrepancies between detection rates obtained with different screening assays led to the adoption of an improved approach and recommendations for future bat coronavirus surveillance studies were made.

Opsomming

Koronavirusse is RNA virusse wat vier genera insluit. Die alfa- en betakoronavirusse word algemeen geassosieer met minder ernstige siektes by mense. Uitbrakings van ernstige respiratoriese siektes in 2002 en 2012 het egter gelei tot die identifisering van nuwe hoogs patogeniese menslike koronavirusse, SARS- en MERS-koronavirus, onderskeidelik. Vlêrmuise, orde Chiroptera, word beskou as die reservoir gasheer, wat tot die oorsprong van alle soogdierkoronavirusse gelei het.

Tot dusver is min studies oor koronavirusse in Suid-Afrikaanse vlêrmuise gepubliseer. Met min kennis van die diversiteit en voorkoms van vlêrmuiskoronavirusse in hierdie streek; het hierdie studie ten doel om die bestaande koronavirusdiversiteit binne Suid-Afrikaanse vlêrmuispopulasies asook faktore wat vlêrmuis-koronavirus-ekologie kan beïnvloed, te ondersoek en beskryf. Nege verskillende koronavirus spesies, agt alfakoronavirusse en een betacoronavirus, is in tien verskillende vlêrmuis spesies geïdentifiseer. Die studie het nie net gedemonstreer dat diverse koronavirusse in verskillende vlêrmuise van Suider-Afrika voorkom nie, maar ook addisionele ondersteuning aan 'n deurlopende verspreiding van MERS-verwante betakoronavirusse in Suid-Afrikaanse vlêrmuise verleen, met uiteenlopende variante wat in twee verskillende vespertilioniedvlêrmuise aangetref word.

'n Spesiespesifieke waarneming van *Neoromicia capensis* (Kaapse serotien) vlêrmuise het drie verskillende koronavirus spesies opgespoor en genetiese diversiteit in verskillende geografiese streke opgemerk.Verskeie gevalle van meervoudige infeksies met twee verskillende koronavirusse is opgemerk, wat die potensiaal vir rekombinasie aantoon, wat kan lei tot 'n nuwe koronavirus wat soönotiese potensiaal kan hê. Hierdie studie het getoon dat beide gasheer- en omgewingsfaktore koronavirus ekologie kan beïnvloed. Vroulike *Neoromicia capensis* vlêrmuise wat voorkom in laagliggende areas in die Woud bioom het die hoogste waarskynlikheid om koronavirus positief te wees. Afwykings tussen opsporingsyfers wat met verskillende siftingsmetodes verkry is, het gelei tot die aanvaarding van 'n verbeterde benadering en aanbevelings vir toekomstige vlêrmuis-koronavirus toesigstudies is gemaak.

Personal Acknowledgements

I have had the privilege of spending the last few years completing my PhD within the Division of Medical Virology. During this time I have gained invaluable research skills and experience. I have learned how to work independently and how to be adaptable in the face of change as is inevitably to be expected when conducting research. For this I am most grateful.

Thank you to the staff and students in the Division of Medical Virology for every friendly smile, chit-chat, and cup of tea shared. Your support and encouragement is appreciated.

Thank you to our collaborators, Prof. Corrie Schoeman (University of KwaZulu-Natal) and Dr Leigh Richards (Durban Natural History Museum), and their respective field teams, for collecting study samples and for accommodating my many requests.

A special thank you must go to Karmistha Poovan for her friendship and constant support. Thank you for putting up with my endless questions, moments of crises, and requests to "quickly run something by you", you are the epitome of a true friend!

To Dr Ndapewa Ithete, thank you for your willingness to serve as a co-supervisor on this study. I have not only learned a great deal of practical knowledge from you but also how to teach and guide others with patience and kindness.

A huge thank you must go to Prof. Corrie Schoeman for your co-supervision. Thank you for exposing me to the world of bats and ecology and for broadening my understanding of the wider subject within which my study falls. Your expertise has been invaluable and I am grateful for all that I have learned from you. Thank you for convincing me to get out into the field - the fieldwork expeditions certainly created some of my most memorable memories!

An enormous thank you must go to Prof. Wolfgang Preiser for supervising me. Your guidance, encouragement, input, and support have been invaluable to the success of this study. Thank you for sharing your wealth of knowledge and expertise with me – I have never not left having learned something new after chatting to you. You have greatly facilitated my growth as a scientific researcher during this journey and for that I am most grateful.

To my dear parents, Wayne and Erica, thank you for your unending love and support, and for always believing in me. You have continuously inspired and motivated me to follow my dreams - I could not have asked for better parents, thank you. Friends and family, near and afar, your support during this journey is greatly appreciated – I could not have done this without all the encouragement along the way.

Lastly, thanks must go to my loving husband, Ludi, for your unwavering love, support and motivation – you continue to be my number one fan! Thank you for all the encouragement and patience throughout the madness of it all. Thank you for keeping me well-fed and happy – your uncanny ability to make me laugh each day has certainly made the challenging days more bearable. The journey has been a long one but so much more enjoyable with you by my side.

Formal Acknowledgements

Individual	Institute	Contribution
Dr Leigh Richards	Durban Natural History Museum	collection and provision of loaned sample material
Dr Ndapewa Ithete	Division of Medical Virology, Stellenbosch University	provision of RGU_2c primers
Dr Tasnim Suliman	Division of Medical Virology, Stellenbosch University	provision of primers for extended amplification of lineage C betacoronaviruses
Mr Quartus Laubscher	Private	monthly collection of samples from a bat colony on his property
Prof. Corrie Schoeman	School of Life Sciences, University of KwaZulu-Natal	collection and provision of study samples; provision of R code for biogeographic analyses
	Central Analytical Facility: DNA Sequence Unit, Stellenbosch University	sequencing electrophoresis
	South African Weather Service	provision of requested weather data.
	German Research Trust (DFG)	provision of project funding
	National Health Laboratory Service (NHLS) Research Trust	provision of project funding
	Harry Crossley Foundation	provision of project funding
	Poliomyelitis Research Foundation (PRF)	provision of project funding and scholarships
	National Research Foundation (NRF)	provision of scholarships and travel bursaries
	Stellenbosch University	provision of scholarships and travel bursaries

"Nature is a very strange affair, and the strangeness already encountered by our friends the physicists are banalities compared to the queer things being glimpsed in biology and the much queerer
things that lie ahead."

- Lewis Thomas (1985)

Table of Contents

List of Figures

List of Tables

List of abbreviations

Abbreviations	Meaning
aa	amino acid
ACE2	angiotensin converting enzyme II
AIC	Akaike Information Criterion
AICc	corrected Akaike Information Criterion
ASF	African swine fever
BIC	Bayesian Information Criterion
BLAST	Basic Local Alignment Sequence Tool
blastn	Standard Nucleotide BLAST
blastx	Translated BLAST
bp	base pair
BS	bootstrap
BSA	bovine serum albumin
BtCoV	bat coronavirus
cDNA	complementary DNA
COI	cytochrome oxidase I
CoV	coronavirus
cyt *b*	cytochrome b
ddNTPs	dideoxynucleotides
DDP4	dipeptidyl peptidase 4
dNTPs	deoxynucleotides
DEPC	diethylpyrocarbonate
DNA	deoxyribonucleic acid
E	envelope protein
EIDs	emerging infectious diseases
EPT	Emerging Pandemic Threats Program
ERGIC	endoplasmic reticulum-Golgi intermediate compartment
FEC	feline enteric CoV
FIPV	feline infectious peritonitis virus
FMI	forearm mass index
GPS	Global Positioning System
HE	hemagglutinin-esterase
HIV	Human Immonodeficiency Virus
HS	Hot Start
ICTV	International Committee on Taxonomy of Viruses
IOM	Institute of Medicine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IUCN	International Union for Conservation of Nature
kb	Kilobase
LB	Luria-Bertani
M	Membrane protein
MCS	multiple cloning site
MEGA	Molecular Evolutionary Genetics Analysis
MERS	Middle East Respiratory Syndrome
ML	Maximum Likelihood method
MSA	multiple sequence alignment
N	nucleocapsid protein
NCBI	National Centre for Biotechnology Information
NeoCoV	*Neoromicia* MERS-related betacoronavirus (Genbank ID KC869678)
NJ	Neighbour joining method
nsp	non-specific protein
nt	nucleotide

oligo	oligonucleotide
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEDV	porcine epidemic diarrhoea virus
pp	polyprotein
R_0	basic reproductive number
RdRp	RNA dependent RNA polymerase
RefSeq	online NCBI reference sequence database
RGU	RdRp-based grouping unit
RNA	ribonucleic acid
RSA	Republic of South Africa
RT	reverse transcriptase
RT-PCR	reverse transcription PCR
S	spike protein
S1/2	spike protein subunit 1/2
SANBI	South African National Biodiversity Institute
SARSr	SARS-related
SB	sodium boric acid
ssRNA	single stranded RNA
TAE	Tris-acetic acid
Tm	melting temperature
TRS	transcription regulatory sequence
UAE	United Arab Emirates
UK	United Kingdom
USA	United States of America
USAID	United States Agency for International Development
UV	ultraviolet
WHO	World Health Organization

Chapter 1 Introduction

This dissertation is divided into five chapters that detail the introduction, literature review, methods, results, discussion and conclusion, respectively.

Chapter one provides an introduction to the study described and discussed in this dissertation. A brief background will set the context for the rationale driving the research question. The formulated hypotheses and resulting objectives are listed along with a brief overview of the significance of this study.

1.1 Brief background

The research question posed in this dissertation is concerned with emerging infectious diseases (EIDs), specifically viruses with zoonotic potential; i.e. the potential to be naturally transmitted from animals to humans through direct or indirect contact via an intermediate host (Morse, 1995; Taylor *et al.*, 2001; Woolhouse and Gowtage-Sequeria, 2005).

A number of animals, such as rodents, wild birds, and bats, have been identified as reservoirs for EIDs (Reed *et al.*, 2003; Calisher *et al.*, 2006; Mackenzie and Jeggo, 2013; Han *et al.*, 2015b). This study focuses on bats, flying mammals belonging to the order Chiroptera, as reservoirs of EIDs. With more than 1200 different bat species recorded to date, bats form the second largest mammalian group. With a number of characteristics enabling the maintenance of virus lifecycles within bat populations, these animals have gained increased recognition as reservoir hosts for zoonotic viruses such as rhabdoviruses, paramyxoviruses, and coronaviruses (Calisher *et al.*, 2006; Monadjem *et al.*, 2010; Drexler *et al.*, 2012; Han *et al.*, 2015b).

Coronaviruses (CoVs), the viruses of interest in this study, are positive sense single-stranded RNA viruses, widely recognised for their genetic diversity due to their large genome size and unique random template switching mechanism employed during replication (Cavanagh and Britton, 2008). CoVs encompass four genera, namely *Alpha*-, *Beta*-, *Gamma*-, and *Deltacoronavirus*, usually associated with mild disease in humans. However, in 2002 and 2012, outbreaks of severe respiratory disease led to the identification of highly pathogenic human CoVs, Severe Acute Respiratory Syndrome (SARS)-CoV and Middle Eastern Respiratory Syndrome (MERS)-CoV, respectively (Drosten *et al.*, 2003; Ksiazek *et al.*, 2003; Zaki *et al.*, 2012).

Subsequent research efforts have implicated bats as the likely original hosts for all mammalian CoVs including the less pathogenic human CoVs, HCoV-229E and HCoV-NL63 (Vijaykrishna *et al.*, 2007; Woo *et al.*, 2012b; Corman *et al.*, 2015; Tao *et al.*, 2017). Recent studies have shown alpha- and betaCoVs circulating in various bat species with a wide geographic distribution (Gloza-

1

Rausch *et al.*, 2008; Pfefferle *et al.*, 2009; Anthony *et al.*, 2013; Corman *et al.*, 2013; Góes *et al.*, 2013; Lelli *et al.*, 2013; Drexler *et al.*, 2014; Razanajatovo *et al.*, 2015; Smith *et al.*, 2016; Wang *et al.*, 2017). With the majority of bat CoVs (BtCoVs) considered to be bat host species-specific rather than region-associated and with a great known diversity of bat species, this group of mammals creates a species richness capable of driving novel CoV emergence (Chu *et al.*, 2006; Gloza-Rausch *et al.*, 2008; Turmelle and Olival, 2009; Reusken *et al.*, 2010; Anthony *et al.*, 2017b).

Despite the wealth of evidence supporting the idea that bats are the reservoir host from which all mammalian CoVs emerged, very little is known about the mechanisms of their maintenance and amplification in bats. First studies looking at the ecology of CoVs within bat populations have indicated that the age and reproductive status of bats may play a role in the amplification and transmission of BtCoVs (Gloza-Rausch *et al.*, 2008; Drexler *et al.*, 2011) . Furthermore, not only has coinfection of different bat species with two or more CoVs been demonstrated, but the transmission of CoVs between different bat species has also been reported (Tang *et al.*, 2006; Lau *et al.*, 2012a; Ge *et al.*, 2016). These cross-species transmission events and instances of coinfection may drive recombination events that could lead to the emergence of novel BtCoVs with zoonotic potential (Tao *et al.*, 2017). These findings indicate important factors that might influence the emergence of zoonotic CoVs from this reservoir.

1.2 Rationale

A number of extrinsic and intrinsic changes, such as season and age, have been found to be potential predictors of virus infections in bats (Drexler *et al.*, 2011; Plowright *et al.*, 2014; Anthony *et al.*, 2017b; Seltmann *et al.*, 2017). With increased globalisation and urbanization leading to encroachment onto wildlife habitats, changes in agricultural and domestic animal farming practices, and enormous increases in the trade of animals and animal products, contact between humans and wildlife virus reservoirs, such as bats, has greatly increased (Lederberg, 1993; Taylor *et al.*, 2001; McMichael, 2004). This increased contact drives the likelihood of pathogen spillover into the human population that could lead to outbreaks, epidemics, or even pandemics as seen with the outbreak of SARS-CoV (WHO, 2004; de Wit *et al.*, 2016). Surveillance is therefore an important defence strategy to detect the early emergence of new pathogens. It is also important to understand the ecology behind the emergence of new pathogens and this requires a multi-disciplinary approach (Meslin, 1992; Morse, 2012; Koopmans, 2013; McNamara *et al.*, 2013).

To date, few studies have been published on CoVs in South African bats providing very little information regarding the diversity of BtCoVs found in this region and how prevalent these viruses are (Müller *et al.*, 2007; Geldenhuys *et al.*, 2013; Ithete, 2013; Ithete *et al.*, 2013). Evidence of CoVs in South African bats was first published in 2007 with the detection of antibodies in bats reactive to SARS-CoV, while the first CoV sequences representing three different bat alphaCoVs were published in 2013 (Müller *et al.*, 2007; Geldenhuys *et al.*, 2013). During co-supervisor Dr

Ithete's PhD study, eleven alpha- and one betaCoV sequence(s) were obtained from three different bat species in the Western Cape and KwaZulu-Natal provinces of South Africa, providing additional evidence for the existence of CoVs circulating in South African bat populations (Ithete, 2013).

The detected betaCoV was identified as a MERS-related CoV, termed NeoCoV, and was isolated from a *Neoromicia capensis* bat in South Africa (Ithete *et al.*, 2013). Comparative analysis of NeoCoV with MERS-CoV isolates from humans and camels has demonstrated that these viruses belong to the same viral species (Corman *et al.*, 2014a). The bat species, *N. capensis*, is thought to have the most widespread distribution in Southern Africa and these bats could therefore play an important role in maintaining and disseminating potentially zoonotic CoVs in this region (Jacobs *et al.*, 2008).

The preliminary findings noted above drove the need to assess BtCoVs in South African bat populations on three different levels. Firstly, general surveillance of all accessible bat species to broadly describe the diversity of BtCoVs across a wide range of Southern African bat species and environments. Secondly, species-specific surveillance of *N. capensis* bats to assess with limitations, CoV diversity, prevalence, and association with ecological factors within a specific bat species; species-specific surveillance of *N. capensis* bats would be limited by the inability to ensure a fully representative sampling of the *N. capensis* population across all sampled regions. Thirdly, longitudinal sampling of a bat colony to assess, where possible, CoV viral shedding patterns and host-pathogen dynamics.

The study discussed in this dissertation, "The diversity of coronaviruses in Southern African bat populations", therefore aims to describe the existing CoV diversity within bat populations across specific provinces and biomes within South Africa as well as factors that might influence bat-CoV ecology and diversity that could lead to the emergence of novel CoVs from bats.

1.3 Research question

Having identified a gap regarding current knowledge available pertaining to CoVs in Southern African bats, the research question as stated below was formulated:

What is the existing diversity of coronaviruses within Southern African bat populations in specific regions and biomes of South Africa, and what host and / or biogeographic factors influences bat-coronavirus ecology?

1.4 Hypotheses

In an attempt to answer the above research question, four hypotheses were proposed:

- Different CoVs can be found within different bat species across specific provinces and biomes of South Africa.

- There is a high diversity of CoVs within South African bats at the individual bat; bat colony; and bat species level across specific provinces and biomes.

- Co-infection with different CoVs, strains and / species, occurs and increases CoV diversity within South African bat populations at the individual bat level, within bat colonies, within bat species, and / or between different bat species across specific provinces and biomes.

- Host and environmental factors such as age, reproductive state, rainfall, and temperature influences CoV diversity and ecology in bat populations.

1.5 Research objectives

A number of objectives were set in an attempt to prove or reject the above stated hypotheses. These objectives are listed below.

1. Detect and identify previously unknown CoVs in South African bat populations across specific provinces and biomes.
2. Investigate and describe CoV diversity within a specific bat species, *N. capensis*, across different geographical locations within South Africa over time.
3. Investigate and describe CoV prevalence and diversity within an established *N capensis.* bat colony in South Africa at different time points across at least one reproductive season.
4. Determine, where possible, full genome sequences of identified unknown CoVs from South African bats.
5. Where possible, characterize identified unknown CoVs from South African bats in terms of phylogeny, genome annotation, recombination analysis, and protein analysis.
6. Investigate if co-infection with different CoVs occurs in individual South African bats, in bat colonies, and within different bat species.
7. In collaboration with zoologists, collect and analyse biological and ecological data on sampled bats at the individual bat; bat species; and bat colony level to better understand pathogen-host dynamics that may drive CoV diversity and co-infection within bat populations across specific provinces and biomes of South Africa.

1.6 Significance of this research study

To date, very little literature is available on bat viruses and host-pathogen ecology from South Africa. Most studies have focused on lyssaviruses and filoviruses with recent investigations focused on detecting viruses such as orthobunya- and orthoreoviruses not in bats but in their associated bat flies (Crick *et al.*, 1982; Markotter *et al.*, 2006, 2008, Paweska *et al.*, 2016; Jansen van Vuren *et al.*, 2016, 2017). Despite the scarcity of published studies on CoVs, the available body of work has revealed the presence of several CoVs within South African bats, signalling the potential for great CoV diversity in this region (Müller *et al.*, 2007; Geldenhuys *et al.*, 2013; Ithete *et*

4

al., 2013). None of these studies have however conducted large-scale surveillance or systematically monitored bat colonies or specific bat species for CoV infections and / or described accompanying biogeographical data. The existence of only four South African publications relating to BtCoVs further indicates the lack of published knowledge regarding CoVs in bat populations of Southern Africa and highlights the knowledge gap pertaining to pre-emptive surveillance of bat populations and the understanding of pathogen-host ecology in a South African setting. Furthermore, the project embraced a multidisciplinary approach by collaborating with zoologists to study bats as reservoirs of potentially emerging zoonotic BtCoVs that to our knowledge is the first of its kind in our region.

This study aimed to add to the existing body of knowledge regarding the diversity of CoVs within their presumably most important wildlife reservoir, bats. The results obtained here have added to the available information relating to CoV diversity and ecology in Southern African bat populations. Knowledge regarding the diversity and distribution of coronaviruses within different bat species across specific provinces and biomes of South Africa gained from this study could further assist in the development of improved wildlife surveillance sampling and screening strategies for better detection of novel BtCoVs in this region.

Chapter 2: Literature Review

This chapter introduces relevant literature related to the topic of this dissertation. The broad themes covered by this literature review include emerging infectious diseases, bats as reservoir hosts, and coronaviruses of bat origin.

2.1 An introduction to emerging infectious diseases and why we should study them

Emerging infectious diseases (EIDs) have been defined as infectious diseases that are either newly appearing in a population or are known infectious diseases with rapidly expanding geographic ranges accompanied by an incidence that has either increased in the last two decades or threatens to increase in the near future due to changes in its underlying epidemiology (Morse and Schluederberg, 1990; IOM, 1992, 2003; Anon, 1994; Taylor *et al.*, 2001; Woolhouse and Dye, 2001). A third category includes newly recognised diseases, EIDs resulting from the recognition of an existing disease that has previously gone undetected, or a known agent that due to adaptive changes has increased in its pathogenicity causing more severe disease (Morse and Schluederberg, 1990; IOM, 1992; Anon, 1994; Dobson and Foufopoulos, 2001). Additionally, EIDs may be recognised as re-emerging when a known infectious disease reappears in a population with an increased incidence following a previous decline in its incidence (IOM, 1992). EIDs is a complex term that has evolved over time, recently a framework was proposed to redefine EIDs into three groups namely, emerging pathogens that cause EIDs with detrimental impacts on susceptible host populations, emerging pathogens that are capable of causing disease with the potential to result in a novel EID, and novel potential pathogens that have as yet no evidence for causing clinical illness and therefore may or may not have the potential to become in an EID (Rosenthal *et al.*, 2015). In the context of EIDs there are two main kinds of hosts to consider. The reservoir host and its associated ecologic system naturally harbours the infectious agent indefinitely, usually with no overt signs of disease while a susceptible population can be infected by the pathogen, usually with overt disease, and in the process of disease emergence can either represent an intermediate or incidental host, or become a new host in which the pathogen establishes itself (Ashford, 1997, 2003; Haydon *et al.*, 2002).

EIDs were first defined during the 1980's following a number of major outbreaks worldwide at a time when it was thought that infectious diseases were a thing of the past or at the very least limited to developing regions of the world (Morse, 1991; Chomel, 1998; Cohen, 2000). This complacency towards infectious disease resulted from the medical advances made in the fields of antibiotics and vaccine development that dramatically reduced the burden of disease. Major

disease outbreaks including the global HIV pandemic and viral haemorrhagic fevers at the time, and in recent years outbreaks of severe acute respiratory disease and encephalitic fevers, coupled with periodic influenza epidemics, indicate however that infectious diseases will remain important causes of disease and death in the human population (Morse, 1991, 1993; Lederberg, 1993; Schrag and Wiener, 1995; Cohen, 2000; Field *et al.*, 2001; Fragaszy and Hayward, 2014; Holmes and Zhang, 2015; de Wit *et al.,* 2016).

EIDs are caused by bacteria, viruses, protozoa, fungi, and helminths to varying degrees and pose a threat to human health, domestic and wildlife animal populations and, although not discussed in this review, crop- and wild-plants (Taylor *et al.*, 2001; Jones *et al.*, 2008). It has been well established that EIDs commonly result from changes in the ecology of the infectious agent's host species (reservoir), the susceptible host population, and / or the pathogen itself (Schrag and Wiener, 1995; Daszak *et al.*, 2001; Woolhouse *et al.*, 2005). These changes influence a complex relationship between humans, domestic animals, wildlife, and the infectious agents they carry as depicted in Figure 2.1 (Daszak *et al.,* 2001). These factors driving disease emergence will briefly be discussed in Section 2.3.
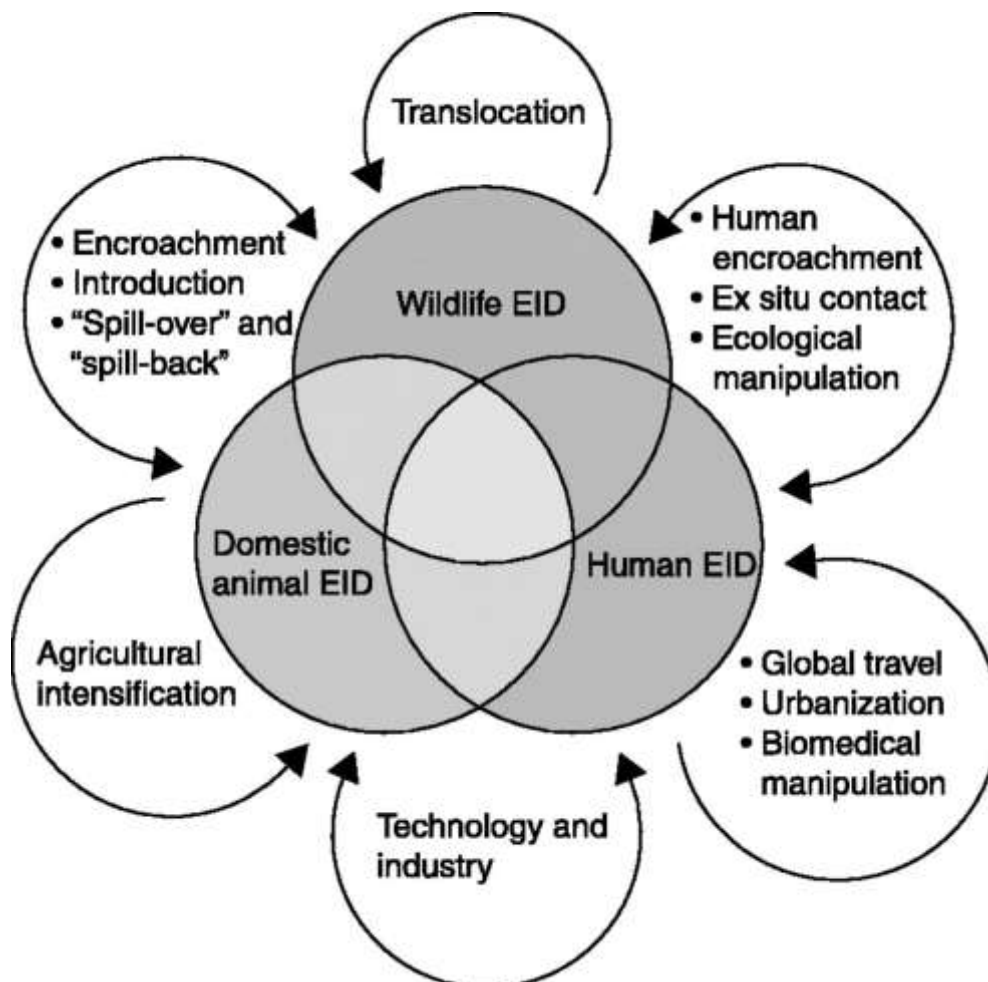
Figure 2.1 The dynamic interaction between hosts and potentially emerging viruses. Disease emergence is driven by a complex of interactions between wildlife, domestic animals, humans, and changes in their ecology. Image adapted from Daszak et al. (2001) *Image used with permission from The American Association for the Advancement of Science.*

7

It is the increased recognition of the importance that this complex network plays in the emergence of new diseases that concepts such as One Health, EcoHealth, and conservation medicine were established. The One World, One Health and Manhattan Principles highlight that humans, animals, and the environment are inextricably linked; humans can no longer be the sole focus in combating disease emergence (Cook *et al.*, 2004; van Helden *et al.*, 2013). EcoHealth aims to ensure the sustainable health of people, animals, and ecosystems through the discovery and understanding of drivers leading to ecosystem and social changes that influences human health and well-being (Wilcox and Kueffer, 2008; Charron, 2012). Conservation medicine is concerned with the interaction between pathogens and disease how they are connected to the interactions between species and their associated ecosystems (Aguiire *et al.*, 2012). These concepts all highlight that cross-discipline collaborations and the understanding of pathogen-host dynamics and environmental interactions are pivotal in assessing the potential for new EIDs to enter the human population (Wilcox and Kueffer, 2008; Aguiire *et al.*, 2012; Cunningham *et al.*, 2017).

Outbreaks of EIDs in domestic animals, particularly those of agricultural value, can result in great economic losses due to mass culling required to limit the spread of disease. In 1983, an outbreak of a virulent influenza A H5N2 strain in chickens in Pennsylvania, USA resulted in the culling of over 17 million birds at an estimated loss of $61 million (Bean *et al.*, 1985; Kawaoka and Webster, 1988). Highly pathogenic avian influenza viruses have continued to emerge since then, affecting South East and East Asia, Europe, Africa, and the Americas (Alexander, 2006; Brown, 2010). Another important disease of domestic animals that has resulted in significant economic losses and threats to food security includes the porcine epidemic diarrhoea virus (PEDV) in the swine industries of the USA and Asia (Song and Park, 2012; Lee, 2015). African swine fever (ASF), causing serious haemorrhagic fever in pigs with a nearly 100% mortality rate, has resulted in great economic losses in many African countries due to mass slaughter of infected herds (Boshoff *et al.*, 2006). It first appeared in Kenya in 1921 and although now largely considered endemic in some regions and even eradicated in others, AFSW continues to spread to previously uninfected countries, posing a threat to pig production (Wozniakowski *et al.*, 2016).

EIDs in wildlife populations are less well monitored and historically have only been deemed important when agriculture or human health appeared at risk (Daszak, 2000). However, EIDs in wildlife populations can lead to losses in biodiversity when endangered animal populations are decimated due to disease, threatening the conservation of global biodiversity (Laurenson *et al.*, 1998; Daszak, 2000; Altizer *et al.*, 2003).

EIDs in domestic and wildlife populations do not only pose a threat to the primarily affected population but can pose further risks when exposure and contact with diseased animals lead to cross species transmission events, pathogen spillover, and subsequent emergence of new pathogens in domestic or wildlife populations (Daszak, 2000; Bengis *et al.*, 2002). The spillover of

canine distemper from domesticated dogs to African wild dogs is one such example and has resulted in significantly decreased populations of African wild dogs with local extinction in some parts of the Serengeti (Alexander and Appel, 1994; Goller *et al.,* 2010).

The spillover of pathogens from animal to human populations is known as zoonosis, a phenomenon highlighted in the literature as one of the most common causes of EIDs affecting the human population (Morse, 1995; Chomel, 1998; Daszak, 2000; Cleaveland *et al.,* 2001). Important zoonotic EIDs in recent decades include HIV, severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), Ebola-, Hendra-, Nipah-, Zika-, and West Nile virus disease (Field *et al.,* 2001; Morse, 2012; de Wit *et al.,* 2016; Singh *et al.,* 2017). For human and public health, natural animal reservoirs therefore serve as an important source of new human disease agents (Chomel, 1998).

Despite the early recognition of EIDs, studies to quantitatively assess host ranges of pathogens along with risk factors of human disease emergence were only published in the twenty-first century (Cleaveland *et al.,* 2001; Taylor *et al.,* 2001; Woolhouse and Gowtage-Sequeria, 2005). Based on the analysis of databases available at the time, the first published studies suggested that zoonotic pathogens were more likely to be associated with disease emergence in humans than non-zoonotic pathogens, and that certain taxonomic groups, particularly RNA viruses, with broad host ranges capable of multi-host transmission were more likely to be classified as causative agents of EIDs (Burke, 1998; Cleaveland *et al.,* 2001; Taylor *et al.,* 2001; Woolhouse, 2001; Woolhouse and Gowtage-Sequeria, 2005; Woolhouse *et al.,* 2005). Multi-host pathogens, such as influenza A and rabies viruses, are encountered by several different host populations, of which some will serve as infection reservoirs, others as dead-end host populations, and some as amplifying hosts responsible for maintaining active transmission of the pathogen (Woolhouse, 2001; Haydon *et al.,* 2002). The first spatiotemporal analysis of EID events by Jones (2008) confirmed the notion that zoonotic agents, particularly those originating from wildlife populations account for the majority of EIDs affecting the human population. With each individual drug-resistant microbial strain considered a unique pathogen, this study concluded that bacteria were more likely to cause EIDs than viruses (Jones *et al.,* 2008). Importantly, this widely cited study highlighted low and middle income countries as EID hotspots, regions where high human population density overlaps with great biological diversity and where surveillance efforts are most lacking (Jones *et al.,* 2008).

These early studies emphasized the importance of understanding interactions between humans, domestic, and wildlife populations with pathogens capable of transmission between multiple hosts of particular interest and importance in the battle against EIDs (Daszak, 2000; Cleaveland *et al.,* 2001). With the great diversity of EIDs, broadly targeted surveillance and applied research efforts to detect, identify, and understand EIDs have been suggested numerous times in the literature as a means to monitor both animal and human populations for potentially novel EIDs, especially when

zoonotic EIDs represent an increasing and significant threat to global health (Morse and Schluederberg, 1990; IOM, 1992; Berkelman *et al.*, 1994; Binder *et al.*, 1999; Cohen, 2000; Daszak, 2000; Woolhouse and Gowtage-Sequeria, 2005; Woolhouse *et al.*, 2005; Jones *et al.*, 2008; Anthony *et al.*, 2017a). Furthermore, research relating to the ecology, pathology, and population biology of host-pathogen systems at the individual, population, and environmental level is needed to fully elucidate the underlying causes of EIDs (Morse and Schluederberg, 1990; Daszak, 2000; Anthony *et al.*, 2017a, 2017b). Surveillance of potential pathogens in wildlife populations is therefore an important step in the prevention of zoonotic disease outbreaks, that together with the assessment of anthropogenic factors that could influence disease emergence, requires a multidisciplinary, One Health, approach where veterinarians, clinicians, epidemiologists, pathologists and public health specialists are all key role players (Chomel, 1998; Daszak, 2000; Wood *et al.*, 2012; Cunningham *et al.*, 2017).

This dissertation focuses on bats as reservoirs for potentially emerging zoonotic viruses, specifically coronaviruses (CoVs), that may be pathogenic to humans. Therefore, the remainder of this literature review will focus predominantly on emerging viruses from vertebrate reservoirs. As briefly indicated, it is however important to keep in mind that a large number of EIDs are not viral in nature and that not all EIDs, including emerging viruses, originate in vertebrate reservoirs.

2.2 Viruses as emerging infectious diseases

Modern medicine still struggles to effectively control viral pathogens. Few have been controlled by vaccination or antiviral therapies highlighting the need not only to understand their replication mechanisms but also to understand the pressures and mechanisms influencing their emergence (Domingo and Holland, 1997). Emerging viruses, like other emerging pathogens, are often not newly evolved organisms but instead are existing viruses invading new host groups or geographic regions due to changes in "viral traffic" patterns, commonly resulting from human intervention (Morse and Schluederberg, 1990; Morse, 1991; Schrag and Wiener, 1995).

Viruses have been noted to have higher mutation rates than other pathogens. This ability to evolve and adapt more quickly to new hosts may explain the greater relative risk of emergence among viruses (Cleaveland *et al.,* 2001; Longdon *et al.*, 2014; Johnson *et al.*, 2015). RNA viruses in particular are known to have a higher mutation rate than DNA viruses resulting from a general absence of proofreading mechanisms during RNA synthesis (Morse and Schluederberg, 1990). Genetic changes within the pathogen's genome that facilitate spillover or zoonosis constitutes host adaptation and may involve anything from mutation of only a few nucleotide substitutions to major genetic changes such as those associated with recombination and genome reassortment (Morse and Schluederberg, 1990; Domingo and Holland, 1997; Burke, 1998; Woolhouse, 2001). Pathogen adaption through genetic change is exemplified by the influenza A virus that with its eight-segmented genome readily undergoes reassortment that has resulted in a number of pandemic

human strains (Oxford, 2000; Morse and Schluederberg, 1990). Influenza A epidemics arise when the virus undergoes minor genetic changes through mutations leading to antigenic drift; in 1983 a single mutation within the avirulent H5N2 influenza A genome resulted in a fatal epidemic amongst chickens in the USA (Bean *et al.*, 1985; Kawaoka and Webster, 1988; Morse and Schluederberg, 1990). These genetic variation mechanisms can lead to a complex, dynamic distribution of diverse virus strains, so-called quasispecies, that are closely related but not identical with the fittest becoming the most dominant one within the population of genomes (Morse and Schluederberg, 1990; Domingo and Holland, 1997).

RNA viruses are more likely infect avian and mammal species suggesting that RNA viruses may hold fundamental characteristics that make them more transmissible and more capable of crossing species barriers than their DNA counterparts (Cleaveland *et al.*, 2001). The variable and adaptable nature of RNA viruses' genetics along with the effects of changes in the environments are conducive to the emergence of new viral pathogens, facilitating host jumping if a new potential host is encountered (Domingo and Holland, 1997).

2.2.1 The process of zoonotic virus emergence (how do viruses emerge?)

It has been suggested by Morse (1991) that the process of emergence requires two events: namely, the introduction of the agent into a new susceptible host population followed by the establishment of infection within the new population that leads to further dissemination. Therefore, the susceptible host species must be exposed to the pathogen, either through interaction with the reservoir host or via an intermediate host. The latter frequently serves as an amplifying host. Common transmission routes for viruses from reservoirs to new hosts involve direct contact either through physical touch or close proximity, indirect contact by means of virus-contaminated food or surfaces, and lastly through vector-borne contact such as biting arthropods (Taylor *et al.*, 2001).

Secondly, the pathogen must be capable of infecting the new host species using cell receptors that are phylogenetically conserved playing an important role in facilitating this step (Woolhouse, 2002; Woolhouse *et al.*, 2005). Where suitable receptors are conserved across a range of potential host species, these species will likely be predisposed to infection by viruses using these specific cell receptors; this is exemplified by the wide host range of foot-and-mouth virus using the integrin vitronectin cell receptor and rabies virus which uses the nicotinic acetylcholine receptor (Woolhouse *et al.*, 2005) Additionally, certain viruses, such as rabies, produce many genetic variants that can infect a broader range of host species with successful variants often becoming associated with a specific host species leading to host specificity (Woolhouse, 2001). This explains why RNA viruses, having a high mutation rate and often existing as a quasispecies of genetic

variants, are capable of infecting a wider host range and are more likely to be zoonotic than their DNA counterparts (Woolhouse, 2001).

It is important to keep in mind that although a number of infectious agents, including viruses, are capable of infecting humans and causing disease, the majority of zoonotic pathogens are not highly transmissible within the human population and do not result in major epidemics (Figure 2.2) (May *et al.*, 2001; Woolhouse and Gowtage-Sequeria, 2005; Pike *et al.*, 2010; Preiser, 2012; Meyer *et al.*, 2015). A number of zoonotic pathogens enter the human population but are unable to sustain human-to-human transmission (Pike *et al.*, 2010).
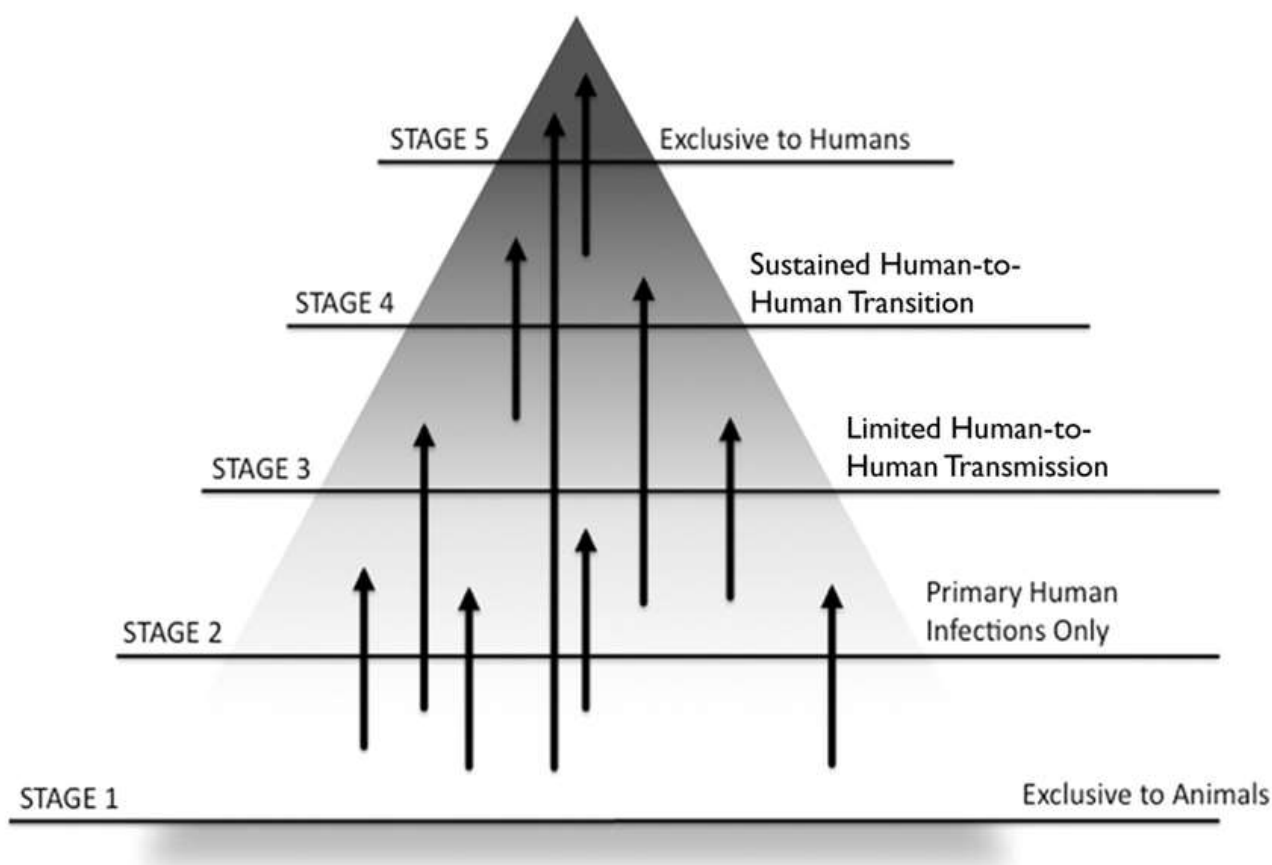
Figure 2.2 The different stages of zoonotic disease emergence. The diagram depicts the different stages required for zoonotic pathogens to move through before being able to cause major epidemics in human populations. Used with permission from Oxford University Press (Pike *et al.*, 2010).

It is not enough for a virus to solely have the ability to infect a population, but rather its ability to be transmitted efficiently within that population is an important factor for emergence; viruses adapted to human transmission are therefore most likely to emerge (IOM, 1992; Pike *et al.*, 2010; Johnson *et al.*, 2015). The infection of a new host population and subsequent widespread transmission within the new host species often do not occur at the same time but rather results from a complex network of environmental and anthropogenic changes that increase the likelihood of dissemination (Figure 2.1) (IOM, 1992).

The potential size of an EID outbreak is determined by the pathogen's basic reproduction number (R_0), i.e., the average number of secondary cases of infection resulting from a single primary introduction into a population (May *et al.*, 2001; Woolhouse *et al.*, 2005). When pathogens have a low transmissibility (R_0 approaches 0) within the human population, the size of the outbreak is influenced mostly by the number of times the pathogen is introduced into the human population; this is exemplified by the rabies and Rift Valley fever viruses (Woolhouse and Gowtage-Sequeria, 2005; Woolhouse *et al.*, 2005). The size of an outbreak caused by a highly transmissible (R_0 exceeds 1 by far) pathogen in contrast is largely dictated by the size of the susceptible population; this is typified by the influenza A virus and measles (Woolhouse and Gowtage-Sequeria, 2005; Woolhouse *et al.*, 2005).

2.2.2 Reservoirs of emerging viruses

A number of animals have been recognised as reservoirs of viruses; these include rodents, birds, and bats (Reed *et al.*, 2003; Wong *et al.*, 2007; Han *et al.*, 2015a). Although the focus in this dissertation is on vertebrate reservoirs of emerging viruses, it should be noted that arthropods play an important role as vectors in the transmission of vector-borne viruses, arboviruses, such as West Nile, Rift Valley fever, Dengue, Yellow Fever, and Sindbis (Morse and Schluederberg, 1990; Gubler, 2001).

Rodents have been viewed as an important mammalian reservoir of infectious diseases. However, the same characteristics that give rodents this recognition also apply to bats; certain species of both bats and rodents are usually found in large, high density aggregates, and certain species of both are commensal with humans (Messenger *et al.*, 2003). Recently, bats have gained increased recognition as a reservoir host for zoonotic viruses (Calisher *et al.*, 2006; FAO, 2011; Kuzmin *et al.*, 2011; Plowright *et al.*, 2014; Moratelli and Calisher, 2015). A recent study investigating mammalian host-virus relationships demonstrated that for a given species, the total number of viruses capable of infecting that species as well as the proportion likely to be zoonotic, are predictable (Olival *et al.*, 2017). The study went on to confirm the findings from a quantitative comparative analysis that found bats to host more zoonotic viruses per species than rodents (Luis *et al.*, 2013; Olival *et al.*, 2017). Bats exhibit a higher degree of sympatry than rodents in that bat species are more likely to overlap in their geographic distribution, where different species may share roosts and feeding grounds, than rodent species. Bats are therefore more likely to share viruses through interspecies transmission than rodents (Luis *et al.*, 2013, 2015). What is particularly concerning is that the pathogenicity of pathogens jumping from bats to humans and the number of zoonotic events have increased over the last twenty years (Dobson, 2005; Olival *et al.*, 2012).

2.3 Bats as virus reservoirs

Bats, order Chiroptera, with more than 1200 currently recognised species, represent the second largest order of mammals and were previously divided into two suborders, Megachiroptera (Old World fruit bats / Megabats) and Microchiroptera (Microbats) (Dobson, 1875). This suborder division was largely based on differences in size, sensory, and feeding characteristics with Microchiroptera bats being small-bodied and having the ability to echolocate and detect insect prey while the Megachiroptera bats were large bodied bats inhabiting the Old World tropics that use sight and olfactory senses to feed on fruit and / or nectar (Dobson, 1875; Simmons, 2005). Recent molecular findings have challenged this traditional monophyletic mega- and microbat division (Simmons, 2005). Molecular findings have led to the most recently accepted taxonomy where the order Chiroptera is divided into two suborders, Vespertilioniformes and Pteropodiformes (Hutcheon and Kirsch, 2006). The Pteropodiformes include the Old World Pteripodidae, Hipposideridae, Rhinolophidae, Megadermatidae, and Rhinopomatidae families while all other families are considered to belong to the Vespertilioniformes suborder (Eick *et al.*, 2005; Hutcheon and Kirsch, 2006; Monadjem *et al.*, 2010).

The characteristics that distinguish bats from other mammals are likely to influence the role played by bats in the maintenance and transmission of zoonotic viruses (Calisher *et al.*, 2006). For example, the ability to fly, coupled with their general abundance, and wide distribution with an often gregarious roosting nature, are possible key attributes to the greater occurrence of viruses found in bats compared to other mammalian groups (Messenger *et al.*, 2003; Calisher *et al.*, 2006; Wong *et al.*, 2007; Luis *et al.*, 2013; Han *et al.*, 2015b; Moratelli and Calisher, 2015). These life history traits increase the probability of intra- and inter-species and long distance transmission of viruses, facilitating virus dissemination and maintenance (Calisher *et al.*, 2006; Turmelle and Olival, 2009; Kuzmin *et al.*, 2011; Olival *et al.*, 2012).

Bats are the only mammals capable of self-powered flight. Daily foraging movements and seasonal migratory patterns can facilitate the dispersal of viruses between different bat colonies and bat species across different geographic locations. For example, rabies disease has been associated with the migratory routes of *Pipistrellus nathussii* in France while an annual migration of fruit bats has been indicated as the likely source of the 2007 outbreak of Ebola in the Democratic Republic of Congo (Brosset, 1990; Calisher *et al.*, 2006; Leroy *et al.*, 2009; Altizer *et al.*, 2011).

The unique immune system of bats is thought to assist in the maintenance of viruses. With the evolution of flight, came the hollowing of bones, influencing components of the immune system that, coupled with elevated metabolic rates and body temperatures, may facilitate the long-term maintenance of viruses without overt disease in these animals (Dobson, 2005; Zhang *et al.*, 2013; O'Shea *et al.*, 2014). Furthermore, the ability of certain bat species to enter a state of torpor and seasonal hibernation, with lower body temperatures and reduced metabolic rates, may further

facilitate the long-term maintenance of viruses by suppressing immune responses and delaying the clearance of viruses (Calisher *et al.*, 2006; George *et al.*, 2011; O'Shea *et al.*, 2014).

The vast number of different bat species documented to date provide a diverse pool of potential reservoirs that can maintain a great viral richness (Wong *et al.*, 2007; Turmelle and Olival, 2009). Furthermore, the long evolutionary history of bats has facilitated the coevolution of viruses, such as lyssaviruses and CoVs, with bats as their natural reservoirs (Badrane and Tordo, 2001; Cui *et al.*, 2007). Viruses that evolved with bats may use cellular receptors and biochemical pathways that are conserved in mammals, enhancing the ability of these viruses to transmit to other mammals (Calisher *et al.*, 2006).

Bats inhabit a range of roosts in both natural and man-made structures, such as foliage, caves, hollow trees, mine shafts, bridges, and roof spaces (Kunz and Lumsden, 2003; Wong *et al.*, 2007; Monadjem *et al.*, 2010). Roosts play a significant role in the practice of mating, hibernating, and rearing of young while catering for complex social interactions that coupled with often large population densities and crowded roosting behaviour can facilitate intra- and interspecies transmission of viruses (Kunz and Peirson, 1994; Kunz and Lumsden, 2003; Calisher *et al.*, 2006). Because many bat species roost near or in human settlements, these animals provide a source of potential zoonotic spillover not only to humans but also to domestic animals. The likelihood of such interaction is further increased by deforestation and changes in land-use (Jones *et al.*, 2013).

With bats identified as likely reservoir hosts for a number of pathogenic zoonotic viruses of interest to public health, such as lyssaviruses, paramyxoviruses, and CoVs, research efforts aimed at detecting novel viruses in bats have increased substantially (Field *et al.*, 2001; Kuzmin *et al.*, 2011; Ge *et al.*, 2013; Wang and Hu, 2013; Hu *et al.*, 2015). A number of viruses of which the zoonotic potential is yet unknown have also been detected in these mammals; of interest is the recent detection of influenza A-like viruses and hantaviruses (Tong *et al.*, 2012; Weiss *et al.*, 2012; Arai *et al.*, 2013; Wu *et al.*, 2014; Xu *et al.*, 2015; Witkowski *et al.*, 2016). As highlighted by Dobson (2005), there is much we still do not know about how bats maintain viruses and therefore it is important to obtain more knowledge on the ecology and immune responses of bats if we are to prevent future EIDs from these animals (Dobson, 2005).

2.4 Factors driving the emergence and transmission of viruses in bats

Jones *et al.* (2008) provided quantitative evidence that EID events have been increasing over time. A number of factors have been identified and described that drive the general emergence of new infectious diseases. These can broadly be divided into six themes that include changes in environment and land use, international travel and commerce, changes in demographics and behaviour, pathogen adaptation and change, changes in technology and industry, and breakdown

15

of public health systems and measures (IOM, 1992; Cohen, 2000; Cleaveland *et al.*, 2001; Woolhouse and Gowtage-Sequeria, 2005; Woolhouse *et al.*, 2005). The most frequent factor amongst these themes is the influence of human intervention and behaviour (Morse and Schluederberg, 1990). The twentieth century saw great changes in technology and society that have continued to accelerate into the twenty-first century driving the increase in EIDs (Cohen, 2000; Jones *et al.*, 2008).

Factors driving the emergence of new viruses do not act in isolation. Anthropogenic changes occur amidst pathogen evolution and may lead to increased transmission between individuals or increased contact between the pathogen and its new host population or influences selective pressure for the dominance of potentially pathogenic strains adapted to the changing conditions (Daszak *et al.*, 2001). The drivers of virus emergence from bats have been linked to those associated with viral richness, such as host diversity and climate variability, and transmission opportunities, such as human population density, increased mobility, bushmeat hunting, and changes in livestock and agriculture practices (Ludwig *et al.*, 2003; Brierley *et al.*, 2015).

2.4.1 Climate variability, agricultural changes, and urbanization

Despite the strong evidence for the role played by genetic characteristics of pathogens and their influence on the emergence of disease, changes of host and pathogen ecology are some of the most frequently recognised factors influencing emergence (Morse, 1995; Schrag and Wiener, 1995). Ecological changes and agricultural development are especially important as factors leading to the introduction of zoonotic viruses that are often associated with high-case fatalities of a previously unrecognised disease (Morse, 1995). Such changes do not only influence human disease emergence but can also lead to interspecies transmission that can have devastating economic consequences when impacting domestic animals and wildstock (Morse, 1997).

With the cultivation of large pig populations using traditional free-range husbandry practices and little or no biosecurity for semi-intensive or intensive farming, agricultural practices have been linked to the spread of ASF in Africa. Poverty, lack of organised pig production and ineffective management of ASF has exacerbated the spread of the disease in the region (Penrith *et al.*, 2013).

The rapid expansion of the human population has increased the need for housing, potable water, food, and numerous other resources, resulting in the destruction of natural habitats and encroachment on areas inhabited by bats leading to an increased interface for potential zoonotic transmission events (Millar and Moore, 2006; Han *et al.*, 2015b). This increased contact leads to a host-parasite network between humans, domestic animals, and wildlife contributing to the emergence of zoonotic disease by facilitating the introduction of a remote virus into a larger susceptible population (Morse, 1997; Chomel, 1998; Daszak, 2000).

With the destruction of natural bat habitats and climatic changes comes a shortage of natural food resources that accompanied by expanding agricultural practices may further encourage the intrusion of bat populations and spread of viruses as demonstrated by the emergence of the Nipah virus. A major anthropogenic change influencing the emergence of the Nipah paramyxovirus in Malaysia was a change in agriculture practice leading to an abundance of fruit trees near pig farms that, accompanied by climatic changes brought about by an El Niño-related drought, facilitated an influx of the reservoir host, pteropid fruit bats, to the pig rearing area resulting in increased opportunities for contact with a susceptible intermediate host population, pigs; the pig population served as an amplifying host that transmitted the virus to humans (Daszak *et al.*, 2001; Field *et al.*, 2001; Yob *et al.*, 2001; Chua, 2003, 2010). The rapid spread of Nipah virus through the Malaysian population was likely exacerbated by an increase in pig trade at the time with augmentation of transmission caused by inter-farm spread via domestic dogs and cats (Daszak *et al.*, 2001; Field *et al.*, 2001; Chua, 2010). As demonstrated with the emergence of Nipah virus, the expansion of agriculture coupled with continual changes in farming practices has allowed humans to alter their surrounding environment, often leading to the introduction of new infections that can be exacerbated by climatic changes (Morse, 1995).

2.4.2 Advancements in trade and travel

The advancement of transport during the twentieth century has greatly increased international travel activities and subsequently has facilitated the global transmission of localized infections to new populations and geographic regions disseminated through human-to-human contact and aerosols where they can lead to major outbreaks (Morse, 1995; Schrag and Wiener, 1995; Soto, 2009).

The increased mobility of people and goods and animals through trade over long distances is an important driver sustaining endemics. For example, the transport of infected pigs and movements of infected abattoir workers greatly contributed to the local and international spread of infection during the outbreak of Nipah virus in Malaysia and Singapore (Paton *et al.*, 1999; Lam and Chua, 2002; Chua, 2010; Han *et al.*, 2015b). Similarly, the transboundary spread of ASF across many African countries has been associated with the movement of infected animals and pig meat (Penrith et al., 2013). More recently, the dissemination of ASF in the former Soviet republic of Georgia is thought to have resulted from importation of porcine products contaminated with the ASF virus by ship from the eastern side of southern Africa (Rowlands *et al.*, 2008).

The outbreak of SARS has been linked to live animal markets trading in palm civets, the intermediate amplifying host of SARS-CoV; live animal markets where a diverse range of animal species are forced into close contact with each other provide a mixing vessel for spillover and emergence of new viruses (Wang and Eaton, 2007; Wang and Crameri, 2014). International air travel further contributed to the spread of SARS from its original outbreak in Guangdong, China to

25 other countries in less than six months (WHO, 2004; Millar and Moore, 2006; Soto, 2009). Similarly, the 2014 - 2016 outbreak of Ebola haemorrhagic fever, the largest outbreak to date, quickly spread to other countries with the travel of infected patients from West Africa (Holmes *et al.*, 2016; Jacobsen *et al.*, 2016).

2.4.3 Changes in human demographics

The rapid transmission of EIDs has been documented several times in regions with civil unrest and is often accompanied with political and socio-economic changes that, coupled with a low resource setting or inadequate health system infrastructure, rapidly fuels the spread of a new emerging virus. This has been noted for the recent Ebola outbreak where changes in human settlement patterns and land use due to political-economic shifts resulted in populations with increased vulnerability to zoonotic events (Holmes *et al.*, 2016; Dzingirai *et al.*, 2017; Singh *et al.*, 2017).

2.5 Detecting emerging viruses

The increased recognition of zoonotic diseases as a threat to human and animal health has led to the establishment of organisations such as PREDICT, a project part of the United States Agency for International Development's (USAID) Emerging Pandemic Threats Program (EPT), that aims to increase the global capacity for the detection and discovery of viruses with pandemic potential (PREDICT Consortium, 2014). Advances in technology have led to improved detection and identification tools (Chomel, 1998). For example, the virus causing Hantavirus pulmonary syndrome was one of the first instances where a human disease of unknown origin was identified using molecular epidemiology by detecting genetic material using the polymerase chain reaction (PCR) (Nichol *et al.*, 1993). With the development of PCR and nucleic acid sequence analysis tools, molecular biology has advanced rapidly, increasing available knowledge greatly (Chomel, 1998). The development of PCR assays that broadly target members of virus families such as the CoVs, paramyxoviruses, and astroviruses, has been especially fruitful in obtaining genomic information of viruses circulating in animal reservoirs such as bats (Poon *et al.*, 2005; de Souza Luna *et al.*, 2007; Chu *et al.*, 2008b; Tong *et al.*, 2008).

Serology-based assays can also be useful in detecting evidence of viral infection in reservoir host populations but obtaining sufficient volumes of blood from small mammals, such as insectivorous bats, can limit such investigations. A number of bat species are threatened and conservation laws often prevent protocols requiring exsanguination with preference for non-invasive sampling protocols (www.iucnredlist.org) (Mickleburgh *et al.*, 2002). The use of serology is further limited by a lack of assays able to detect different viruses within the same virus family or genera, further exacerbated when there is antibody cross-reactivity between closely related viruses.

Cell culture based studies attempting to isolate novel viruses from bats have had varied success; for some viruses such as the CoVs where cell culture has rarely been reported this has been less rewarding than for other viruses such as the paramyxoviruses where a number of novel bat-borne paramyxoviruses have been isolated in cell culture (Barr *et al.*, 2015; Pavri *et al.*, 1971; Chua *et al.*, 2001; Baker *et al.*, 2013; Ge *et al.*, 2013; Yadav *et al.*, 2016). Isolation of bat-borne viruses requires suitable cell lines expressing appropriate cell receptors. For bat viruses not utilising common cell receptors, the requirement for bat-specific cell lines has hindered successful isolation (Olival *et al.*, 2012). Advancements in the development of bat cell lines has greatly improved the success of virus isolation and will continue to prove beneficial for future isolation attempts of bat-borne viruses (Crameri *et al.*, 2009; Eckerle *et al.*, 2014; He *et al.*, 2014b). Isolation attempts are additionally hindered when the reservoir host maintains the virus at low titres, is only transiently infected, or if there is no acute infection at the time of sampling (Mandl *et al.*, 2015).

An alternative to conventional PCR is next generation sequencing that facilitates the rapid generation of thousands to millions of short sequences from a single sample without the detection bias introduced when using primers targeting a specific genome region as with conventional PCR (Behjati and Tarpey, 2013). This technology has increasingly been used in an attempt to decipher the bat virome (Ge *et al.*, 2012; He *et al.*, 2013; Dacheaux *et al.*, 2014; Wu *et al.*, 2016). However, with the resulting large volumes of data, a number of statistical and computational challenges arise, not to mention problems relating to enriching samples for targets of interest such as RNA viruses that are often present in low abundance relative to host and microbial DNA and RNA and the need to confirm findings of interest with Sanger sequencing (Beerenwinkel *et al.*, 2012; Marston *et al.*, 2013). Additionally, the costly computational and storage infrastructure required to setup next generation sequencing facilities is not readily available in all research settings, particularly those with limited resources (Behjati and Tarpey, 2013). With continual advancements in this field of detection technology, next generation sequencing will likely become increasingly more commonplace.

Although conventional PCR is limited by the requirement of a known target region that ultimately biases the detection of viruses, with viruses that have more diverse sequences likely to go undetected, it still provides a cost effective and robust method for conducting surveillance of viruses in bat populations (Yang and Rothman, 2004; Wang and Taubenberger, 2010). Furthermore advancements have led to the development of real-time PCR that allows for viral load quantification and has potential for high-throughput screening, while the development of muliplex PCR has allowed for the simultaneous detection of multiple targets using more than one pair of primers in a single reaction tube; additional methods combining real-time and multiplex PCR have further increased available methods for the surveillance of viruses (Wang and Taubenberger, 2010; Nguyen *et al.*, 2013; Waggoner *et al.*, 2013). In this study, conventional PCR assays were used to conduct surveillance of bat populations for CoVs.

19

2.6 Coronaviruses

2.6.1 A brief history of coronaviruses

During the 1930's a virus was identified that caused respiratory disease in young chickens (Schalk and Hawn, 1931; Bushnell and Brandly, 1933; Beach and Schalm, 1936). This virus was first cultured in 1937 and would later become known as infectious bronchitis virus (IBV), the first of many coronaviruses (CoV) to be identified (Beaudette and Hudson, 1937). Electron microscopy (EM) studies of IBV revealed enveloped pleiomorphic virions with "spike" surface projections that would become recognised as the characteristic morphology of CoVs (Reagan and Hauser, 1948; Reagan and Brueckner, 1952; Berry *et al.*, 1964).

A few years later during the 1950's, a virus causing liver disease in laboratory mice was identified (Gledhill and Andrewes, 1951; Jordan and Mirick, 1951; Nelson, 1952). Named murine hepatitis virus (MHV), it was the second CoV ever described and is still an important pathogen of laboratory mice (Tyrrell *et al.*, 1978; Baker, 1998).

During the 1960's, viruses that didn't fit known virus classification criteria were isolated from human patients with mild respiratory illness (Tyrrell and Bynoe, 1965; Hamre and Procknow, 1966; Tyrrell *et al.*, 1968). During EM studies it was found that two of these human viruses, isolates OC43 and 229E, resembled IBV (Hamre and Procknow, 1966; Almeida and Tyrrell, 1967; McIntosh *et al.*, 1967) being morphologically similar enveloped RNA viruses. These isolates were recognised as the first pathogenic human CoVs (HCoVs) and have formally been named HCoV-229E and HCoV-OC43. Subsequently the *Coronaviridae* virus family was proposed with IBV designated as the prototype virus (Tyrrell *et al.*, 1975; Fenner, 1976).

2.6.2 Coronavirus taxonomy and biology

CoVs belong to the *Coronavirinae* subfamily that along with the *Torovirinae* subfamily form the *Coronaviridae* family within the order *Nidovirales* that hosts positive sense single stranded RNA viruses with genomes that produce a nested set of subgenomic mRNAs during replication (González *et al.*, 2003). Other families within the order *Nidovirales* include the *Arteriviridae, Mesoniviridae,* and *Roniviridae* (ICTV, 2012; Adams *et al.*, 2016).

CoVs were originally classified by serology into three genera namely, Group 1; Group 2; and Group 3 viruses with Group 1 and 2 known to infect mammals and Group 3 known to infect birds (González *et al.*, 2003; ICTV, 2005; Fehr and Perlman, 2015). As the number of CoV full genome sequences increased, phylogenetic analyses served as a driver for the reclassification of CoV into the *Alpha-, Beta-,* and *GammaCoV* genera with the most recent addition of a fourth genus, *DeltaCoV* (Carstens, 2010; Adams and Carstens, 2012; ICTV, 2012).

Within the *Alphacoronavirus* genus are the human CoVs NL63 and 229E as well as several CoVs important to veterinary medicine such as the already mentioned PEDV, transmissible gastroenteritis virus (TGEV), and feline infectious peritonitis virus (FIPV). Additionally, there are several bat CoV in the *Alphacoronavirus* genus that were identified in *Miniopterus*, *Rhinolophus*, *Myotis*, and *Scotophilus* bat species (ICTV, 2012; Adams *et al.*, 2016).

The *Betacoronavirus* genus is further divided into four lineages, A to D, with four CoV known to cause disease in humans: HCoV-OC43 and HCoV-HKU1 in lineage A, SARS-CoV in lineage B, and the most recently identified human CoV, MERS-CoV in lineage C (ICTV, 2012; van Boheemen *et al.*, 2012). The *Betacoronavirus* genus also contains several CoV identified in bats belonging to the *Rousettus*, *Pipistrellus*, *Tylonycteris* bat genera (ICTV, 2012; Adams *et al.*, 2016).

The *Gamma-* and *Deltacoronavirus* generas mostly comprise avian CoV such as the already mentioned IBV and Turkey CoV with the *Gammacoronavirus* genus also containing the mammalian Beluga Whale CoV SW1 that is commonly used as an out group in phylogenetic analyses (ICTV, 2012; Woo *et al.*, 2012b; Adams *et al.*, 2016).

For CoVs to be assigned a genus and species, phylogenetic analyses of full genome sequences are required to assess evolutionary distances based on several family-wide conserved domains within ORF1ab encoding for nsp 5, 12, 13, 14, 15, and 16.

The phylogenetic relationship between different members of the *Coronavirinae* subfamily is depicted in Figure 2.3.
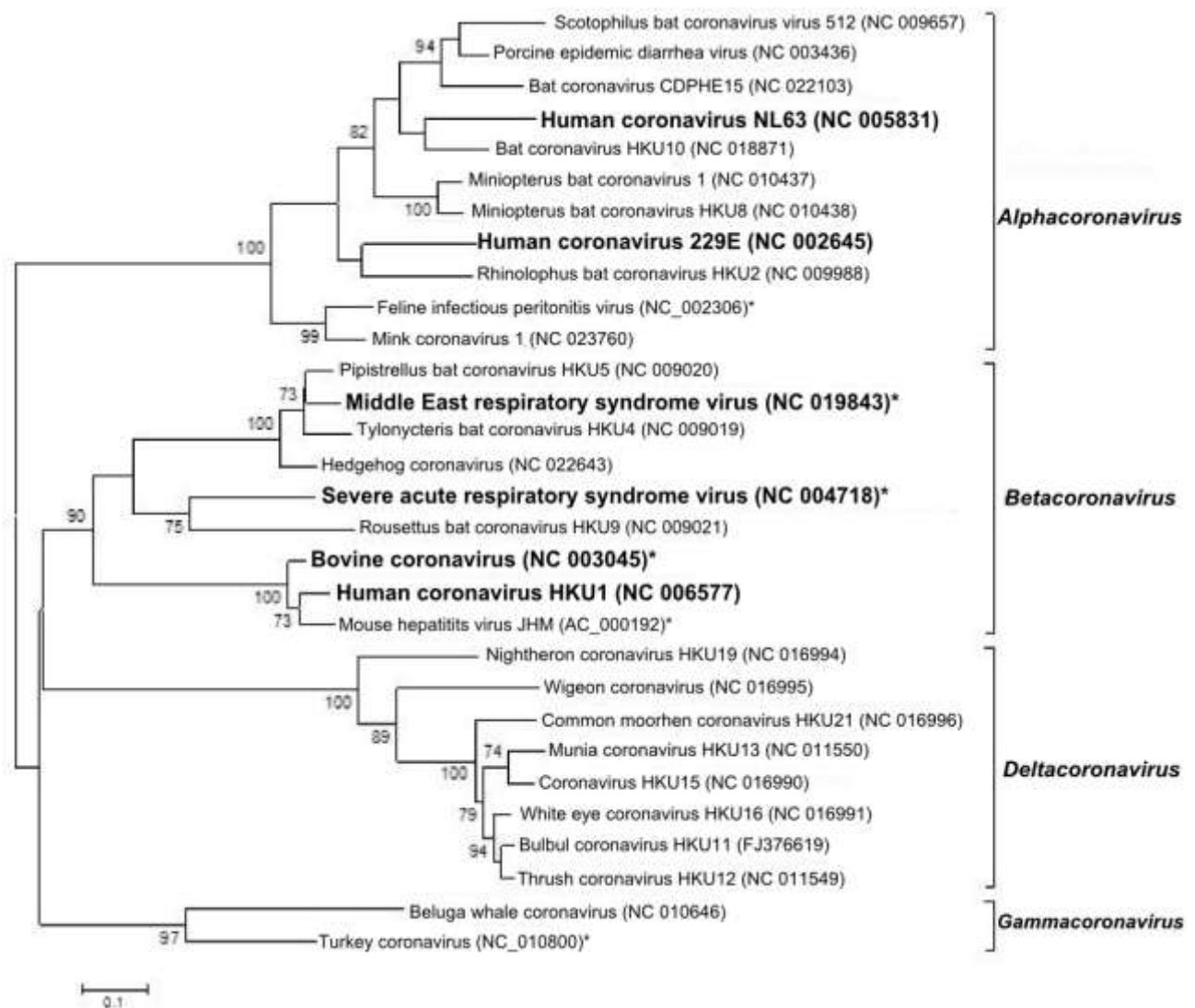
Figure 2.3 Coronavirus phylogeny of ICTV-recognised coronaviruses species. Phylogeny was inferred using the Maximum Likelihood method based on the Le Gascuel model with gamma distributed sites with some invariant site variation allowed (LG + G + I) in MEGA7 based on a partial RNA dependent RNA polymerase amino acid sequence corresponding to nucleotide positions 13898 – 14871 in the bat CoV HKU10 genome (Genbank ID NC_018871) (Le and Gascuel, 2008; Kumar *et al.*, 2016). Statistical support was achieved with 1000 bootstrap replicates; values less than 70% have been omitted from the figure. All gaps and missing data were deleted, there were a total of 272 amino acid positions in the final dataset and the tree was left unrooted. Coronavirus species known to infect humans are indicated in bold font. CoV sequences indicated with ✳ are representative of a CoV species: feline infectious peritonitis virus = Alphacoronavirus 1, bovine CoV = Betacoronavirus 1, SARS-CoV = SARS-related coronaviruses, MERS-CoV = MERS-related coronaviruses, mouse hepatitis virus JHM = murine coronaviruses, and turkey CoV = avian coronaviruses.

The CoV RNA genome along with its viral nucleocapsid protein (N) forms a helical nucleocapsid that is surrounded by a host cell-derived lipid envelope. CoV virions are pleomorphic to spherical with a diameter ranging from 120-160 nm; the virus-specific proteins, membrane glycoprotein (M); envelope protein (E); and spike glycoprotein (S) are embedded across the virion's surface forming large club-shaped 'spike' surface projections that give CoV virions their characteristic crown-like appearance (corona, Latin = crown) in electron micrographs. (Baker, 2008; Cavanagh and Britton, 2008; Peiris and Poon, 2009; ICTV, 2012).

The CoV genomes are the largest RNA genomes identified to date, ranging from 27 to 32 kb. The CoV genome is characterised by multiple open reading frames (ORFs) with two large overlapping ORFs, 1a and 1b, coding the replicase gene that occupies two-thirds of the genome at the 5' end (Baker, 2008; ICTV, 2012). ORFs are regions of DNA encoding for proteins that contain specific initialisation and termination signals for translation, namely start and stop codons. The two overlapping CoV ORFs are translated into polyproteins pp1a and pp1ab with ORF1b translated following a -1 ribosomal frameshift, a unique characteristic of CoV biology. For mammalian CoVs, polyproteins pp1a and pp1ab are post-translationally processed by proteinase cleavage with the papain-like and 3C-like proteases into sixteen mature non-structural proteins (nsp) (ICTV, 2012; van Boheemen *et al.*, 2012; de Wit *et al.*, 2016). The majority of nsp's are enzymes essential for CoV replication while others are required for host-virus interactions (ICTV, 2012). Another unique characteristic of the CoV genome is the encoding of an exoribonuclease function by nsp14 that provides proofreading capability allowing for the maintenance of the very large CoV genome without accumulating harmful mutations (Denison *et al.*, 2011). The remainder of the genome is transcribed into a nested set of subgenomic mRNAs (nido, Latin = nested) that contain a common 5' leader sequence attached to downstream gene sequences during discontinuous synthesis of the minus-sense subgenomic RNAs where subgenomic RNA synthesis is initiated at the 3' end of the genome and terminates once a transcription regulatory sequence (TRS) is encountered (Baker, 2008; Perlman and Netland, 2009; ICTV, 2012). These synthesised strands are then translocated to the complementary leader TRS at the 5' end of the genome through base-pairing interactions and transcription continues through the 5' end of the genome producing a chimeric minus-sense subgenomic RNA that serves as the template for synthesis of viral mRNA (Pasternak *et al.*, 2006; Baker, 2008; Perlman and Netland, 2009; ICTV, 2012; Fehr and Perlman, 2015; de Wit *et al.*, 2016).

The subgenomic mRNA encode the structural proteins and a number of genus-specific accessory proteins for which many the function remains unknown although some are thought to play a role in interfering with the host immune response (Yang *et al.*, 2013c; de Wit *et al.*, 2016). The genes encoding the structural proteins of CoV are located at the 3' end of the genome. The structural proteins order of S-E-M-N with a variable number of accessory proteins scattered in between depending on the respective CoV species; for a subset of betaCoVs, an additional structural

protein encoding hemagglutinin-esterase (HE) is also present (Baker, 2008; Woo *et al.*, 2012a, 2012b). The S protein consists of two subunits that are involved in receptor-binding and fusion respectively, the E protein plays an enhancing role in the budding of assembled virions, the integral M protein is largely involved in virus assembly, and the N protein exists in a complex configuration encapsulating the viral RNA genome.

The receptor-binding region of the S protein facilitates receptor recognition allowing CoV virions to bind to susceptible host cells; the S protein N-terminal end (subunit 1) mediates the receptor attachment while the C-terminal end (subunit 2) facilitates membrane fusion between the virion and the host cell allowing the CoV virion to enter the cell (Baker, 2008; Belouzard *et al.*, 2012). Following the binding of a CoV virion to its respective receptor via the S protein, the virion enters the cell via membrane fusion, releasing its RNA genome where it is uncoated and undergoes translation of the two large polyproteins and transcription of the subgenomic mRNAs in the cell cytoplasm (Cavanagh and Britton, 2008; de Wit *et al.*, 2016). Unlike other enveloped viruses that assemble at the plasma membrane, CoV assemble by budding into the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) (Cavanagh and Britton, 2008). Following transcription, M proteins are targeted towards the ERGIC by independent targeting signals where these proteins interact with N proteins before budding into the ERGIC lumen to form virions. Newly assembled virions are released from cells through the process of exocytosis (Hogue and Machamer, 2008). The genome structure and replication process of MERS and SARS-CoV are depicted in Figure 2.2.
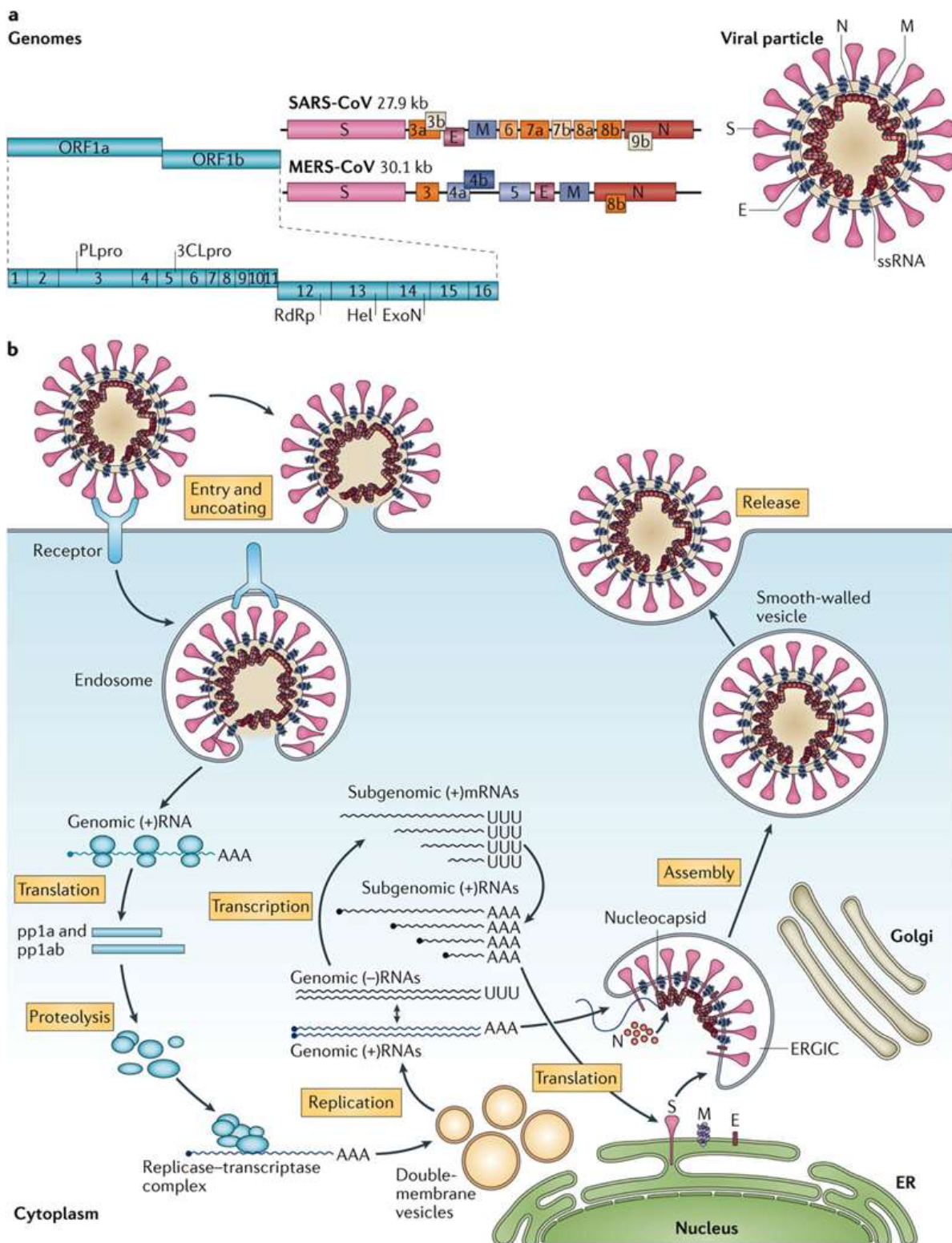
Figure 2.4 The genome organisation and replication process of SARS- and MERS-CoV.
Figure a The severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) genomes each encode two large polyproteins, pp1a and pp1ab, which are proteolytically cleaved into 16 non-structural proteins (nsps). Four structural proteins are encoded namely the envelope glycoprotein spike (S), transmembrane envelope (E) and membrane (M) proteins and the nucleocapsid (N) protein that encapsulates the viral RNA to form the helical nucleocapsid. Figure b Upon cell entry, the viral RNA is uncoated in the cytoplasm. ORF1a and ORF1ab are translated to produce pp1a and pp1ab followed by protolytic cleavage into 16 nsps that form the RNA replicase–transcriptase complex. This complex is found associated with intracellular membranes derived from the rough endoplasmic reticulum (ER) where it drives the production of negative-sense RNAs ((−)RNAs) through both replication and transcription. A subset of subgenomic RNAs is produced through discontinuous synthesis and transcribed into subgenomic (+)mRNAs. The resulting structural proteins are assembled into the nucleocapsid and viral envelope in the ER–Golgi intermediate compartment (ERGIC) lumen before being released from the cell via exocytosis (de Wit *et al.*, 2016). *Used with permission from the Nature Publishing Group.*

2.6.3 Coronaviruses as animal and human pathogens

CoVs have a broad host range, and with their ability to infect pigs, cows, chickens, dogs, and cats, early research efforts were more focussed on veterinary aspects rather than a possible threat to human health. Several CoVs cause disease resulting in economic losses due to detrimental effects on livestock populations. TEGV and PEDV are important viral pathogens of swine where they inflict gastroenteritis leading to substantial morbidity and mortality while the already mentioned IBV continues to affect commercial poultry populations (Doyle and Hutchings, 1946; Cook *et al.*, 2012; Lau and Chan, 2015; Lee, 2015; Promkuntod, 2015). PEDV, endemic in Asia and Europe, (re)emerged in the USA and Asia causing large-scale epidemics despite current vaccination schemes highlighting the importance of CoVs as animal and human pathogens (Lee, 2015). An enteric viral ailment of turkeys, blue comb disease, was first described in the 1950's (Tumlin *et al.*, 1957). Caused by Turkey CoV, it has caused considerable economic losses for the turkey production industry over the years (Panigrahy *et al.*, 1973; Ritchie *et al.*, 1973; Cavanagh, 2005).

In domestic and wild cats, feline CoV exists as two pathotypes namely feline enteric CoV (FEC) and the highly virulent feline infectious peritonitis virus (FIPV) that causes systemic and often deadly disease following the acquisition of mutations during a harmless persistent infection of FEC that changes its pathogenicity (Kipar and Meli, 2014; Tekes and Thiel, 2016). In dogs, canine CoVs traditionally caused mild gastrointestinal ailments but with an increasing number of reports of lethal canine CoV infection accompanied with both gastrointestinal and systemic disease, these viruses are now considered to be an EID of dogs (Licitra *et al.*, 2014; Priestnall *et al.*, 2014).

Bovine CoVs occur worldwide in cattle causing three distinct clinical syndromes namely calf diarrhoea, winter dysentery with haemorrhagic diarrhoea in adults, and respiratory disease in cattle of all ages (Saif, 2010). It is thought that a zoonotic transmission event between cattle and humans led to the emergence of HCoV-OC43 (Vijgen *et al.*, 2005, 2006).

As previously mentioned, the first CoV causing disease in humans to be identified were HCoV-229E and HCoV-OC43, and are associated with mild respiratory and enteric disease in young children, the elderly, and immunocompromised individuals. Until the emergence of SARS and MERS, these were the only two CoV known to infect humans, accounting for 15 to 30% of common cold cases (Cavanagh and Britton, 2008).

The third HCoV to be identified was SARS-CoV following an outbreak of severe acute respiratory disease in the Guangdong province in China during November 2002 (Drosten *et al.*, 2003; Ksiazek *et al.*, 2003; Rota *et al.*, 2003). SARS-CoV emerged as a highly pathogenic HCoV that rapidly spread through the human population via the respiratory route causing a total of 8096 cases and 774 fatalities (WHO, 2004; Cheng *et al.*, 2007). Following the identification of this highly pathogenic

CoV, interest in human CoV was piqued leading to the identification of two additional human CoV, HCoV-NL63 and HCoV-HKU1 (Berger *et al.*, 2004; van der Hoek *et al.*, 2004; Woo *et al.*, 2005b).

HCoV-NL63 was first isolated from an infant suffering from respiratory disease in the Netherlands. Since then it has been recognised as an important etiologic agent in respiratory infections worldwide (Fouchier *et al.*, 2004; van der Hoek *et al.*, 2004; Abdul-Rasool and Fielding, 2010). HCoV-HKU1 was isolated from adult patients in Hong Kong suffering from mild respiratory disease similar to that reported for HCoV-229E, OC43, and NL63 (Woo *et al.*, 2005a, 2005b).

The sixth and most recent HCoV to be identified, MERS-CoV, was isolated in 2012 from a patient with unexplained severe acute respiratory disease in Saudi Arabia (Bermingham *et al.*, 2012; Zaki *et al.*, 2012). MERS-CoV is associated with severe community-acquired pneumonia as index cases, with onward nosocomial infections accounting for a large proportion of total cases. The outbreak is ongoing and has been largely limited to the Arabian Peninsula, except for the recent outbreak with many cases in Korea (WHO, 2015, 2017). Isolated cases have been documented in other countries with infected individuals having a travel history to the Arabian Peninsula (Chowell *et al.*, 2015; WHO, 2017). The largest MERS-CoV outbreak outside of the Arabian Peninsula took place in Korea 2015 (WHO, 2015; Cho *et al.*, 2016).

HCoVs infect all age groups and apart from SARS- and MERS-CoV, often present with mild and subclinical disease with more severe lower respiratory tract infections in infants and the elderly and those with comorbidities or immunosuppression; a recent publication reported the first HCoV-229 associated death in a malnourished infant (Trombetta *et al.*, 2016; Konca *et al.*, 2017).

A zoonotic transmission event has been linked to the emergence of both SARS- and MERS-CoV, with bats implicated to have played a role in the evolutionary history of both these viruses. The emergence of SARS- and MERS-CoV is discussed in the following sections.

2.6.4 Emerging coronaviruses

Before the emergence of SARS- and MERS, CoV were considered to have narrow host ranges with greater importance as veterinary rather than human pathogens. The twenty-first century has seen the emergence of two highly pathogenic CoV affecting the human population, SARS- and MERS-CoV.

2.6.4.1 Severe acute respiratory syndrome coronavirus

SARS-CoV is a lineage B betaCoV that uses the angiotensin converting enzyme II (ACE2) cell receptor to infect humans. It caused severe acute respiratory syndrome (SARS) which first appeared in China in November 2002 and was soon followed by major outbreaks in mainland China and Hong Kong in 2003 (Drosten *et al.*, 2003; Fouchier *et al.*, 2003; Ksiazek *et al.*, 2003; Lee *et al.*, 2003; Zhong *et al.*, 2003). The global spread SARS-CoV was greatly assisted by travel

with cases reported in Vietnam, Canada, Singapore, Philippines, Australia, England, Sweden, France and several other countries (Breiman *et al.*, 2003; Hsu *et al.*, 2003; Parry, 2003; Poutanen *et al.*, 2003; Vu *et al.*, 2004; WHO, 2004). The enforcement of stringent infection control measures greatly curbed the spread of SARS-CoV and in July 2003 the pandemic was declared over after 8096 reported cases that resulted in 774 deaths (Hawkey, 2003; Seto *et al.*, 2003; WHO, 2004).

The behaviour of this pathogen was markedly different from previously recognised human CoVs and respiratory pathogens with genomic characterisation studies revealing SARS-CoV to represent a unique CoV only moderately related to other known CoVs (Hawkey, 2003; Marra *et al.*, 2003; Rota *et al.*, 2003).

The detection of highly similar sequences in palm civets and raccoon dogs at a live-animal market in China confirmed the notion that SARS-CoV emerged from an animal reservoir and through spillover events accompanied by mutational adaptive changes, gained the ability to infect humans (Guan *et al.*, 2003; Kan *et al.*, 2005). However subsequent studies indicated that although palm civets and raccoon dogs may have been the immediate source of the SARS-CoV outbreaks, they were not the natural reservoir host but instead served as amplifying intermediate hosts (Tu *et al.*, 2004; Kan *et al.*, 2005).

Analysis of patient sequences throughout the pandemic indicate that zoonotic transmission into the human population from palm civets and raccoon dogs was a rare event with human-to-human transmission, exacerbated by 'super-spreaders', and nosocomial infections, particularly before aerosol and fomites were identified as risks for the spread of SARS-CoV, accounting for the majority of reported cases (Booth *et al.*, 2003; Hawkey, 2003; Lee *et al.*, 2003; Zhao *et al.*, 2003; Chowell *et al.*, 2015).

A number of genetically diverse SARS-related CoVs have been detected in *Rhinolophus* sp. bats from China suggesting these mammals to be the natural reservoir host of SARS-like CoVs (Lau *et al.*, 2005; Li *et al.*, 2005; Ren *et al.*, 2006; Yuan *et al.*, 2010; Yang *et al.*, 2013b). However, with considerable differences within S1 indicating an inability for these SARS-like CoVs to use ACE2, these CoVs were found not to represent the immediate precursor to civet and human SARS-CoV. The recent reports however of SARS-like viruses that have experimentally demonstrated their ability to utilise the ACE2 receptor to infect human and other mammalian cell lines provide confirmatory evidence that SARS-CoV may have jumped the species barrier directly into humans from bats (Ge *et al.*, 2013; Menachery *et al.*, 2015; Yang *et al.*, 2015). Along with the detection of SARS-related CoVs in other countries such as Bulgaria, Taiwan, and Kenya, these studies demonstrate an existing potential for the re-emergence of SARS-CoV from *Rhinolophus* sp. bats (Tong *et al.*, 2009; Drexler *et al.*, 2010; Ge *et al.*, 2013; Wang and Hu, 2013; Menachery *et al.*, 2015, 2016; Chen *et al.*, 2016).

2.6.4.2 Middle East respiratory syndrome coronavirus

MERS-CoV, a lineage C betaCoV using the dipeptidyl peptidase 4 (DDP4) receptor, emerged in 2012 in the Arabian Peninsula. It caused disease similar to SARS but with a lower transmissibility within the human population (van Boheemen *et al.*, 2012; Zaki *et al.*, 2012; Hijawi *et al.*, 2013; Raj *et al.*, 2013). As with SARS-CoV, travel has facilitated the spread of the outbreak and similarly, nosocomial infections have accounted for the majority of cases. A limited number of cases of MERS-CoV infection have been reported from 27 countries including Germany, Egypt, the USA and the United Kingdom, with approximately 80% of cases reported from the Kingdom of Saudi Arabia (WHO, 2017). The largest outbreak of MERS outside of the Arabian Peninsula occurred in South Korea in 2015 with 186 laboratory confirmed cases including 36 deaths reported, following importation of the infection by an individual traveller returning from the Arabian Peninsula (WHO, 2015; Cho *et al.*, 2016). As of 6 July 2017, 2040 laboratory-confirmed cases of MERS-CoV infection with at least 712 related fatalities have been reported to the WHO from 27 countries (WHO, 2017). With serological and genomic studies providing strong evidence that dromedary camels are the source of human MERS-CoV infections, similar to the emergence of SARS-CoV, bats have been implicated as a likely ancestral host rather than a direct source of the virus (Reusken *et al.*, 2013, 2016; Azhar *et al.*, 2014; Meyer *et al.*, 2014).

A number of betaCoVs of bat-origin show relatedness to MERS-CoV confirming a likely ancestral reservoir host role for bats in the evolutionary process leading to the emergence of MERS-CoV; *Tylonycteris* BtCoV HKU4, identified in China, has been shown to utilise the DDP4 receptor, providing further evidence for bats as the natural reservoir hosts of MER-related CoVs (Wang *et al.*, 2014; Anthony *et al.*, 2017a). Although the genome of MERS-CoV showed relatedness to *Tylonycteris* HKU4 and *Pipistrellus* HKU5 BtCoVs, the sequences differed significantly with calculated pairwise evolutionary distances across several replicase domains less than 90% to known CoVs warranting the classification of MERS-CoV as a separate *Betacoronavirus* lineage C species (Woo *et al.*, 2007, 2009, 2012a; van Boheemen *et al.*, 2012; Adams *et al.*, 2016). The virus most closely related to MERS-CoV at the time of the outbreak was identified in a *Taphazous perforatus* bat trapped near the home of an infected patient; the sequence showed 100% nucleotide identity to a human MERS-CoV isolate but the finding was greatly limited by the short sequence of only 184 nucleotides detected in only one individual and no subsequent confirmatory studies (Memish *et al.*, 2013; Chan *et al.*, 2015). Other studies have reported lineage C betaCoVs in bats from Mexico, Thailand, Netherlands, Italy, and China, and even in hedgehogs (Woo *et al.*, 2007; Anthony *et al.*, 2013; Corman *et al.*, 2014b; Lelli *et al.*, 2013; Wacharapluesadee *et al.*, 2013; Reusken *et al.*, 2014). The closest full genome sequence available from a bat, sharing 85.6% nucleotide identity with human and camel MERS-CoV sequences, was detected in a *Neoromicia capensis* bat in South Africa (Ithete *et al.*, 2013; Corman *et al.*, 2014a). Recently, a closely related lineage C betaCoV sequence was obtained from a *Pipistrellus hesperidus* in Uganda strengthening

the hypothesis that vespertilionid bats may be the reservoir host for lineage C betaCoVs. Further investigations are required to understand their role in the emergence of MERS-CoV (Corman *et al.*, 2014a; Anthony *et al.*, 2017a).

2.6.5 Coronaviruses and bats

There are currently nine BtCoVs species recognised by the International Committee on Taxonomy of Viruses (ICTV). These include six alphaCoVs, BtCoV CDPHE15, BtCoV HKU10, *Miniopterus* BtCoV 1, *Miniopterus* BtCoV HKU8, *Rhinolophus* BtCoV HKU2, and *Scotophilus* BtCoV 512 and three betaCoVs, *Pipistrellus* BtCoV HKU5, *Rousettus* BtCoV HKU9, and *Tylonycteris* BtCoV HKU4. Additional BtCoV sequences have also been indicated to fall within the *SARS-* and *MERS-related CoV* species (ICTV, 2012; Ge *et al.*, 2013; Anthony *et al.*, 2017a).

The first publication regarding CoVs in bats was published in 2005, following the detection of SARS-CoV in Himalayan palm civets (*Paguma larvata*) and raccoon dogs (*Nyctereutes procyonoides*) that suggested an interspecies transmission event prior to its emergence in the human population, which prompted a surveillance study of wildlife animals in Hong Kong (Guan *et al.*, 2003; Poon *et al.*, 2005). Twelve different bat species were sampled with CoVs detected in three *Miniopterus* sp., namely *M. magnater*, *M. pusillus*, and *M. schreibersii,* while no CoVs were detected in any of the other wildlife species sampled that included mammals, reptiles, and birds (Poon *et al.*, 2005). The CoV sequences, identified as alphaCoVs and named BtCoV 1, detected across three different geographic regions and in three different *Miniopterus* sp., were highly similar indicating potential interspecies transmission between these three bat species (Poon *et al.*, 2005; Chu *et al.*, 2006). With no CoV sequences detected in cohabiting *Myotis* sp. bats, this study provided the first indication that the host range of BtCoVs may be restricted at the genus or species level (Poon *et al.*, 2005).

A subsequent surveillance study focussing on *Miniopterus* sp. bats in Hong Kong identified additional BtCoV 1 sequences that following phylogenetic analyses were found to be present as two different strains, named 1A and 1B (Chu *et al.*, 2006). Following full genome sequencing and characterisation, these viruses were assigned to the *Miniopterus* BtCoV 1 species (Chu *et al.*, 2006, 2008a; ICTV, 2012).

Another surveillance study conducted in Hong Kong at the time revealed the presence of several different BtCoVs in six different bat species (Woo *et al.*, 2006, 2007). The study detected novel alpha- and betaCoVs that would later be classified by the ICTV as novel CoV species. These included *Rhinolophus* BtCoV HKU2 and *Miniopterus* BtCoV HKU8 alphaCoVs detected in *R. sinicus* bats and *M. pusillus* bats and betaCoV belonging to the *Rousettus* BtCoV HKU9, *Tylonycteris* BtCoV HKU4 and *Pipistrellus* BtCoV HKU5 species detected in *Rousettus*

leschenaulti, *Tylonycteris pachypus* and *Pipistrellus abramus* bats respectively (Woo *et al.*, 2006, 2007; ICTV, 2012).

All the miniopterid-borne CoV sequences obtained during these studies appear to have descended from a common ancestor implying that these CoV may have evolved along with their hosts, resulting in a restricted host range (Chu *et al.*, 2006, 2008a; Woo *et al.*, 2006).

At the time of identifying BtCoV HKU10, this BtCoV was detected in two very different bat species, *Rousettus leschenaulti* in China and *Hipposideros pomona* in Hong Kong (Woo *et al.*, 2007; Lau *et al.*, 2012a). Full genome characterisation and phylogenetic analyses indicated that the CoV sequences obtained from these two different bat species were highly similar across the genome except in the S region where amino acid identities between the two sequences dropped to 60.5%. Despite this difference, according to CoV classification criteria these two sequences are representative of the same CoV species that likely infects different bat species (ICTV, 2012; Lau *et al.*, 2012a; Adams *et al.*, 2016). The differences in the S gene along with molecular clock analysis indicates a likely recent inter-species transmission event from *R. leschenaulti* to *H. pomona* with subsequent adaptation to the new host species (Lau *et al.*, 2012a). This publication further marked the first report of interspecies BtCoV transmission between bats belonging to different suborders (Lau *et al.*, 2012a).

Scotophilus BtCoV 512 was identified during a surveillance study in China where it was detected in *S. kuhli* bats commonly found in the region; its genome organization most closely resembles that of PEDV (Tang *et al.*, 2006). Subsequent genomic studies have indicated that *Scotophilus* BtCoV 512 may have played a role in the emergence of PEDV through a spillover event from bats to swine (Huang *et al.*, 2013). During the same surveillance study, BtCoV 133 was identified and although it has not been classified as a species on its own, this BtCoV represents one of the earliest identified lineage C betaCoVs (Tang *et al.*, 2006).

A recently recognised BtCoV species, BtCoV CDPHE15, was isolated from a *Myotis lucifugus* bat in the USA; there is currently little information available about this BtCoV but from its phylogeny (Figure 2.2) similarly to *Scotophilus* BtCoV 512, it appears to be associated with CoVs identified in domestic animals (Genbank ID NC_022103, unpublished study by Town *et al.,* 2013)(Adams *et al.*, 2016).

Bat surveillance studies have been conducted around the world including countries such as Australia, Cambodia, Thailand, Italy, Bulgaria, Brazil, Mexico, Japan, and China to name a few, and have revealed a great diversity of bat CoVs (Miura and Kitaoka, 1977; Carrington *et al.*, 2008; Drexler *et al.*, 2010; August *et al.*, 2012; Yang *et al.*, 2013b; Góes *et al.*, 2013; Lelli *et al.*, 2013; Lima *et al.*, 2013; Chen *et al.*, 2014; Wacharapluesadee *et al.*, 2015; Smith *et al.*, 2016; Lacroix *et al.*, 2017).

A study that conducted CoV surveillance in an abandoned mineshaft inhabited by at least six different bat species reported the detection of several alpha- and betaCoVs. Coinfection with different CoVs was reported in all bat species sampled as well as the detection of high interspecies host diversity with the same CoV species detected in different bat species and / or families (Ge *et al.*, 2016). With CoVs known to undergo recombination, reports such as this indicate that new CoVs are likely to emerge continually. With a plethora of alpha- and betaCoVs detected in bats and many still likely to be identified it is not surprising that bats are thought to be the ancestral reservoir host from which all mammalian CoVs emerged (Woo *et al.*, 2012b).

As mentioned, bats and their associated CoVs have been implicated in the emergence of both SARS- and MERS-CoV, two highly pathogenic HCoVs. Other studies have additionally found evidence for BtCoVs to have played a role in the emergence of both HCoV-229E and HCoV-NL63 (Pfefferle *et al.*, 2009; Huynh *et al.*, 2012; Corman *et al.*, 2015; Tao *et al.*, 2017).

A recent large-scale study to assess global CoV diversity patterns detected 91 different CoVs from 12333 bats (Anthony *et al.*, 2017b). From this study it was estimated that each bat species probably hosts 2.67 different CoV species indicating that with approximately 1200 recorded bat species there is an estimated 3204 different BtCoVs of which many have not been identified (Anthony *et al.*, 2017b). The findings from this study indicate that the diversity of BtCoVs is vast with many still to be identified and described and much more still to be learned about the ecology of bats and CoVs.

2.6.5.1 African bat coronavirus studies

To date, very little has been published on BtCoV from the African continent. The first study looking at evidence for CoVs in bats on the African continent was reported in 2007. This was a serology-based study that detected the presence of antibodies against SARS-CoV in both frugivorous and insectivorous bats, particularly *Rousettus aegyptiacus* and *Mops condylurus*, from South Africa and the Democratic Republic of Congo (Müller *et al.*, 2007). No evidence of viral nucleic acids was detected, likely due to the use of serum, and virus neutralisation tests were negative indicating that SARS-related CoVs from Africa may utilise a different cell receptor (Müller *et al.*, 2007).

The first bat CoV sequences from the African continent were published in 2009 by Tong *et al.* This surveillance study sampled 221 bats from Kenya. CoVs were detected in eleven different bat species with an overall prevalence of 18.5% (Tong *et al.*, 2009). The majority of CoV sequences were obtained from *Miniopterus* sp. bats, namely *M. africanus*, *M. inflatus*, *M. minor*, and *M. natalensis*. Interestingly, this study also reported the detection of CoVs in fruit bats, *Eidolon helvum* and *Rousettus aegyptiacus* (Tong *et al.*, 2009). Despite only obtaining 121 bp sequence fragments, this study provided the first sequence data for alpha- and betaCoVs including SARS-like CoVs in *Chaerephon* bats from the African continent (Tong *et al.*, 2009). Later work generated

extensive sequencing data for a subset of viruses identified during this study and revealed the presence of seven distinct CoVs of which three likely represent novel alphaCoVs and one a novel betaCoV (Tao *et al.*, 2012).

In the same year, a surveillance study of twelve different bat species in Ghana was published (Pfefferle *et al.*, 2009). Testing of 123 insectivorous bats resulted in an overall prevalence of 9.76%, nearly half that reported by Tong *et al.* (2009), with no positive frugivorous bat samples found, despite testing 212 faecal samples (Pfefferle *et al.*, 2009). All positive samples were obtained from *Hipposideros* sp. bats. Using primers designed on initial screening results, fragments of 817 bp and 1221 bp were obtained for phylogenetic analyses indicating the presence of both alpha- and betaCoVs. Of particular interest was the detection of bat alphaCoV sequences shown to share a common ancestor with HCoV-229E and the detection of SARS-like betaCoVs in a non-*Rhinolophus* bat species (Pfefferle *et al.*, 2009). Further investigation of a large sample of African bats provided conclusive genetic evidence that HCoV-229E has an evolutionary origin in *Hipposideros* sp. bats with animal and human viruses showing close genetic relatedness, ultimately forming a single viral species (Corman *et al.*, 2015).

HCoV-229E-related CoVs have additionally been detected in Kenya (Tao *et al.*, 2017; Waruhiu *et al.*, 2017). A recent study conducted over several years provided further evidence for HCoV-229E to be of bat origin and further suggested that HCoV-NL63 resulted from a recombination event between NL63-like CoVs in *Triaenops afer* bats and 229E-like viruses *Hipposideros* sp. bats with the breakpoint located in the S protein (Tao *et al.*, 2017).

In 2010, a report of a betaCoV detected in a *Hipposideros* bat from Nigeria was published (Quan *et al.*, 2010). Although found to be phylogenetically related to other SARS-like CoVs, the genome organisation of the virus detected in the study was unique with three overlapping ORFs between the M and N genes not previously observed leading to the proposal for a second betaCoV subgroup (Quan *et al.*, 2010). Phylogenetic analysis demonstrated a relationship between this virus, named Zaria BtCoV, and the SARS-like virus detected in *Hipposideros* bats in Ghana, forming a separate clade within the lineage B SARS-like betaCoVs (Pfefferle *et al.*, 2009; Quan *et al.*, 2010).

The first bat CoV sequences reported from South Africa were from alphaCoVs detected in *Neoromicia* sp., *Miniopterus* sp., and *Mops midas* bats (Geldenhuys *et al.*, 2013). The sequence from a *Miniopterus* sp. bat demonstrated phylogenetic relatedness to other *Miniopterus* sp.-borne alphaCoVs, the *Neoromicia* sp.-derived sequence showed phylogenetic relatedness to CoV sequences in *Nyctalus* sp. bats, and the sequence from a *M. midas* bat demonstrated phylogenetic relatedness to an alphaCoV sequence obtained from a *Chaerephon* sp. bat in Kenya (Dominguez *et al.*, 2007; Tong *et al.*, 2009; Drexler *et al.*, 2010; Reusken *et al.*, 2010; Geldenhuys *et al.*, 2013). This study provided the first evidence of bat CoV RNA in South African bat populations and with

three different alphaCoVs detected in three different geographic regions, namely Makhada in Limpopo, Taung in the North West, and Irene in the Gauteng province, provided the first evidence that BtCoVs in South Africa may be diverse and widely distributed (Geldenhuys *et al.*, 2013).

In 2013, alphaCoVs showing high similarity to those previously reported from South Africa were detected in bats sampled at Greyton and Table Mountain in the Western Cape and at Ndumo and Vryheid in KwaZulu-Natal (Geldenhuys *et al.*, 2013; Ithete, 2013; Ithete *et al.*, 2013). Of greater importance and interest was the detection of a betaCoV from a vespertilionid *Neoromicia capensis* bat sampled at Phinda in KwaZulu-Natal with high sequence identity to the pathogenic human MERS-CoV; this full genome sequence has since represented the closest known non-human and non-camel relative to MERS-CoV (Ithete *et al.*, 2013; Corman *et al.*, 2014a). Additional phylogenetic analysis of the full NeoCoV genome revealed marked differences in subunit 1 of the spike gene containing the receptor-binding domain, indicating that NeoCoV does not represent the direct ancestor of MERS-CoV and that the pathogenic human MERS-CoV has likely resulted from a non-recent recombination event between as yet unknown role players, a similar scenario to the emergence of SARS-CoV (Corman *et al.*, 2014a). With NeoCoV gene sequences holding a basal sister relationship across all open reading frames (ORFs) phylogenetically analysed, it was proposed that NeoCoV could be used to infer the root of the MERS-related CoV phylogeny (Corman *et al.*, 2014a). A recent study has reported the detection of a betaCoV in a *Pipistrellus hesperidus* bat from Uganda that shows close genetic relatedness to NeoCoV and human MERS-CoV, further evidence supporting an ancestral origin for MERS-CoV in bats (Anthony *et al.*, 2017a).

2.6.6 Ecology of coronaviruses in bats

With environmental and life history traits thought to be useful in inferring predictors of viral richness, understanding the ecology of CoVs and how it relates to extrinsic biogeographical factors such has habitat type, rainfall, or altitude, and intrinsic life history factors such as age, reproductive state, or body condition, could allow for improved surveillance efforts if factors predicting CoV positivity were known (Turmelle and Olival, 2009).

A recent large-scale study investigating the global diversity of CoVs identified a number of predicting factors of CoV positivity, concluding that CoV diversity in bats is primarily driven by the host's ecology (Anthony *et al.*, 2017b). As commonly reported, sample types containing faecal material i.e. faecal pellets or faecal swabs, were significantly more likely to produce a positive result than other sample types (Lau *et al.*, 2005, 2007; Pfefferle *et al.*, 2009; Fischer *et al.*, 2016; Anthony *et al.*, 2017b). The study additionally found bat species to be a significant predictor with Pteropodidae bats in Africa more likely to be CoV positive, and for the subset of samples from Asia, Pteropodidae, Miniopteridae, and Vespertilionidae bats showed significant association with CoV detection.

The study additionally found age to be a predictor of CoV positivity, with juvenile bats more likely to be CoV positive than adult individuals (Anthony *et al.*, 2017b). This finding has been reported elsewhere in the literature and is likely a reflection of infection dynamics within bat colonies where viral shedding patterns are likely associated with birth pulses and young pups presenting as a susceptible population for virus transmission leading to acute infections (Gloza-Rausch *et al.*, 2008; Drexler *et al.*, 2011; August *et al.*, 2012; Annan *et al.*, 2013; Plowright *et al.*, 2014). Similarly, sex and reproductive status have also been indicated as predictors of virus infection in bats. In the global bat meta-study, male bats from Africa were more likely to be CoV positive than female bats (Anthony *et al.*, 2017b). Pregnant or lactating bats are more prone to virus infection not only for CoVs but also for other viruses such as the Hendra virus and a recently identified novel bocaparvovirus (Plowright *et al.*, 2008; Annan *et al.*, 2013; Lau *et al.*, 2016a).

Most bat studies report collecting samples from apparently healthy bats. Despite detecting a number of different viruses in these mammals, it appears that infection does not cause overt disease. Several studies have reported on using measures of body condition, such as weight, to investigate whether or not viruses have a negative impact on their host (Lau *et al.*, 2010, 2012a; Maganga *et al.*, 2014b; Seltmann *et al.*, 2017). The detection of BtCoV HKU10 and SARS-related BtCoVs in *Hipposideros pomona* and *Rhinolophus sinicus* bats respectively was associated with lower body weights (Lau *et al.*, 2010, 2012a). A study comparing shedding patterns of astroviruses and CoVs found low body weight to be associated with astrovirus infection but not with CoV infection (Seltmann *et al.*, 2017).

A number of geographic and environmental factors, such as temperature, humidity, rainfall, and elevation, have been noted to be predictors of positivity for other viruses and their animal hosts, such as influenza in birds (Si *et al.*, 2013; Magee *et al.*, 2014). Seasonality has been identified as a possible predictor of virus positivity in bats. Recent studies suggested that sampling of bats during the drier season were more likely to yield CoV positive results (Lau *et al.*, 2010; Anthony *et al.*, 2017b). However, a study investigating astroviruses and CoVs found astroviruses to have a seasonal association but not CoVs (Seltmann *et al.*, 2017). Seasonality likely corresponds to the host's reproductive cycle, dependent on when mating and birthing takes place (Plowright *et al.*, 2008; Lau *et al.*, 2010; Drexler *et al.*, 2011).

Habitat type has not been widely investigated as a possible predictor of CoV infection. The recent global study on CoV diversity patterns compared different sampling areas according to broadly defined animal-human interface categories and found that bats sampled in areas where animal use, such as markets, hunting, trade, zoos, sanctuaries, handling by veterinarians or researchers, and wild animal farms, occurred frequently, were more likely to be CoV positive (Anthony *et al.*, 2017b). A study assessing the effect of habitat fragmentation on the general viral richness of bats found that host distribution size and shape significantly influences viral richness where larger

distribution areas and fragmentation of the bat host species were associated with increased viral richness (Maganga *et al.*, 2014b).

Biodiversity loss has been linked to the emergence of new diseases when host diversity decreases and a dominant species capable of amplifying the virus thrives leading to increased disease transmission; it remains important to consider however that areas with high biodiversity may serve as a pool of pathogens from which new zoonotic viruses could emerge (Ostfeld, 2009; Keesing *et al.*, 2010). Biodiversity richness is one of South Africa's greatest natural assets and results from several ecosystems that roughly translate into nine biomes. Biomes are generally associated with regions defined on climate patterns as moisture and temperature strongly dictate vegetation establishment and survival leading to characteristic groupings of plant and animal species accompanied by a general appearance based on shape of vegetation and landscape features, found in the region (Rutherford *et al.*, 2006).

The nine biomes of South Africa are Grassland, Savanna, Forest, Albany Thicket, Indian Ocean Coastal Belt, Fynbos, Desert, Nama-Karoo, and Succulent Karoo (Mucina *et al.*, 2006; Rutherford *et al.*, 2006). The Grassland and Savanna biomes are both characterised by a summer rainfall pattern and are identified by a grassy ground layer. Where the Savanna biome is further characterised by the presence of a distinct upper layer of woody plants, trees are absent in the Grassland biome (Rutherford *et al.*, 2006). The Fynbos biome is characterised by two vegetation types namely, Fynbos and Renosterveld. Together with the Succulent Karoo biome that represents a species-rich semidesert, these biomes are characterised by a winter rainfall pattern and have a notably rich floral diversity (Linder and Hardy, 2004; Rutherford *et al.*, 2006). The South African Forest biome is the smallest biome in the region and despite being highly fragmented, these forest have some of the highest tree species richness of all temperate forests worldwide (Silander, 2001). An extant of coastal enclaves of forests represents the Indian Ocean Coastal Belt represents a unique biome of South Africa (Rutherford *et al.*, 2006). The Albany Thicket biome consists of mixed transitional vegetation types usually found in the Savanna, Nama-Karoo, and Forest biomes creating an unusual environment with diverse vegetation types. In contrast the Nama-Karoo and desert biomes have a considerably lower species-richness of all the biomes (Rutherford *et al.*, 2006).

With great biodiversity and a changing landscape due to growing population numbers coupled with a limited body of literature on bat viruses, South Africa represents an understudied region in the surveillance effort of bat viruses and understanding the host-virus ecology that might influence their emergence.

Chapter 3: Materials and Methods

3.1 Ethics and Permits

The necessary ethical clearance and permits for handling and sampling bats were obtained from all relevant authorities as detailed below.

3.1.1 Ethics

The nature of this research study relied solely on the use of non-invasively collected animal-derived material. The necessary ethical clearance was obtained from Stellenbosch University's Animal Care and Use committee (Appendix A).

2012 - 2015 (PI: Prof. Wolfgang Preiser) Stellenbosch University Research Ethics Committee: Animal Care and Use SU-ACUM12-0001: Catching and non-invasive sampling of bats. Renewed annually by submission of a project report.

2016 – 2017 (PI: Dr Ndapewa Ithete) Stellenbosch University Research Ethics Committee: Animal Care and Use SU-ACUD16-00008: Investigation of novel bat-borne viruses with zoonotic potential in South Africa

3.1.2 Permits

The non-invasive collection of faecal pellets from bats in their natural habitat between January 2014 and October 2016 was conducted in collaboration with co-supervisor on this study, Prof. Corrie Schoeman from the University of KwaZulu-Natal, and Dr Leigh Richards, from the Durban Natural Sciences Museum, along with their own research teams.

Bat faecal pellets collected by Prof. Corrie Schoeman and his team were obtained under collection permits for the Western Cape: Cape Nature (0056-AAA041-00079, 0056-AAA041-00091, 0056-AAA041-00135), Eastern Cape: Department of Economic Affairs, Environment, and Tourism (CRO 163/13CR, CRO 164/13CR, CRO 15/15CR, CRO 16/15CR), Northern Cape: Department of Environment and Nature Conservation (FAUNA 1541/2014), KwaZulu-Natal: Ezemvelo KZN Wildlife (OP 4189/2013, OP 3899/2015), and Limpopo: samples collected on private land with permission from the landowners.

Bat faecal pellets collected by Dr Leigh Richards and her team were provided by the Durban Natural History Museum. The provided samples were sourced ethically and under collecting permits issued for the Northern Cape: Department of Environment and Nature Conservation (FAUNA 1578/2015), and KwaZulu-Natal: Ezemvelo KZN Wildlife (OP4361/2015).

3.2 Selection of bat trapping sites

Sampling across different environments enables the sampling of a higher richness of bat species, important for general surveillance (Lourenço *et al.*, 2010). For this study, general surveillance for a diverse range of coronaviruses (CoVs) from different bat species was conducted in the four provinces: KwaZulu-Natal, Eastern Cape, Western Cape, and Northern Cape. During this study, opportunities arose to additionally sample sites in the Limpopo province; a subset of these samples was included in this study.

Neoromicia capensis bats were selected as a target species for a species-specific sampling effort to assess the effect of broad scale biogeography on CoVs in bats. The distribution of *N. capensis* is vast and may be the most widespread of all bat species in Southern Africa; its southern distribution covers most of South Africa (Jacobs *et al.*, 2008; Monadjem *et al.*, 2010). This geographic range lends itself particularly well to a broad range sampling effort. Furthermore, the recent identification of a MERS-related betaCoV in this bat species, along with several alphaCoVs, highlighted *N. capensis* as a species of great interest for investigating the diversity of CoVs in South Africa at the bat species-level (Ithete, 2013; Ithete *et al.*, 2013; Corman *et al.*, 2014a).

Neoromicia capensis bats were sampled from populations located across much of its distribution range in South Africa across several major biomes in the region. These biomes included the Nama-Karoo, Fynbos, Forest, Albany Thicket, Grassland, Indian Ocean Coastal Belt, and Savanna. The characteristics of South African biomes were described in Chapter 2 Section 2.5.7. Figure 3.1 depicts the distribution of *N. capensis* across the different provinces and biomes of South Africa.

In an attempt to study CoV infection dynamics longitudinally, a *N. capensis* colony with CoV infected individuals was identified on a farm, Cloeteskraal, in the Western Cape.

1A

1B
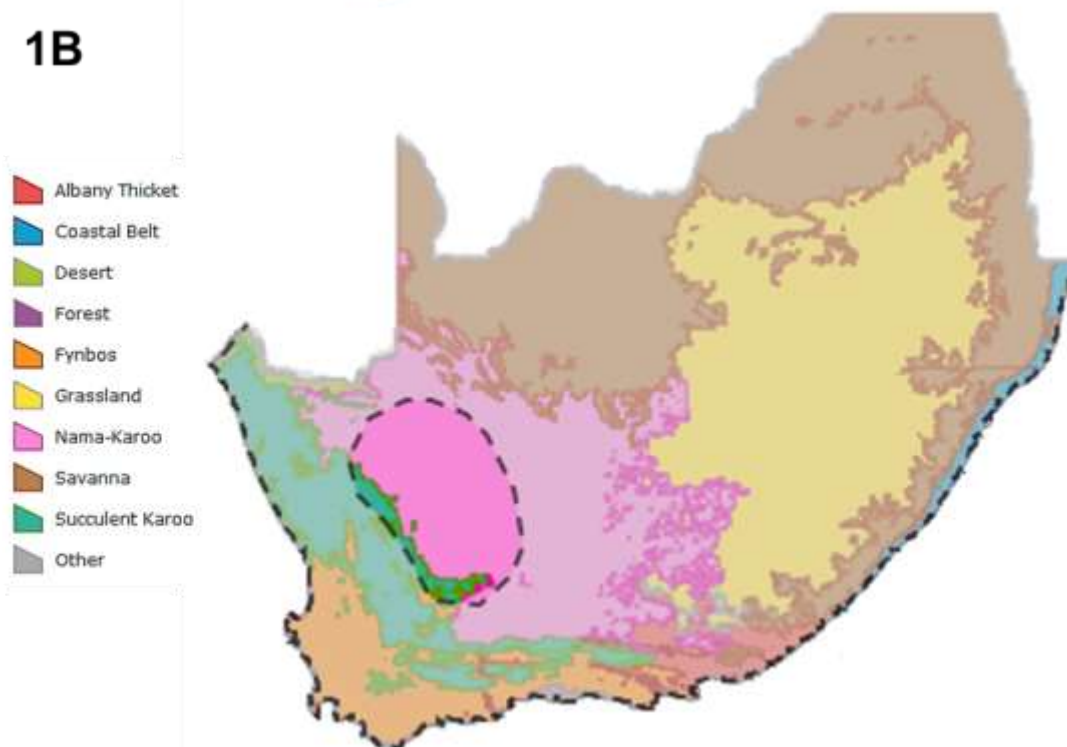
Figure 3.1 *Neoromicia capensis* distribution across South African provinces and biomes. Figure 1A depicts the distribution of *N. capensis* across the provinces of South Africa. Provinces where bat trapping took place have been labelled. Figure 1B depicts the distribution of *N. capensis* bats across the different biomes of South Africa with each biome represented by a different colour as indicated in the legend. In each figure the distribution of *N. capensis* bats has been indicated by the area between the country perimeter and the oblong area demarcated by the dotted line, i.e. the entire country excluding the oblong area situatredin the Northern Cape. The maps were created using the ArcGIS® software online tool by Esri ©1995-2017 (www.esri.com/arcgis). The *N. capensis* distribution data was obtained from the IUCN Red List (www.iucnredlist.org), the biome distribution data was obtained from SANBI (www.bgis.sanbi.org), and the geographic information system data for the provincial base map was obtained from Map Library (www.maplibrary.org) (Mucina et al., 2006; Rutherford et al., 2006; Jacobs et al., 2008; IUCN, 2012)**.**

3.3 Bat trapping

Bats were captured humanely using traps designed to hold the animals without causing any harm or excessive stress (Sikes and Gannon, 2011). All efforts were made to ensure captured bats were protected from predation and stress by placing individuals in separate cloth bags hung in sheltered and well-ventilated areas in sight of the field team. Processing was carried out in the shortest amount of time possible.

Appropriate biosafety precautions were taken during all bat trapping events. All individuals handling bats wore appropriate protective equipment such as gloves, were vaccinated against rabies, and received bat handling training to minimise risk of bites and scratches.

Traps were generally set up at dusk and checked regularly until one hour after the moon had fully risen. At other times, harp traps were left overnight and checked regularly before taking them down at dawn (Kunz *et al.*, 2009; Sikes and Gannon, 2011).

Two methods of capturing bats are widely used, mist nets and harp traps. There is an apparent variation in the susceptibility of certain species to be captured in either a mist net or harp trap, and along with differences in situational suitability, the selection of capture method is an important consideration in the field (Francis, 1989; Kunz *et al.*, 1996; Larsen, 2007). It has been indicated that in general, harp traps may be more efficient for bat sampling but factors such as surrounding vegetation and structures may lead to a preference for mist nets (Francis, 1989). During this study mist nets, harp traps, and hand nets were used as terrain, environment, and field experience dictated.

3.3.1 Mist nets

Mist nets are manufactured from fine nylon netting and are usually held taught between two poles with shelves of netting that have loose pockets for holding trapped bats (Mitchell-Jones, 2004). Mist netting is widely used and is often the preferred method of capture due to its high success rate (Kunz *et al.*, 1996; Marques *et al.*, 2013). During this study 6m, 9m, and 12 m mist nets (Ecotone, Poland) with mesh size of 20 mm were used.

A point of concern when using mist nets is the risk of predation during monitoring intervals and the potential for harming individuals during extrication following their entanglement in the net (Mitchell-Jones, 2004). Additionally, insectivorous bats have small teeth with sharp cutting edges that enable them to chew threw nets with ease and escape if not removed from the net shortly after capture, affirming the need to regularly check nets during sampling efforts (Francis, 1989; Sikes and Gannon, 2011). To minimise any risk of escape, mortality, or injury to captured bats, erected mist nets were monitored at regular intervals, every ten to fifteen minutes during the time period of high bat activity, in accordance with good sampling practice (Sikes and Gannon, 2011).

The repeated use of mist nets in the same locality can lead to net avoidance by bats. This can affect sampling by reducing both the number of bats sampled and the number of species detected (Marques *et al.,* 2013). During fieldtrips, the same site was never sampled more than twice consecutively. For repeated sampling of sites across different seasons, net avoidance is not an apparent issue as it has been indicated that bats forget the location of nets with a recommended interval of three weeks (Marques *et al.*, 2013).

3.3.2 Harp trap

The harp trap consists of a frame supporting banks of carefully spaced and tensioned vertically strung fine lines into which bats fly and slide down into a cloth collection bag attached at the base (Kunz *et al.*, 1996; Mitchell-Jones, 2004). The use of a harp trap is well-suited for situations where either the target bats are at high density or where they are restricted to narrow flyways such as outside the entrance of a cave or mine or in an overgrown forest path (Mitchell-Jones, 2004).

Without the possibility of entanglement, this trapping method minimises the risk of harm to the individual bats (Mitchell-Jones, 2004). Due to the terrain and environment at bat trapping sites, these traps were used less frequently than mist nets and when used, were regularly checked and guarded against predators. During this study, a three-bank harp trap was used (Faunatech, Australia).

3.3.3 Hand nets

Round circular nets with fine mesh were used at times to capture bats roosting in low hanging roofs or during day time collections from mines or caves when most individuals were in torpor. Nets were used statically i.e. not swayed to catch bats in flight, to avoid any harm to individuals (Kunz *et al.*, 1996; Sikes and Gannon, 2011). Care was taken to avoid any noise or disturbance to prevent early emergence of roosting bats (Mitchell-Jones, 2004). Nets were held just below the exit of the roosting site to catch individuals emerging at dusk or were used to collect individuals in torpor from the roof of caves or mines.

3.4 Collection of morphological, physiological and biogeographic data

During each bat trapping event, individual bats were examined externally to obtain morphological and physiological data. To allow for biogeographic analyses, data on the surrounding environment at sampling sites was collected as described in the following sections.

3.4.1 Physical measurements and physiological assessments

To allow for morphometric-based identification of bats captured in the field, various measurements, such as weight and forearm length, were taken as per widely used protocols (Parnaby, 1999; Mitchell-Jones, 2004; Monadjem *et al.*, 2013). These measurements along with a number of other characteristics, such as pelage and ear shape, used in conjunction with available species reference data and a dichotomous key, usually enabled accurate species identification in the field (Monadjem *et al.*, 2010). The species of a number of specimens, including all that were found to be CoV positive, was additionally confirmed by molecular methods (Section 3.8.5).

3.4.1.1 Weight and forearm length

The weight of each bat was determined by weighing the bat while still in the cotton cloth bag using a spring scale (PESOLA, Switzerland) to the closest 0.5 g and then subtracting the weight of the empty cotton cloth bag following the removal of the bat for handling and sample collection.

Forearm length is widely used as a proxy of body size. Using a caliper tool, as depicted in Figure 3.2, the maximum forearm length to the nearest 0.5 mm was measured for each bat as the distance between the elbow and wrist with the wing in a folded position (Parnaby, 1999; Mitchell-Jones, 2004; Schoeman and Jacobs, 2008).

Using the mass and forearm length values, the forearm mass index (FMI) for each bat was determined. FMI, similar to body mass index (BMI) but using the forearm length instead of the full body length, was used as a measure of body condition (Meng *et al.*, 2016). It was calculated using the formula:

$$FMI = \frac{mass\ (kg)}{[forearm\ length\ (m)]^2}$$

Figure 3.2 Visual depiction of a forearm measurement being taken. The image depicts the forearm of a *Miniopterus natalensis* bat being measured with a caliper tool. © N Cronjé 2016.

3.4.1.2 Age

Age, demarcated as adult or juvenile was determined by transilluminating the wing membrane with a headlight to examine the epiphyseal cartilage of the fourth metacarpal-phalangeal joint. Ossification is an indication of aging with joints becoming more knuckle-like as cartilage is replaced with bone, and this criterion has been widely used to provide reliable age estimation (Kunz and Anthony, 1982; Parnaby, 1999; Mitchell-Jones, 2004; Brunet-Rossinni and S., 2009).

3.4.1.3 Sex and reproductive status

The sex and reproductive state of each captured individual was recorded when possible. Male bats were identified by the presence of a visible penis while female bats were identified by the presence of a pair of mammary glands and nipples (Mitchell-Jones, 2004). Determining the reproductive status of bats is a difficult task. With bats generally considered seasonal breeders, examining the developmental state of the testes can provide an indication of a male's reproductive status. Based on the position of the testes, either descended into the scrotum or not, male bats were identified

either as scrotal, indicating that individuals were sexually active at the time of sampling, or non-scrotal. Female bats were identified as non-pregnant, pregnant, lactating, or post-lactating with lactation identified by the presence of enlarged nipples often with fur missing in the surrounding area due to suckling and post-lactation identified by darkened and keratinised nipples (Haarsma, 2008; Racey, 2009).

3.4.1.4 Other data

Wing punches (wing membrane biopsies) were taken with a sterilised 3 mm biopsy punch, of each bat as part of a different study our collaborators are undertaking and stored in ethanol. Although not relevant to this study, these wing punches facilitated the useful identification of recaptures during the sampling period. Wing punches cause no ill-effect to the bat and usually heal within three weeks (Mitchell-Jones, 2004).

3.4.2 Geographic and weather data

Geographic details regarding each bat trapping site were documented. The global positioning system (GPS) co-ordinates and altitude of each sampling site were recorded. The biome was determined using the GPS co-ordinate of each bat trapping site and the South African National Biodiversity Institute's (SANBI) 2011 biome delineation map (http://bgis.sanbi.org/) based on the National Vegetation map (Mucina *et al.*, 2006; Rutherford *et al.*, 2006).

Weather data was obtained from the South African Weather Service (www.weathersa.co.za) for all bat trapping sites corresponding to the month during which sampling took place. The month's maximum and minimum temperatures, and total monthly rainfall was obtained from the nearest weather station where data was available.

3.4.3 Collection of faecal pellets

Faecal pellets, when obtainable, were either collected directly from the bat or from the cotton bag in which the bat was placed during processing. Collected pellets were placed in cryovials (Greiner Bio-One) containing 1 ml of RNAlater™ (Ambion, USA) and labelled with the corresponding sample identifier. The sample identifier used detailed the date the sample was obtained, bat trapping site code as designated by the field team, species code per in-field species identification, and sample number per the species sampled. For example, the sample identifier **20140923_LNR_NC3** details that on 23 September 2014 at the bat trapping site Lotheni Nature Reserve, this sample was obtained from the third *N. capensis* bat captured.

For the longitudinal study of a *N. capensis* CoV-infected colony, pooled faecal pellet samples were collected monthly. Once a month, a plastic sheet was placed underneath the roosting colony at

dusk. The next morning ten pooled samples each consisting of five faecal pellets in RNAlater™ were collected for processing.

3.5 Extraction of nucleic acids

CoVs are single stranded RNA viruses. To detect their presence in sample material by molecular biology techniques, high quality and relatively pure viral RNA is required. Successful nucleic acid extraction relies on the effective disruption of virus-containing cells followed by the denaturation of nucleoprotein complexes while inactivating nucleases and preventing contamination (Tan and Yiap, 2009).

Commercially available kits for the extraction of viral RNA from cell-free biological fluids were used either manually or on the QIAcube (Qiagen, Germany) automated platform with cell-free supernatant from homogenised faecal pellet samples. The selection between manual and automated protocols was dependent on the availability of the QIAcube, with the automated platform the preferred protocol of choice due to its higher throughput capabilities, efficiency, and minimised effect of human error. Conventional manual nucleic acid extraction protocols are susceptible to contamination and inter / intra operator variability (Kim *et al.*, 2014). The QIAcube system is not fully automated in that some degree of sample preparation is required before the automated protocol is followed. This enables the QIAcube to be more versatile as almost all manual extraction kits available from Qiagen and Machery-Nagel (MN, Germany) can be adapted for use on the QIAcube platform regardless of sample type (Germi *et al.*, 2012).

Nucleic acid extraction kits used during this study were all spin-column based. Spin-column based extraction methods work on the principle that nucleic acids, once released from lysed cells will bind to a silica membrane with the appropriate pH allowing proteins, salts, and other contaminants to be washed away. Nucleic acids are easily eluted from the silica membrane with a low salt buffer or water.

The NucleoSpin® RNA Virus (Machery-Nagel, Germany) or QIAamp® Viral RNA Mini (Qiagen, Germany) extraction kit was used for the extraction of nucleic acids from faecal pellets, following manufacturer's protocols.

Regardless of which kit or protocol was followed, DNase treatment was not carried out due to collected sample material not being for the exclusive use of this project. Additionally, omitting the DNase treatment step permitted extracted nucleic acid material to be used directly in PCR assays targeting host DNA for species-confirmation by molecular methods. Carrier RNA was used for the extraction of viral RNA from all sample types as it enhances the yield of target molecules and minimises the likelihood of RNA degradation (Kim *et al.*, 2014).

3.5.1 Homogenisation of faecal pellets

Before the nucleic acid extraction protocol was carried out, faecal pellet material was homogenised to disrupt all cells in the sample. For each sample four 2.4 mm RNase & DNase free metal beads (Omni International, USA) were placed in a 2 ml centrifuge tube (Greiner Bio-One, Austria) containing 600 µl of 1× phosphate buffered saline (PBS) (Lonsa, USA) to which a single faecal pellet was added. Cell disruption was carried out on the TissueLyser LT (Qiagen, Germany) for 7 minutes at 50 Hz before centrifugation at 11000 × g for 2 minutes to pellet the debris. For pooled faecal pellet samples, the protocol above was followed with the minor adjustment that five pellets were homogenised in 1 ml PBS.

As starting material, 140 - 150 µl of cell-free supernatant was used for extraction either on the QIAcube or manually according to the manufacturer's instructions. Remaining sample homogenate was stored at -80 °C for future use. For both nucleic acid extraction kits, the extracted nucleic acids were eluted in 50 µl, aliquoted into single-use volumes, and stored at -80°C until needed.

3.5.2 NucleoSpin® RNA Virus kit

To the initial 150 µl starting sample volume, 600 µl of the provided lysis buffer containing carrier RNA was added, well mixed and incubated at 56 °C for 5 minutes. Before stepwise loading of the NucleoSpin® RNA Virus Column, 600 µl 100% ethanol (Sigma Aldrich, Germany) was added and vortexed. Binding of the nucleic acids was achieved by centrifuging the spin-column for 1 minute at 8000 × g and discarding the flow-through. The bound nucleic acids were washed by adding 500 µl of the provided wash buffer, RAW, to the spin-column and centrifuging for 1 minute at 8000 × g before discarding the flow-through. A second wash step with 600 µl of RAV3 followed before a final wash step with 200 µl RAV3 and centrifugation for 5 minutes at 11000 × g. To remove residual ethanol and to dry the silica membrane before elution of the nucleic acids, the spin-column was placed in a clean collection tube and centrifuged for 5 minutes at 11000 x g. Viral nucleic acids were eluted by adding 50 µl nuclease free water pre-heated to 70 °C, incubating at room temperature for 2 minute and centrifuging for 1 minute at 11000 × g.

3.5.3 QIAamp® Viral RNA Mini extraction kit

To the initial 140 µl starting sample volume, 560 µl of the provided lysis buffer containing carrier RNA was added, well mixed and incubated at room temperature for 10 minutes. To this, 560 µl 100% ethanol was added before vortexing for 15 seconds. The QIAamp® Mini Column was loaded step-wise starting with 630 µl of solution and centrifuging for 1 minute at 6000 × g before discarding the flow-through. The wash steps followed with the addition of 500 µl buffer AW1 and centrifugation at 6000 × g for 1 minute before discarding the flow-through. A second wash step followed with 500 µl buffer AW2 and centrifugation at 20000 × g for 3 minutes. Before elution of the

viral nucleic acids, the spin-column was placed in a clean collection tube and centrifuged for 1 minute at 20000 × *g*. Nucleic acids were eluted with 50 µl of buffer AVE following a 1 minute incubation at room temperature for 1 minute before centrifuging at 6000 × *g* for 1 minute.

To assess the quality of extracted nucleic acids, randomly selected samples from each batch were subjected to spectrophotometric analysis with the NanoDrop 1000™ Spectrophotometer v3.1.0 (Thermo Fisher Scientific, USA). Due to the presence of carrier RNA in the extracted sample RNA, this assessment was merely qualitative and could not serve as a quantitative assessment.

3.6 Spectrophotometric analysis

The pedestal of the NanoDrop 1000™ Spectrophotometer v3.1.0 (Thermo Fisher Scientific, USA) was cleaned with 70% ethanol and nuclease-free water before applying 1 µl nuclease-free water to the pedestal for initialisation of the instrument per the manufacturer's instructions. The spectrophotometer was blanked with 1 µl of the solution used to elute the nucleic acids. For each sample, 1 µl was applied to the pedestal and measurements taken at wavelength 260 nm to calculate the concentration of nucleic acids in the sample and at 280 nm to provide an estimate of the extracted nucleic acid's purity based on the OD_{260}/OD_{280} ratio where values of 1.8 and 2.0 were considered pure for DNA and RNA samples respectively. The pedestal was wiped clean with ethanol between each sample to ensure no carry-over occurred between readings.

3.7 Reverse transcription

To detect the presence of CoV RNA from extracted sample nucleic acids by standard molecular techniques, RNA was reverse transcribed to complementary DNA (cDNA) using a reverse transcriptase (RT) enzyme. Depending on availability, either the RevertAid RT (Thermo Fisher Scientific, USA) or Maxima RT (Thermo Fisher Scientific, USA) kit was used with random hexamers, either Random Primers (Thermo Fisher Scientific, USA) or Random Hexamer Primers (Bioline, UK) depending on availability. Greater specificity of PCR assays can be obtained when performing reverse transcription with gene-specific primers as the generated cDNA is enriched with the target region, particularly useful when the target is present at a low copy-number (O'Connell, 2002). However, since the cDNA generated here would be used for multiple assays, the use of general hexamers was more feasible and cost-effective.

A 20 µl reaction volume was used to avoid multiple freeze-thaws and possible contamination during downstream assays. Reverse transcription products were at -20 ˚C for short-term storage or -80 ˚C for long-term storage. All thermocycling for reverse transcription was conducted on either a GeneAmp® PCR Systems 9700 (Applied Biosystems, USA) or SimpliAmp (Applied Biosystems, USA). thermal cycler.

3.7.1 Reverse transcription with RevertAid reverse transcriptase

Initially, 5 µl of extracted nucleic acids was added to 7.5 µl of a pre-mix containing diethylpyrocarbonate (DEPC) -treated water and 100 pmol random hexamers followed by a 5 minute incubation at 60 °C. The reactions were immediately cooled on ice before the addition of 7.5 µl of a pre-mix containing the RT reaction buffer at a 1× final concentration, deoxynucleotides (dNTPs) (Bioline, UK) at a final concentration of 1 mM for each dNTP, 20 U RiboLock RNase Inhibitor (Thermo Fisher Scientific, USA), and 200 U RevertAid RT enzyme. The reactions were incubated at 25 °C for 10 minutes, 42 °C for 60 minutes, and 70 °C for 10 minutes.

3.7.2 Reverse transcription with Maxima reverse transcriptase

For reverse transcription with Maxima RT (Thermo Fischer Scientific, USA), 5 µl of extracted nucleic acids was added to 9.5 µl of a pre-mix containing DEPC-treated water, 100 pmol random hexamers, and dNTPs (Bioline, UK) at a final concentration of 0.5 mM for each dNTP. These reactions were incubated for 5 minutes at 60 °C before being cooled on ice for 2 minutes. To this, 5.5 µl of a pre-mix containing RT reaction buffer at a 1× final concentration, 20 U RiboLock RNase Inhibitor (Thermo Fisher Scientific, USA), and 200 U Maxima RT was added. The reactions were incubated at 25 °C for 10 minutes, 50 °C for 30 minutes, and 85 °C for 5 minutes.

3.8 PCR assays used during this study

All thermocycling for PCR assays used during this study was conducted on either a GeneAmp® PCR Systems 9700 (Applied Biosystems, USA) or SimpliAmp (Applied Biosystems, USA) thermal cycler. All PCR products were visualised using horizontal DNA gel electrophoresis (Section 3.9). Samples identified as positive underwent subsequent PCR product purification (Section 3.10) followed by Sanger sequencing (Section 3.11) for confirmation and generation of sequencing data for phylogenetic analyses.

3.8.1 Primers

Primers used during this study were synthesised either by Integrated DNA Technologies (USA) or Inqaba Biotec™ (RSA). The Basic Local Alignment Sequence Tool (BLAST), found on the National Centre for Biotechnology Information (NCBI) website (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and the Primer Mapping Tool in Geneious R8 and R10 (Biomatters Ltd, New Zealand) were used to confirm that primers used in PCR assays aligned correctly to target regions. Where new primers were designed, additional *in silico* testing against nucleotide alignments of viruses of interest was performed in Geneious R8 and Molecular Evolutionary Genetics Analysis Version 6.0 or 7.0 (MEGA6, MEGA7) (Tamura *et al.,* 2013; Kumar *et al.,* 2016). When required, the online primer melting temperature (Tm) calculator on the Thermo Fisher Scientific website (www.thermofisher.com) was used to determine the recommended annealing temperature

according to the respective DNA polymerase to be used with the primer pair in question i.e. *Taq*-based, or Platinum SuperFi DNA polymerase. Primers used in each PCR protocol are listed by name with the corresponding oligonucleotide (oligo) sequence with each PCR assay.

3.8.2. Pan-CoV PCR assay

A PCR assay designed by de Souza Luna *et al.* (2007) capable of detecting the full spectrum of CoVs using a generic reverse transcription PCR (RT-PCR) was used to screen samples during this study. The design approach of this PCR assay is commonly referred to as a pan-PCR assay, indicating the use of an alignment of existing genome sequences of a species, or in this instance virus family, to identify conserved gene sets that can be broadly targeted using so-called universal primers (Yang *et al.,* 2013a). PCR products generated by this method require further confirmatory tests to identify exactly which strain or species was detected.

To maintain sensitivity for the detection of a broad range of CoVs, the PCR assay consisted of two rounds of amplification with the product of the first round PCR serving as template for the second round PCR, a nested PCR (de Souza Luna *et al.*, 2007). The PCR targets a region of the RNA dependent RNA polymerase (RdRp) with the resulting nested PCR product corresponding to positions 14417 – 14859 in *Rousettus* bat CoV (BtCoV) HKU10, GenBank number: NC_018871.

For this study the original protocol using a one-step RT-PCR enzyme mix (Qiagen, Germany) for first round PCR and Platinum *Taq* (Invitrogen, Germany) for second round PCR was adjusted for use with TrueStart Hot Start (HS) *Taq* DNA Polymerase (Thermo Fisher Scientific, USA) (de Souza Luna *et al.*, 2007; Ithete, 2013). A PCR product approximately 450 base pairs (bp) in size indicated a positive result.

Briefly, for first round PCR, each 25 µl reaction consisted of 5 µl sample template cDNA, nuclease-free water, PCR reaction buffer with a 1× final concentration, 1 ng Bovine Serum Albumin (BSA; Sigma-Aldrich, USA), 0.625 U TrueStart HS *Taq,* and a final concentration of 2.5 mM $MgCl_2$, 0.25 mM for each dNTP, and 0.2 µM for each primer; PC2S2 (equimolar solution of PCS1 and PCS2) and PC2As2 (equimolar solution of PCAs1, PCAs2, and PCAs3; Table 3.1).

Since the template cDNA was derived from nucleic acids extracted from faecal material, BSA was added as an enhancing agent. BSA is a globular protein widely used in biochemical assays due to its stability and lack of interference within biological reactions (Kreader, 1996; Farell and Alexandre, 2012). Its use in PCR assays serves to reduce the effect of inhibitors that may be present in nucleic acids from sample material likely to contain many impurities, such as faecal material (Kreader, 1996; Farell and Alexandre, 2012).

The first round PCR thermocycling started with 2 minutes at 95 °C followed by 10 touchdown cycles where the initial annealing temperature is gradually reduced with each cycle to optimise

primer binding conditions. These touchdown cycles consisted of 20 seconds of denaturing at 94 °C, 30 seconds of annealing starting at 60 °C with a decrease of 1 °C per cycle, and 40 seconds of elongation at 72 °C. The touchdown cycles were followed by 30 cycles of 20 seconds at 95 °C, 30 seconds at 54 °C, and 40 seconds at 72 °C. A final extension of 5 minutes at 72 °C followed by holding at 4 °C concluded the thermocycling (de Souza Luna *et al.*, 2007; Ithete, 2013).

For the second round PCR, a 50 µl reaction was set up as before but excluding BSA and replacing primers PC2S2 and PC2As1 with primers PCS (equimolar solution of PCS3 and PCS4) and PCNAs (Table 3.1). Briefly, each 50 µl reaction consisted of 2 µl PCR product from the first round PCR, nuclease-free water, PCR reaction buffer with a 1× final concentration, 1.25 U TrueStart HS *Taq* polymerase, and a final concentration 2.5 mM MgCl$_2$, 0.25 mM for each dNTP, and 0.3 µM of each primer PCS and PCS4.

The second round PCR thermocycling conditions consisted of 2 minutes at 95 °C followed by 30 cycles of 20 seconds at 95 °C, 30 seconds at 50 °C, and 40 seconds at 72 °C. A final extension of 5 minutes at 72 °C with holding at 4 °C concluded the protocol (de Souza Luna *et al.*, 2007; Ithete, 2013).

Table 3.1 Oligonucleotide sequences of primers used in the Pan-CoV screening PCR assay for the detection of all *Coronavirinae* members.

Oligo name	Oligo name	5' → 3' Oligo sequence
PC2S2	PCS1	TTA TGG GTT GGG ATT ATC
	PCS2	TGA TGG GAT GGG ACT ATC
PC2As2	PCAs1	TCA TCA CTC AGA ATC ATC A
	PCAs2	TCA TCA GAA AGA ATC ATC A
	PCAs3	TCG TCG GAC AAG ATC ATC A
PCS	PCS3	CTT ATG GGT TGG GAT TAT CCT AAG TGT GA
	PCS4	CTT ATG GGT TGG GAT TAT CCC AAA TGT GA
PCNAs	PCAs4	CAC ACA ACA CCT TCA TCA GAT AGA ATC ATC A

3.8.3 Extended RdRp PCR assay

Despite the successful identification of BtCoVs in samples using the protocol designed by de Souza Luna *et al.* (2007), the generated CoV sequence of approximately 395 bp after trimming the primer sequences has been found to be inefficient for extensive phylogenetic analyses with reliable resolution (Gloza-Rausch *et al.*, 2008; Drexler *et al.*, 2010). In 2010, Drexler *et al.* published their work on a PCR assay generating longer partial RdRp sequences to improve the classification of CoVs.

Lineage-specific forward primers targeting different CoV species were used with the Pan-CoV PCR assay's reverse primer, PCNAs, in separate PCR reactions. Forward primers designed for this assay included those specific for the detection of alphaCoVs, named Gr1Sp, and those specific for the detection of lineage B SARS-related betaCoVs, named SARSr (Drexler *et al.*, 2010). During this study, SARSr forward primers were not used in combination but instead in separate PCR

50

reactions. Forward primers, named RGU_2c, based on the same target region identified by Drexler *et al.* (2010) for extending the partial RdRp sequence were designed for the detection of lineage C MERS-related betaCoVs by co-supervisor Dr N Ithete using an alignment of closely related lineage C betaCoV sequences. These primers are listed in Table 3.2.

Table 3.2 Oligonucleotide sequences of lineage-specific primers used with reverse primer PCNAs to amplify an extended region of the coronavirus RNA dependent RNA polymerase.

Target CoV species	Oligo name	5' → 3' Oligo sequence
alphaCoVs with Gr1Sp primers	Gr1Sp F1	TTC TTT GCA CAG AAG GGT GAT GC
	Gr1Sp F2	CTT TGC ACA AAA AGG TGA TGC W*GC
lineage C MERS-related	RGU 2c F1	TTY* GCD* CAA GAT GGA M*AT GCT GC
betaCoVs with RGU_2c primers	RGU 2c F2	TTY* GCD* CAA GAT GGT M*AT GCT GC
lineage B SARS-related	SP3080 F1	CTT CTT CTT TGC TCA GGA TGG CAA TGC TGC
betaCoVs with SARSr primers	SP3195 F2	ATA CTT TGA TTG TTA CGA TGG TGG CTG
	SP3374 F3	CTA TAA CTC AAA TGA ATC TTA AGT ATG C

*Degenerate bases: W = A/T, Y= C/T, D = A/G/T, M = A/C

This assay was adapted for use with GoTaq® DNA Polymerase (Promega, USA) or Maxima HS *Taq* DNA Polymerase (Thermo Fisher Scientific, USA) depending on availability. Briefly, each 25 µl reaction consisted of 2.5 µl sample cDNA, nuclease-free water, PCR reaction buffer with a 1× final concentration, 1.25 U *Taq* polymerase, and a final concentration of 0.2 mM for each dNTP, and 0.5 µM for each primer.

The thermocycling started with 2 minutes at 95 °C when GoTaq® DNA Polymerase was used or 4 minutes at 95 °C when Maxima HS *Taq* DNA Polymerase was used, followed by 10 touchdown cycles of 20 seconds at 94 °C, 50 seconds at 60 °C with each subsequent cycle 1 °C lower than the previous, and 1 minute at 72 °C. The touchdown cycles were followed by 50 cycles of 20 seconds at 95 °C, 50 seconds at 54 °C, and 1 minute at 72 °C. A final extension of 5 minutes at 72 °C with holding at 4 °C concluded the protocol (Ithete, 2013). PCR products approximately 974 bp in size were considered positive when the RGU_2c, Gr1Sp, and SARSr SP3080 specific primers were used, while PCR products 861 bp and 682 bp in size were considered positive when the SARSr SP3195 and SP3374 primers were used, respectively.

3.8.4 Screening protocol

The initial screening protocol followed during this study for the detection of CoVs in bat samples involved screening all samples using the Pan-CoV screening PCR (section 3.8.2). For positive samples detected using this assay, sequencing results presented an indication of what CoV species was present in the sample. These positive samples then underwent additional amplification with the Extended RdRp PCR assay (section 3.8.3) using the corresponding lineage-specific forward primer set. However, after screening 600 bat faecal samples, the detected diversity was considerably lower than reported in recent literature findings (Ge *et al.*, 2016; Góes *et al.*, 2016; Kim *et al.*, 2016). A subset of samples (n = 334) was therefore subjected to repeat

screening using the Extended RdRp PCR assay with all three lineage-specific primer sets used in separate parallel reactions. For the SARSr primer set, only the SARSr F3 primer was used.

3.8.5 Confirmation of host species identity with molecular methods

In recent decades the number of bat species identified has rapidly increased with the resolution of species limits becoming increasingly more challenging when overlapping morphological features are at play (Monadjem *et al.*, 2013). The molecular analysis of mitochondrial DNA cytochrome *b* (cyt *b*) and cytochrome oxidase I (COI) coupled with morphological features greatly improves the accuracy and reliability of species identification. To confirm the species identity of collected samples by molecular methods, the cyt*b* and / or COI sequences were evaluated using the online Basic Local Alignment Tool (BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi) (described in section 3.11.5). The respective primers used in each assay are detailed in Table 3.3.

Table 3.3 Oligonucleotide sequences of primers used to amplify the cytochrome *b* and cytochrome oxidase I gene for species confirmation by molecular methods.

Target	Oligo name	5' → 3' Oligo sequence
cytochrome *b*	L14724	TGA Y*AT GAA AAA Y*CA TCG TTG
	H15915R	CAT TTC AGG TTT ACA AGA C
cytochrome oxidase I	LCO1490	GGT CAA CAA ATC ATA AAG ATA TTG G
	HC02198	TAA ACT TCA GGG TGA CCA AAA AAT CA

*Degenerate bases: Y= C/T

3.8.5.1 Cytochrome *b* PCR assay

Primers targeting cyt *b* were designed for studies on the evolution of mammals (Irwin *et al.*, 1991). These same primers were used successfully in studies used to identify new bat species and are listed in Table 3.3 (Monadjem *et al.*, 2013).

The assay was adapted for use with GoTaq® DNA Polymerase (Promega, USA) or Maxima HS *Taq* DNA Polymerase (Thermo Fisher Scientific, USA) depending on availability. Reactions were run in 30 µl volumes consisting of 5 µl sample DNA, nuclease-free water, Green PCR reaction buffer at a 1× final concentration, 1.25 U *Taq* polymerase, and a final concentration of 1.5 mM MgCl$_2$, 0.2 mM for each dNTP, and 0.4 µM for each primer.

Thermocycling conditions followed that optimised by Bastos *et al.* (2011). Thermocycling commenced with 2 minutes at 95 °C when GoTaq® DNA Polymerase was used or 4 minutes at 95 °C when Maxima HS *Taq* DNA Polymerase was used. This was followed by two cycles of 95 °C for 12 seconds, 52 °C for 30 seconds, and 70 °C for 1 minute. Three cycles followed of 95 °C for 12 seconds, 50 °C for 30 seconds, and 70 °C for 1 minute. After these touchdown-like cycles, 45 cycles of 95 °C for 12 seconds, 48 °C for 30 seconds, and 70 °C for 1 minute were run. A final

extension of 5 minutes at 72 °C with holding at 4 °C completed the thermocycling (Bastos *et al.*, 2011). PCR products 1200 bp in size were considered a positive result.

3.8.5.2 Cytochrome oxidase I PCR assay

Where cyt *b* amplification was unsuccessful, primers targeting the COI gene, developed by Folmer *et al.* (1994) were used as a secondary molecular tool for species confirmation.

Briefly 30 µl reaction volumes consisted of 5 µl sample DNA, nuclease-free water, Green PCR reaction buffer at a 1× final concentration, 1.25 U *Taq* polymerase, and a final concentration of 1.5 mM MgCl$_2$, 0.2 mM of each dNTP, and 0.4 µM of each primer; LCO_1490 and HCO_2198 (Table 3.3).

The cycling parameters consisted of a 2 minute denaturation at 95 °C when Go *Taq* polymerase was used and a 4 minute denaturation at 95 °C when Maxima HS *Taq* polymerase was used followed by 5 cycles of 30 seconds at 95 °C, 40 seconds at 50 °C, 1 minute at 70 °C; followed by 40 cycles of 30 seconds at 95 °C, 40 seconds at 55 °C, 1 minute at 70 °C; with a final extension of 10 minutes at 72 °C. A product approximately 710 bp in size was considered a positive result (Folmer *et al.*, 1994).

3.8.6 Extended sequencing of novel betacoronaviruses

During this study lineage C MERS-related betaCoVs showing similarity to NeoCoV, previously identified by Dr Ndapwea Ithete as described in Chapter 1 and Chapter 2, were identified (Ithete, 2013; Ithete *et al.*, 2013; Corman *et al.*, 2014a). Full genome amplification of a subset of these betaCoVs was attempted. With the full-length genome of CoVs notoriously difficult to obtain, various enzymes were used to amplify different genome regions. Each enzyme with its corresponding general protocol is described in the following subsections. For ease of reference, the enzyme and primers used to obtain each fragment has been tabulated in Appendix B.

3.8.6.1 *NSeq* betacoronavirus PCR assay

Following initial detection of betaCoVs displaying phylogenetic relatedness to the MERS-related NeoCoV using the Extended RdRp PCR assay (section 3.8.3), the assay developed by Corman *et al.* (2012) was used to amplify a portion of the nucleocapsid (N) protein. This assay was designed as part of a set of assays used for confirmatory testing for human MERS cases (Corman *et al.*, 2012).

The assay was designed using a nested approach of two rounds of amplification using the Superscript III one step RT-PCR system with Platinum™ *Taq* Polymerase (Invitrogen, USA). With this system, reverse transcription and PCR amplification are performed in a single tube with gene-specific primers for increased sensitivity.

For the first round PCR the 25 µl reaction consisted of 5 µl extracted sample nucleic acids, nuclease-free water, PCR reaction buffer with a 1× final concentration, 1 µl of SuperScript™ III RT / Platinum™ *Taq* Mix, and a final concentration of 0.8 mM $MgSO_4$, 0.4 µM of each primer; *NSeq*-Fwd and *NSeq*-Rev (Table 3.4). A PCR product approximately 291 bp in size was considered a positive result.

Table 3.4 Oligonucleotide sequences of primers used to amplify a fragment of the betacoronavirus nucleocapsid protein

Oligo name	5' → 3' Oligo sequence
NSeq Fwd	CCT TCG GTA CAG TGG AGC CA
NSeq Rev	GAT GGG GTT GCC AAA CAC AAA C
NSeq Fnest	TGA CCC AAA GAA TCC CAA CTA C

First round PCR thermocycling started with a 20 minute incubation at 50 °C allow first strand synthesis followed by 3 minutes at 95 °C and 45 cycles of 15 seconds at 95 °C, 15 seconds at 56 °C, and 30 seconds at 72 °C. A final extension of 2 minutes at 72 °C and holding at 4 °C concluded the thermocycling (Corman *et al.*, 2012).

The second round PCR protocol was adapted for use with the Expand™ High Fidelity[PLUS] PCR System (Sigma-Aldrich, USA) instead of Platinum *Taq* Polymerase Kit (Invitrogen). Briefly, the 50 µl reaction consisted of 1 µl first round PCR product, nuclease-free water, PCR reaction buffer without $MgCl_2$ with a 1x final concentration, 1 U of Expand High Fidelity[PLUS] Enzyme Blend, and a final concentration of 0.2 mM $MgCl_2$, 0.4 µM of each primer; *NSeq*-Fnest and *NSeq*-Rev (Table 3.4).

Second round PCR thermocycling commenced with 3 minutes at 94 °C followed by 45 cycles of 15 seconds at 94 °C, 15 seconds at 56 °C, and 30 seconds at 72 °C. A final extension of 2 minutes at 72 °C and holding at 4 °C concluded the thermocycling (Corman *et al.*, 2012).

3.8.6.2 NeoCoV full genome amplification approach

With the *NSeq* betaCoV RT-PCR Assay producing positive results the protocol designed to amplify the full genome of NeoCoV consisting of a panel of nested PCRs was used in an attempt to obtain the full genome of the identified novel betaCoVs in this study (Corman *et al.*, 2014a). The PCR assay was originally designed as a nested PCR using the SuperScript™ III One-Step RT-PCR System with Platinum™ *Taq* DNA (Invitrogen, USA) for cDNA synthesis and first round PCR and Platinum™ *Taq* Polymerase Kit (Invitrogen, USA) for second round PCR (Corman *et al.*, 2014a). When primers originally designed by Corman *et al.* (2014) were unsuccessful in amplifying regions of the betaCoVs identified during this study, new primers were designed based on the NeoCoV genome (GenBank: KC869678) and already amplified regions of the newly identified betaCoVs.

For the first round PCR, the SuperScript™ III One-Step RT-PCR System with Platinum™ *Taq* DNA Polymerase (Invitrogen, USA) was used. Briefly, a 25 µl reaction volume consisted of 5 µl extracted sample nucleic acids, nuclease-free water, PCR reaction buffer with a 1× final concentration, 1µl SuperScript™ III RT / Platinum™ *Taq* Mix, 1 µg BSA, and a final concentration of 0.8 mM $MgSO_4$, and 0.4 µM of each primer (Appendix B).

The thermocycling consisted of a 20 minute incubation at 50 °C for first strand cDNA synthesis followed by a 3 minute denaturation at 95 °C before 45 cycles of 15 seconds at 95 °C, 20 seconds at an annealing temperature deemed appropriate for the primers used, and 60 seconds at 72 °C. A final extension of 3 minutes at 72 °C with holding at 4°C concluded the thermocycling (Corman *et al.*, 2014a).

When the first round PCR did not generate visible or suitable bands by horizontal DNA gel electrophoresis, a second round PCR was attempted. The second round PCR was adapted for use with the Expand™ High Fidelity[PLUS] PCR System (Sigma-Aldrich, USA). The 50 µl reaction volume consisted of 1 µl of the first round PCR product nuclease-free water, PCR reaction buffer with a 1× final concentration, 0.2 µl Expand High Fidelity[PLUS] Enzyme Blend, and a final concentration of 2.5 mM $MgCl_2$, 0.2 mM for each dNTP, and 0.6 µM for each primer (Appendix B).

Second round PCR thermocycling consisted of 3 minutes at 95 °C followed by 45 cycles of 15 seconds at 95 °C, 20 seconds at 56 °C, and 60 seconds at 72 °C with a final extension time of 3 minutes at 72 °C and holding at 4 °C (Corman *et al.*, 2014a).

When amplification was unsuccessful with the described method, an alternative thermocycling approach was attempted for the first round PCR with the second round PCR the same as described above. The first strand cDNA synthesis was followed by a 3 minute incubation at 95 °C followed by 10 touchdown cycles consisting of 15 seconds at 95 °C, 20 seconds at 60 °C with each subsequent cycle decreasing by 1 °C, and 60 seconds at 72 °C. These touchdown cycles were followed by 40 cycles of 15 seconds at 95 °C, 20 seconds at 50 °C, and 60 seconds at 72 °C. A final extension of 3 minutes at 72 °C with holding at 4 °C concluded the thermocycling.

3.8.6.3 Attempted full genome amplification using SuperFi™

When amplification was not achieved following the NeoCoV genome amplification approach, amplification with Platinum™ SuperFi™ DNA Polymerase (Invitrogen, USA), was attempted. This is a high-fidelity polymerase with DNA proofreading ability for high specificity and increased yields of amplified product. For each primer set attempted, the annealing temperature and extension time were adjusted per the manufacturer's recommendations. The general protocol used a 25 µl reaction volume consisting of 2 µl sample cDNA synthesised using Maxima RT (Thermo Fischer Scientific, USA), nuclease-free water, Green PCR reaction buffer with a 1× final concentration,

0.25 µl Platinum™ SuperFi™ DNA Polymerase, a final concentration of 0.2 mM for each dNTP, and 0.5 µM for each primer (Appendix B).

The general thermocycling consisted of 30 seconds at 98 °C followed by 45 cycles of 10 seconds at 98 °C, 10 seconds at the appropriate annealing temperature, and 72 °C for the duration determined by the length of the region to be amplified based on an extension time of 44 seconds per kb. A final extension of 5 minutes at 72 °C and holding at 4 °C concluded the thermocycling. Where successful amplification was not immediately obtained, a temperature gradient was used to adjust the annealing temperature.

3.9 DNA gel electrophoresis

Horizontal DNA gel electrophoresis is a widely-used tool to separate DNA fragments by size. Through the use of an electric current, negatively charged DNA fragments will separate out by size when migrating through a gel medium, such as agarose, with smaller fragments moving faster through the gel than their larger counterparts (Weaver, 2008). DNA electrophoresis is made possible using an electrolytic buffer, commonly Tris-acetic acid (TAE) buffer.

Instead of TAE buffer, sodium boric acid (SB) was used as the conductive medium for DNA electrophoresis during this study. SB provides better resolution at higher voltages than TAE and generates less heat, avoiding problems such as poor gel integrity associated with TAE gels, facilitating more efficient DNA electrophoresis (Brody and Kern, 2004).

SB buffer was prepared as a 20× stock solution using 8 g sodium hydroxide with 45 g boric acid dissolved in 1 L Milli-Q® (Merck Millipore, Germany) distilled water and used as a 1× working solution with a final concentration of 10 mM sodium hydroxide and pH of 8.5 (Brody and Kern, 2004).

For expected PCR product sizes of less than 1 kb, 2% (m/v) agarose gels were prepared while 1.5% (m/v) agarose gels were used for expected PCR product sizes greater than 1 kb. An alternative safe nucleic acid stain to mutagenic ethidium bromide, GR Green (Excellgen, USA) or Pronasafe (Laboratorios Conda, Spain), depending on availability, was used as an in-gel stain to visually detect DNA PCR products following electrophoresis.

Briefly, Milli-Q® water was added to either 1.5 g or 2 g of Lonza® LE agarose (BioWhitaker, USA) or TopVision™ Agarose (Thermo Fisher Scientific, USA) depending on the expected PCR product size and available agarose, to a volume of 100 ml. The mixture was heated in a microwave to boiling point until a clear solution was obtained and then allowed to cool while stirring to approximately 60 °C. The nucleic acid stain was added to a final concentration of 1× and the solution poured into an electrophoresis tray with a 1 mm gel comb. The gel was allowed to set for

approximately 30 minutes before the comb was removed and SB added to sufficiently cover the gel for efficient electricity conduction during electrophoresis.

For certain PCR *Taq* polymerase kits such as Super FI and GoTaq®, a Green PCR reaction buffer including a density reagent and tracking dyes to facilitate direct loading of samples for DNA gel electrophoresis was used. In these instances, 5 µl of sample was directly loaded into an agarose gel well. Samples not amplified with a Green PCR reaction buffer, such as TrueStart HS *Taq* DNA Polymerase (Thermo Fisher Scientific, USA) and Maxima HS *Taq* DNA Polymerase (Thermo Fisher Scientific, USA), were mixed with 6× Orange Loading Dye (Thermo Fisher Scientific, USA) to visually track their migration. Briefly, 5 µl of sample was mixed with 1 µl of loading dye before loading.

Depending on the expected size of the target fragments, 3 µl of either ready-to-use GeneRuler™ 100 bp (Thermo Fisher Scientific, USA) or O'GeneRuler™ 1 kb (Thermo Fisher Scientific, USA) DNA ladder was loaded alongside samples as a sizing reference. To ensure sound resolution of fragments, DNA gel electrophoresis was conducted at 90 V for 40 minutes followed by visualisation with the UV (ultraviolet) transilluminator at 254 nm of the UVIprochemi D-77LS-26M (UVItec, UK) gel documentation imaging system. UVIsoft Gel Analysis Software's UVIBand package (UVItec, UK) v. 12.11 was used to optimise the gel image.

3.10 PCR product purification

Positive PCR reactions underwent PCR product purification. Depending on availability, kits such as Wizard® SV Gel and PCR Clean-Up System (Promega, USA), Rapid PCR Enzyme Cleanup Set (New England Biolabs, UK), or MinElute® PCR Purification (Qiagen, Germany) were used to remove excess nucleotides and primers before Sanger sequencing and / or cloning (described in section 3.11 and 3.12). To assess the quality and DNA concentration, all purified PCR amplification products underwent spectrophotometric measurement using the NanoDrop 1000™ as described in section 3.6.

3.10.1 Wizard® SV Gel and PCR Clean-Up System

The spin-column based Wizard® SV Gel and PCR Clean-Up System (Promega, USA) was used following the manufacturer's centrifugation protocol.

Briefly, an equal volume of membrane binding solution was added to the sample PCR product volume remaining after DNA electrophoresis and well mixed. The solution was transferred to the provided spin-column and centrifuged for 1 minute at 16000 × *g* before discarding the flow-through and adding 700 µl of the provided ethanol-based membrane wash buffer. The column was centrifuged at 16000 × *g* for 5 minutes and the flow-through discarded before a 1 minute centrifugation to allow any residual ethanol that might interfere with downstream applications to

evaporate. The column was transferred to a 1.5 ml collection tube and the purified PCR product DNA eluted with 50 µl of nuclease-free water following a 1 minute incubation at room temperature and centrifugation at 16000 × g for 1 minute.

3.10.2 MinElute® PCR Purification Kit

The MinElute® PCR Purification Kit (QIAGEN, Germany) was used per the manufacturer's recommendations with all centrifugation steps carried out at 17900 × g.

Briefly, 5 volumes of the provided binding buffer was added to 1 volume of PCR product. The mixture was transferred to the provided MinElute® column and centrifuged for 1 minute before discarding the flow-through. A wash step with of 750 µl of the provided wash buffer and centrifugation for 1 minute followed before discarding the flow-through. To ensure maximum recovery of the purified DNA, the column was centrifuged again for 1 minute to facilitate the evaporation of any traces of ethanol present in the wash buffer. To elute the purified DNA, 10 µl of provided elution buffer was applied to the centre of the column, incubated for 1 minute and centrifuged for 1 minute.

3.10.3 NucleoSpin Gel® and PCR Clean Up Kit

Following gel electrophoresis PCR reaction volumes were adjusted to 60 µl for PCR purification using the NucleoSpin® Gel and PCR Clean-up Kit (Machery-Nagel, Germany). To each sample, twice the PCR reaction volume of the provided buffer NTI was added followed by centrifugation at 11000 × g for 30 seconds. The silicone membrane with bound sample DNA was washed twice by adding 700 µl of the provided buffer NT3 and centrifuged at 11000 × g for 30 seconds. Prior to elution, the spin-columns were centrifuged for 1 minute at 11000 × g and incubated at 70 °C for 5 minutes to dry the membrane and allow any residual ethanol to evaporate. Purified PCR product DNA was eluted using 15 – 20 µl of the provided buffer NE, following a 1 minute incubation at room temperature and centrifugation at 11000 × g for 1 minute.

3.10.4 Rapid PCR Enzyme Cleanup Set

The Rapid PCR Enzyme Cleanup Set (New Biolabs, UK) is not spin-column based but instead utilises two recombinant enzymes namely Exonuclease I and Shrimp Alkaline Phosphatase to facilitate full enzymatic degradation of excess primers and input material present in the PCR reaction. The enzyme set dephosphorylates any excess dNTPs and allows for direct downstream applications.

Briefly, 1 µl of each enzyme was added to 5 µl of PCR reaction while working on ice. Following a brief vortex and centrifugation to concentrate all components at the bottom of the tube, the reactions were incubated at 37 °C for 5 minutes and inactivated at 80 °C for 10 minutes.

3.11 Sanger sequencing

Sanger sequencing was developed in 1977 to determine the DNA sequence of PCR products of interest. It relies on the selective incorporation of chain-terminating dideoxynucleotides (ddNTPs) by DNA polymerase (Sanger *et al.*, 1977). The Sanger sequencing protocol consists of a specialised sequencing PCR, removal of excess sequencing reagents, and sequencing electrophoresis as described below.

3.11.1 Sequencing PCR

For sequencing PCR reactions, the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) for use with the ABI 3130*xl* Genetic Analyzer was used. This sequencing kit utilises a pre-mixed ready reaction format requiring only the addition of template and a template-specific primer for fluorescence-based cycle sequencing. Samples were sequenced bidirectionally, separate reactions for each forward and reverse primer, to produce contiguous sequences.

Sequencing reactions were optimised for 10 µl reaction volumes consisting of 3 µl of the provided 5× PCR reaction buffer, 1 µl provided Ready-Reaction mix, 2 µl nuclease-free water, 2 µl of 2.5 µM primer, with 2 µl of template DNA. For sequencing of PCR products, purified PCR product DNA was diluted to provide a total of 15 – 25 ng DNA per reaction while for the sequencing of cloned plasmids (described in section 3.12), purified plasmid DNA was diluted to provide a total of 150 - 300 ng of double stranded DNA.

Thermocycling commenced with 1 minute at 96 ˚C, followed by 30 cycles of 96 ˚C for 10 seconds, 54 ˚C for 5 seconds, and 60 ˚C for 4 minutes, followed by holding at 4 ˚C to conclude the thermocycling. Reactions were briefly centrifuged before commencing sequencing reaction clean-up.

3.11.2 Sequencing reaction clean-up

The BigDye® XTerminator™ Purification Kit (Applied Biosystems, USA) was used to remove salts, charged particles, and unincorporated BigDye terminators that may lead to dye blob formation and poor sequencing data during sequencing electrophoresis. Briefly, for each sample, 45 µl of the provided SAM™ solution and 10 µl of the provided XTerminator™ solution was added before vortexing for 30 min.

3.11.3 Sequencing electrophoresis

Capillary-based sequencing electrophoresis is the process whereby the fluorescent ddNTPs incorporated during the sequencing PCR are excited and their emitted signals recorded to determine which ddNTP was incorporated, ultimately allowing the sequence to be determined. The data recorded during sequence electrophoresis consist of a series of peaks displaying fluorescent intensities in a chromatogram. The DNA sequence of the target fragment is read from the peaks of these chromatograms in software such as Geneious R8 and R10 (Biomatters Inc., New Zealand). During this study, cleaned sequencing reactions were centrifuged before undergoing sequencing electrophoresis on the ABI 3130*xl* Genetic Analyser (Applied Biosystems, USA). At times it was necessary to outsource the sequencing electrophoresis to the Central Analytical Facilities' DNA Sequencing Unit at Stellenbosch University.

Where poor sequencing data was obtained, indicated by low quality read scores or the presence of multiple ambiguous bases indicating the possible presence of a quasispecies, when directly sequencing purified PCR amplified products, these fragments were cloned into a vector and sequenced with M13 primers as described in section 3.12.

3.11.4 Contiguous sequences assembly and basic sequence editing

Raw sequencing data chromatogram files were imported into Geneious R8 or R10 (Biomatters Inc., New Zealand). Forward and reverse sequences were selected by sample and *de novo* contig assembly performed with primers trimmed to produce contiguous sample sequences. Sequence quality was assessed and ambiguous bases were resolved to a single nucleotide where possible by reviewing the chromatogram data at the relevant positions. Ambiguous bases occur when overlapping fluorescence signals appear at a specific position in the sequence. This could be due to the presence of more than one sequence in the sample i.e. double peak formation due to sequence variations or at times it could be due to a so-called blob formation during the sequencing electrophoresis.

When required, contiguous sequences were translated into amino acid sequences in Genious R8 or R10 using the correct reading frame as indicated by the BLAST result (described in section 3.11.5). Reading frames divide nucleotide sequences into consecutive non-overlapping triplets representing codons that translate into stop signals or amino acids that make up proteins encoded by the gene of interest. To determine the identity of the resulting contiguous sequence, the online BLAST tool was used as described in section 3.11.5.

3.11.5 Basic Local Alignment Sequence Tool (BLAST)

The Basic Local Alignment Sequence Tool (BLAST), available on the National Centre for Biotechnology Information (NCBI) website (https://blast.ncbi.nlm.nih.gov/Blast.cgi), was used to

determine if the sample sequences obtained matched the targeted fragments of the various PCR assays used. In general, the Standard Nucleotide BLAST (blastn) function was used employing the nucleotide collection dataset with the search optimised for somewhat similar sequences. The Translated BLAST (blastx) with the non-redundant protein sequence database was used at times to determine the correct reading frame of sample sequences for translation and / or to confirm the blastn result.

3.12 Cloning

The cloning of a DNA fragment into a plasmid vector is common practice in molecular biology laboratories. Cloning was used during this study when it was difficult to obtain high quality sequencing data directly from the purified PCR product and to establish positive plasmid controls to generate *in vitro* transcribed CoV RNA for use in RT-PCR screening assays.

Commonly, DNA ligase facilitates the covalent linkage of a compatible DNA fragment with linearized plasmid ends. DNA polymerases such as TrueStart Hot Start (HS) *Taq* DNA Polymerase (Thermo Fisher Scientific, USA) and Maxima HS *Taq* DNA Polymerase (Thermo Fisher Scientific), as used in this study, are known to add additional non-template A nucleotides to the 3'- ends of blunt-ended double-stranded DNA fragments, so-called poly-A tailing. This generates a 3'-A overhang at both ends enabling cloning through direct ligation of the PCR product into a linearised T vector with 3'-T overhangs, otherwise known as TA cloning (Zhou and Gomez-Sanchez, 2000).

The InsTAclone PCR Cloning Kit (Thermo Fisher Scientific, USA) provides a TA cloning system with cloning vector pTZ57R/T for efficient ligation. Additionally, the cloning vector contains a T7 promoter sequence required for *in vitro* transcription (described in section 3.13). The pTZ57R/T cloning vector (Figure 3.3) facilitates colony selection by ampicillin resistance and blue / white screening methods and enables easy excision of the insert or sequencing with commonly used restriction enzymes and M13 / pUC primers.
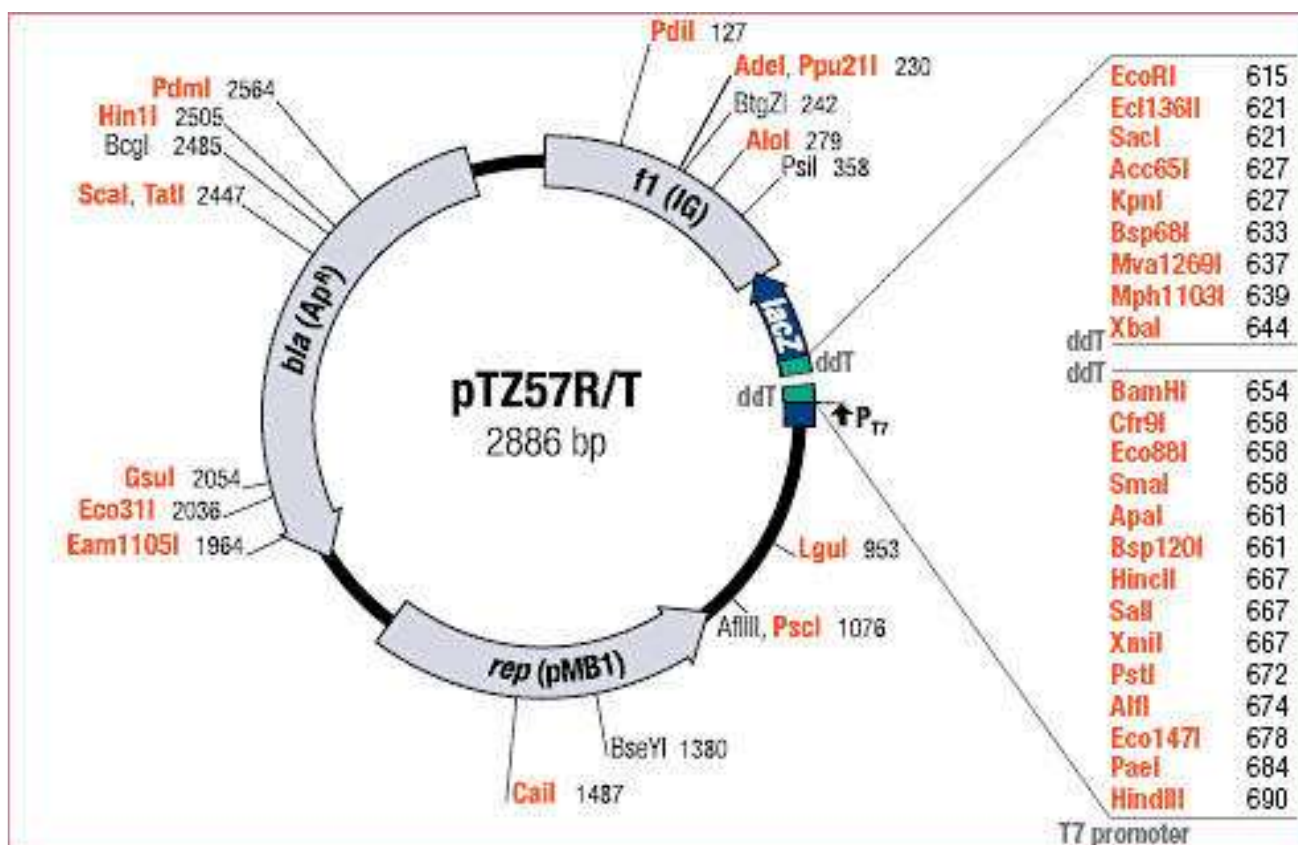
Figure 3.3 Map of cloning vector pTZ57R/T (Thermo Fisher Scientific, USA). This diagram depicts the cloning vector pTZ57R/T from the InsTAclone PCR Cloning Kit (catalogue number #K1213, #K1214) used with unique restriction sites, selection genes, and position of the T7 promoter indicated. (www.thermofisher.com).

The presence of an ampicillin resistance gene in the cloning vector allows for the selection of transformed cells when plated on agar containing ampicillin against which resistance is conferred. To reduce the likelihood of falsely identifying colonies as transformed, a second mechanism, blue / white screening, was incorporated. The cloning vector pTZ57R/T contains a lacZ α-peptide with a multiple cloning site (MCS) while the chemically competent *E. coli* cells used for transformation contain the lacZ deletion mutant (*lacZΔM15*). Successful transformation results in the formation of a functional β-galactosidase enzyme that can digest the lactose analogue X-Gal, producing a bright blue colour in the presence of Isopropyl β-D-1-thiogalactopyranoside (IPTG). When a DNA fragment of interest is successfully ligated into the MCS of the cloning vector, the lacZ α-peptide is disrupted and β-galactosidase is not produced resulting in the formation of white colonies.

Fragments cloned during this study were purified PCR products obtained from relevant PCR reactions and used for the purposes of either generating positive controls or to improve sequencing data. The cloning of all DNA fragments followed the general protocol of ligation, transformation, colony picking, and plasmid DNA purification as detailed. All cloned fragments were verified with Sanger sequencing following plasmid DNA purification.

3.12.1 Positive controls

To ensure that the pan-CoV and extended RdRp PCR assays (Section 3.8) were running optimally, positive plasmid DNA controls were generated that could be used directly as DNA positive controls or *in vitro* transcribed to RNA (described in section 3.13). These positive controls are briefly described below.

3.12.1.1 Pan-CoV PCR assay positive controls

A ~974 bp partial fragment of the *RdRp* gene of a *Miniopterus natalensis*-derived alphaCoV was amplified from a bat faecal RNA sample, specimen code MSTM2. This sample was previously identified as positive by Dr Ithete during her PhD (Ithete, 2013). The fragment for cloning was amplified using the extended RdRp PCR assay with the RGU_2c primer set.

The amplified PCR product underwent PCR product purification using the Wizard® SV Gel and PCR Clean-Up System (Promega, USA) as described in Section 3.10. The resulting positive control plasmid DNA underwent *in vitro* transcription (Section 3.13) to generate positive control RNA.

To assess the sensitivity of the PCR assay, a serial dilution of the quantified positive control RNA (Section 3.13) underwent reverse transcription and subsequent amplification in duplicate with the Pan-CoV PCR assay.

3.12.1.2 Extended RdRp PCR assay lineage B betacoronavirus positive control

Since no lineage B betaCoV positive samples were previously identified, a positive control was required for use in the extended RdRp PCR assay using the lineage B betaCoV SARSr primers. With no positive sample or PCR products to work with, a 992 bp gBlocks gene fragment (Integrated DNA Technologies, USA) based on a SARS-related BtCoV (NC_014470: BM48-31/BGR/2008) was synthesised. The PCR product generated using this gBlock gene fragment with the SARSr F1 primer in the extended RdRp PCR assay (Section 3.8) was purified with the MinElute® (QIAGEN, Germany) PCR product purification system before commencing cloning. The resulting positive control plasmid DNA was used directly in PCR assays at a 1:20 dilution and did not undergo *in vitro* transcription.

3.12.2 Ligation

Ligation for cloning was performed according to the manufacturer's recommendations with an adjusted incubation protocol. The recommended amount of PCR product to be used in the ligation reaction was established from the manufacturer's guidelines with approximately 0.172 ng of purified DNA required per bp for the length of the fragment to be cloned. Based on the spectrophotometric measurement of the purified PCR product DNA, the appropriate dilutions using nuclease-free water were made. The 30 µl ligation reaction consisted of nuclease-free water,

ligation buffer with a final concentration of 1×, 3 µl of the pTZ57R/T vector, the recommended amount of PCR DNA for the respective fragment, and 1 µl T4 DNA ligase. The ligation reactions were briefly vortexed and centrifuged before undergoing incubation at 25 °C for 2 hours, 4 °C for 16 hours, and 5 minutes at 85 °C, an incubation protocol followed in our laboratory (Fisher 2016).

3.12.3 Transformation

Transformation was performed using either prepared *Escherichia coli* strain JM109 cells as per the InsTAclone PCR Cloning kit protocol (Thermo Fisher Scientific, USA) or with the Mix & Go (Zymo Research, USA) JM109 chemically competent cells according to the manufacturer's instructions. Luria-Bertani (LB) broth agar plates and LB liquid media were used for the growth of transformed bacteria. These were prepared prior to transformation as described below.

3.12.3.1 Preparation of Luria broth agar and liquid media

Since the cloning system used here is conducive to antibiotic selection and blue / white screening, LB agar plates were prepared with ampicillin and IPTG / X-Gal. Briefly 10 g of LB and 7.5 g of agar were weighed and dispensed into a 1 L Pyrex glass bottle. Milli-Q® (Merck Millipore, Germany) distilled water was added to a final volume of 500 ml and well mixed before sterilisation by autoclaving for 30 minutes at 121 °C and 100 kPa above atmospheric pressure in a steam sterilizer autoclave (Huxley, Taiwan). The LB agar was allowed to cool to approximately 60 °C before adding ampicillin to a final concentration of 50 - 100 µg/ml, IPTG (Sigma-Aldrich, USA) and X-Gal (Thermo Fisher Scientific, USA) were added to a final concentration of 20 mM and 80 µg/ml respectively. Approximately 25 ml of LB agar was poured into each plastic Petri dish (NEST, USA) and stored inverted at 4 °C once set.

LB liquid medium was prepared as above but omitting the addition of agar, IPTG, and X-Gal. Only ampicillin was added to the LB liquid medium once cooled.

3.12.3.2 Competent cells

For the preparation of chemically competent *E. coli* cells as per the InsTAclone PCR Cloning Kit's (Thermo Fisher Scientific, USA) protocol, the provided C-medium was heated to 37 °C and seeded with JM109 cells (Promega, USA) in a 2 ml volume for overnight culture at 37 °C on a shaker. On the day of transformation, 150 µl of the overnight culture was added to culture tubes containing 1.5 ml of pre-warmed C-medium and incubated at 37 °C for 20 minutes on a shaker. The culture was centrifuged, the supernatant discarded, and the pelleted cells re-suspended in 300 µl of provided T-solution before being incubated on ice for 5 minutes. Centrifugation was repeated, the supernatant discarded, and the pelleted cells re-suspended in a smaller volume of 120 µl T-solution and again incubated on ice for 5 minutes. A volume of 2.5 µl ligation reaction was

aliquoted into a microcentrifuge tube to which 50 µl of this prepared JM109 cell culture was added before being gently mixed.

For the Mix & Go (Zymo Research, USA) JM109 cells, 2 - 4 µl of the ligation reaction mixture was added to a single-use 50 µl aliquot of competent cells that had been thawed on ice.

For both prepared JM109 cells and Mix & Go JM109 cells, the transformation reactions were incubated on ice for 5 minutes before being immediately plated on pre-warmed LB-ampicillin X-Gal-IPTG agar plates and incubated overnight at 37 °C.

3.12.4 Colony selection

Following overnight incubation, the LB-ampicillin X-Gal-IPTG agar transformation plates were examined for the presence of white colonies. The prepared LB broth with ampicillin was warmed to 37 °C and 4 ml added to 14 ml round bottomed polypropylene culture tubes (Falcon™ Corning, USA). With a sterile inoculation loop or pipette tip, a single individual white colony was picked from the agar plate and swirled into the LB medium of a culture tube. This was repeated for each colony picked. Culture tubes were incubated overnight at 37 °C on a shaker at 190 - 210 rpm. Plasmid DNA from these overnight cultures was purified using a mini prep protocol (Section 3.12.6).

3.12.4.1 Glycerol stocks

For long-term preservation of transformed bacterial cultures containing plasmids of interest, bacterial glycerol stocks were made. The use of glycerol serves to stabilize the bacterial cells at low temperatures preventing damage to cell membranes and preserving cell viability. For plasmid DNA generating the expected band size during PCR, before plasmid DNA purification commenced, 500 µl of overnight liquid culture was well mixed with 500 µl sterile 50% (v/v) glycerol solution in a 2 ml screw top tube or cryovial (Corning, USA), labelled, and stored at -80 °C (Addgene, no date).

To recover transformed bacterial cells from the glycerol stock, a sterile inoculation loop was used to scrape off some frozen bacterial culture and plated on an LB agar plate for overnight incubation growth. Colonies were then picked and cultured as needed.

3.12.4.2 Bulk culturing

When a large amount of plasmid DNA was required for example for *in vitro* transcription (Section 3.12.7), transformed bacterial cells recovered from glycerol stocks were cultured in 4 ml LB broth with ampicillin at 50 – 100 mg/ml at 37 °C for 8 hours or overnight on a shaker at 190 – 210 rpm. The bulk culture was created by adding 300 µl of the small overnight culture to 100 ml LB broth with ampicillin at 50 - 100 mg/ml and incubated at 37 °C on a shaker at 190 – 210 rpm overnight. Plasmid DNA from bulk cultures was purified using a midi prep purification protocol.

3.12.5 Plasmid DNA purification

To undertake Sanger sequencing of cloned fragments or to *in vitro* transcribe cloned fragments of interest, plasmid DNA was purified from the bacterial cells. Depending on the volume of bacterial cell culture from which the plasmid DNA was to be purified, either a mini- or midiprep was performed.

3.12.5.1 Miniprep plasmid DNA purification

Plasmid DNA purification for bacterial cell culture volumes of less than 10 ml was carried out using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, USA) per the manufacturer's instructions with centrifugation at 12000 × g. Briefly, the overnight culture was centrifuged for 4 minutes and the supernatant discarded. The pelleted cells were re-suspended in 250 µl of the provided re-suspension buffer and the cell suspension transferred to a microcentrifuge tube. To release the plasmid DNA from the cells, 250 µl of the provided lysis buffer was added and gently mixed by inverting the tubes 4 - 6 times before adding 350 µl of the provided neutralisation buffer. Centrifugation for 5 minutes pelleted the precipitated cell debris and chromosomal DNA. The clear supernatant was carefully transferred to a supplied GeneJET column. The column was centrifuged for 1 minute and the flow-through discarded before adding 500 µl of the provided wash buffer containing ethanol. The column was centrifuged for 1 minute and the flow-through discarded before repeating the wash step. To ensure the removal of any residual ethanol from the wash steps, the column was centrifuged again for 1 minute. The GeneJET column was transferred to a clean 1.5 ml microcentrifuge tube and 50 µl of the provided elution buffer carefully added to the centre of the column membrane and incubated for 2 minutes at room temperature before eluting the plasmid DNA by centrifugation for 2 minutes.

3.12.5.2 Midiprep plasmid DNA purification

For large bacterial cell cultures, 50 - 250 ml in volume, plasmid DNA purification was carried out using the PureYield™ Plasmid Midiprep System (Promega, USA) according to the manufacturer's instructions. Before commencing the protocol, large overnight bacteria cell cultures grown in flasks were transferred to 50 ml centrifuge tubes (Corning, USA). Briefly, bacterial cells were pelleted by centrifugation at 5000 × g for 10 minutes and the supernatant discarded before the bacterial cell pellet was re-suspended in the provided respective cell re-suspension buffer before following the respective kit's protocol.

The cells were lysed with 3 ml of the provided cell lysis buffer and the tube inverted to gently mix the lysate before incubating at room temperature for 3 minutes. Cell lysis was stopped with the addition of 5 ml of the provided neutralisation buffer and gently mixing by inverting the tube 5 – 10 times. The solution was centrifuged for 15 minutes at 15000 × g before decanting the clear supernatant into the provided PureYield™ Clearing Column stacked onto a PureYield™ Binding

Column attached to a vacuum manifold. The vacuum was applied until all the liquid had passed through both columns before removing the PureYield™ Clearing Column. Following the addition of 20 ml of the provided wash buffer, the vacuum was applied to draw the liquid through. To dry the membrane before elution, the vacuum was applied for 1 minute. The PureYield™ Binding Column was removed from the vacuum manifold and a 1.5 ml microcentrifuge tube (Eppendorf, Germany) placed at the bottom of the Eluator™ Vacuum Elution Device (Promega, USA) before inserting the PureYield™ Binding Column into the Eluator™ Vacuum Elution Device and attaching the whole setup to the vacuum manifold. For elution of the plasmid DNA into the attached 1.5 ml microcentrifuge tube, 400 µl of nuclease-free water was passed through the column by applying the vacuum for 1 minute.

The purified plasmid DNA from cloned PCR products for the purposes of obtaining improved sequencing data underwent Sanger sequencing (Section 3.11) using M13 primers, designed for the sequencing of cloned fragments. Purified plasmid DNA to be used to generate positive controls underwent further processing to generate *in vitro* transcribed RNA as detailed below.

3.13 *In vitro* transcribed coronavirus RNA

Positive control plasmid DNA as detailed in section 3.12.1 was *in vitro* transcribed to generate RNA. Using RNA positive controls instead of only positive plasmid control DNA additionally allows one to determine if reverse transcription prior to the screening PCR assay was successful.

3.13.1 Restriction enzyme digestion

To initiate the process of *in vitro* transcription, the purified plasmid DNA was linearised downstream of the inserted sequence of interest and the T7 promoter sequence (Figure 3.3). A restriction enzyme digest of 1 – 2 µg plasmid DNA was performed with *EcoRI* (New England Biolabs, UK) according to the manufacturer's instructions. For 1 µg plasmid DNA, a 50 µl reaction consisting of nuclease-free water, reaction buffer with a final concentration of 1×, and 1 µl *EcoRI* was incubated for 37 °C for 1 hour followed by heat inactivation at 65 °C for 20 minutes.

3.13.2 DNA concentration, desalting and enzyme removal

Prior to *in vitro* transcription, linearised plasmid DNA was concentrated, desalted and excess enzymes were removed using the spin-column based MinElute® Reaction Cleanup Kit (QIAGEN, Germany) according to the manufacturer's protocol. The protocol followed was identical to that described in section 3.10.2 with the exception that instead of adding 5 volumes of the provided binding buffer to the samples, 300 µl of the provided buffer ERC was added before proceeding as previously described.

3.13.3 *In vitro* transcription

To synthesise RNA molecules based on a DNA sequence, *in vitro* transcription exploits the RNA polymerase's requirement for a highly specific promoter sequence that is commonly found in bacteriophage systems (Beckert and Masquida, 2011). With the widely used T7 promoter sequence, derived from the *E. coli* T7 phage, contained upstream of the cloning vector pTZ57R/T's MCS (Figure 3.3), the ligated DNA sequence of interest underwent an *in vitro* transcription reaction with the TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific, USA). The manufacturer's protocol was followed with approximately 1 µg of linearised DNA.

Briefly, a 20 µl reaction consisted of 4 µl of the provided reaction buffer, 8 µl dNTPs with a stock concentration of 100 mM each, 1 µg of linearised template DNA, and 2 µl TranscriptAid Enzyme Mix. The reaction was briefly centrifuged to concentrate the reagents at the bottom of the tube before incubating for 2 hours at 37 °C.

3.13.4 DNase treatment

To remove any residual template DNA following *in vitro* transcription, a DNase treatment was performed. For each 1 µg of linearised DNA *in vitro* transcribed, 2 U of DNase I (Thermo Fisher Scientific, USA) was added to the *in vitro* transcription reaction mixture and incubated at 37 °C for 15 minutes. The reaction was inactivated by commencing RNA purification immediately after.

3.13.5 DNase inactivation and RNA purification

To extract and purify the *in vitro* transcribed RNA following DNase treatment, the PureLink® RNA Mini Kit (Thermo Fisher Scientific, USA) protocol for purifying RNA from *in vitro* transcription reactions. Following RNA purification, the eluted RNA concentration was measured with the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA) as described in Section 3.13.6 and the copy number determined.

With the PureLink® RNA Mini Kit (Thermo Fisher Scientific, USA), one volume of the provided lysis buffer, containing 2-mercaptoethanol, and one volume of 100% ethanol was added to one volume of sample and mixed by vortexing. The sample was applied to a provided spin-column and centrifuged for 15 seconds at 12000 × *g*. Two wash steps were performed with the addition of 500 µl of the provided wash buffer and centrifuged as before. The membrane was dried by centrifuging the empty column for 1 minute at 12000 x *g* before transferring the column to an elution tube. Thirty-five µl of RNAse free water was applied to the centre of the column, incubated for 1 minute at room temperature, and centrifuged for 2 minutes at 12000 × *g* to elute the RNA.

3.13.6 RNA quantification and determination of insert copy number

To estimate the limit of detection of the screening Pan-CoV PCR assay, the copy number of the *in vitro* transcribed positive control fragment RNA was determined. The concentration of the purified RNA was quantified on the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA) using the Qubit RNA HS Assay (Thermo Fisher Scientific, USA) for low-abundance RNA samples with concentrations between 250 pg/µl and 100 ng/µl without quantitating DNA, protein, or free nucleotides.

During RNA quantification with the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA), 0.5 ml tubes (Axygen®, USA) were used as per the manufacturer's recommendations. The Qubit® working solution was prepared by diluting the provided Qubit® RNA HS Reagent 1:200 in Qubit® RNA HS Buffer. For calibration, two standards were prepared in 200 µl volumes consisting of 190 µl of prepared Qubit® working solution and 10 µl standard. Samples were prepared for measurement by mixing 1 - 20 µl of sample RNA with Qubit® working solution for a total volume of 200 µl. Reactions were briefly vortexed before incubating for 2 minutes at room temperature before quantification on the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA). The standards were measured before running the samples with measurements taken in triplicate to determine an average reading. Since the samples were diluted in Qubit® working solution, the actual concentration was determined using the following equation:

$$concentration\ of\ sample = QF\ value \times \frac{200}{x}$$

The **QF value** is the reading given by the Qubit Fluorometer and **x** is the volume of sample added in µl. To calculate the copy number of ssRNA fragments per µl of sample, an online tool found at www.endmemo.com/bio/dnacopynum.php was used. The fragment size along with the concentration value as determined using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA) was used as input values for the calculation.

3 .14 Phylogenetic analyses

Phylogenetics is the study of evolutionary relatedness between organisms. Using sequence alignments from related organisms evolutionary changes that have taken place are inferred and represented graphically by phylogenetic trees (Sleator, 2013). In essence, the tree depicts the evolutionary relationship between taxa, groups of organisms, where the tree tips represent a group of closely related organisms descending from a common ancestor represented by the tree nodes. An out group is often included when constructing phylogenetic trees and represents a taxon less related to the group of interest in order to orientate the tree according to the bigger tree of life (Anon, 2017).

3.14.1 Multiple sequence alignments

For phylogenetic analysis of sample sequences, closely related viral sequences were downloaded from GenBank, an annotated collection of publically available DNA sequences, on the NCBI website (https://www.ncbi.nlm.nih.gov/genbank/). Multiple sequence alignments (MSA) of nucleotide or amino acid sequences were performed either using the in-program ClustalW algorithm with MEGA6 or MEGA7 or using the online MSA tool, MAFFT v.7 (http://mafft.cbrc.jp/alignment/server/), with default parameters (Tamura *et al.*, 2013; Kumar *et al.*, 2016).

3.14.2 Sequence datasets for phylogenetic analyses

For phylogenetic analysis of sequences obtained from general surveillance screening, two sequence data sets were established; a dataset of nucleotide sequences corresponding to the short 395 bp fragment obtained from the Pan-CoV PCR and a dataset of amino acid sequences corresponding to the translated nucleotide fragments obtained with the Extended RdRp PCR.

All BtCoV sequences corresponding to the 395 bp sequences obtained using the Pan-CoV PCR assay were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) as well as all CoV sequences in the NCBI's Reference Sequence (RefSeq) database (https://www.ncbi.nlm.nih.gov/refseq/) as available on 01 May 2017. Similarly, all available BtCoV sequences were downloaded and trimmed to correspond to the 816 bp RdRp-based grouping unit (RGU) as proposed by Drexler *et al.* (2010) for tentative CoV classification. The RGU fragment corresponds to nucleotides 14781 - 15596 in SARS-CoV strain Frankfurt 1 (GenBank: AY291315) (Drexler *et al.*, 2010). These 816 bp sequences were translated into amino acid sequences using the correct reading frame (Section 3.11.4). Nucleotide and amino acid sequences were aligned and a pairwise distance matrix generated in MEGA7. Redundant identical non-RefSeq sequences were removed as well as closely related sequences with pairwise distances ranging from 0.97 – 0.99 for amino acid sequences and 0.80 – 0.95 for nucleotide sequences depending the dataset. Sample sequences obtained during this study were aligned to this dataset for phylogenetic analyses.

3.14.3 Putative coronavirus classification

Tentative classification of CoVs detected during this study was based on sequences obtained using the Extended RdRp PCR assay (section 3.8.3). The criterion used was based on that published by Drexler *et al.* (2010). Pairwise distance matrices were generated in MEGA7 using amino acid MSA. Based on the generated pairwise distance matrix, values greater than 0.048 (4.8 %) indicated different species for alphaCoVs while values greater than 0.51 (5.1 %) indicated different species for betaCoVs (Drexler *et al.* 2010, Drexler *et al.*, 2014). Pairwise distance values

greater than 0.32 (32.0 %) were used to distinguish between different CoV genera (Drexler *et al.*, 2010).

3.14.4 Phylogenetic substitution model selection

To determine the best suited model for phylogenetic inference of nucleotide and amino acid sequence alignments, a model test was performed using JModelTest 2.1.10 and the online ProtTest 2.4 server (darwin.uvigo.es/software/prottest_server.html), respectively (Abascal *et al.*, 2005; Darriba *et al.*, 2012). Both model tests utilise the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) to compare available models (Fabozzi *et al.*, 2014). The model indicated as the best model by the majority of criterion scores was selected when performing subsequent phylogenetic analyses.

3.14.5 Phylogenetic inference methods

Methods employed to infer phylogenetic relatedness can generally be categorized as either distance- or character-based. Distance-based methods use an algorithm with an evolutionary model to generate a distance matrix based on the number of differences between each pair of sequences (Sleator, 2013). The neighbour-joining (NJ) method is a distance-based method and was used during this study (Saitou and Nei, 1987). Being a distance-based method of phylogenetic inference, NJ has the disadvantages that different trees may result from the same MSA depending on the order in which sequences are added to the alignment and that instead of a consensus tree generated from several trees, only a single tree is produced (Sleator, 2013). Despite these disadvantages, the NJ method is useful for the analysis of shorter nucleotide sequences.

The alternative character-based methods such as maximum likelihood (ML) and Bayesian inference methods used during this study use an algorithm to determine the most probable tree for each MSA analysed. The characters at each position in the MSA are assessed individually to determine a tree score representative of the log-likelihood value and the posterior probability for ML and Bayesian inference trees respectively (Sleator, 2013). Bayesian inference is a probabilistic method that employs the Markov chain Monte Carlo algorithm to determine the most likely phylogenetic tree for the given MSA, with the advantage of being able to account for phylogenetic uncertainty (Yang and Rannala, 1997; Mau *et al.*, 1999). The ML and Bayesian inference methods generally result in trees with improved phylogenetic accuracy.

3.14.5.1 Phylogenetic trees

NJ and ML trees were produced in MEGA7 using nucleotide MSA with the pairwise distance model to infer phylogenetic relatedness (Kumar *et al.*, 2016). All alignment positions with gaps or missing data were deleted. To validate the accuracy of the consensus tree generated, 1000 bootstrap (BS) replicates were performed.

Bootstrapping is a statistical measure assessing the internal support for clades to determine the phylogenetic accuracy of the generated tree. BS replicates randomly re-sample a subset of sequences from the MSA in question several times and report how frequently each clade from the generated tree was observed. The default number of BS replicates is 1000 with a BS value of 70% or more representative of phylogenetic accuracy (Sleator, 2013).

ML trees were constructed using the best-fit model as determined by the model selection test to infer phylogenetic relatedness of nucleotide and amino acid MSA. All BS values less than 70% were omitted in figures depicting phylogenetic trees.

Bayesian inference was conducted in Mr Bayes 3.2.6 using amino acid MSA with standard parameters according to the evolutionary model indicated by the model selection test (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; Ronquist *et al.*, 2012). Run settings were adjusted for 2 million generations with the Markov chain to be sampled every 1000 trees. Chain information was printed every 1000 samples and the first 25% of sampled trees were discarded from each chain. Bayesian inference trees were viewed and edited in Figtree 1.4.1 and all probability values less than 0.9 were omitted in figures depicting Bayesian phylogenetic trees.

3.15 Ecological analyses using logistic regression

To test the effects of extrinsic and intrinsic factors on the prevalence of CoV infection in *N. capensis* bats, binomial logistic regressions were conducted in R (v. 3.3.1 R Core Team, 2016). Logistic regressions were conducted using the glm with the logit function. All models were run through the intercept.

First, life history trait and biogeographic data collected at the study sites were assessed for autocorrelation. Where variables were highly correlated to each other ($R^2 > 0.7$), only one variable was used in subsequent analyses. The small number of juveniles present in the dataset additionally led to the exclusion of age as a variable due to its significant weakening of the model. Four intrinsic and five extrinsic variables were included in the logistic regression analyses (Table 3.5). The best model was identified with the corrected Akaike Information Citerion (AICc) using the glmulti package in R (Calcagno and de Mazancourt, 2010).

Table 3.5 Extrinsic and intrinsic variables included in logistic regression analyses.

Intrinsic variables	Extrinsic variables
forearm length	altitude
forearm mass index (FMI)	total monthly rainfall
sex	biome
reproductive status	maximum monthly temperature
	minimum monthly temperature

Model accuracy, model support, goodness of fit, strength of association, and significance of individual predictors were evaluated using bootstrapping, assessment of the residual deviance, McFadden's pseudo R^{2}, chi-square test, and Wald's Chi-square tests, respectively.

Chapter 4: Results

This Chapter is set out to reflect the outline of Chapter 3: Methods where possible. The chapter starts with the bat trapping sites that were used during the study followed by the physiological and biogeographic data collected from trapped bats. Each of these sections is reported according to the three different sampling arms, namely general surveillance, species-specific surveillance of *Neoromicia capensis* bats, and longitudinal surveillance of *N. capensis* bats. The screening results are additionally divided into results obtained for all samples using the Pan-CoV PCR assay and results obtained for the subset of samples that underwent repeat screening with the Extended RdRp assay. The chapter goes on to report the findings of the molecular assays for species confirmation before detailing the extended genome amplification results of detected betaCoVs. The results from the phylogenetic analyses of CoV sequences follows on from this and is divided into the phylogenetic analyses results of general surveillance sequences, *N. capensis* species-specific sequences, and additional phylogenetic analyses from the extended genome amplification sequences. The chapter ends with the description of the results following biogeographic analyses to discern if any extrinsic or intrinsic factors, such as biome or sex, respectively, may serve as potential predictors of CoV infection in South African *N. capensis* bat populations.

For ease of reference, a list of bat species codes is given in Appendix C.

4.1 Bat trapping sites

Between January 2014 and October 2016, bat trapping took place at 53 different sites across 5 provinces: KwaZulu-Natal (n = 21), Eastern Cape (n = 10), Western Cape (n = 12), and Northern Cape (n = 6), as well as in Limpopo (n = 4). All bat trapping sites, with abbreviated site codes, along with the corresponding GPS co-ordinates in decimal point form are listed in Appendix C Table 1. For each site listed in Appendix C Table 1 it has been indicated if samples collected there were used for the general, species-specific, or longitudinal surveillance part of this study. For two sites, namely Hopefield Farm (HFP) in the Northern Cape and Babanango Valley (BVL) in KwaZulu-Natal, bats were sampled at multiple locations at the site and therefore depending on the data displayed, at times these locations may be presented collectively.

Samples for the general surveillance aspect of this study were obtained from 45 sites while samples for the species-specific surveillance of *N. capensis* bats were collected at 20 different sites. For the longitudinal study of *N. capensis* colonies, a single CoV-infected colony where regular monthly sampling could take place was identified. Samples for this part of the study were therefore collected at a single site, Cloeteskraal Farm (CCK), located in Velddrif, Western Cape.

4.2 Bat trapping

As part of the general and species-specific surveillance aims of this study, 600 individual bats were trapped and sampled. In an attempt to investigate CoV infection patterns in the *N. capensis* bat colony longitudinally, individual bats were, except for an initial sampling effort, not trapped and instead pooled data was passively collected for eight months (Chapter 3 Section 3.4.3). The morphological and biogeographic data collected for the three different sampling arms are detailed below under their respective headings.

4.3 Morphological, phyisiological and biogeographic data

4.3.1 General surveillance

Faecal pellets from 404 individual bats across 20 different bat species from 5 different bat families were collected. These individuals were trapped and sampled across 45 different sites (Appendix C Table 1). The sample distribution by species is shown in Table 4.1 along with the number of each species sampled in each province.

Table 4.1 Geographic distribution of samples collected as part of the general surveillance effort. The table lists the number of each species sampled across the different provinces. Bat species are grouped by family.

Bat family	Bat species	Eastern Cape	KwaZulu-Natal	Limpopo	Northern Cape	Western Cape	Total number sampled
Hipposideridae	*Hipposideros caffer*	0	**3**	0	0	0	**3**
Miniopteridae	*Miniopterus fraterculus*	0	**6**	0	0	0	**6**
	Miniopterus natalensis	**6**	**12**	**1**	**3**	**8**	**30**
Molossidae	*Chaerephon pumilus*	0	**15**	0	0	0	**15**
	Mops midas	0	0	**5**	0	0	**5**
	Tadarida aegyptiaca	0	**4**	0	0	**4**	**8**
Nycteridae	*Nycteris thebaica*	0	1	0	0	0	**1**
Rhinolophidae	*Rhinolophus capensis*	**31**	0	0	0	**21**	**52**
	Rhinolophus clivosus	**9**	**57**	0	**24**	**21**	**111**
	Rhinolophus darlingi	0	0	0	**13**	0	**13**
	Rhinolophus denti	0	0	0	**14**	0	**14**
	Rhinolophus simulator	0	**15**	0	0	0	**15**
	Rhinolophus swinnyi	**5**	**2**	0	0	0	**7**
Vespertilionidae	*Neoromicia nana*	0	**29**	0	0	**12**	**41**
	Pipistrellus hesperidus	**9**	**25**	**7**	**1**	**7**	**49**
	Kerivoula lanosa	**1**	0	0	0	0	**1**
	Laephotus botswanae	**3**	0	0	0	0	**3**
	Myotis bocagii	0	**3**	**1**	0	0	**4**
	Myotis tricolor	**5**	**10**	0	0	**1**	**16**
	Scotophilus dinganii	**1**	**8**	**1**	0	0	**10**
Total		**70**	**190**	**15**	**55**	**74**	**404**

4.3.1.1 Weight and forearm length

As detailed in Chapter 3 Section 3.4, forearm length and weight were used as species identifier tools in conjunction with a number of other characteristics to identify bats in the field. The range, indicated by the maximum and minimum measurement taken, for weight and forearm length measurements by each species sampled is indicated in Table 4.2. All listed ranges were in accordance with those designated to each respective species for identification purposes (Monadjem *et al.*, 2010).

Table 4.2 The weight and forearm measurement ranges of sampled bats according to species.

Bat family	Bat species	Weight (g)		Forearm length (mm)	
		Min	Max	Min	Max
Hipposideridae	*Hipposideros caffer*	7.5	7.9	46.5	47.5
Miniopteridae	*Miniopterus fraterculus*	7.0	9.0	43.0	44.0
	Miniopterus natalensis	7.0	14.5	42.0	48.0
Molossidae	*Chaerephon pumilus*	8.0	11.0	36.0	40.0
	Mops midas	33.0	44.0	61.0	65.0
	Tadarida aegyptiaca	13.0	18.0	45.0	50.0
Nycteridae	*Nycteris thebaica*[1]	11.0	11.0	50.0	50.0
Rhinolophidae	*Rhinolophus capensis*	9.0	15.0	47.0	52.0
	Rhinolophus clivosus	12.0	27.5	48.0	58.0
	Rhinolophus darlingi	8.0	12.0	47.0	53.0
	Rhinolophus denti	5.5	9.0	41.5	43.5
	Rhinolophus simulator	6.0	10.5	47.0	43.0
	Rhinolophus swinnyi	7.0	8.5	45.0	41.0
Vespertilionidae	*Kerivoula lanosa*[1]	5.0	5.0	33.0	33.0
	Laephotus botswanae	7.0	9.0	36.0	37.0
	Myotis bocagii	8.0	9.0	39.0	39.0
	Myotis tricolor	9.0	17.0	48.0	51.0
	Neoromicia nana	3.0	4.5	29.0	33.0
	Pipistrellus hesperidus	3.5	8.0	30.0	35.0
	Scotophilus dinganii	21.0	34.0	51.0	55.0

[1]Only one individual sampled

4.3.1.2 Age, sex, and reproductive status

All bats trapped and sampled for general surveillance were identified as adult individuals. The 404 samples were collected from 180 (44.5%) female, 222 (55.0%) male, and 2 (0.5%) individuals that were not sexed due to the individual bats having escaped before being processed.

The reproductive status of 402 of the 404 sampled bats was known; the two individual bats for which no sex was recorded similarly had no reproductive status data. Of the 180 female bats from which samples were collected, 143 (79.4%) were not pregnant, 16 (8.8%) were pregnant at the time of sampling, 11 (6.1%) were lactating, 8 (4.4%) were found to be post-lactating and 2 (1.1%) female bats were trapped with their pups which were not sampled nor included in the study. Of the 222 male bats, 110 (49.5%) were scrotal and 112 (50.5%) were non-scrotal where scrotal indicated male bats to be sexually active at the time of sampling.

4.3.1.3 Geographic data

The 404 samples for the general surveillance effort were collected across the 8 different biomes described in Chapter 2 and 3 namely, Albany Thicket (n = 31, 7.7%), Indian Ocean Coastal Belt (n = 64, 15.8%), Forest (n = 13, 3.2%), Fynbos (n = 25, 6.2%), Grassland (n = 32, 7.9%), Nama-Karoo (n = 1, 0.3%), Savanna (n = 214, 53%), and Succulent Karoo (n = 24, 5.9%). Bat trapping took place across an altitude gradient ranging from 10 to 1496 m above sea level. The altitude and biome for each bat trapping site is listed in Appendix C Table 1.

Weather data was obtained from the South African Weather Service (www.weathersa.co.za) for all sampling sites corresponding to the month during which sampling took place. Appendix C Table 2 lists the weather stations from which data was obtained for each bat trapping site, together with the accompanying weather data.

4.3.2 Species-specific surveillance

For the species-specific surveillance of *N. capensis* bats, 196 individual faecal pellet samples were collected from 20 different bat trapping sites (Appendix C Table 1) in the Eastern Cape (n = 43, 21.9%), KwaZulu-Natal (n = 39, 19.9%), Northern Cape (n = 40, 20.4%), and Western Cape (n = 74, 37.8%) provinces. At certain bat trapping sites namely ABA, LNR, HFP, CDK, FEK, and CGC, sampling took place more than once during the study period.

4.3.2.1 Weight and forearm length

As described in Chapter 3 Section 3.4, weight and forearm length measurements were used as species identifier tools during sampling events. The weight of the trapped *N. capensis* bats ranged from 4.0 to 12.0 g with forearm lengths ranging from 30.5 to 38 mm. The generally accepted weight range for *N. capensis* bats is 3.4 to 10.1 g (Monadjem *et al.*, 2010). Here the higher weight range is due to the presence of pregnant females in the sample subset.

4.3.2.2 Age, sex, and reproductive status

Of the 196 *N. capensis* samples, 188 (95.9%) were identified as adult and 8 (4.1%) were identified as juveniles. Of the juvenile bats, 2 were male and 6 were female. Overall, 52 male (26.5%) and 144 (73.5%) female *N. capensis* bats were sampled.

Reproductive status data was available for all 196 *N. capensis* bats. Of the 144 female bats, 33 (22.9%) were identified as pregnant and 12 (8.3%) as post-lactating. For male bats, 19 (36.5%) were scrotal and 33 (63.5%) were non-scrotal at the time of sampling.

4.3.2.4 Geographic data

Overall 196 individual *N. capensis* bats were sampled across six biomes namely, Albany Thicket (n = 20, 10.2%), Forest (n = 5, 2.5%), Fynbos (n = 69, 35.2%), Grassland (n = 52, 26.5%), Nama-Karoo (n = 4, 2.1%), and Savanna (n = 46, 23.5%). The *N. capensis* bat trapping sites across the different biomes are indicated on the map in Figure 4.1. The demarcation of biomes was based on data obtained from the South African National Biodiversity Institute (SANBI).

The sampling took place across an altitudinal gradient ranging from 6 to 1568 m above sea level. Appendix C Table 1 details the sites where *N. capensis* bats were trapped along with each site's corresponding altitude, and biome.

Weather data was obtained from the South African Weather Service (www.weathersa.co.za) for all sampling sites corresponding to the month during which sampling took place. Appendix C Table 2 lists the weather stations from which data was obtained for each bat trapping site, together with the accompanying weather data.
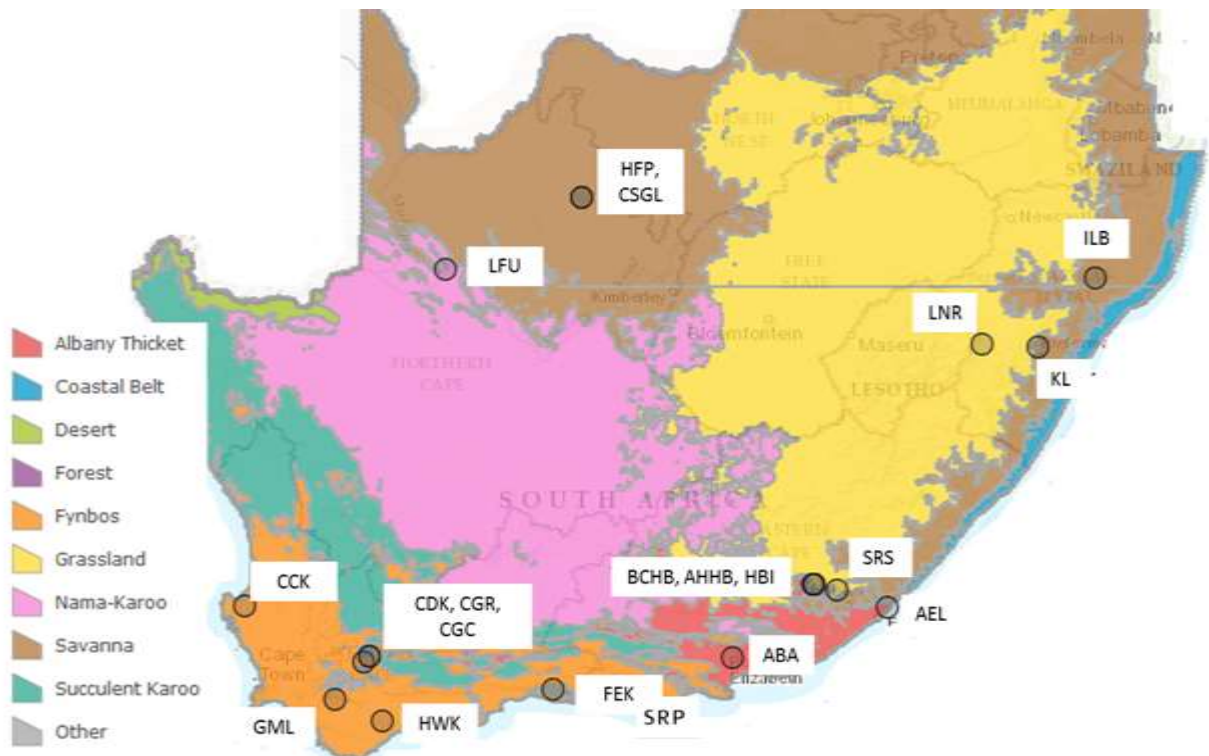
Figure 4.1 Distribution of *Neoromicia capensis* bat trapping sites across the different biomes of South Africa. The map was created using the ArcGIS® software online tool by Esri © 1995-2017 (www.esri.com/arcgis). The biome distribution data was obtained from SANBI (www.bgis.sanbi.org) based on the 2006 vegetation map (Mucina et *al*,. 2006; Rutherford et *al*., 2006).

4.3.3 Longitudinal surveillance of a *Neoromicia capensis* colony

In 2015 a *N. capensis* colony with the potential to conduct a longitudinal study of CoV prevalence and amplification patterns was identified on Cloeteskraal Farm situated in Velddrif in the Western Cape. At the time of identifying the colony, 5 individual *N. capensis* bats were caught and sampled. The data collected for these individuals has been included in Section 4.3.2 and is briefly summarised below.

4.3.3.1 Weight and forearm

As indicated in Section 4.3.2, the weight and range of these 5 individual bats fell within the expected range for the species. The weight ranged from 8.0 to 9.0 g and the forearm length measurements ranged from 37.0 to 38.0 mm.

4.3.3.2 Age, sex, and reproductive status

All 5 bats were identified as adults, 4 non-pregnant females and 1 scrotal male.

4.3.3.4 Geographic data

Cloeteskraal Farm, Veldrif, Western Cape province, is located in the Fynbos biome at an altitude of 17 metres above sea level.

Monthly collection of pooled sampling commenced with the aim to monitor CoV viral loads over time. Ten pooled samples of 5 pellets each were collected each month in 2015 from March to May, and again from September to December. No samples were collected for the months of June, July, and August in 2015 when winter conditions led to temporary migration of the bat colony. Five pooled samples of 5 pellets each were collected in 2016 during the months of January and February. The reduced number of pooled samples collected for these months and the subsequent suspension of sampling occurred following prolonged drought conditions leading to greatly reduced bat numbers in the colonies that have subsequently not recovered (Personal communication, Cloeteskraal farm owner Mr Q Laubscher).

4.4 Coronavirus screening results

All collected faecal samples described in the previous section underwent nucleic acid extraction and reverse transcription before being subjected to the various PCR assays described in Chapter 3 Section 3.8. PCR results were determined by horizontal DNA gel electrophoresis and confirmed with Sanger sequencing and BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) following PCR product purification. When necessary, purified PCR products underwent cloning to obtain improved sequence data as described in Chapter 3 Section 3.12. The PCR results obtained during this study are detailed in the following sections according to the different sampling arms of this study. A combined summary of results obtained for the two different screening approaches are detailed in Sections 4.4.1.3 and 4.4.2.2 for general surveillance species-specific results respectively.

4.4.1 General surveillance

4.4.1.1 Pan-coronavirus PCR assay results

4.4.1.1.2 By species

All 404 faecal pellet samples collected for general surveillance were screened with the Pan-CoV PCR assay as described in Chapter 3 Section 3.8 (de Souza Luna *et al.*, 2007). The Pan-CoV PCR assay detected 10 (2.5%) positive samples from 3 *Miniopterus natalensis,* 1 *Rhinolophus capensis*, 2 *R. clivosus*, 1 *R. simulator*, 1 *Scotophilus dinganii,* and 2 *Pipistrellus hesperidus* bat samples (Table 4.3). Partial CoV RdRp sequences corresponding to positions 14417 – 14859 in *Rousettus* bat CoV (BtCoV) HKU10, GenBank number: NC_018871 were obtained from all 10 positive samples with blastn results indicating all to belong to the alphaCoV genus except for 2 sequences obtained from *P. hesperidus* samples shown to be betaCoVs.

81

Two sequences obtained from *M. natalensis* bats showed highest sequence identity (99.0%) to alphaCoV sequences obtained previously from a *M. natalensis* bat in South Africa (Genbank ID KF843851) while the third sequence from a *M. natalensis* bat showed highest sequence identity (89%) to an alphaCoV sequence from a *M. schreibersii* bat in Australia (Genbank ID EU834955) (Ithete, 2013; Smith *et al.*, 2016). The three sequences from *R. capensis* and *R. clivosus* bats showed highest sequence identity (88 – 89%) to alphaCoV sequences from *Rhinolophus* sp. bats in Kenya (Genbank ID GU065417 and GU065418, unpublished study by Conrardy *et al.*, 2009). The sequence from the *S. dinganii* bat showed highest sequence identity (96%) to an alphaCoV sequence from a *S. heathii* bat in Thailand (Genbank ID KJ020603) (Wacharapluesadee *et al.*, 2015).

Two betaCoV sequences obtained from *P. hesperidus* bats showed highest sequence identity (94%) to MERS-CoV sequences from a dromedary camel in the United Arab Emirates (UAE, Genbank ID KT751244) (Wernery *et al.*, 2015).

Appendix D Table 1 details the blastn results for all sequences obtained by screening of general surveillance samples. The table indicates which sequences were obtained with both assays and those that were only obtained with the Extended RdRp PCR assay. The Genbank accession number of the closest related sequence as ranked by percentage identity is indicated as well as the species and country from which this sequence originated.

4.4.1.1.2 By morphological and physiological characteristics

The 10 CoV-positive samples originated from 6 female and 4 male bats. At the time of sampling, 1 female bat was lactating, 1 female bat was post-lactating, and 2 male bats were scrotal (Table 4.4).

4.4.1.1.3 By geographic characteristics

The 10 CoV positive bats were trapped in the Eastern Cape (n =1), Limpopo (n =1), KwaZulu-Natal (n = 4), and Western Cape (n = 4) provinces. The bat trapping sites in these provinces were located in the Succulent Karoo (n = 1), Forest (n = 3), Savanna (n = 4), Albany Thicket (n = 1), and Indian Ocean Coastal Belt (n = 1) biomes (Table 4.5).

4.4.1.2 Extended RdRp PCR assay

4.4.1.2.1 By species

As detailed in Chapter 3 Section 3.8 repeat screening of a subset of samples was undertaken using the Extended RdRp PCR assay (Drexler *et al.*, 2010). Of the 404 samples collected as part of the general surveillance effort, 250 (62.0%) underwent repeat screening. Using the Extended RdRp assay, 17 (6.8%) of these samples were identified as positive, including all 8 samples in the subset that had previously been identified as positive using the Pan-CoV screening assay (Table 4.5). The additional 9 positive samples identified by the Extended RdRp assay were from 1

Chaerephon pumilus, 3 *M. natalensis*, 3 *P. hesperidus*, and 2 *Neoromicia nana* bats. BLAST results indicated all sequences obtained to belong to the alphaCoV genus except for those from *P. hesperidus* samples that were shown to be betaCoVs.

The longer partial RdRp sequences obtained using the Extended RdRp PCR assay corresponds to an 816 bp region, position 13898 to 14870 in the *Rousettus* BtCoV HKU10 (Genbank ID NC_018871), which can be used for provisional CoV species classification. This assay has not been as widely used as CoV screening assays generating shorter sequences resulting in limited availability of published and submitted sequences in the Genbank database for comparison.

The sequence obtained from *C. pumilus* showed highest sequence identity (96%) to an alphaCoV sequence from a *Chaerephon* sp. bat in Kenya (Genbank ID HQ728486) (Tong *et al.*, 2009). Two of the three sequences from *M. natalensis* showed highest sequence identity (91 – 96%) to alphaCoV sequences from *M. natalensis* bats in Kenya (Genbank ID GU065410 and GU065411, unpublished study by Conrardy *et al.,* 2009) while the third *M. natalensis*-derived sequence showed highest similarity (98%) to an alphaCoV sequence from an *M. natalensis* bat in South Africa (Genbank ID KF843851) similar to the *M. natalensis*-derived sequences obtained with the Pan-CoV PCR assay (Ithete, 2013). Two *N. nana*-derived sequences both showed highest sequence identity to *N. capensis*-derived AlphaCoV sequences from South Africa. One of the *N. nana*-derived sequences had highest sequence identity (96%) to an *N. capensis*-derived sequence (Genbank ID KF843854) that has been shown to be closely related to other *N. capensis*-derived sequences (Ithete, 2013). The second *N. nana*-derived sequence, 20160411DC63_NN, showed highest sequence identity (94%) to an *N. capensis*-derived sequence (Genbank ID KF843858) that has been shown not to cluster phylogenetically with other *N. capensis*-derived alphaCoV sequences (Ithete, 2013).

Two of the three betaCoV sequences from *P. hesperidus* bats showed highest sequence identity (91 – 94%) to dromedary camel-derived MERS-CoV sequences from the UAE (Genbank ID KT751244) similar to the betaCoV sequences obtained using the Pan-CoV PCR assay (Wernery *et al.*, 2015). The third betaCoV sequence showed highest sequence identity (98%) to a betaCoV sequence, named NeoCoV, previously detected in a *N. capensis* bat in South Africa (Genbank ID KC869678) that was shown to be a MERS-related CoV (Ithete *et al.*, 2013; Corman *et al.*, 2014a).

4.4.1.2.2 By morphological and physiological characteristics

The subset of samples that underwent repeat screening consisted of samples from 113 (45.2%) female and 137 (54.8%) male bats. Of the positive samples detected in the subset, 10 (58.8%) were obtained from female bats and 7 (41.2%) were obtained from male bats (Table 4.4).

4.4.1.2.3 By geographic characteristics

The additional positive samples identified using the Extended RdRp PCR assay were obtained from the Eastern Cape (n = 1), KwaZulu-Natal (n = 5) and Western Cape (n = 3). The bat trapping sites where these samples were obtained were located in the Albany Thicket (n = 1), Indian Ocean Coastal Belt (n = 4), Savannah (n = 1), Fynbos (n = 2), and Nama-Karoo (n = 1) as detailed in Table 4.5.

4.4.1.3 Collective overview of general surveillance screening results

With the Extended RdRp PCR assay only being used on a subset of 250 samples, the overall results from both PCR assays is summarised by species in Table 4.3, morphological characteristics in Table 4.4 and by geographic factors in Table 4.5.

From Table 4.3 it can be seen that for the subset of samples that underwent repeat screening, the Extended RdRp PCR assay detected more than twice as many positive samples compared to the Pan-CoV PCR assay. The Pan-CoV PCR failed to detect CoVs in two different bat species, *N. nana* and *C. pumilus*, that were found to carry CoVs when using the Extended RdRp PCR. Using the Extended RdRp PCR increased the detection of CoVs in samples from *M. natalensis* and *P. hesperidus* bats.

Table 4.3 Summarised results obtained using the Pan-CoV and Extended RdRp assays by bat species sampled. Results from screening the complete general surveillance sample dataset as well as the results for the subset of samples that underwent repeat screening with the Extended RdRp PCR assay are detailed by bat species. The original Pan-CoV PCR assay screening results for samples included in the subset for repeat screening are provided for comparative purposes.

Bat family	Bat species	Total sampled	Full dataset positive Pan-CoV results	Repeat screened subset		
				Total in subset for repeat screening	Original Pan-CoV positive results for subset	Ext RdRp positive results for subset
Hipposideridae	*Hipposideros caffer*	3	0	0	0	0
Miniopteridae	*Miniopterus fraterculus*	6	0	7	0	0
	Miniopterus natalensis	29	3 (10.3%)	18	2 (11.1%)	5 (27.8%)
Molossidae	*Chaerephon pumilus*	15	0	14	0	1(7.1%)
	Mops midas	5	0	5	0	0
	Tadarida aegyptiacus	8	0	6	0	0
Nycteridae	*Nycteris thebaica*	1	0	0	0	0
Rhinolophidae	*Rhinolophus capensis*	52	1 (1.9%)	27	0	0
	Rhinolophus clivosus	111	2 (1.8%)	71	2 (2.8%)	2 (2.8%)
	Rhinolophus darlingi	13	0	9	0	0
	Rhinolophus denti	14	0	8	0	0
	Rhinolophus simulator	15	1 (6.6%)	11	1 (9.1%0	1 (9.1%)
	Rhinolophus swinnyi	7	0	1	0	0
Vepertilionidae	*Neoromicia nana*	41	0	23	0	2 (8.7%)
	Pipistrellus hesperidus	49	2 (4.1%)	32	2 (6.3%)	5 (15.6%)
	Kerivoula lanosa	1	0	0	0	0
	Laephotus botswanae	3	0	0	0	0
	Myotis bocagii	4	0	4	0	0
	Myotis tricolor	16	0	5	0	0
	Scotophilus dinganii	11	1 (9.1%)	9	1 (11.1%)	1 (11.1%)
Total		**404**	**10 (2.5%)**	**250**	**8 (3.2%)**	**17 (6.8%)**

During this study for general surveillance purposes, more samples were obtained from male bats than female bats, 55.0% compared to 44.5%. A similar ratio of male to female samples was included in the subset for rescreening. From Table 4.4 it can be seen that positive samples were mostly obtained from female bats. The difference in the number of CoVs between male and female

individuals was however not found to be significant regardless of which screening PCR assay was used.

Table 4.4 Summarised results obtained using the Pan-CoV and Extended RdRp assays by bat morphological characteristics, sex and reproductive status. Results from screening the complete general surveillance sample dataset as well as the results for the subset of samples that underwent repeat screening with the Extended RdRp PCR assay are detailed by sex and reproductive status. The original Pan-CoV PCR assay screening results for samples included in the subset for repeat screening are provided for comparative purposes.

Sex and reproductive status	Total sampled	Full dataset positive Pan-CoV results	Total in subset for repeat screening	Original Pan-CoV positive results for subset	Ext RdRp positive results for subset
Female	**180 (44.5%)**	**6 (3.3%)**	**113 (45.2%)**	**5 (4.4%)**	**10 (8.9%)**
Pregnant	16	0	4	0	2 (12.5%)
Not pregnant	143	4 (2.8%)	97	3 (3.1%)	6 (6.2%)
Lactating	11	1 (0.9%0	10	1 (10.0%)	1 (10.0%)
Post lactating	8	1 (12.5%)	2	1 (50.0%)	1 (50.0%)
With pup	2	0	0	0	0
Male	**222 (55.0%)**	**4 (1.8%)**	**137 (54.8%)**	**3 (2.2%)**	**7 (5.1%)**
Scrotal	110	2 (1.8%)	76	1 (1.3%)	2 (2.6%)
Non-scrotal	112	2 (1.8%)	61	2 (3.3%)	5 (8.2%)
Unknown	**2 (0.5%)**	**0**	**0**	**0**	**0**

From Table 4.5 it can be seen that for general surveillance samples, no CoV positive bats were identified in the grassland and Nama-Karoo biomes. Using the Extended RdRp PCR assay, additional CoV positive bats were identified in the Albany Thicket, Fynbos, Indian Ocean Coastal Belt, and Savanna biomes.

Table 4.5 Summarised results obtained using the Pan-CoV and Extended RdRp assays by geographic characteristics of the bat trapping site where positive individuals were sampled. Results from screening the complete general surveillance sample dataset as well as the results for the subset of samples that underwent repeat screening with the Extended RdRp PCR assay are detailed by the biomes in which bat trapping took place. The original Pan-CoV PCR assay screening results for samples included in the subset for repeat screening are provided for comparative purposes.

Biome	Total sampled	Full dataset positive Pan-CoV results	Repeat screened subset		
			Total in subset for repeat screening	Original Pan-CoV positive results for subset	Ext RdRp positive results for subset
Albany thicket	31	1 (3.2%)	9	0	1 (11.1%)
Forest	13	3(23.1%)	11	3 (27.3%)	3 (27.3%)
Fynbos	25	0	17	0	2 (11.8%)
Indian Ocean Coastal Belt	64	1 (1.6%)	47	1 (2.1%)	5 (10.6%)
Grassland	32	0	16	0	0
Savanna	214	4 (1.9%)	126	4 (3.2%)	5 (4.0%)
Nama-Karoo	1	0	1	0	0
Succulent Karoo	24	1 (4.2%)	23	0	1 (4.4%)
Total	**404**	10 (2.5%)	**250**	8 (3.2%)	17 (6.8%)

4.4.2 Species-specific surveillance of *Neoromicia capensis* bats

4.4.2.1 Pan-CoV PCR assay results

All 196 faecal pellets collected from *N. capensis* bats were screened for the presence of CoV using the Pan CoV PCR assay as described in Chapter 3 Section 3.8.2 (de Souza Luna *et al.*, 2007). The PCR assay detected 46 (23.5%) positive samples. Sequences were obtained for 44 of the positive samples with blastn results indicating all of these sequences to be closely related to members of the *AlphaCoV* genus. These alphaCoVs had the closest sequence identity (96 – 100%) to other closely related alphaCoV sequences previously obtained from South African *N. capensis* bats; Genbank accession ID KF843855 – KF843857, KF843859 – KF843862, and JQ19818 (Geldenhuys *et al.*, 2013; Ithete, 2013). These published South African *N. capensis* bats have shown close relatedness to alphaCoVs previously detected in *Nyctalus* sp. bats from China (Genbank ID KJ473809), Spain (Genbank ID HQ184055) and the Netherlands (Genbank ID GQ259963) (Reusken *et al.*, 2010; Falcón *et al.*, 2011; Geldenhuys *et al.*, 2013; Ithete, 2013; Wu *et al.*, 2016).

One alphaCoV sequence obtained from a bat at Drie Kuilen Nature Reserve (20150106CDK_NC1) showed high sequence identity (99%) to an alphaCoV sequence (Genbank ID KF843858) detected in a *N. capensis* bat in South Africa that was shown not to cluster phylogenetically with other *N. capensis*-derived alphaCoV sequences (Ithete, 2013). The second highest sequence identity (79%) was to a sequence obtained from a *Myotis daubentonii* bat (KF569975.1) from China (He *et al.*, 2014a). This blastn result was similar to that recorded for the sequence from sample 20160411DC63_NN obtained during the screening of general surveillance samples using the Extended RdRp PCR (Appendix D).

For two samples, sequences could not be obtained from the PCR products generated with the Pan-CoV PCR assay. Sequences were later obtained using the Extended RdRp PCR assay which blastn results indicated both to be betaCoVs with high sequence identity (94%) to NeoCoV (Genbank KC869678) (Ithete *et al.*, 2013; Corman *et al.*, 2014a).

For one sample, 20140923LNR_NC7, collected at the Lotheni Nature Reserve (LNR) in KwaZulu-Natal, examination of the sequence chromatogram revealed the presence of double peaks that indicated possible coinfection. When this sequence was compared against 7 sequences from bats sampled at the same site on the same day, it appeared that this sample contained 2 alphaCoV sequence strains. Within the 395 bp sequence, 8 polymorphic nucleotide positions were identified. Figure 4.2 depicts these nucleotide positions across the different sequence strains and the sequence from sample 20140923LNR_NC7.
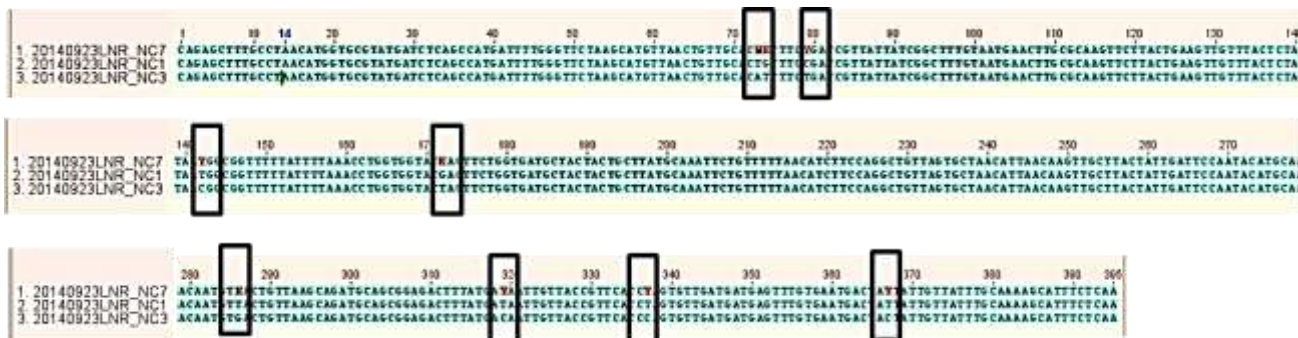
Figure 4.2 AlphaCoV sequence strains detected in *Neoromicia capensis* bats sampled during the same sampling event demonstrating the presence of two sequence variants and the presence of both these sequences in a single sample, 20140923LNR_NC7. Sequence 20140923LNR_NC1 and 20140923LNR_NC3 are representative of the two variants detected within the sample group. Alignment image adapted from Geneious R10 (Biomatters Inc., New Zealand).

These 8 nucleotide sites result in 1 non-synonymous and 7 synonymous differences on the amino acid level. The non-synonymous difference at nucleotide position 74, G ⇆ T resulted in an amino acid switch of alanine ⇆ serine at residue 25 when translated in the second reading frame. Alanine is an aliphatic non-polar hydrophobic amino acid while serine is a polar hydroxyl hydrophilic amino acid. For the purposes of reference for presentation of results and discussion, the different strains have arbitrarily been designated as strain 2014/1 and strain 2014/2 with strain 2014/1 having the alanine at residue 25.

This site was sampled twice during this study. Following the identification of the different strains, assessment of the 7 sequences obtained from the first sampling event in September 2014 indicated that 2 samples had the strain 2014/1 sequence, 4 had the strain 2014/2 sequence, and 1, 20140923LNR_NC7, had both strains. The site underwent repeat sampling in January 2016 and 4 sequences were obtained. On the amino acid level, all 4 sequences contained the alanine amino acid at residue position 25 as observed in strain 2014/1. Comparison of these sequences to those obtained in September 2014 on the nucleotide level revealed the presence of neither strain 2014/1 or strain 2014/2 but instead appeared to be a combination of the two 2014 strains resulting in strain 2016/1. Additionally, 2 nucleotide positions namely 328 and 391 differed in the 2016 strain to both 2014 strains that resulted in synonymous changes at the amino acid level.

The nucleotide differences observed between the different strains identified are indicated in Table 4.6.

Table 4.6 Nucleotide differences between the three *Neoromicia capensis*-derived sequence strains identified from the Lotheni Nature Reserve bat trapping site based on the 395 bp fragment obtained with the Pan-CoV PCR assay. The nucleotide positions in the 395 bp fragment where differences were observed are indicated. The nucleotide position where the difference results in a non-synonymous amino acid change is indicated in bold font.

Strain	Nucleotide position									
	73	**74**	142	172	286	319	328	337	367	391
2014/1	T	**G**	T	G	T	T	C	T	T	C
2014/2	A	**T**	C	T	G	C	C	C	C	C
2016/1	A	**G**	T	G	G	C	T	T	C	T

The results following the clonal sequencing of the PCR product from the coinfected sample, 20140923LNR_NC7, to further investigate the diversity of alphaCoV sequences within this sample are detailed later in Section 4.7.

4.4.2.1.1 By morphological and physiologic characteristics:

AlphaCoV sequences were detected in 5 juvenile bats and 39 adult bats. The 2 betaCoV sequences were detected in samples from 1 male and 1 female adult bat. Of the 196 *N. capensis* samples screened using the Pan-CoV PCR assay, samples from 37 of 144 (25.7%) female and 9 of 52 (17.3%) male bats were identified as positive. The CoV sequences obtained from female bats came from 4 pregnant, 4 post-lactating, and 29 individuals that were not pregnant. The CoV sequences obtained from male bats came from 7 non-scrotal and 2 scrotal individuals (Table 4.7).

4.4.2.1.1 By geographic characteristics

The 46 CoV positive samples as identified using the Pan-CoV PCR assay were detected in bats sampled in the Eastern Cape (n = 8), KwaZulu-Natal (n = 11), and Western Cape (n = 27). No samples collected from bats in the Northern Cape were found to be positive using this assay. The two betaCoV sequences were obtained from bats sampled at different sites in the Eastern and Western Cape. By biome, these 46 samples were collected from bats across the Albany Thicket (n = 7), Fynbos (n = 23), Forest (n = 4), and Grassland (n = 12) and were distributed across the altitude range of 6 to 1568 metres above sea level. No positive samples were detected in the Nama-Karoo and Savanna biomes (Table 4.8).

As detailed in Section 4.3.2, repeat sampling took place at 6 bat trapping sites. The Pan-CoV PCR assay results from all sites where *N. capensis* surveillance samples were collected are detailed in Appendix D Table 2. For site Addo Backpackers (ABA) in the Eastern Cape, 5 of 11 (45.5%) samples were positive and found to have alphaCoV sequences during the first sampling event in October 2014. The second sampling event in July of the following year detected 1 positive bat from 7 (14.3%) sampled; the sequence from this sample was identified as a betaCoV.

Repeat sampling at Lotheni Nature Reserve (LNR) produced similar results across the 2 sampling events in September 2014 and January 2016. During the September 2014 trip, 7 of 19 (36.8%) bats were positive for alphaCoV while in January 2016, 4 of 10 (40.0%) bats were positive for alphaCoVs.

A decrease in the number of positive bats sampled between January and September 2015 was observed at Drie Kuilen Reserve (CDK) where 4 of 12 (33.3%) bats were positive for alphaCoVs in January and only 3 of 17 (17.6%) bats were positive in September. A similar observation was made at a nearby site, Gecko Rock Cottage (CGC), where 2 of 5 (40.0%) sampled bats were positive for alphaCoV in January but only 3 of 12 (25.0%) were positive in September with 1 sequence identified as a betaCoV rather than an alphaCoV.

4.4.2.2. Extended RdRp PCR assay results

As previously mentioned, repeat screening of a subset of samples was undertaken using the Extended RdRp PCR assay (Drexler *et al.*, 2010). Of the *N. capensis* samples collected, 84 underwent repeat screening. The Extended RdRp PCR assay detected 47 (56.0%) CoV positive samples. These positive samples included all 32 samples in the subset previously indicated to be positive using the Pan-CoV PCR assay. The Extended RdRp PCR assay detected an additional 15 CoV positive samples that blastn results indicated to be 1 alphaCoV and 14 betaCoV sequences. Additionally, the Extended RdRp PCR detected 10 coinfected samples. These samples had previously been found to only contain alphaCoV sequences using the Pan-CoV PCR assay. BLAST results of the sequences obtained for these coinfected samples indicated the presence of an alphaCoV and betaCoV sequence in each of these 10 samples.

BLAST results of the additional alphaCoV sequence showed similar results to that for sample 20150106CDK_NC1 and 20160411DC63_NN2 as described in Section 4.4.1.2 and Section 4.4.2.1 respectively. This additional alphaCoV sequence was obtained from sample 20150815HFP_PH5nc from Hopefield Farm (HFP) in the Northern Cape. The sequence showed highest sequence identity (99%) to the alphaCoV sequence (Genbank ID KF843858) detected in a *N. capensis* bat in South Africa shown not to cluster phylogenetically with other *N. capensis*-derived alphaCoV sequences (Ithete, 2013).

The additional 24 betaCoV sequences, 14 from newly identified positive samples and 10 from samples previously identified as alphaCoV positive but found to be coinfected with a betaCoV using the Extended RdRp PCR assay, produced similar blastn results to those already reported. All betaCoV sequences showed highest sequence identity (91 – 98%) to NeoCoV (Genbank KC869678) and a dromedary camel in the UAE (Genbank ID KT751244) (Corman *et al.*, 2014a; Wernery *et al.*, 2015).

4.4.2.2.1 By morphological and physiologic characteristics

The subset of samples that underwent repeat screening consisted of samples from 62 female and 22 male bats. Of the 47 positive samples identified using the Extended RdRp PCR assay, 38 (81.0%) were from female bats and 9 (19.0%) were from male bats. The 10 samples identified to be coinfected with an alpha- and betaCoV were obtained from 7 female and 3 male bats.

The subset of *N. capensis* samples that underwent repeat screening consisted of samples from 6 pregnant, 2 post-lactating, and 34 non-pregnant female bats as well as 5 scrotal and 17 non-scrotal male bats. The 47 positive samples identified using the Extended RdRp PCR assay were obtained from 3 pregnant, 1 post-lactating, and 34 non-pregnant females, and from 2 scrotal and 7 non-scrotal males (Table 4.8). One post-lactating and 6 non-pregnant female bats, and 1 non-scrotal and 2 scrotal male bats were identified as being alpha- and betaCoV coinfected.

Five of the 8 juvenile *N. capensis* bats were included in the subset for repeat screening. All 5 had previously been identified as alphaCoV positive when using the Pan-CoV PCR assay. The extended RdRp assay not only detected these 5 alphaCoV sequences but also detected 1 individual to be coinfected with a betaCoV.

4.4.2.2.1 By geographic characteristics

The 47 samples identified as CoV positive using the Extended RdRp PCR assay were detected in bats from the Eastern Cape (n = 7), KwaZulu-Natal (n = 12), Western Cape (n = 20) and Northern Cape (n = 8). Previously with the Pan-CoV PCR assay, no positive samples had been detected in bats from the Northern Cape. The 8 sequences obtained from the 7 bats from 2 sites in the Northern Cape were identified as 1 alphaCoV and 7 betaCoVs (1 case of co-infection).

The positive samples as identified by the Extended RdRp PCR assay were obtained from sites located in the Albany Thicket (n = 6), Fynbos (n = 17), Forest (n = 3), Grassland (n = 10), Nama-Karoo (n = 1), and Savanna (n = 10) biomes (Table 4.9). The samples were obtained across an altitudinal gradient of 17 to 1568 metres above sea level.

Of the 20 sites sampled for species-specific surveillance of *N. capensis* bats, samples from only 17 sites were included in the subset for repeat screening. The 3 sites that were not included in the subset were Sandile's Rest (SRS) and Arminel Hotel Hogsback (AHHB) in the Eastern Cape and Silver Ranch Plettenberg Bay (SRP) in the Western Cape. In the subset of samples that underwent repeat screening, samples from multiple trapping events were available for sites LNR, CGC, CDK, and ABA. The Extended RdRp PCR assay results from all sites where *N. capensis* surveillance samples were collected are detailed in Appendix D Table 2.

4.4.2.3 Collective overview of *Neoromicia capensis* surveillance screening results

With the Extended RdRp PCR assay only being used on a subset of 84 samples, the overall results from both PCR assays is summarised by CoV genus in Table 4.7, physiological and morphological characteristics in Table 4.8 and by geographic factors in Table 4.9.

Table 4.7 Collective screening results using the Pan-CoV and Extended RdRp PCR assays for species-specific surveillance of *N. capensis* bats by CoV genus detected.

PCR Assay	No. screened	No. samples positive	No. alphaCoV sequences	No. betaCoV sequences	No. samples alpha|betaCoV coinfected
Pan-CoV All	196	46 (23.5%)	44 (22.5%)	2 (1.0%)	0 (0.0%)
Pan-CoV Subset	84	32 (38.8%)	30 (35.7%)	2 (2.4%)	0 (0.0%)
Extended RdRp	84	47 (55.9%)	31 (36.9%)	26 (30.9%)	10 (11.9%)

From Table 4.7 it can be seen that for the subset of samples that were rescreened using the Extended RdRp PCR assay, 30.9% of samples were found to harbour betaCoV sequences compared to 2.4% using the Pan-CoV PCR assay. For the subset of samples that underwent repeat screening, similar detection rates for alphaCoVs were observed for the Pan-CoV and Extended RdRp PCR assays, 36.9% and 35.7% respectively. Instances of alpha- and betaCoV coinfection were only detected when using the Extended RdRp PCR assay. There was no significant difference in the detection of alphaCoVs between the Pan-CoV and Extended RdRp PCR assay for the subset of samples. There was however a significant difference ($p < 0.05$) in the detection of betaCoVs and instances of coinfection for the subset of samples using the two different PCR assays.

Table 4.8 Summarised results obtained using the Pan-CoV and Extended RdRp assays by bat physiological and morphological characteristics, sex and reproductive status, and age. Results from screening the full *Neoromicia capensis* dataset as well as the results for the subset of samples that underwent repeat screening with the Extended RdRp PCR assay are detailed by sex, reproductive status, and age. The original Pan-CoV PCR assay screening results for samples included in the subset for repeat screening are provided for comparative purposes.

| | | Total sampled | Full dataset positive Pan-CoV results | Repeat screened subset | | |
				Total in subset for repeat screening	Original positive Pan-CoV results for subset	Ext RdRp positive results for subset
Sex	**Female**	**144 (73.5%)**	**37 (25.7%)**	**62 (73.8%)**	**24 (38.7%)**	**38 (61.3%)**
	Pregnant	**33**	4 (12.1%)	**6**	1 (16.7%)	3 (50.0%)
	Post-lactating	**12**	4 (33.3%)	**3**	1 (33.3%)	1 (33.3%)
	Not pregnant	**99**	29 (29.3%)	**53**	22 (41.5%)	34 (64.1%)
	Male	**52 (26.5%)**	**9 (17.3%)**	**22 (26.2%)**	**8 (36.4%)**	**9 (40.9%)**
	Scrotal	**19**	2 (10.5%)	**5**	2 (40.0%)	2 (40.0%)
	Non-scrotal	**33**	7 (21.2%)	**17**	6 (35.3%)	7 (41.2%)
Age	Adult	**188**	41 (21.8%)	**79**	27 (34.2%)	42 (53.0%)
	Juvenile	**8**	5 (62.5%)	**5**	5 (100.0%)	5 (100.0%)

From Table 4.8 it can be seen that CoVs were predominantly detected in non-pregnant female bats. There was however no significant difference in the detection of CoVs between male and female individuals, between scrotal and non-scrotal males, nor across the different female reproductive states. The difference in the detection of CoVs between adult and juvenile bats was significant ($p < 0.05$), the sample size of the juveniles was however notably smaller.

As indicated in Table 4.9, using the Extended RdRp PCR assay to repeat screen a subset of samples detected CoVs in samples from two additional biomes, namely the Nama-Karoo and Savanna. The majority of CoV positive samples were obtained from bats trapped in the Fynbos biome. Albeit a small sample size, the Forest biome had the highest detection rate.

Table 4.9 Summarised results obtained using the Pan-CoV and Extended RdRp assays by geographic characteristics of the bat trapping site where individuals were sampled. Results from screening the full Neoromicia capensis dataset as well as the results for the subset of samples that underwent repeat screening with the Extended RdRp PCR assay are detailed by the biomes in which bat trapping took place. The original Pan-CoV PCR assay screening results for samples included in the subset for repeat screening are provided for comparative purposes.

Biome	Total sampled	Full dataset positive Pan-CoV results	Total in subset for repeat screening	Original Pan-CoV positive results for subset	Ext RdRp positive results for subset
Albany thicket	20	7 (35.0%)	11	5 (45.5%)	6 (54.6%)
Forest	5	4 (80.0%)	4	3 (75.0%)	3 (75.0%)
Fynbos	69	23 (33.3%)	26	15 (57.5%)	17 (65.4%)
Grassland	52	12 (23.1%)	17	9 (52.9%)	10 (58.8%)
Nama-Karoo	4	0	4	0	1 (25.0%)
Savanna	46	0	22	0	10 (45.5%)
Total	**196**	46 (23.5%)	**84**	32 (38.0%)	47 (56.0%)

The sites where CoV positive *N. capensis* bats were trapped across the different provinces and biomes are indicated on the map in Figure 4.3.
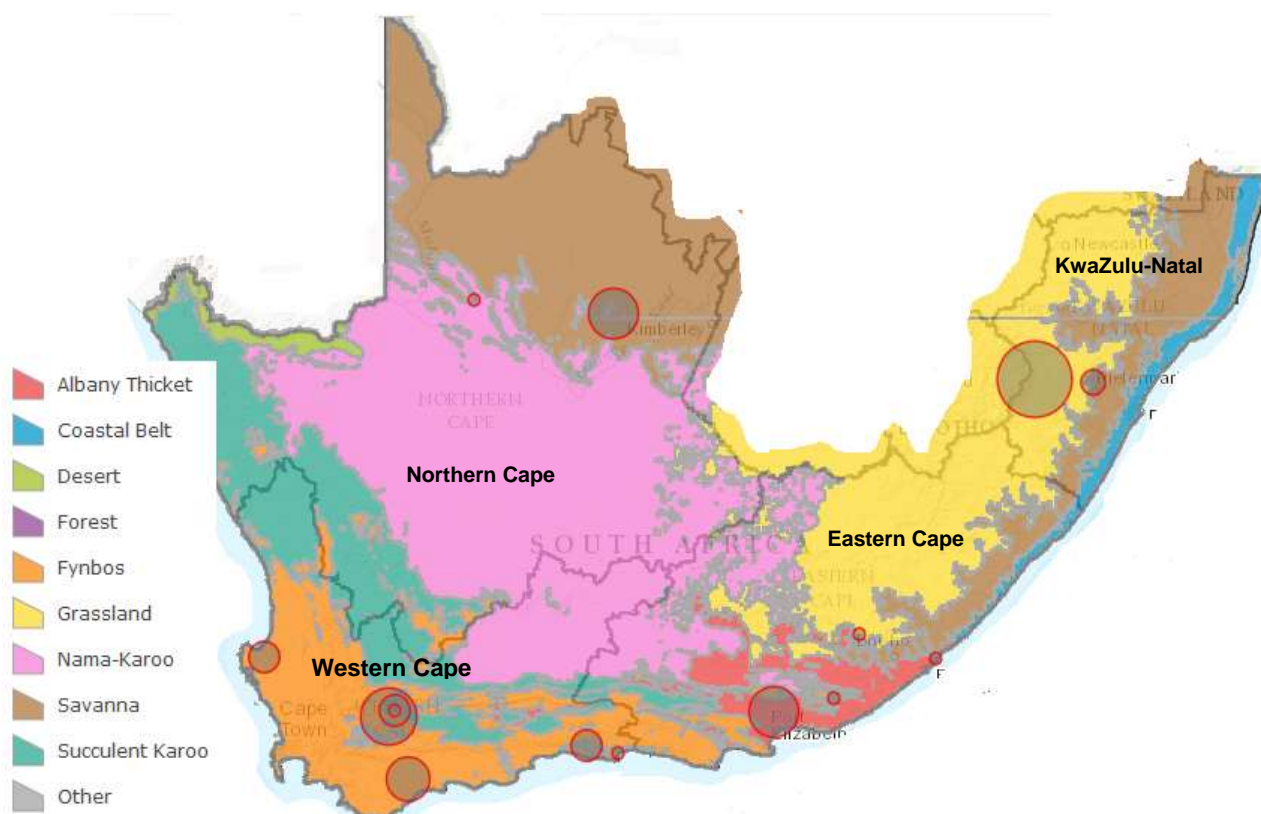
Figure 4.3 The map indicates the sites where coronavirus positive *Neoromicia capensis* bats were sampled across the different biomes and provinces of South Africa. The size of the symbol is representative of the number of positive samples obtained at each site. The map was created using the ArcGIS® software online tool by Esri ©1995-2017 (www.esri.com/arcgis). The *N. capensis* distribution data was obtained from the IUCN Red List (www.iucnredlist.org), the biome distribution data was obtained from SANBI (www.bgis.sanbi.org), and the geographic information system data for the provincial base map was obtained from Map Library (www.maplibrary.org) (Mucina *et al.*, 2006; Rutherford *et al.*, 2006; Jacobs *et al.*, 2008; IUCN, 2012).

4.4.3 Longitudinal study

4.4.3.1 Pan-coronavirus PCR assay

As detailed in Section 4.3.3, 5 individual bats from an *N. capensis* bat colony on Cloeteskraal Farm (CCK) in the Western Cape were sampled prior to commencing passive collection of pooled samples for the longitudinal surveillance of this colony. Initial screening of these individually collected samples using the Pan-CoV PCR assay yielded 3 (60.0%) positive results. BLAST results of the sequences obtained indicated the presence of an alphaCoV in the colony.

4.4.3.2. Extended RdRp PCR assay

The individually collected samples underwent repeat screening with the Extended RdRp PCR assay. An additional 2 sequences were obtained that blastn results indicated to be related to sequences from the *Betacoronavirus* genus. These 2 sequences were obtained from a male bat coinfected with the alphaCoV and the second from a female bat previously identified as negative with the Pan-CoV PCR assay. This resulted in overall 4 of 5 bats being CoV positive (80.0%). Since these were *N. capensis* bats, these results have been included in Section 4.4.2. (Appendix D Table 2).

4.4.3.2 Screening of pooled samples

Although evidence of CoV infection of bats within the colony was found, subsequent testing of collected monthly pooled samples using both the Pan-CoV and Extended RdRp PCR assay all yielded negative results. Testing of individual faecal pellets from a subset of pooled samples also yielded negative results. Real-Time PCR quantitative amplification was attempted by Dr Ithete that too yielded negative results for all collected pooled samples. No further analyses could be performed for this aspect of the study.

4.5 Host species confirmation by molecular methods

As detailed in Chapter 3 Section 3.8.5, molecular methods were used to confirm the species identity allocated to samples in the field by morphometric-based assessment. For this purpose, the cyt *b* and / or COI genes were amplified and the resulting sequences compared to available voucher speciemen and reference sequences on Genbank using BLAST. Molecular confirmation of species identity was carried out for 266 samples (Monadjem *et al.*, 2010, 2013; Bastos *et al.*, 2011). This was done at random, for all CoV positive samples, and for samples from bat species with overlapping morphological features, such as *N. capensis* and *P. hesperidus*, that make in-field species identification by morphometric assessment difficult (Monadjem *et al.*, 2010, 2013).

Molecular confirmation was in agreement with in-field species identification for 210 samples and conflicted with in-field species identification for 56 samples.

The species most commonly misidentified in the field was *N. capensis* with 39 trapped individuals identified in the field as *P. hesperidus*. A group of 24 *P. hesperidus* samples collected during a bat trapping expedition in the Eastern Cape during October 2014 were found to be *N. capensis* by molecular methods. These samples were consistently incorrectly identified across all sites during this specific field trip. Two additional samples originally identified as *P. hesperidus* that were found to be *N. capensis* were from sites, Arminel Hotel Hogsback (AHHB) and Addo Backpackers (ABA), sampled during this expedition but during different sampling events.

A group of 8 *N. capensis* samples were incorrectly identified as *P. hesperidus* across two sampling events at the Hopefield Farm (HFP) bat trapping site in the Northern Cape. Three bats identified as *P. hesperidus* during a sampling event at Lotheni Nature Reserve (LNR) were also found to be *N. capensis*. Following molecular confirmation, 2 additional *P. hesperidus* samples were found to be *N. capensis*-derived; 1 from the à La Fugue Guesthouse in Upington (LFU), and 1 from Forest Edge in Knysna (FEK).

A second species that was misidentified in the field was *R. simulator*. A group of 10 samples identified as *R. swinnyi* at Mooiplaas Gold Mine (MPG) in KwaZulu-Natal was found to be *R. simulator* by molecular methods across different sampling events.

A third species that was misidentified in the field was *M. fraterculus*. A group of samples obtained from 11 bats trapped in the Hilton train tunnel (HTT) that were all identified in the field as *M. natalensis* bats were found by molecular analysis of the respective cyt *b* gene sequences to instead be samples from 4 *M. natalensis* bats and 7 *M. fraterculus* bats.

All results presented in this chapter have used the corrected species as confirmed by mitochondrial gene sequencing. From the data presented in Table 4.2 in Section 4.3.3.1, these species confirmation results are in accordance with the weight and forearm measurement ranges used as species-identifier tools for each of the respective bat species (Monadjem *et al.*, 2010). Where applicable, these species changes have been indicated in the sample name with the addition of the confirmed species code in lowercase letters as shown in the following example: **20141027ABA_PH6nc** indicates that the bat identified as *P. hesperidus* (PH) in the field was found by molecular means to be *N. capensis* (nc). Table 3 Appendix D lists the misidentified samples with the correct species as confirmed by molecular methods.

4.6 Extended genome amplification of betacoronavirus sequences

As indicated in Section 4.4, several betaCoVs were detected during this study. All 31 betaCoV sequences that were obtained showed the highest sequence identities to MERS-related CoV. With only a single published bat-borne MERS-related betaCoV sequence from South Africa, NeoCoV, available at the onset of this study, an attempt was made at full genome amplification of the betaCoV from six betaCoV positive samples. Full genome sequencing was not successful but in

addition to the partial RdRp sequences obtained from both the Pan-CoV and Extended RdRp screening PCR assays, additional partial genome sequences were obtained from 20140122FEK_PH1, 20140122FEK_PH4, 20150720ABA_NC1, 20160303FEK_NC1, 20160304FEK_NC1, and 20160304FEK_NC2 for additional phylogenetic analyses (de Souza Luna *et al.*, 2007; Drexler *et al.*, 2010). The partial genome sequences obtained and their sequence identities to other known betaCoVs are detailed in the following sections.

4.6.1 *NSeq* betacoronavirus assay

The *NSeq* betacoronavirus RT-PCR assay was used to confirm the detection of MERS-related CoV by generating a partial sequence of the nucleocapsid (N) protein (Corman *et al.*, 2012). The assay successfully amplified partial N sequences and BLAST results confirmed that the sequences obtained were from MERS-related CoVs. Partial N sequences from *P. hesperidus* bat samples, 201401122FEK_PH1 and 20140122FEK_PH4, had highest sequence identity (89%) to a MERS-related betaCoV detected in a *P. hesperidus* bat from Uganda (Genbank ID KX574227) (Anthony *et al.*, 2017a).

4.6.1 NeoCoV genome amplification and SuperFi assays

The PCR protocol used to amplify the full genome of NeoCoV was used with some success to obtain additional partial betaCoV genome sequences from samples 20140122FEK_PH1, 20140122FEK_PH4, and 20150720ABA_NC1 (Corman *et al.*, 2014a). With partial RdRp and N sequences from the detected betaCoV showing high sequence identity to NeoCoV, some of the primers designed by Corman et al (2014) were modified to better target NeoCoV-like genomes. Additional primers targeting different regions of the NeoCoV genome were kindly provided by Dr T Suliman (SU: Medical Virology) (Appendix B).

Following difficulty in amplifying additional partial BetaCoV genome sequences using the NeoCoV full genome amplification approach, a protocol using Platinum™ SuperFi™ DNA Polymerase (Invitrogen, USA) was established and used with better success. Once a number of partial genome regions were amplified, additional primers were designed based on the sequences generated to fill gaps where possible. The basic protocol was followed as described in Chapter 3 Section 3.8.6.

The following sections detail the partial genome sequences obtained. The South African NeoCoV betaCoV sequence (Genbank ID KC869678) was used as the reference sequence when mapping generated sequences. For reference, all partial genome sequences obtained are listed along with the regions covered in the corresponding NeoCoV genome in Appendix D Table 4.

4.6.2 Partial and full gene sequences obtained for phylogenetic analyses

Using the De Novo Assemble tool in Geneious R8 and R10 (Biomatters Inc., New Zealand), partial genome contiguous sequences were obtained for each of the six betaCoV positive samples that underwent additional genome amplification. Using the NCBI's online ORFfinder tool (https://www.ncbi.nlm.nih.gov/orffinder/) along with the reference mapping tool, Map to Reference, in Geneious R8 and R10, partial and full gene sequences for phylogenetic analyses were identified for each sample.

Identical sequences were obtained from samples caught at the same site on the same date and likely represent betaCoVs from the same bat colony. When sequences from the same genome region were obtained for 20140122FEK_PH1 and 20140122FEK_PH4, sequence alignments indicated these sequences to be identical to each other. Similarly, sequences obtained from the same genome region for 20160303FEK_NC1, 20160304FEK_NC1, and 20160304FEK_NC2 were identical to each other. For these identical sequences, a single representative sequence was used in subsequent analyses, referred to as 20140122FEK_PH and 201603FEK_NC respectively. Sequences from the three sampling events (20150720ABA, 20140122FEK, and 201603FEK) were found to be non-identical not only to each other but also to NeoCoV (Genbank ID KC869678) (Ithete *et al.*, 2013; Corman *et al.*, 2014a).

To summarise Appendix D Table 4, the partial genome sequences for the *N. capensis*-borne betaCoV, 20150720ABA_NC1 and 201603FEK_NC, had blastn results with highest sequence identities (92 – 99%) to NeoCoV (Genbank ID KC869678) (Ithete *et al.*, 2013; Corman *et al.*, 2014a). For the *P. hesperidus*-derived betaCoV, 20140122FEK_PH, blastn results did not consistently indicate highest sequence identities to NeoCoV. Instead, for several sequences including the membrane (M) protein, highest sequence identities were found to dromedary camel (Genbank ID KU740200, KX108943, KT751244) and human (Genbank ID KU851859, KX154694, unpublished study by Assiri *et al.,* 2016) MERS-CoV sequences and to a MERS-related betaCoV (Genbank ID KX574227) recently detected in a *P. hesperidus* bat from Uganda (Wernery *et al.*, 2015; Kandeil *et al.*, 2016; Lau *et al.*, 2016b; Anthony *et al.*, 2017a).

Full gene sequences were obtained from all three sampling events for the M and envelope (E) proteins, and ORF3, ORF4a, ORF4b, and ORF5. The full gene sequence for the N protein was obtained from samples 20140122FEK_PH and 20150720ABA_NC1 but not for 201603FEK_NC. Partial gene sequences for the ORF1a, ORF1b, and Spike (S) proteins were obtained for all three betaCoV genomes. The S protein is subdivided into to two subunits. Subunit one (amino acid residues 1 – 747 in NeoCoV) contains the RNA binding domain while subunit two (amino acid residues 747 – 1344) is involved in the fusion of cellular and viral membranes; partial S sequences from both S subunits were obtained from all three betaCoVs.

Currently little is understood about the function and role of ORFs 3, 4a, 4b, and 5. For this reason subsequent phylogenetic analyses focused on the functional proteins E, M, N, S, and ORF1ab. The results of the phylogenetic analyses of these sequences are detailed in Section 4.11.

4.7 Cloning

4.7.1 Positive control plasmid

As described in Chapter 3, positive plasmid DNA controls were established to ensure that the pan-CoV and extended RdRp PCR assays were functioning optimally.

The positive control for the Pan-CoV PCR assay was a 917 bp partial RdRp sequence of an alphaCoV obtained from an *M. natalensis* faecal sample (Ithete, 2013). The cloned fragment underwent *in vitro* transcription and was used in its RNA form, included in all CoV screening processes from the reverse transcription step.

To assess its detection in the PCR assay, a serial dilution of the positive control RNA underwent reverse transcription and subsequent amplification in the Pan-CoV PCR assay. The lowest concentration of plasmid RNA following reverse transcription that was detected was 1.1×10^3 copies per µl. This *in vitro* transcribed RNA was used as a positive control at 1.1×10^3 copies per µl in subsequent Pan-CoV PCR assays as well as in the Extended RdRp PCR assay where it was detectable with the RGU_2c primers (Drexler *et al.*, 2010). Figure 4.4 depicts the DNA electrophoresis image following amplification of the serially diluted RNA positive control with the Pan-CoV PCR assay (de Souza Luna *et al.*, 2007).
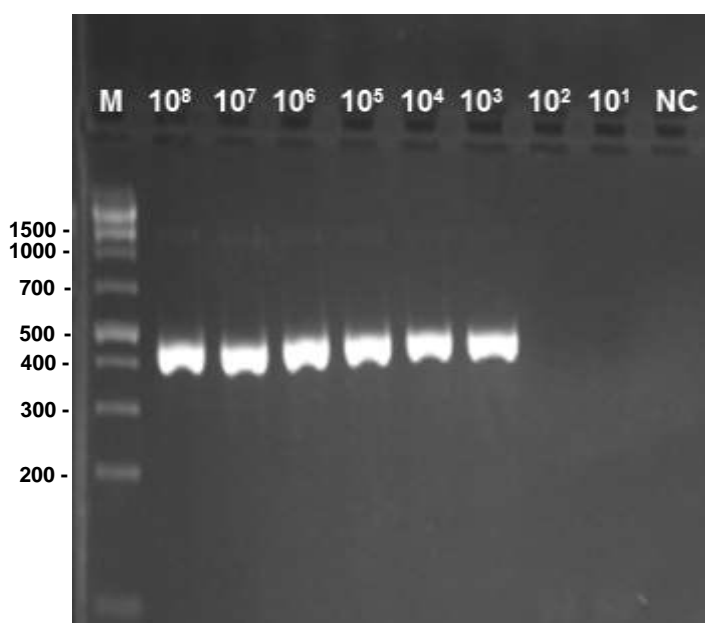
Figure 4.4 DNA electrophoresis image of serially diluted positive control RNA amplified with the Pan-CoV PCR assay following reverse transcription. A 10× serial dilution of the positive control RNA with a starting concentration of 1.1×10^8 copies per µl was reverse transcribed and amplified using the Pan-CoV PCR assay. NC = negative control M = GeneRuler 1kb Plus DNA marker (Thermo Fisher Scientific, USA). Marker bands 75 to 1500 bp indicated to left of DNA marker.

A second positive control plasmid consisting of a cloned 992 bp partial *RdRp* gene gBlock sequence (Integrated DNA Technologies, USA) from a SARS-related BtCoV (NC_014470) served as an additional control for the Extended RdRp PCR assay with the primers targeting SARS-related CoVs and was used directly in its DNA form.

4.7.3 Clonal sequencing for the assessment of diversity within a coinfected sample

As detailed in Section 4.4.2.1 a sample, 20140923LNR_NC7, collected at the KwaZulu-Natal Lotheni Nature Reserve (LNR) site, produced positive PCR results using the Pan-CoV PCR assay (de Souza Luna *et al.* 2007). Following Sanger sequencing, coinfection was suspected due to the presence of multiple double peaks present in the sequence data chromatogram. Comparison of this sequence against the six sequences obtained from other CoV positive samples collected at LNR on the same day, revealed two alphaCoV strains, 2014/1 and 2014/2, within this colony of *N. capensis* bats.

To assess the diversity of alphaCoV variants within the individual bat sample, 20140923LNR_NC7, shown to contain both strain 2014/1 and strain 2014/2 sequences, the partial RdRp sequence amplified with the Pan-CoV PCR assay was cloned per the protocol described in Chapter 3 Section 3.12. Forty colonies were picked for Sanger sequencing (de Souza Luna *et al.*, 2007), of which 39 yielded adequate plasmid DNA concentrations for sequencing reactions. Of the 39 colonies sequenced, 35 clonal sequences were obtained for analyses with three lost due to failed sequencing reactions and 1 sequence excluded due to the presence of double peaks in the sequence chromatogram indicating that a pure colony had not been picked. At the nucleotide level, 16 (45.7%) of the 35 clonal sequences obtained were identical to strain 2014/1 and five (14.3%) were identical to strain 2014/2. Of the remaining 14 clonal sequences, none were identical to each other. The 2016/1 strain nucleotide sequence was not present in any of the 35 clonal sequences obtained from the coinfected sample, 20140923LNR_NC7.

At the amino acid level, 20 of the 35 sequences (57.1%) had the same alanine-containing amino acid sequence as strain 2014/1 and 8 (22.9%) had the same serine-containing amino acid sequence as strain 2014/2. The remaining 7 sequences contained additional amino acids.

4.9 General surveillance phylogenetic analyses

For this section, sequences obtained from the species-specific surveillance of *N. capensis* bats have been included in addition to sequences obtained from CoV positive samples identified as part of the general surveillance effort. These *N. capensis*-derived sequences have been included to allow for putative species classification purposes and to provide a full representation of the different CoV detected following the screening of 600 bat faecal samples. More in-depth

phylogenetic analyses to assess the diversity of CoV at the species and individual bat level were additionally performed on the 395 bp partial RdRp sequences from *N. capensis*-derived sample sequences, detailed in Section 4.10. To avoid redundancy, a representative 395 bp sequence of the *N. capensis*-derived sample sequences that showed highest sequence identity to other South African alphaCoV with relatedness to *Nyctalus*-borne alphaCoVs was included (Geldenhuys *et al.*, 2013; Ithete, 2013). The inclusion of this representative sequence provides an indication of where these *N. capensis*-derived viruses cluster in the bigger phylogenetic schematic of alphaCoVs and adds to the description of the diversity of BtCoVs detected. The sequence that was randomly selected was from sample 20150106CDK_NC5 and has been indicated in relevant figures as "NC-borne representative" in the sequence name. AlphaCoV 395 bp sequences from other *N. capensis* bats that blastn results (Section 4.4) indicated to not be closely related to these sequences were similarly included for reference in the full alphaCoV phylogeny. For provisional species classification based on the RdRp Grouping Unit (RGU) criterion, the non-identical 272 aa sequences from *N. capensis* samples were also included in the following phylogenetic analyses (Drexler *et al.*, 2010).

Due to its short length, phylogenetic analyses with the 395 bp sequence datasets were conducted at the nucleotide level, providing an indication of evolutionary diversity. For the 816 bp partial RdRp sequences corresponding to the RGU as defined by Drexler *et al.* (2010), phylogenetic analyses were conducted at the amino acid level using the resulting 272 aa translated sequences for provisional CoV species classification and to provide an indication of evolutionary diversity at the more conserved functional protein level.

4.9.1 Sequences obtained for phylogenetic analyses from CoV screening PCR assays

From the CoV screening PCR assays used in this study, partial RdRp sequences either 395 bp or 816 bp in length were obtained from CoV positive samples. Since the two partial RdRp sequences do not overlap exactly, depending on the PCR assay generating the positive result and the quality of sequences obtained, both partial RdRp sequences were not obtained for all positive samples in this study.

Collectively, including the *N. capensis* samples, 90 CoV sequences were identified from the 600 samples screened during this study. Of the 90 CoV sequences, 59 were identified using BLAST as alphaCoVs and 31 as betaCoVs. The number of 395 bp and 816 bp nucleotide sequences obtained is indicated in Table 4.10. The table further indicates the number of each partial RdRp sequence obtained per the detected CoV genus and indicates the number of samples from which either only the 395 bp, only the 816 bp, or both the 395 bp and 816 bp sequences were obtained.

Table 4.10 Partial RdRp sequences obtained for phylogenetic analyses using the Pan-CoV and Extended RdRp PCR assays.

Sequence lengths obtained	Overall	Number of alphaCoVs	Number of betaCoVs
395 bp only	30	23	7
816 bp only	9	1	8
both 395 bp and 816 bp	48	32	16
neither 395 bp nor 816 bp	3	3	0

As indicated in Table 4.10, for 3 samples insufficient sequencing data was obtained for their inclusion in either the 395 bp or 816 bp sequence dataset. These sequences were all from *N. capensis*-derived samples.

For samples 20150918CDK_NC15, 20150107CGC_NC1, and 20160112LNR_NC2, only short sequence fragments of 294 bp, 315 bp, and 353 bp were obtained, respectively. BLAST results indicated all three sequences to be alphaCoV with sequence identity values of 93 – 96% to alphaCoV previously detected in South African *N. capensis* bats (GenBank ID KF843856 and KF843855) (Ithete, 2013). Sequence alignments of these sequences showed them to be identical to alphaCoV sequences obtained from the same sampling event. These samples could therefore be excluded without compromising the dataset. All partial RdRp gene sequences generated in this study that were included in subsequent datasets for phylogenetic analyses were submitted to Genbank (https://www.ncbi.nlm.nih.gov/genbank/), the corresponding Genbank accession numbers are listed in Appendix G.

4.9.2 Sequences excluded from datasets

For improved phylogenetic resolution, sequence length is important. For this reason, three South African BtCoV sequences obtained by Geldenhuys *et al.* (2013), each only 275 bp in length, were excluded from the sequence datasets used in the following phylogenetic analyses. A phylogenetic tree depicting all currently known alphaCoV sequences detected in South Africa, including these three shorter sequences and the sequences obtained during this study is depicted in Appendix E Figure 1 for reference purposes. The three shorter South African sequences were found to cluster with other sequences from this study and their phylogenetic clustering is noted as follows:

For a sequence from an *N. capensis* bat (Genbank ID JQ5198) previously reported to be related to *Nyctalus*-borne alphaCoV, the phylogenetic tree depicted in Appendix E Figure 1 shows it to cluster with other South African *N. capensis*-derived alphaCoV (Geldenhuys *et al.*, 2013; Ithete, 2013). The sequence from an *M. natalensis* bat (Genbank ID JQ519817) was found to cluster with another *M. natalensis*-derived sequence from this study, both showing relatedness to the ICTV reference sequence for *Miniopterus* BtCoV 1A (Geldenhuys *et al.*, 2013). The third sequence, from a *Mops condylurus* bat, was shown to cluster with another South African *M. condylurus*-derived

103

sequence and two other alphaCoV sequences from *C. pumilus* and *R. simulator* identified during this study (Geldenhuys *et al.*, 2013; Ithete, 2013).

The sequence from sample 20140923LNR_NC7 that was shown to be coinfected with two alphaCoV variants, resulting in several ambiguous bases in its sequence, was omitted from the sequence dataset. Since phylogenetic analyses were conducted with the deletion of all alignment positions containing missing data or ambiguities, the inclusion of this sequence would result in a shorter sequence alignment. Additionally, since sequences from both strains present in the sample were included in the datasets, this sample's sequence was sufficiently represented in the dataset.

Similarly, the 395 bp betaCoV nucleotide sequences from sample 20150815HFP1 NC1 and 20150920CGC_NC4 were excluded from the analyses due to the presence of several ambiguous bases. Comparisons with sequences obtained from the same bat trapping sites and sampling events across non-ambiguous positions showed pairwise-distances of 0 – 1.8% allowing for their elimination without compromising the dataset.

For ease of viewing and reading of resulting phylogenetic trees, only non-identical sample sequences were aligned to the sequences datasets. Appendix E Figures 2 - 4 depict three phylogenetic trees displaying the phylogenetic relatedness for all alphaCoV 395 bp sequences (20140923LNR_NC7 excluded), all betaCoV 395 bp sequences (excluding 20150816HP1_NC1 and 20150920CGC_NC4), and all alpha- and betaCoV 272 aa sequences along with the ICTV-recognised *Alpha-* and *Betacoronavirus* species sequences. Sequences that are only represented by a single partial RdRp sequence, either 395 nt or 272 aa, have been highlighted in each figure. CoV species as recognised by the ICTV (Figure 2.3, Chapter 2) were included in all phylogenetic analyses for reference purposes and have consistently been indicated in large bold oblique font along with other RefSeq sequences of interest.

4.9.3 Sequence datasets

As described in Chapter 3 Section 3.14.2, sequence datasets of 395 bp nucleotide and 272 aa protein sequences against which sample sequences could be aligned and phylogenetically compared to were established. The sequence datasets included all RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) CoV sequences. All other CoV sequences in Genbank (https://www.ncbi.nlm.nih.gov/genbank/) that corresponded to either the 395 bp or 272 aa sequence were aligned and pairwise distances calculated in MEGA7 to eliminate identical and closely related sequences (Kumar *et al.*, 2016).

4.9.3.1 Alphacoronavirus 395 bp sequence datasets

For the 395 bp alphaCoV nucleotide sequence dataset, four datasets containing non-RefSeq Genbank sequences excluding sequences that showed pairwise distances of less than 10%, 15%, 18%, and 20%, respectively, were established. Comparing the preliminary NJ trees showed little

variation for the clustering of sample sequences across the four datasets except for the *M. natalensis*-derived sequences.

From the phylogenetic tree depicted in Appendix E Figure 2, where all 395 bp alphaCoV sequences are depicted with ICTV sequences, it appears that the sequences obtained from *M. natalensis* bats cluster in two groups, one related to *Miniopterus* BtCoV HKU8 species and one to *Miniopterus* BtCoV 1A. Only one putative *Miniopterus* BtCoV 1A related sequence was obtained during this study that was shown to cluster with another South African CoV sequence from an *M. natalensis* bat (Appendix E Figure 1) (Geldenhuys *et al.,* 2013). Unfortunately, the longest partial RdRp sequence obtained for sample 20150922CDK_MN1, despite attempting extended amplification with the Extended RdRp PCR assay, was 623 bp in length and could not be included in the 816 bp sequence dataset for confirmation of this clustering and provisional species classification based on the RGU criteria (Drexler *et al.,* 2010).

A pairwise distance matrix using this longer 623 bp fragment in its translated form also indicated the sequence from 20150922CDK_MN1 to be more closely related *to Miniopterus* BtCoV 1A than to *Miniopterus* BtCoV HKU8. The pairwise distance between 20150922CDK_MN1 and *Miniopterus* BtCoV 1A was 3.6% compared to 5.5% between 20150922CDK_MN1 and *Miniopterus* BtCoV HKU8.

This relationship is reflected in the 395 bp alphaCoV sequence dataset that excludes non-RefSeq Genbank sequences with pairwise distances of less than 10% but not in the other three datasets where instead it appears that all *M. natalensis*-derived CoV sequences are *Miniopterus* BtCoV HKU8-related. It was found however that if two Genbank sequences from *M. natalensis* bats from China and Kenya (Genbank ID KF294270, unpublished study from Wang *et al.,* 2013, and GU065410, unpublished study from Conrardy *et al.,* 2009) were included in these three sequence datasets, the resulting preliminary phylogenetic trees correctly reflected the putative relationship between 20150922CDK_MN1 and *Miniopterus* BtCoV 1A.

Since the dataset that excluded non-RefSeq Genbank sequences with pairwise distances of less than 10% contained a large number of sequences that made viewing the resulting phylogenetic tree difficult, it was decided to use the sequence dataset where non-RefSeq Genbank sequences with pairwise distances less than 18% were excluded with the exception of the two sequences (Genbank ID KF294270 and GU065410) to correctly reflect the relationship between the different *M. natalensis*-derived alphaCoV sequences.

4.9.3.2 Betacoronavirus 395 bp sequence dataset

For the 395 bp betaCoV nucleotide sequence dataset for which fewer non-RefSeq Genbank sequences were available, two sequence datasets that excluded non-RefSeq Genbank sequences showing pairwise distances less than 10% and 15% were established. Comparison of the

preliminary NJ trees showed no change in the clustering of sample sequences and the dataset excluding sequences with pairwise distances of less than 10% was selected for subsequent phylogenetic analyses.

4.9.3.3 Extended partial RdRp 272 aa sequences

For the 272 aa alpha- and betaCoV sequence datasets, two datasets were created for each that excluded non-RefSeq Genbank sequences with pairwise distances of less than 1% or 2%, respectively. The dataset excluding non-RefSeq Genbank sequences with pairwise distances of less 2% was selected for the alphaCoV phylogenetic analyses while the dataset excluding non-RefSeq Genbank sequences with pairwise distances of less 1% was selected for the phylogenetic analyses of betaCoV sequences purely based on visual preferences as the relationship between study sequences and dataset sequences did not appear to significantly change across the different datasets.

4.9.4 General surveillance phylogenetic trees

The results from the phylogenetic analyses of alpha- and betaCoV sequences are presented separately. For non-study sample BtCoV sequences represented on phylogenetic trees, the country of origin has been indicated along with an abbreviated code representing the bat species as detailed in Appendix C.

4.9.4.1 General surveillance: alphacoronaviruses

During this study 59 alphaCoV positive samples were detected. These alphaCoV sequences were detected in 8 different bat species (Section 4.4) including *N. capensis*. The pairwise distances across these nucleotide sequences, including the representative *N. capensis*-derived sequence, of this conserved genome region averaged 13.0% across the non-identical amino acid sequences, indicating the presence of different *Alphacoronavirus* species based on the RGU criterion that different *Alphacoronavirus* species show pairwise distances greater than 4.8% (Drexler *et al.*, 2010, Drexler *et al.*, 2014).

Figure 4.5 and Figure 4.6 depict the phylogenetic trees resulting from the analyses of the 395 bp and 272 amino acid sequence datasets respectively. For ease of reference, eight positions where study samples cluster on the *Alphacoronavirus* phylogenetic trees have been indicated by number. The numbers on both trees represent the same cluster of study samples. In Figure 4.6, position one is not indicated due to the lack of an available 272 aa sequence for this cluster of study samples as represented in Figure 4.5. The corresponding phylogenetic tree for the 272 aa dataset using Bayesian inference in Mr Bayes is depicted in Appendix E Figure 5 and showed similar phylogenetic clustering of sequences as depicted by the ML phylogenetic tree depicted in Figure 4.7.

At position one in Figure 4.5, the sequence from sample 20150922CDK_MN1 is seen to cluster with other *Miniopterus* sp. derived sequences from Kenya (Genbank ID GU0654, unpublished study by Conrardy *et al.,* 2009) and China (Genbank ID KF294270, unpublished study by Wang *et al.,* 2013) that show apparent phylogenetic relatedness to the ICTV-recognised *Miniopterus* BtCoV 1A species. The pairwise distances between sample 20150922CDK_MN1 and the three *Miniopterus* sp. derived sequences at position 1 ranged from 9.2 – 17.1% indicating diversity across this short partial sequence of the relatively conserved *RdRp* gene. Unfortunately, as noted in Section 4.9.3.1, the longer (272 aa) partial RdRp sequence could not be obtained from sample 20150922CDK_MN1 for inclusion in the amino acid phylogenetic analysis depicted in Figure 4.6 that would allow for putative species classification of this sequence and confirmation of its phylogenetic relatedness to the other *Miniopterus* sp.-derived CoV sequences. However, a longer partial RdRp sequence of 185 aa was obtained with an amino acid pairwise distance of 3.6% indicating this sequence to likely be most closely related to *Miniopterus* BtCoV 1A.

The results from a phylogenetic analysis of an alignment of this 185 aa 20150922CDK_MN1 sample sequence against the *Miniopterus* sp.-derived and related CoV sequences corresponding to position 2 and 3 in Figure 4.6 are depicted in Figure 4.7. This phylogenetic tree further confirms the likelihood that the sequences from 20150922CDK_MN1 putatively belong to the *Miniopterus* BtCoV 1A species.

Based on an alignment of short 275 bp sequences, the sequence from 20150922CDK_MN1 showed close phylogenetic relatedness (Figure 4 of Appendix 4) to another *M. natalensis*-derived sequence (Genbank ID JQ519817) from South Africa for which only a short sequence was available (Geldenhuys *et al.*, 2013). Based on the observation made here, it is likely that this sequence (Genbank ID JQ519817) also belongs to the *Miniopterus* BtCoV 1A species.

At position two in Figure 4.5, a sequence from a *M. natalensis* sample, 20161022DC253, appears not to cluster with other known CoV but is positioned between two clusters of *Miniopterus* sp.-borne CoV species, *Miniopterus* BtCoV 1A and *Miniopterus* BtCoV HKU8. Similarly, in the corresponding amino acid phylogenetic tree in Figure 4.6, no other sequences are seen to cluster directly with this study sequence. The closest amino acid pairwise distance observed across this alignment was 5.5% to *Miniopterus* BtCoV HKU8. This amino acid pairwise distance is greater than the recommended RGU criterion 4.8% cut-off distance between *Alphacoronavirus* species, signifying that this study sample sequence putatively belongs to a novel BtCoV species. For reference purposes, this CoV will be referred to as unclassified *Miniopterus* BtCoV.
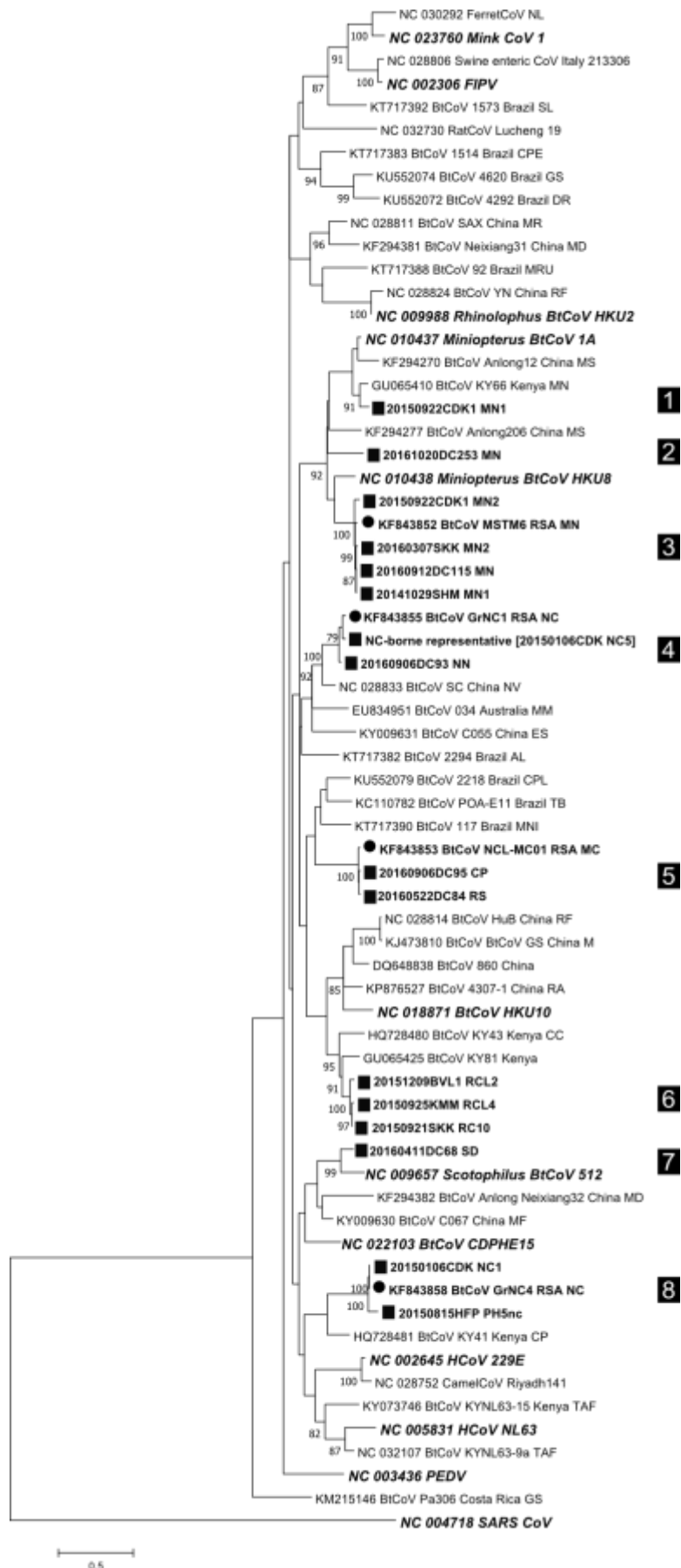
Figure 4.5 Phylogenetic analysis of partial (395 bp) alphacoronavirus RNA dependent RNA polymerase sequences. Phylogenetic inference was determined by the Maximum Likelihood method with 1000 bootstrap (BS) replicates in MEGA 7 (Kumar *et al.*, 2016). The General Time Reversal substitution model with gamma distributed rate variation and invariant sites was used (GTR+G+I). The tree was midpoint rooted with SARS-CoV (NC_004718). KEY: bold oblique font = ICTV-recognised coronavirus species sequences and RefSeq sequences of interest; ■ = study sequences; ● = non-study South African sequences.

108
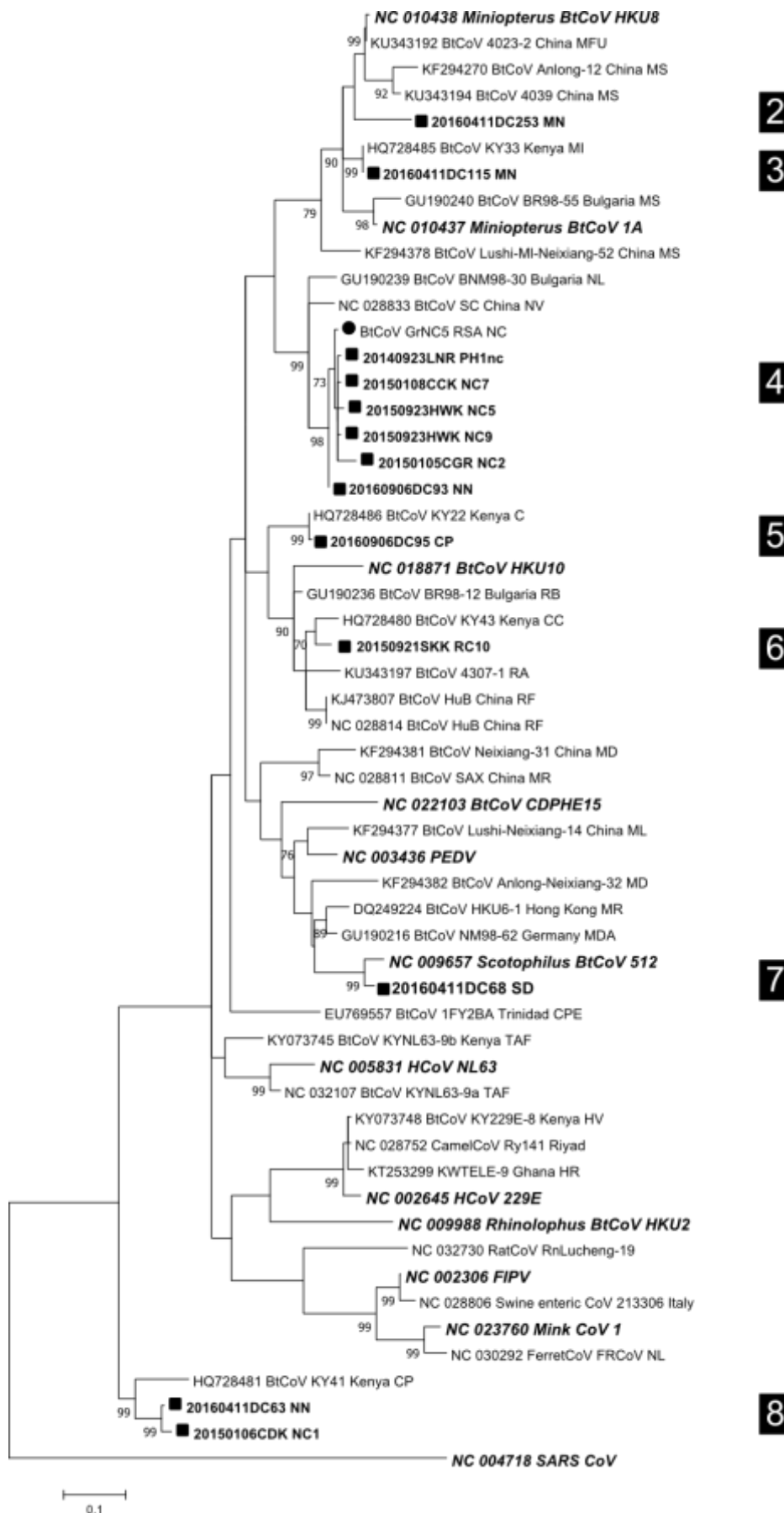
Figure 4.6 Phylogenetic analysis of partial (272 aa) alphacoronavirus RNA dependent RNA polymerase sequences. Phylogenetic inference was determined by the Maximum Likelihood method with 1000 bootstrap replicates in MEGA 7. The Le Gascuel model with gamma distributed rate variation and invariant sites was used (LG+G+I). The tree was midpoint rooted with SARS-CoV (NC_004718). KEY: bold oblique font = ICTV-recognised coronavirus species sequences and RefSeq sequences of interest; ■ = study sequences; ● = non-study South African sequences (Le and Gascuel, 2008; Kumar *et al.*, 2016).

109
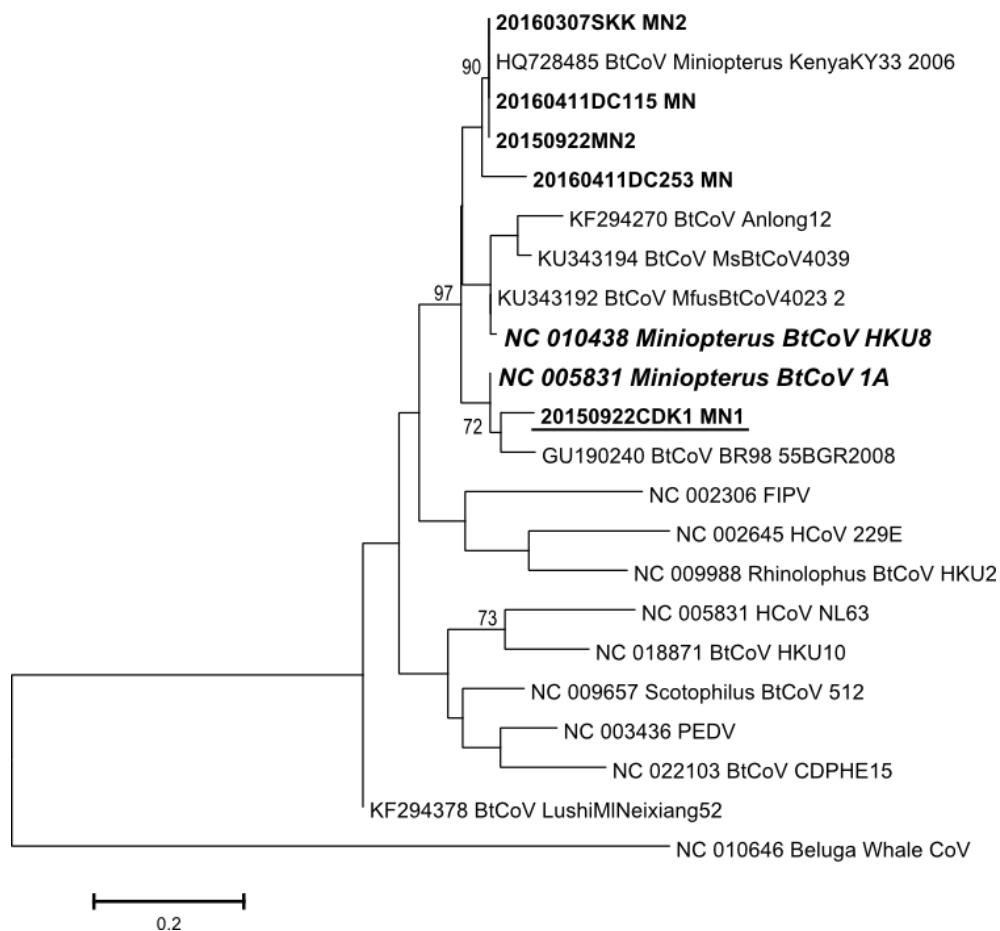
Figure 4.7 Phylogenetic analysis of partial (185 aa) RNA dependent RNA polymerase sequences indicating the putative relationship between *Miniopterus* sp.-derived alphacoronaviruses. Phylogenetic inference was determined by the Maximum Likelihood method with the Le Gascuel substitution model with gamma distributed rates (LG+G) and 1000 bootstrap replicates in MEGA7. There were 185 amino acid positions in the final alignment (Le and Gascuel, 2008; Kumar *et al.*, 2016). The tree was midpoint rooted with Beluga Whale CoV (NC_010646). KEY: bold oblique font = *Miniopterus* sp.-derived ICTV-recognised *Alphacoronavirus* species sequences of interest, bold font = *Miniopterus* sp-derived study sequences; underlined = study sequence of interest: 20150922CDK1 MN1.

At position three in Figure 4.5, four study sequences from *M. natalensis* bats cluster with another *M. natalensis* sequence (Genbank ID KF843852) from South Africa (Ithete, 2013). Collectively these five sequences show phylogenetic relatedness to the *Miniopterus* BtCoV HKU8 species. Only one study sample sequence of 272 aa, 20160411DC115, was obtained for inclusion in the amino acid phylogenetic tree in Figure 4.6. In Figure 4.6 sequence 20160411DC115 is seen to cluster with a sequence from a *Miniopterus inflatus* bat from Kenya (Genbank ID HQ728485) (Tao *et al.*, 2012). Pairwise amino acid distances indicated these two sequences to be identical to each other on the amino acid level. Amino acid pairwise distances of 4.1% and 5.9% were observed between these sequences and *Miniopterus* BtCoV HKU8 and *Miniopterus* BtCoV 1A respectively. This signifies that based on the RGU criterion these sequences likely belong to the *Miniopterus* BtCoV HKU8 species (Drexler *et al.*, 2010).

At position four in Figure 4.5, the study sample sequences from the representative 395 bp *N. capensis* sequence and a *N. nana* sequences are seen to cluster with a South African *N. capensis*-borne sequence (Genbank ID KF843588). As seen in Appendix E Figure 1 and as noted in Section 4.9.2, several South African alphaCoV sequences obtained from *N. capensis* bats cluster phylogenetically within this group of sequences (Geldenhuys *et al.*, 2013; Ithete, 2013). As reported previously, these sequences show relatedness to a *Nyctalus* sp.-borne sequences, represented in Figure 4.5 by a sequence from a *N. velutinus* bat from China (Genbank ID NC_028833, unpublished study from Yan *et al.,* 2014) (Geldenhuys *et al.*, 2013). From Figure 4.5 it is evident that these sequences do not cluster near any ICTV-recognised CoV species. Phylogenetic analysis of non-identical amino acid sequences as depicted in Figure 4.6, similarly show these *N. capensis*-derived sequences to cluster together. The amino acid pairwise distance between the *N. capensis*-derived sequences and the *N. nana*-derived sequence was 1.5%, signifying a strong likelihood that they belong to the same CoV species. Amino acid pairwise distances between these *Neoromicia* sp.-derived sequences and the seemingly related *Nyctalus* sp.-derived sequences was greater than 4.8% indicating that these *Neoromicia* sp.-derived sequences putatively represent a novel CoV species that for reference purposes will be referred to as unclassified *Neoromicia* BtCoV 1. Based on the observations made earlier regarding other closely related South African *N. capensis*-derived sequences as seen in Appendix E Figure 1, these sequences likely all belong to the unclassified *Neoromicia* BtCoV 1 species.

At position five in Figure 4.5 two study sequences, 20160906DC95 and 20160522DC84, show strong phylogenetic relatedness to a South African BtCoV sequence (Genbank ID KF843853) from a *M. condylurus* bat that was previously demonstrated to show close relatedness to a short sequence from another *M. condylurus* bat from South Africa (Geldenhuys *et al.*, 2013; Ithete *et al.*, 2013). These three South African sequences are seen to cluster with sequences from three different bat species in Brazil (Genbank ID KU552079, KC110782, KT717390) (Asano *et al.*, 2016; Góes *et al.*, 2016). All six sequences represented in Figure 4.5 were obtained from different bat

111

species implying low bat-species specificity for these CoV. From Figure 4.5, these sequences do not show close phylogenetic relatedness to any ICTV-recognised CoV species. Longer sequences were not available for all of these sequences to be included in the corresponding amino acid phylogenetic analysis to assess their relatedness. In Figure 4.6, at position 5, the sequence from a *C. pumilus* sample, 201609DC95, is shown to cluster with a sequence (Genbank ID HQ728486) obtained from a *Chaerephon* sp. bat in Kenya (Tao *et al.,* 2012). This sequence (Genbank ID HQ728486) was reported to represent a putative unclassified alphaCoV. The amino acid pairwise distance between sample sequence 201609DC95 and sequence HQ728486 was 0.4% signifying that by the RGU criterion, these sequences belong to the same CoV species. For reference purposes this CoV species will be referred to as unclassified *Chaerephon* BtCoV KY22. Based on the phylogenetic tree in Figure 4.5, it is therefore likely that the sequences previously reported from the two *M. condylurus* bats as well as the sample sequence from a *R. simulator* bat, 20160522DC84, all belong to the putative unclassified *Chaerephon* BtCoV KY22 species.

At position six in Figure 4.5, three *Rhinolophus* sp.–borne study sequences show phylogenetic relatedness to two sequences from Kenya (Genbank ID HQ728480 and GU065425, unpublished study from Conrardy *et al.,* 2009) (Tao *et al.*, 2012). Sequence HQ728480 was obtained from a *Cardioderma cor* bat while the bat species from which sequence GU065425 was obtained is unknown. Sequence HQ728480 was previously reported to represent a putative novel CoV species (Tao *et al.*, 2012). Phylogenetic analysis of available amino acid sequences (Figure 4.6) similarly shows the sequence from sample 20150921SKK_RC10 to cluster with sequence HQ728480 with a corresponding amino acid pairwise distance of 4.5%. According to the RGU criterion, this indicates that sample sequence 20150921SKK_RC10 belongs to the putative novel alphaCoV represented by sequence HQ728480. Based on the phylogenetic tree depicted in Figure 4.5, it is therefore likely that the study sample sequences at position 6 all belong to the same CoV species, demarcated here for reference purposes as unclassified *Cardioderma* BtCoV KY43.

At position seven in Figures 4.5 and 4.6, a study sequence detected in a *S. dinganii* bat shows strong phylogenetic relatedness to the ICTV-recognised *Scotophilus* BtCoV 512. The amino acid pairwise distance between these two sequences was 3.5% confirming that this sample sequence likely to belongs to the *Scotophilus* BtCoV 512 species (Adams *et al.*, 2016).

At position eight, two study sequences obtained from *Neoromicia* sp. bats show strong phylogenetic relatedness to a South African *N. capensis*-borne sequence (Genbank ID KF843858) that was previously shown not to cluster with the unclassified *Neoromicia* BtCoV sequences (Ithete, 2013). Sequences available for inclusion in the 395 bp phylogenetic tree depicted in Figure 4.5 show these sequence to have some relatedness to a sequence from a *C. pumilus* bat from Kenya (Genbank ID HQ728481) (Tao *et al.* 2012). Sequence HQ728481 was previously reported to represent a putative novel unclassified alphaCoV (Tao *et al.,* 2012). The phylogenetic tree in

Figure 4.5 depicts these sequences as clustering within the branches of the alphaCoV phylogeny. However, in Figure 4.6, these sequences can still be seen clustering together but are distinctly positioned outside the alphaCoV phylogenetic tree. Amino acid pairwise distances between the sample sequences and sequence HQ728481 exceeded 8.0% signifying that these sample sequences do not belong to the same unclassified putative novel alphaCoV species as sequence HQ728481. Instead these sequences, obtained from *N. capensis* and *N. nana* bats, represent a putative novel unclassified alphaCoV of their own that for reference purposes will be referred to as unclassified *Neoromicia* BtCoV 2.

Overall eight different *Alphacoronavirus* species sequences were detected in samples obtained from the general surveillance sampling effort and two *Alphacoronavirus* species were detected in the *N. capensis* samples collected as part of the species-specific surveillance effort.

4.9.4.2 General surveillance: betacoronaviruses

During this study 31 betaCoV positive samples were detected from which twenty-three 395 bp and twenty-four 816 bp (272 aa) partial RdRp sequences were obtained. These sequences were detected from 25 *N. capensis* and 6 *P. hesperidus* bats. The average amino acid pairwise distance across non-identical study sequences was 1.2%, strongly suggesting that by the RGU criterion, these sequences all belong to the same *Betacoronavirus* species (Drexler *et al.*, 2010, Drexler *et al.*, 2014). Figures 4.8 and 4.9 depict the phylogenetic trees resulting from the analyses of the 395 bp and 272 amino acid sequence datasets respectively. The corresponding phylogenetic tree for the 272 aa dataset using Bayesian Inference in Mr Bayes is depicted in Appendix E Figure 6 and was depicted similar phylogenetic clustering of sequences as demonstrated with the ML phylogenetic tree depicted in Figure 4.9.
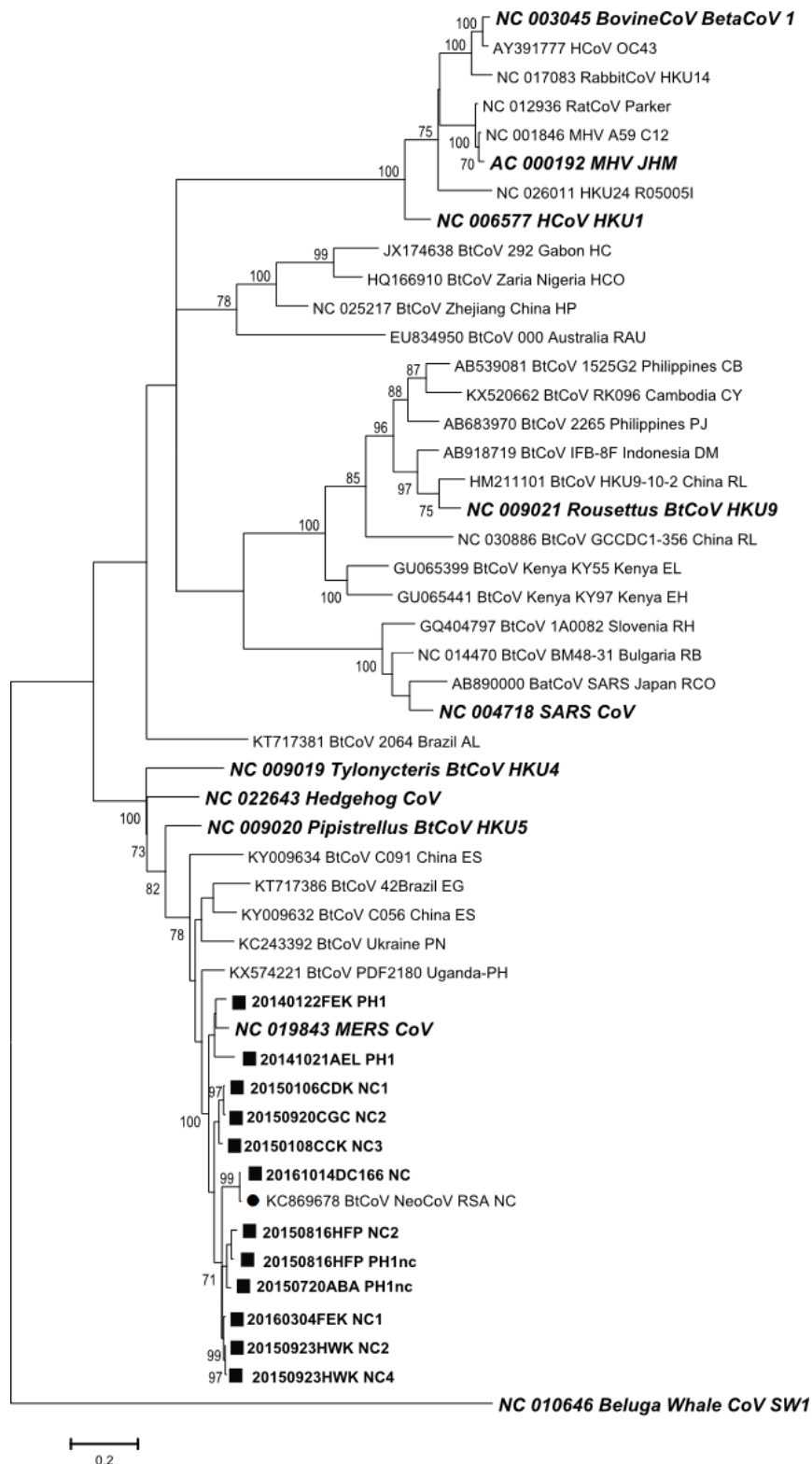
Figure 4.8 Phylogeny of partial (395 bp) betacoronavirus RNA dependent RNA polymerase gene sequences. Phylogenetic inference was determined by the Maximum Likelihood method with 1000 bootstrap replicates in MEGA7 (Kumar *et al.*, 2016). The General Time Reversal model with gamma distributed rate variation with invariant sites was used (GTR+G+I). The tree was midpoint rooted with Beluga Whale CoV SW1 (NC_010646). KEY: bold oblique font = ICTV-recognised coronavirus species sequences and RefSeq sequences of interest; ■ = study sequences; ● = non-study South African sequences.
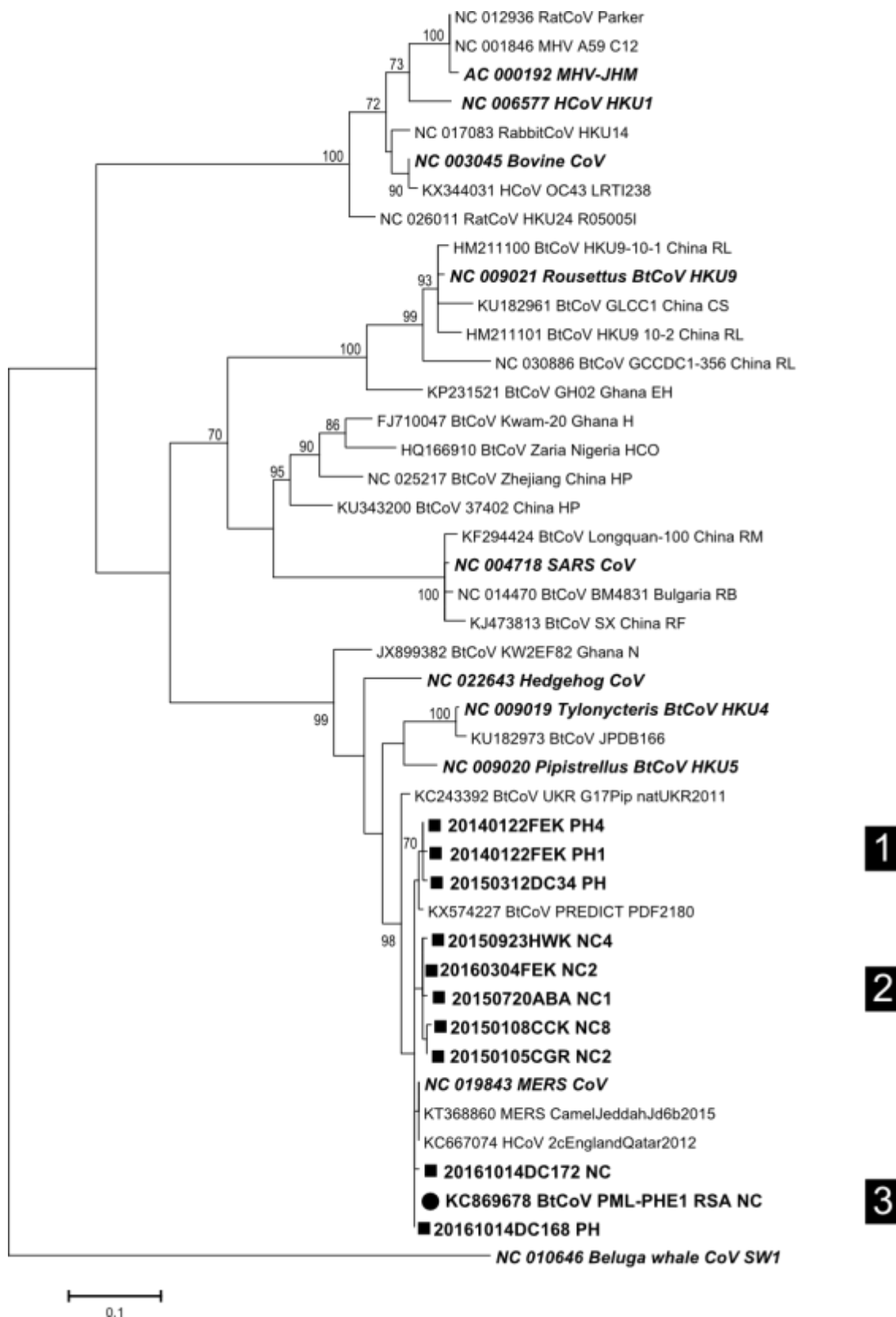
Figure 4.9 Phylogeny of partial (272 aa) betacoronavirus RNA dependent RNA polymerase gene sequences. Phylogenetic inference was determined by the Maximum Likelihood method with 1000 bootstrap replicates in MEGA 7. The Le Gascuel substitution model with gamma distributed rate variation and invariant sites was used (LG+G+I). The tree was midpoint rooted with Beluga whale CoV SW1 (NC_010646). KEY: bold oblique font = ICTV-recognised coronavirus species sequences and RefSeq sequences of interest; ■ = study sequences; ● = non-study South African sequences (Le and Gascuel, 2008; Kumar *et al.*, 2016).

From both Figure 4.8 and 4.9 all betaCoV sequences from study samples show close phylogenetic relatedness to the ICTV-recognised MERS-related CoV, confirming the blastn results noted in Section 4.4. From Figure 4.8, sample sequences can be seen clustering in two groups. betaCoV sequences from *N. capensis* study samples are seen clustering with NeoCoV while the two *P. hesperidus*-derived sample betaCoV sequences are seen clustering with MERS-CoV (Genbank ID NC_019843). This apparent clustering of betaCoV sample sequences by bat species may indicate the presence of bat species-specific variants within the MERS-related CoV species. The recently described MERS-related CoV sequence from a *P. hesperidus* bat in Uganda is seen to cluster outside the *N. capensis*-derived group of betaCoV sample sequences (Anthony *et al.*, 2017a). This sequence was reported to show higher similarity to the South African NeoCoV sequence than to MERS-CoV for this genome region (Anthony *et al.*, 2017a).

Phylogenetic analyses of the available non-identical 272 aa sequences as seen in Figure 4.9 echoes the observations noted in Figure 4.8 to some extent. Again, it appears that with one exception, betaCoV sample sequences cluster by bat species. Except for sample sequence 20161014DC168 that is seen to cluster with NeoCoV, the *P. hesperidus*-derived sample sequences, at position 1 on Figure 4.8, cluster with the betaCoV sequence form a *P. hesperidus* bat in Uganda (Genbank ID KX574227) (Anthony *et al.*, 2017a). Pairwise amino acid distances across these *P. hesperidus*-derived sample sequences reveal equal distances to NeoCoV and the Ugandan *P. hesperidus* betaCoV. In Figure 4.9, the *N. capensis*-derived betaCoV sample sequences appear to cluster in two groups, indicated at position 2 and 3. Pairwise amino acid distances across the two groups indicate the cluster at position 3 to have the shortest pairwise amino acid distances (0.4 – 0.7%) to the human MERS-CoV sequence (Genbank ID NC_019843) representing the MERS-related CoV species. The sequences at position 2 showed pairwise amino acid distances of 0.7 – 1.1% to NeoCoV and 1.1 – 1.5% to MERS-CoV (Genbank ID NC_019843) and the Ugandan *P. hesperidus*-derived betaCoV (Genbank ID KX574227) (Corman *et al.*, 2014a; Anthony *et al.*, 2017a).

4.9.5 Summary of bat coronaviruses detected as part of the general surveillance effort

This section briefly summarises the results from the phylogenetic analyses pertaining to the putative classification of BtCoVs detected during this study as a representation of their diversity.

Based on sequence alignments, phylogenetic trees, and pairwise distances of amino acid sequences corresponding to the RGU region for putative CoV classification, 8 putative *Alphacoronavirus* species and 1 *Betacoronavirus* species were detected during this study (Drexler *et al.*, 2010). Of the 8 putative *Alphacoronavirus* species detected, 3 are recognised as CoV species by ICTV. Samples containing sequences seemingly belonging to *Miniopterus* BtCoV

HKU8, *Miniopterus* BtCoV 1A, and *Scotophilus* BtCoV 512 were detected. Of the 5 remaining putative *Alphacoronavirus* species detected, 3 appear to be novel with pairwise amino acid distances all > 4.8% to the closest related ICTV-recognised species sequences available. The remaining two *Alphacoronavirus* species detected contain sequences related to an unclassified BtCoV KY22 and unclassified BtCoV KY43 (Tao *et al.*, 2012). All 31 betaCoV sequences detected were shown by phylogenetic analyses and pairwise distances to belong to a single ICTV-recognised CoV species, MERS-related CoV.

Table 4.11 details the different BtCoV species detected during this study. Numbers of each detected CoV species and of samples from each bat species in which each CoV species was detected are detailed.

Table 4.11 Bat coronavirus species detected during this study. This table details the five putative and four recognised coronavirus species detected from nine different bat species during this study.

Genus	CoV species	Number detected	Bat species detected in	
Alphacoronavirus	Unclassified *Neoromicia* BtCoV 1	44	NC = 43	NN = 1
	Unclassified *Neoromicia* BtCoV 2	3	NC = 2	NN = 1
	Miniopterus BtCoV HKU-8	4	MN = 4	
	Miniopterus BtCoV 1A	1	MN = 1	
	Scotophilus BtCoV 512	1	SD = 1	
	Unclassified *Cardioderma* BtCoV KY43	3	RCL = 2	RC = 1
	Unclassified *Miniopterus* BtCoV	1	MN = 1	
	Unclassified *Chaerephon* BtCoV KY22	2	CP = 1	RS = 1
Betacoronavirus	MERS-related CoV	31	NC = 26	PH = 5

4.10 Species-specific phylogenetic analyses of coronaviruses detected in *Neoromicia capensis* bats

As described in Section 4.9.4, three different CoV species were detected in the *N. capensis* samples obtained as part of a focussed surveillance effort of this bat species. With the noted discrepancy in detection values between the Pan-CoV and Extended RdRp PCR assays (Section 4.4), and with only a subset of samples undergoing repeat screening, prevalence values could not be accurately estimated. Overall, 61 CoV positive N. capensis individuals were identified of which 10 were found to be coinfected with 2 different BtCoVs. From these positive individuals, 43, 2, and 26 unclassified *Neoromicia* BtCoV 1 and 2, and MERS-related CoV sequences were identified respectively.

As mentioned, during this study 10 *N. capensis* samples were found to be coinfected; 9 were coinfected with a MERS-related CoV and unclassified *Neoromicia* BtCoV 1, and 1 individual was coinfected with a MERS-related CoV and unclassified *Neoromicia* BtCoV 2. The coinfected bats

117

were all sampled from bat trapping sites in the Western Cape, namely sites FEK (n = 3), HWK (n = 2), CGC (n = 2), CDK (n = 1), CCK (n = 1), and CGR (n = 1).

It was noted in Table 4.11 in Section 4.9.5 that, collectively from the 196 *N. capensis* samples screened for the presence of CoVs, 43 unclassified *Neoromicia* BtCoV 1*,* 2 unclassified *Neoromicia* BtCoV 2, and 26 MERS-related CoV sequences were obtained for phylogenetic analyses. Since only 2 unclassified *Neoromicia* BtCoV 2 sequences were obtained, additional phylogenetic analyses to assess CoV diversity at the species level only included unclassified *Neoromicia* BtCoV 1 and MERS-related CoV sequences.

4.10.1 The diversity of unclassified *Neoromicia* BtCoV 1 in *N. capensis* bats

To assess the existing diversity of unclassified *Neoromicia* BtCoV 1 in *N. capensis* bats, all available 395 bp *N. capensis*-derived sequences, excluding those as indicated in Section 4.9, underwent phylogenetic analysis. The average pairwise distance between non-identical sequences was found to be 4.4%. The evolutionary differences between these partial sequences of the relatively conserved *RdRp* gene were inferred with the NJ method using p-distances in MEGA7 (Saitou and Nei, 1987; Kumar *et al.*, 2016).

From Figure 4.10 it can be seen that sequences from the same bat trapping site are likely to cluster together and that there is some evidence of geographic clustering, particularly for sequences obtained in the Fynbos biome of the Western Cape. For some bat trapping sites, HWK, CDK, CGC, and the South African sequences from Greyton (Gr), no sequence variation was observed even between different sampling events while for sequences from other bat trapping sites, such as previously noted for sequences form bat trapping site LNR, some variation was observed not only between different sampling events, but also for sequences obtained from the same sampling event.

Position one in Figure 4.10 refers to a cluster of sequences obtained from three different bat trapping sites in the Eastern Cape. Of interest are sequences obtained from site ABA that can be seen to cluster in three different positions (indicated by ◆) within this phylogenetic tree.

Position two in Figure 4.10 refers to a cluster of sequences shown to be closely related. These sequences were obtained from three different sites, CDK, CGC, and CCK, located in the Western Cape, providing some evidence for geographic clustering. Although these sequences are closely related, they group distinctly according to bat trapping site. Of interest is the sequence from bat trapping site CGR (indicated by ■) that clusters outside the group of sequences from bat trapping site HWK located near Swellendam in the Western Cape. The sequence, 20150105CGR_NC2, was obtained from a bat trapped at the main house on the same property as those caught at

nearby guest cottages, designated as bat trapping site CGC, on the same property. This sequence, 20150105CGR_NC2, does not appear to share the same sequences similarities as sequences from CDK, CGC, and CCK, as observed at position one.
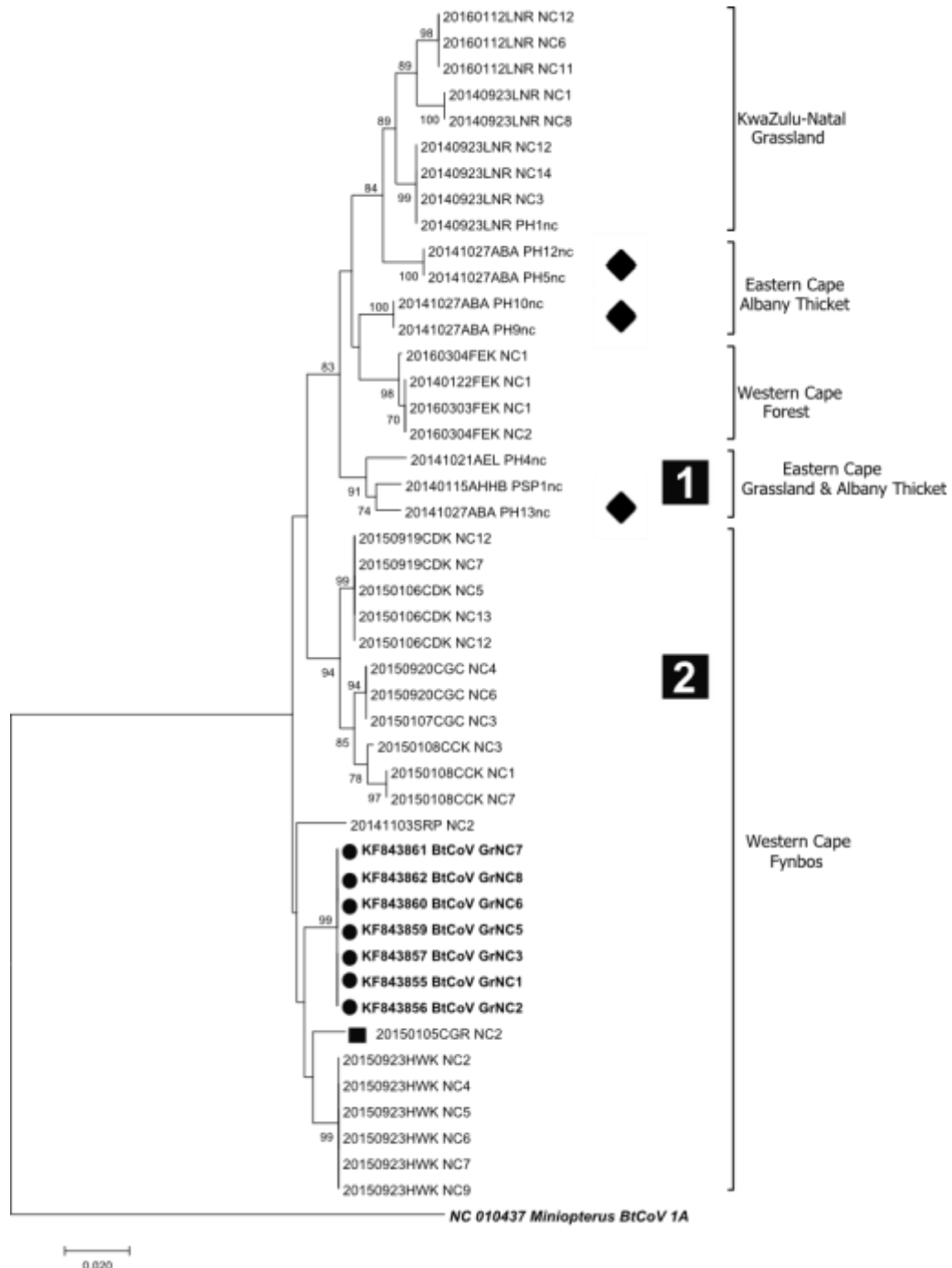
Figure 4.10 Phylogeny of unclassified *Neoromicia* BtCoV 1 in *Neoromicia capensis* bats. Phylogenetic relatedness was inferred using the neighbour-joining method with p-distances and 1000 bootstrap (BS) replicates in MEGA7 (Saitou and Nei, 1987; Kumar *et al.*, 2016). The province and biome in which sequences were obtained is indicated. There were 395 nucleotide positions in the final alignment. KEY: ● with bold font = non-study South African sequences, ◆ and ■ = study sequences of interest. The tree was midpoint rooted with NC_010437 *Miniopterus* BtCoV 1A.

4.10.2 The diversity of unclassified *Neoromicia* BtCoV 1 in an individual bat

As noted in sections 4.4.2 and 4.7.3, two unclassified *Neoromicia* BtCoV 1 strains, 2014/1 and 2014/2, were detected in a sample from an individual bat captured at bat trapping site LNR. Sequences from CoV positive bats trapped at the same site in January 2016 revealed the presence of a new strain, 2016/1, that differed from strains 2014/1 and 2014/2. This can be seen in the different clustering of sequences from the LNR bat trapping site in Figure 4.10 in Section 4.10.1.

Clonal sequences obtained from sample 20140923LNR_NC7 coinfected with strains 2014/1 and 2014/2 were aligned with other sequences obtained at the LNR bat trapping site and phylogenetic relatedness inferred with the NJ method. Figure 4.11 depicts the resulting phylogenetic tree. Strain 2014/1 and strain 2014/2 account for the majority of clonal sequences with 14 additional non-identical sequence variants interspersed within the phylogenetic tree. From Figure 4.11, it is evident that no clonal sequences detected in the sample obtained in September 2014 were identical to the 2016/1 strain sequences obtained nearly 16 months later.

An additional phylogenetic analysis was conducted to assess if any of the clonal sequences detected in samples 20140923LNR_NC7 were identical to other unclassified *Neoromicia* BtCoV 1 sequences obtained during this study. Non-identical clonal sequences were aligned with non-identical *N. capensis*-derived unclassified *Neoromicia* BtCoV 1 sequences and phylogenetic relatedness inferred using the NJ Method. The resulting phylogenetic tree is depicted in Appendix E Figure 7. Clonal sequences from sample 20140923LNR_NC7 were found not to be identical to any other *N. capensis*-derived unclassified *Neoromicia* BtCoV 1 sequences detected during this study and were only found to cluster with sequences obtained at the LNR bat trapping site.
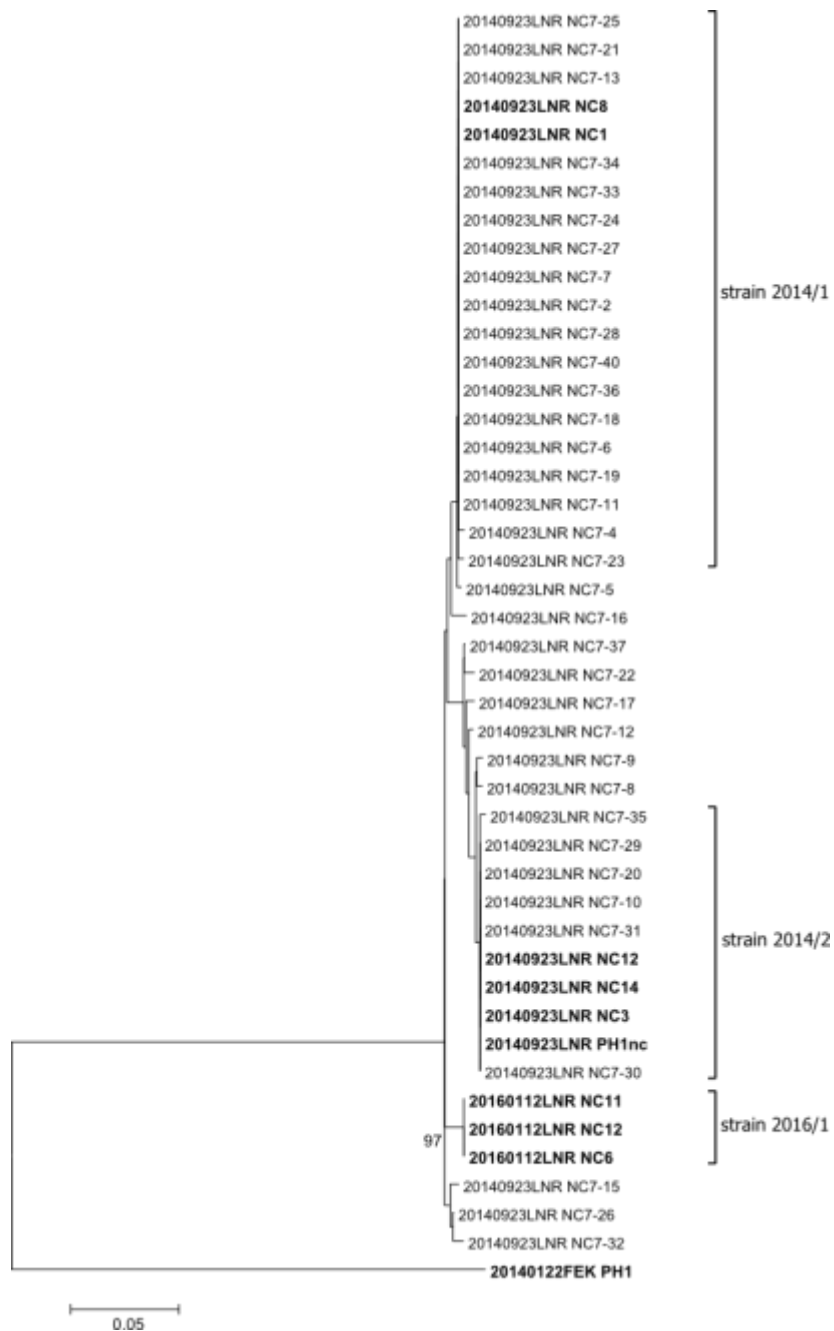
Figure 4.11 Phylogenetic analysis of clonal sequences from sample 20140923LNR_NC7 and other LNR-derived sequences. Phylogenetic relatedness was inferred using the Neighbour Joining method and p-distance with 1000 bootstrap replicates in MEGA7(Saitou and Nei, 1987; Kumar *et al.*, 2016). The tree was midpoint rooted with a betacoronavirus study sample sequence, 20140122FEK_PH1. KEY: non-clonal study sequences indicated in bold font.

4.10.3 MERS-related CoV diversity in *Neoromicia capensis* bats

To assess the existing diversity of MERS-related CoVs in *N. capensis* bats, all available *N. capensis*-derived betaCoV sequences, excluding those as indicated in Section 4.9.2, underwent phylogenetic analysis. To enable sequences from all bat trapping sites to be represented in the analyses, the alignment was shortened to 358 nt due to differences in overlapping regions between the 395 bp and 816 bp sequences. The pairwise distances between non-identical sequences across this 358 nt alignment was found to be 4.8%. The evolutionary relatedness between these partial sequences of the relatively conserved *RdRp* gene was inferred with the NJ method.
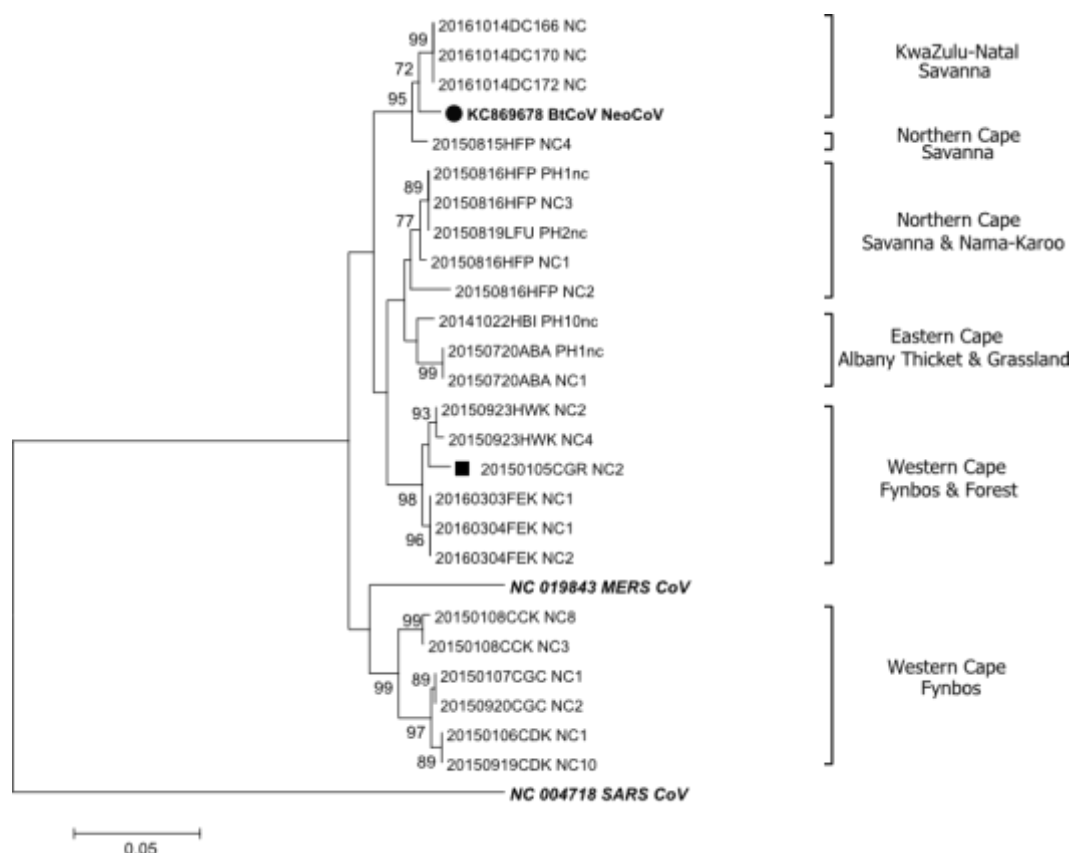
Figure 4.12 The phylogeny of MERS-related CoV in *Neoromicia capensis* bats. Phylogenetic relatedness was inferred using the Neighbour Joining Method with p-distances and 1000 bootstrap (BS) replicates in MEGA7 (Saitou and Nei, 1987; Kumar *et al.*, 2016). BS values less than 70% have been omitted. There were 358 nucleotide positions in the final alignment. The province and biome in which sequences were obtained is indicated. KEY: ● with bold font = non-study South African sequences, ■ = study sequence of interest. The tree was midpoint rooted with NC 004718 SARS-CoV.

Figure 4.12 shows MERS-related CoV sequences obtained from *N. capensis* bats to cluster geographically. The *N. capensis*-derived MERS-related CoV sequences are seen clustering into groups according to bat trapping site regions, namely KwaZulu-Natal, Eastern Cape, Western Cape, and Northern Cape with one exception; the sequence from sample 20150515HFP_NC4 is

122

positioned outside the cluster of sequences from KwaZulu-Natal instead of clustering with sequences from the Northern Cape.

As described in section 4.10.1, sequences from three sites, CCK, CGC, and CDK, in the Western Cape were shown to cluster together as before. Again, of interest is the sequence obtained from sample 20150105CGR_NC2 (indicated by ■), that is positioned outside a group of sequences from the Swellendam bat trapping site HWK in the Western Cape. Both the alpha- and betaCoV sequence obtained from this coinfected sample are positioned near sequences from the HWK bat trapping site rather than the CGC bat trapping site located on the same property as bat trapping site CGR.

4.11 Phylogenetic analyses of partial and complete betacoronavirus gene sequences

As described in Section 4.3.9.3, additional complete and partial gene sequences were obtained representing three different MERS-related betaCoVs detected during this study, 20150720ABA_NC, 201603FEK_NC, and 20140122FEK_PH. The complete E, M, and N gene sequences were obtained while partial sequences were obtained for the ORF1a, ORF1b, and S protein subunits 1 and 2. Except for the E protein, phylogenetic analysis was carried out on the amino acid alignments of these sequences to assess evolutionary differences at the functional protein level. The E protein sequence is only 84 amino acids in length and adequate phylogenetic resolution could not be achieved using ML and Bayesian inference methods; phylogenetic analysis of the E protein was therefore conducted with the corresponding nucleotide sequences. As detailed in Section 4.6, additional partial and full sequences were obtained for ORF3, 4a, 4b, and 5. Since little is known about the functions of these putative genes, as in other publications, they have not been included in the phylogenetic analyses detailed in this section (Corman *et al.*, 2014a; Anthony *et al.*, 2017a).

Sequences included in the dataset for phylogenetic analysis of lineage C betaCoVs were based on sequences included previously in alignments used in publications reporting on the relatedness of NeoCoV and a *P. hesperidus*-derived betaCoV to MERS-CoV for comparative purposes (Corman *et al.*, 2014a; Anthony *et al.*, 2017a). Two additional camel-derived sequences (Genbank ID KX108943 and KU740200), highlighted by BLAST results as detailed in Appendix D Table 4, were included in the dataset.

Phylogenetic trees were constructed as detailed in Chapter 3. Phylogenetic trees constructed using the ML method are depicted in the following sections with the corresponding phylogenetic trees constructed using Bayesian inference depicted in Appendix E Figures 8 – 11.

4.11.1 Phylogenetic analyses of complete gene sequences: E, M, and N

The complete gene sequences of the E, M, and N genes were obtained during this study for phylogenetic analysis. The percentage amino acid identity that each of the study sequences shares with a select group of lineage C prototype betaCoVs, as detailed in Table 4.12, were determined in MEGA7 using pairwise deletions as per published studies (Corman *et al.*, 2014a; Kumar *et al.*, 2016).

Table 4.12 Amino acid sequence pairwise distance comparison with prototype lineage C betacoronaviruses. The pairwise distances between study sample amino acid sequences for the full envelope (E), membrane (M), and nucleocapsid (N) proteins and prototype lineage C betacoronaviruses are detailed below. Prototype sequences: MERS-CoV (NC_019843), NeoCoV (KC869678), PDF-2180 (KX574227), HKU4 (NC_009019), HKU5 (NC_009020), EriCoV (NC_022643).

Sample sequence	ORF	No. amino acids	% amino acid identity of study betaCoV sequences with:					
			MERS	NeoCoV	PDF 2180	HKU4	HKU5	EriCoV
20140133FEK_PH		82	91,5	96,3	92,7	70,7	72	75,6
20150720ABA_NC1	E	82	89	97,6	91,5	73,2	72	74,4
201603FEK_NC		82	91,5	97,6	93,9	73,2	72	97,56
20140122FEK_PH		219	95,9	93,6	95,9	80,4	83,1	79,4
20150720ABA_NC1	M	219	94,5	99,1	94,5	82,2	82,6	80,3
201603FEK_NC		219	96,3	98,6	95,9	82,2	83,6	79,4
20140122FEK_PH		413	90,8	93,2	90,8	75,6	74,8	72,4
20150720ABA_NC1	N	414	92,1	99,2	90	75,6	74,8	71,4

From Table 4.12, it can be seen that for all three proteins, the study sample sequences show highest identity to NeoCoV with one exception; the M protein sequence from study sample 20140122FEK_PH showed highest identity to MERS-CoV (Genbank ID NC_019843) and PDF-2180 (Genbank ID KX574227) instead of NeoCoV (Genbank ID KC869678) (Corman *et al.*, 2014a; Anthony *et al.*, 2017a). For all three proteins the study sample sequences derived from *N. capensis* bats showed higher identity to NeoCoV than the *P. hesperidus* derived study sequence indicating divergence between sequences from the two different bat host species.

Figure 4.13 depicts the ML phylogenies of lineage C betaCoVs including study sample sequences for the E and M protein. For the E gene it can be seen that the study sample sequences cluster together with NeoCoV. The corresponding phylogenetic tree constructed using Bayesian inference in Appendix E Figure 8 similarly depicts study samples clustering with NeoCoV. In Figure 4.15, for the M gene, the apparent divergence between sequences from different bat host species is demonstrated with the two *P. hesperidus*-derived sequences, study sample 20140122FEK_PH and PDF 2180, forming a cluster with human and camel-derived MERS-CoV sequences while the *N. capensis*-derived study sequences are seen clustering together with NeoCoV. This confirms the previously reported blastn (Section 4.6) and amino acid pairwise distance results (Table 4.12).

124
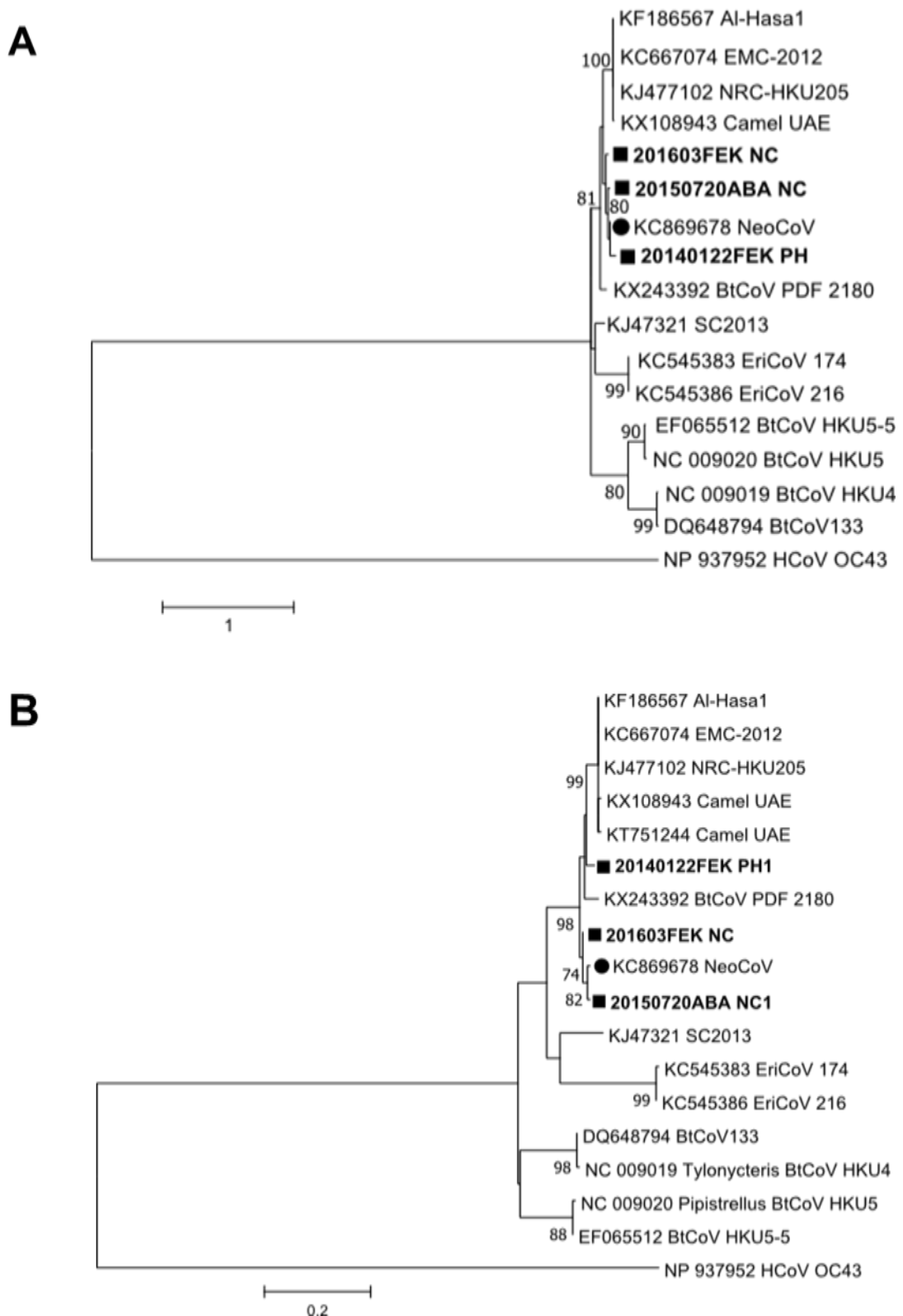
Figure 4.13 Maximum Likelihood phylogeny of the complete envelope (E) and membrane (M) gene sequences of lineage C betacoronaviruses. Maximum Likelihood phylogenies were constructed in MEGA7 with statistical support measured with 1000 bootstrap (BS) replicates (BS values less than 70% have been omitted) (Kumar *et al*., 2016).All positions containing gaps and missing data were eliminated. Study sample sequences are indicated in bold with ■ and NeoCoV (Genbank ID KC869678) is indicated with ●. The tree was midpoint rooted with human coronavirus OC43 (hCoV OC43) (Genbank ID NP 937952). Figure A depicts the E gene ML phylogeny based on the Tamaru-Nei model with gamma distributed evolutionary rates and invariable variation allowed (T93+G+I) (Tamura and Nei, 1993). There were a total of 246 nucleotide positions in the final dataset. Figure B depicts the ML phylogeny of the M gene based on the Le Gascuel model with gamma distributed evolutionary rates (LG+G) (Le and Gascuel, 2008). There were a total of 218 amino acid positions in the final dataset.

Lastly in Figure 4.14, for the N gene, both study sample sequences represented here are seen clustering with NeoCoV. Together, this cluster of South African-borne sequences can be seen forming a cluster with human and camel-derived MERS-CoV sequences while the *P. hesperidus*-derived sequence, PDF 2180 (Genbank ID KX574227), holds a basal position to the rest of the MERS-related CoV sequences. Here the divergence between sequences from different bat host species is less apparent and may rather be indicative of geographic divergence between South African-derived MERS-related BtCoVs and the Ugandan-derived MERS-related BtCoV, PDF 2180.
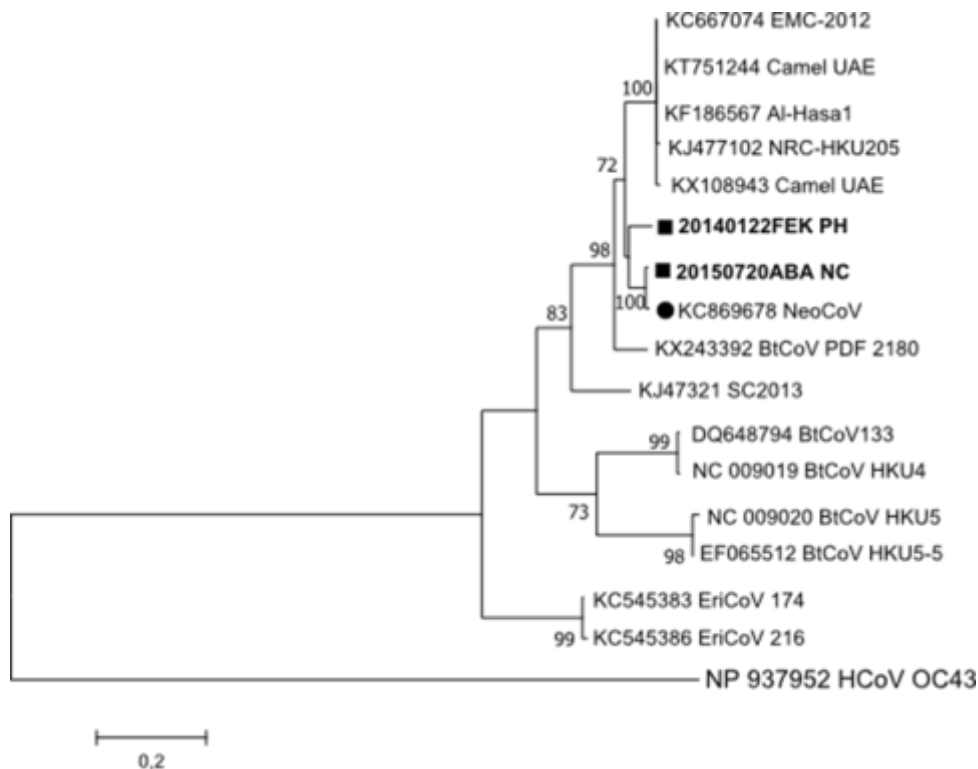
Figure 4.14 Maximum Likelihood phylogeny of complete nucleocapsid (N) gene sequences of lineage C betacoronaviruses. The Maximum Likelihood phylogeny was constructed in MEGA7 based on the Le Gascuel model with gamma distributed evolutionary rate (LG+G) (Le and Gascuel, 2008; Kumar *et al.*, 2016). Statistical support was measured with 1000 bootstrap (BS) replicates (BS values less than 70% have been omitted). All positions containing gaps and missing data were eliminated. There were a total of 381 amino acid positions in the final dataset. Study sample sequences are indicated in bold with ■ and NeoCoV (Genbank ID KC869678) is indicated with ●. The tree was midpoint rooted with human coronavirus OC43 (hCoV OC43) (Genbank ID NP 937952).

4.11.2 Phylogenetic analyses of partial S gene sequences

Partial S1 sequences were obtained for 20140122FEK_PH and 20150720ABA_NC. These sequences however did not overlap significantly and are therefore represented by separate phylogenetic trees in Figure 4.15. Both study sample sequences can be seen forming a cluster with NeoCoV and the *P. hesperidus*-derived MERS-related CoV from Uganda, PDF 2180 (Genbank ID KX574227) (Anthony *et al.*, 2017a). These bat-derived MERS-related CoVs do not cluster with other MERS-related CoVs but instead form a basal cluster to all lineage C betaCoVs as previously reported (Corman *et al.*, 2014a; Anthony *et al.*, 2017a).

126
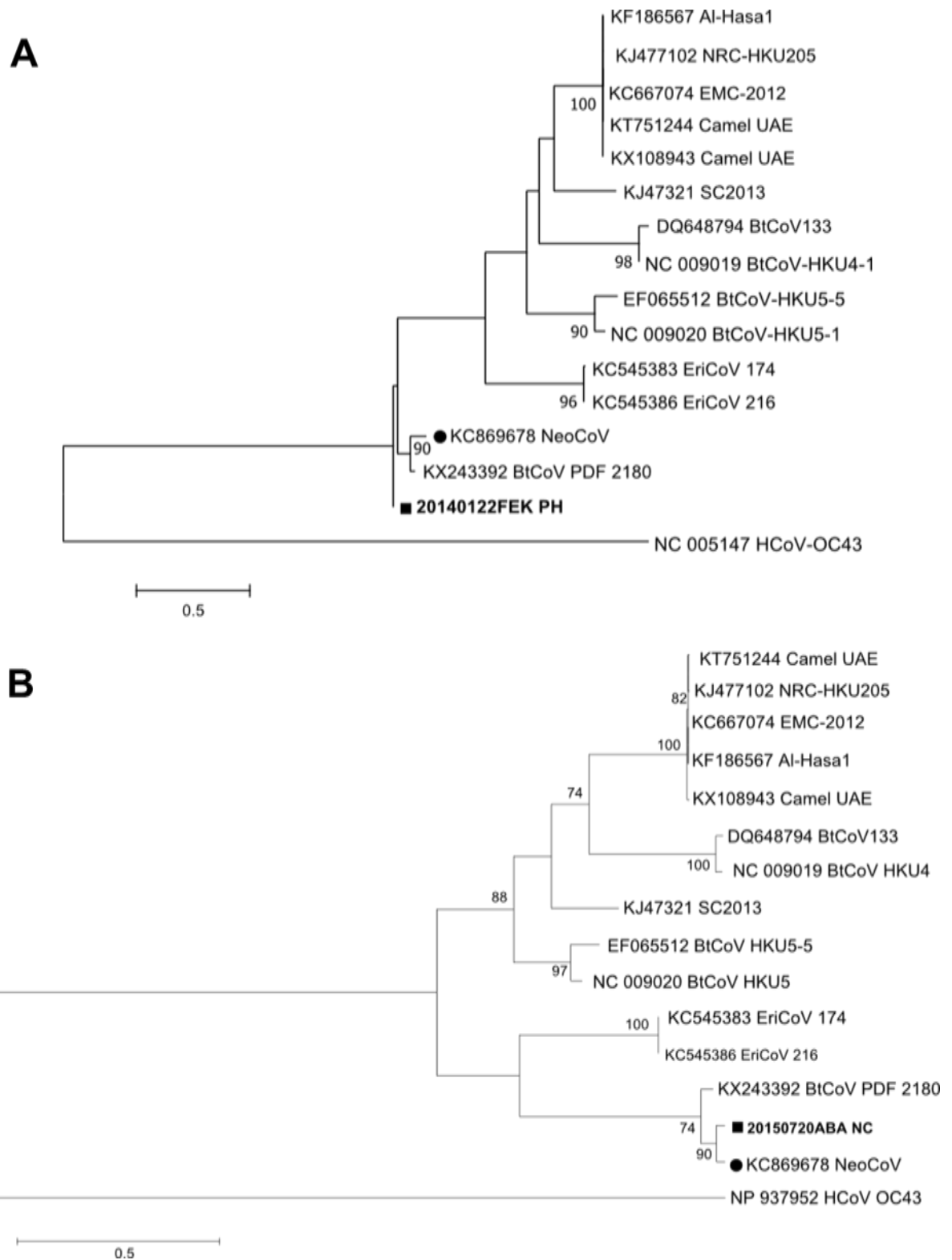
Figure 4.15 Maximum Likelihood phylogeny of partial Spike (S) subunit 1 sequences of lineage C betacoronaviruses. The Maximum Likelihood phylogenies were constructed in MEGA7 based on the Whelan and Goldman model with gamma distributed evolutionary rates (WAG+G) (Whelan and Goldman, 2001; Kumar *et al.*, 2016). Statistical support for each tree was measured with 1000 bootstrap (BS) replicates (BS values less than 70% have been omitted). All positions containing gaps and missing data were eliminated. Study sample sequences are indicated in bold with ■ and NeoCoV (Genbank ID KC869678) is indicated with ●. The trees were midpoint rooted with human coronavirus OC43 (hCoV OC43) (Genbank ID NP 937952). Figure A depicts the ML phylogeny of partial S1 sequences based on an alignment including the partial S1 sequence from study sample 20140122FEK_PH. There were a total of 315 amino acid positions in the final dataset. Figure B depicts the ML phylogeny of partial S1 sequences based on an alignment including the partial S1 sequence from study sample 20150720ABA_NC1. There were 415 amino acid positions in the final dataset.

127

Partial S2 sequences were obtained from 20140122FEK_PH and 201603FEK_NC for phylogenetic analysis. The sequence from 201603FEK_NC was markedly shorter and with only 64 amino acid positions, was excluded from subsequent analyses. The ML phylogeny of the partial S2 sequence of study sample 20140122FEK_PH with prototype lineage C betaCoVs is depicted in Figure 4.16. The study sample sequence 20140122FEK_PH can be seen clustering with a basal sister relationship to the other MERS-related CoVs with a similar clustering pattering observed in the corresponding phylogenetic tree constructed using Bayesian inference in Appendix E Figure 11.
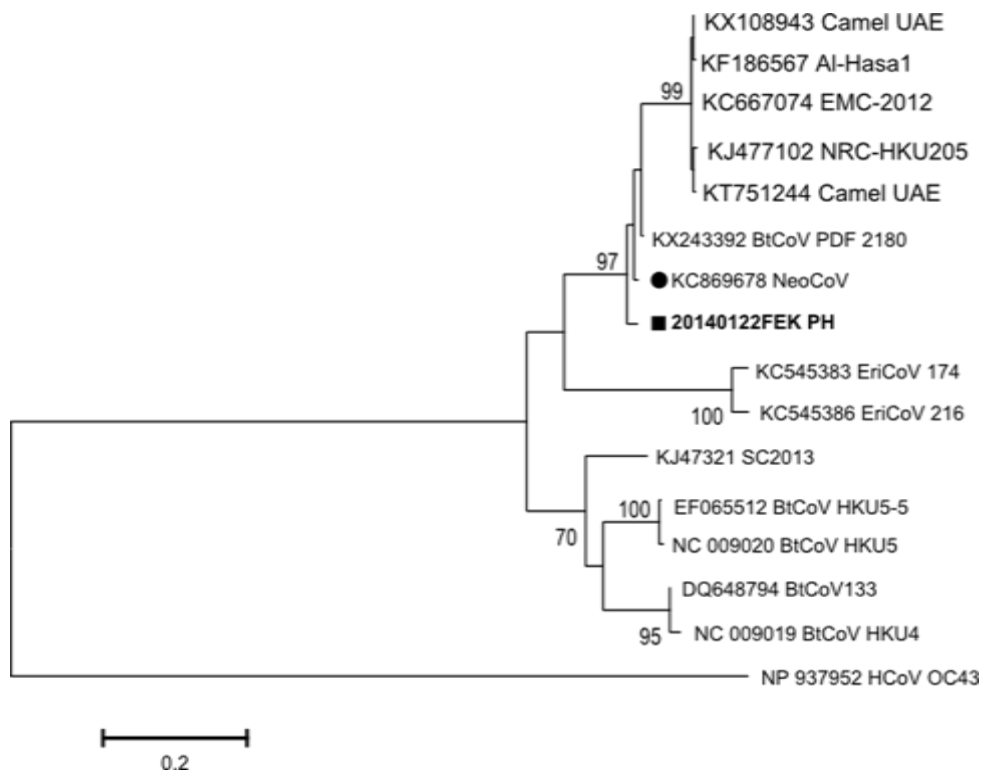
Figure 4.16 Maximum Likelihood phylogeny of partial Spike subunit 2 sequences of lineage C betacoronaviruses. The Maximum Likelihood phylogeny was constructed in MEGA7 based on the Le Gascuel model with gamma distributed evolutionary rates (LG+G). Statistical support was measured with 1000 bootstrap (BS) replicates (BS values less than 70% have been omitted). All positions containing gaps and missing data were eliminated. There were a total of 455 amino acid positions in the final dataset. Study sample sequences are indicated in bold with ■ and NeoCoV (Genbank ID KC869678) is indicated with ●. The tree was midpoint rooted with human coronavirus OC43 (hCoV OC43) (Genbank ID NP 937952) (Le and Gascuel, 2008; Kumar *et al.*, 2016).

4.11.3 Phylogenetic analyses of partial ORF1a and ORF1b gene sequences

Partial sequences were obtained for both ORF1 and ORF1b. The largest overlapping sequences were obtained from study samples 20140122FEK_PH and 20150720ABA_NC1 and these sequences were used for ML and Bayesian inference. The two resulting phylogenetic trees are depicted in Figure 4.17.
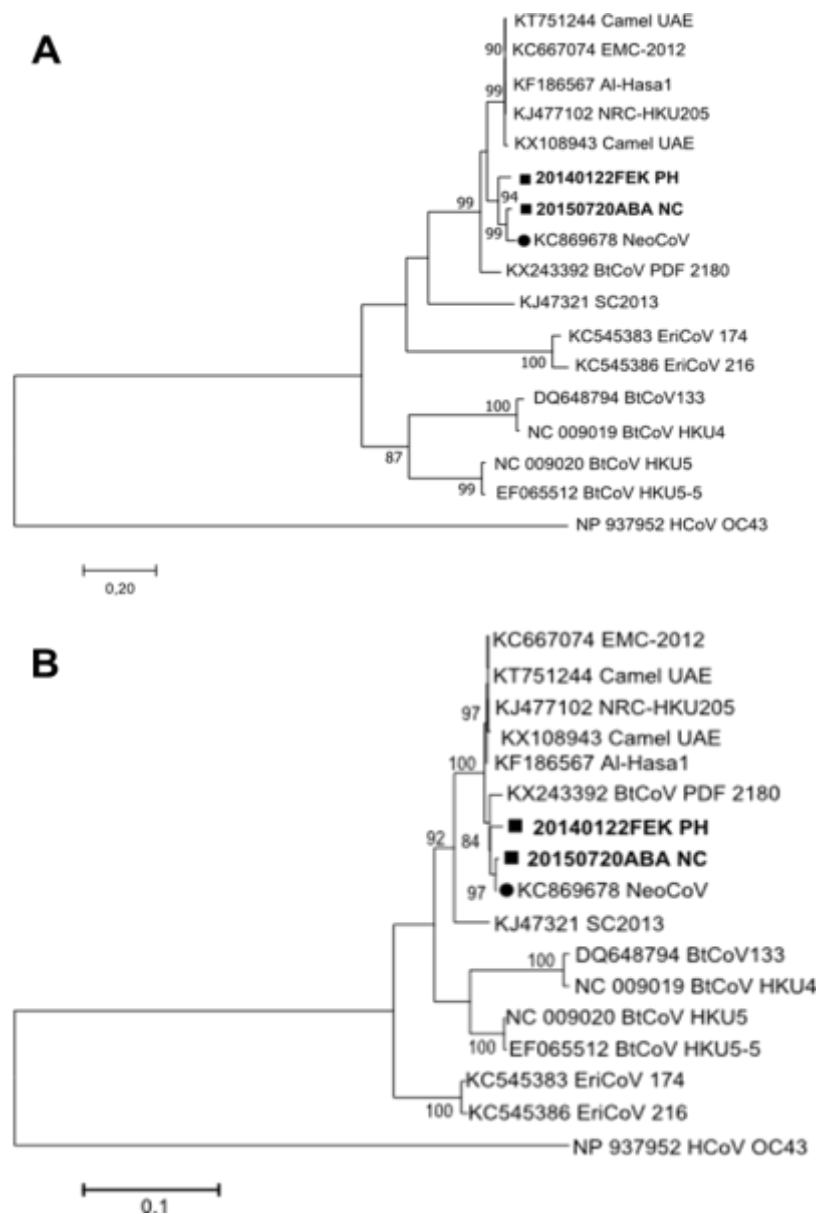
Figure 4.17 Maximum Likelihood phylogeny of partial ORF1a and ORF1b sequences of lineage C betacoronaviruses. The Maximum Likelihood phylogenies were constructed in MEGA7 based on the Le Gascuel model with gamma distributed evolutionary rates and empirical base frequencies (LG+G+F) (Le and Gascuel, 2008; Kumar *et al.*, 2016). Statistical support was measured with 1000 bootstrap (BS) replicates (BS values less than 70% have been omitted). Study sample sequences are indicated in bold with ■ and NeoCoV (Genbank ID KC869678) is indicated with ●. The tree was midpoint rooted with human coronavirus OC43 (hCoV OC43) (Genbank ID NP 937952). Figure A depicts the ML phylogeny of partial ORF1A sequences. There were a total of 866 amino acid positions in the final dataset. Figure B depicts the ML phylogeny of partial ORF1B sequences. There were 1735 amino acid positions in the final dataset.

129

For both the ORF1a and ORF1b phylogenetic trees, the bat-derived MERS-related CoV sequences, NeoCoV, PDF 2180, and study sample sequences 20140122FEK_PH and 20150720ABA_NC1 cluster together. In both instances, the *N. capensis*-derived study sample sequence, 20150720ABA_NC1, demonstrates a closer relationship to NeoCoV than to the *P. hesperidus*-derived sequences. The *P. hesperidus*-derived study sample sequence, 20140122FEK_PH, demonstrates a closer relationship to the South African sequences than to the Ugandan *P. hesperidus* sequence, PDF2180 (Genbank ID KX574227). The phylogenetic trees inferred using Bayesian inference similarly depict this relationship in Appendix E Figure 10. This indicates an apparent genetic divergence between bat host species as well as a genetic divergence linked to geographic distribution.

4.12 Ecological analyses using logistic regression

As described in Chapter 3, logistic regression was used to evaluate potential effects of intrinsic and extrinsic factors on CoV prevalence in *N. capensis* bats.

Because discrepancies in detection rates between the Pan-CoV and Extended RdRp PCR assays were noted during this study, the dataset selection for use in logistic regressions was an important point for consideration. The Extended RdRp PCR yielded a markedly improved detection rate compared to the Pan-CoV PCR assay. However, this assay was used as a screening tool on only a subset of 84 *N. capensis* bats. This markedly smaller sample size rendered the accuracy, support and goodness of fit of logistic regression outputs low (significant residual deviance p = 0.02, poor model fit McFadden R^2 = 0.17). The dataset used for logistic regressions was therefore derived from the *N. capensis* screening results using the Pan-CoV PCR assay across all 196 samples. For ease of reference, unclassified *Neoromicia* BtCoV 1 will be referred to as NeoBtCoV1 in this section.

The best model based on AICc values included the biome, sex, altitude, and maximum temperature for the month in which sampling took place at the bat trapping site. The top five models identified along with their corresponding AICc and weighted values, representing each model's relative likelihood, are reported in Table 4.13. The AICc values of the top five models were considerably similar.

Table 4.13 Model fitting results. The top five models identified in the logistic regression analyses with associated AICc and weighted values.

	Model	AICc value	Weighted AICc value
1	NeoBtCoV1 ~ 1 + biome + sex + altitude + maximum temperature	191.27	0.067
2	NeoBtCoV1 ~ 1 + biome + sex + altitude + maximum temperature + FMI	191.30	0.066
3	NeoBtCoV1 ~ 1 + biome + sex + altitude	191.32	0.065
4	NeoBtCoV1 ~ 1 + biome + sex	192.10	0.044
5	NeoBtCoV1 ~ 1 + biome + sex + altitude + FMI	192.38	0.038

The test of residual deviance for the best model using the chi-squared test was not significant ($p = 0.91$) indicating no evidence against this logistic regression model (i.e. no overdispersion). Bootstrapping showed that the error rate for the best model was 0.23, indicating a fair level of accuracy. In the best model, two of the five variables, biome and sex, significantly reduced the residual deviance (Table 4.14), with an overall decrease in deviance from 211.20 in the null model to 161.50. The McFadden R^2 of 0.33 indicated the best model to be an excellent fit (McFadden 1974).

Table 4.14 Analysis of deviance

	df	Deviance Residual	Df	Residual deviance	p
NULL model			195	211.20	
biome	5	35.82	190	170.42	<0.0001 ***
sex	1	5.09	189	165.33	0.024 *
altitude	1	2.17	188	163.16	0.14
maximum monthly temperature	1	1.67	186	161.50	0.20

[1]Significance codes: 0 *** 0.001 ** 0.01 * 0.05 •

For the best model, sex was the most significant predictor and the Forest biome the second most significant predictor for NeoBtCoV1 infection (p-values <0.05), with altitude close to statistically significant (p-value = 0.073) (Table 4.15). The odds ratio was determined for each significant variable. All determined values are detailed in Table 4.15.

Table 4.15 Logistic regression results for the best model used to identify predictors of NeoBtCoV 1 in *Neoromicia capensis* bats.

Predictor	Coefficient	Std error	Wald's χ^2	df	p-value	odds ratio
Biome:Fynbos	0.53	0.74	0.51	1	0.47	1.70
Biome: Forest	2.68	1.29	4.30	1	0.038 *	14.55
Biome: Grassland	1.47	1.7	1.50	1	0.19	4.34
Biome: Nama-Karoo	-0.17	3181.00	2.80×10^{-5}	1	0.99	5.04×10^{-8}
Biome: Savanna	-0.17	948.90	3.1×10^{-4}	1	0.99	5.11×10^{-8}
Sex: male	-1.22	0.53	5.30	1	0.021 *	0.29
Altitude	-1.21×10^{-3}	6.73×10^{-4}	3.20	1	0.073 •	0.99
Maximum temperature	0.07	0.06	1.60	1	0.20	1.07

[1]Significance codes: 0 *** 0.001 ** 0.01 * 0.05 •
[2]Coefficient = (change in the log odds of the outcome for a one unit increase in the predictor variable)

The negative coefficient for sex suggests that all other variables being equal, male bats were less likely to be NeoBtCoV1 infected with a 1.55 decrease in log odds than female bats. *N. capensis*

131

bats trapped in the Forest biome were more likely to be NeoBtCoV1 positive than those trapped in other biomes with a 2.68 increase in log odds. The collective biome variable was not significant (χ^2 = 5.5, p = 0.36). Altitude had a negative association with NeoBtCoV1 infection i.e. as altitude increased, *N. capensis* bats were less likely to be NeoBtCoV1 infected.

In summary, female *N. capensis* bats and bats trapped at low elevation sites in the Forest biome had the highest probability of being infected with NeoBtCoV1.

Chapter 5: Discussion and Concluding Remarks

5.1. Detection discrepancies between CoV screening PCR assays and recommendations for improved surveillance

The screening protocol used at the start of this study involved screening samples with the widely used Pan-CoV PCR assay developed by De Souza Luna *et al.* (2007) that broadly targets all members of the CoV family (Gloza-Rausch *et al.*, 2008; Pfefferle *et al.*, 2009; Ge *et al.*, 2013; Maganga *et al.*, 2014b). Other commonly used PCR assays to detected CoVs in bats include assays developed by Woo *et al.* (2005) and Watanabe *et al.* (2010) that target a similar region of the genome as the De Souza Luna *et al.* (2007) assay and the PCR assay published by Quan *et al.* (2010) that targets a region of the CoV nucleocapsid protein.

The Pan-CoV PCR was developed in 2007 as a generic reverse transcription PCR (RT-PCR) assay that could broadly detect all members of the CoV family with primers based on CoV sequences known at the time (de Souza Luna *et al.*, 2007). Despite focussing on the detection of human CoV, the primers were designed to target regions of the CoV genome, namely motifs A and C of the RdRp, that are highly conserved at the amino acid level across all known CoVs (Stephensen *et al.*, 1999; de Souza Luna *et al.*, 2007). The initially high sensitivity observed with RNA standards (single RNA copies detected per assay) was greatly reduced when clinical samples from human patients were spiked with RNA standards, resulting in the need for a nested PCR protocol that improved assay sensitivity (de Souza Luna *et al.*, 2007).

Based on sequences obtained from positive samples detected with this PCR assay, CoV lineage-specific primers were then used to extend the sequence read lengths. Using the Extended RdRp PCR assay developed by Drexler *et al.* (2010) allowed for improved phylogenetic resolution and the putative classification of detected CoV. The Extended RdRp PCR assay was developed in response to the need for a provisional classification system, following the finding that the ~400 bp sequences generated by PCR assays broadly targeting the CoV family such as the Pan-CoV PCR assay, are too short for reliable phylogenetic resolution (Drexler *et al.*, 2010). Typically, full CoV genome sequences coupled with the analyses of seven conserved domains within the ORF1ab replicase polyprotein are required by the ICTV for formal classification (ICTV, 2012). Obtaining full CoV genome sequences from field material such as faecal pellets that contain large quantities of background nucleic acids and PCR inhibitors however, is notoriously difficult (Pfefferle *et al.*, 2009; Drexler *et al.*, 2010; Osborne *et al.*, 2011). Additionally, hardly any BtCoVs have been isolated in

133

cell culture, further hindering the generation of sufficient material for full genome sequence amplification (Ge *et al.*, 2013). The Extended RdRp PCR assay generates an 816 bp sequence of the RdRp gene that is used to classify CoV into RdRp Grouping Units (RGU) based on an amino acid pairwise distance criteria that can be used for putative CoV species classification (Drexler *et al.*, 2010).

This protocol, screening followed by extension, was selected for use in this study due to its previous success in our research laboratory and by other groups (Drexler *et al.*, 2010; Annan *et al.*, 2013; Corman *et al.*, 2014b 2013; Ithete, 2013). However, after screening 600 bat faecal samples the Pan-CoV screening PCR assay had only detected 56 (9.3%) positive samples, mostly the alphaCoV *Neoromicia* BtCoV 1 (n = 44) suggesting a possible bias towards the detection of this alphaCoV and an inadequate ability to detect more diverse CoVs in our setting. Other studies using the same Pan-CoV PCR assay have reported similar findings with detection rates ranging from <1.0% to 9.8% with the majority of positive samples being obtained from the same bat species and often infected with the same BtCoV that in most instances was shown to belong to the alphaCoV genus (Gloza-Rausch *et al.*, 2008; August *et al.*, 2012; Maganga *et al.*, 2014b; Fischer *et al.*, 2016). However, recent publications using the same screening protocol have in contrast reported the detection of multiple CoVs across different bat species with increased detection rates of 40.5 – 50.0% and prevalence values for different CoV species ranging from 0.2 – 20.2%, warranting further investigation (Drexler *et al.*, 2010; Ge *et al.*, 2016).

With the Pan-CoV PCR assay's apparent bias towards the detection of alphaCoV in our setting, it was decided to undertake repeat screening of a subset of samples with the Extended RdRp PCR assay and its lineage-specific primers (Drexler *et al.*, 2010). Drexler's alpha- and SARS-related betaCoV specific primers were supplemented by MERS-related betaCoV primers designed in our laboratory as described in Chapter 3 (Drexler *et al.*, 2010). A random subset of 334 samples, of which 40 had been identified as positive using the Pan-CoV PCR assay, underwent repeat screening with the Extended RdRp PCR assay. In addition to identifying all 40 previously Pan-CoV PCR assay-positive samples, the Extended RdRp PCR assay also identified an additional 24 samples as CoV positive. Overall 64 positive samples were identified in the subset, resulting in a detection rate of 19.2%. These included ten coinfected samples, generating a total of 74 sequences from the repeat screening using the Extended RdRp PCR assay. Even more important than the overall increased detection rate was the increased detection of betaCoVs and sequences with more sequence variability, at least on the nucleotide level, than detected with the Pan-CoV PCR assay.

The increased detection rate is likely a result of using lineage-specific primers with increased sensitivity. Three studies reported using the Pan-CoV PCR for first round screening followed with repeat screening of negative samples using more specific primers based on sequences obtained

during initial screening, resulting in improved detection rates (Pfefferle *et al.,* 2009; Drexler *et al.,* 2010; Corman *et al.,* 2013). Although this improved detection rate is welcome for more accurate prevalence values, the improved detection is only possible for CoVs detected during the first round of screening for which more specific primers were designed and likely indicates that more diverse CoVs go undetected by this approach. The studies mentioned here provide evidence that the sensitivity of the Pan-CoV PCR assay may not match the sensitivity reported in the original publication in all settings (de Souza Luna *et al.,* 2007). This is further supported by a study where first round screening conducted with the Pan-CoV PCR assay only detected 4 of 146 positive samples later identified with more specific primers; estimated viral loads with a real-time PCR assay of all positive samples indicated a high mean viral load of 79 000 copies per µl (Corman *et al.,* 2014b). During the study described in this dissertation, a serial dilution of positive control RNA reflecting a sequence related to the alphaCoV *Miniopterus* BtCoV HKU8 indicated a detection limit of approximately 1100 copies per µl with the Pan-CoV PCR assay (Chapter 4 Section 4.7.1).

From an extensive literature search it appears that no study to date has investigated the detection limit of the Pan-CoV PCR, or any other screening PCR broadly targeting the CoV family for that matter, in different settings where bat samples are screened for the presence of CoVs. Furthermore, of fifteen published studies specifically making use of the Pan-CoV PCR assay to screen bat samples, not one has reported on the optimisation of this assay for use as a detection tool for CoVs in bats in their respective study settings (de Souza Luna *et al.,* 2007; Gloza-Rausch *et al.,* 2008; Pfefferle *et al.,* 2009; Drexler *et al.,* 2010, 2011; August *et al.,* 2012; Annan *et al.,* 2013; Ge *et al.,* 2016; Corman *et al.,* 2013, 2014a, 2015; Ge *et al.,* 2013; Maganga *et al.,* 2014b; Goffard *et al.,* 2015; Fischer *et al.,* 2016). Despite published studies reporting the detection of a number of different BtCoVs using the Pan-CoV PCR assay, with the limit of detection not reported in these published bat studies it is difficult to know if the use of this assay is optimal and whether results thus obtained accurately reflect the prevalence and diversity of CoVs in bats (Pfefferle *et al.,* 2009; Drexler *et al.,* 2010; Corman *et al.,* 2013; Ge *et al.,* 2016). This is further complicated due to numerous variations of this PCR assay, from two step RT-PCRs, different thermocycling parameters, to different enzyme reagents. Coupled with inter-laboratory variations it would be naïve to assume then that following a published PCR assay protocol with modifications will produce the same sensitivity as reported in all settings, especially when the published protocol was optimised for clinical human samples, not bat faecal material that likely contains an abundance of PCR inhibitors (de Souza Luna *et al.,* 2007).

Using a two-step RT-PCR approach with random-primed cDNA instead of a one-step RT-PCR approach may explain some of the reduced sensitivity observed during this study due to less target-specific cDNA being generated. Due to a need for sample cDNA to be screened for various other viruses as part of a larger research project in our laboratory, the one-step RT-PCR approach was not feasible. The influence of different sampling and analyses techniques on CoV RNA

detection has received very little attention in the literature with only one note on this topic found in a published supplemental information document (Osborne *et al.*, 2011). In this note, the authors reported discrepancies in detection rates being affected by multiple variables such as the RNA extraction kit used, or whether or not PCR inhibitors were removed (Osborne *et al.*, 2011). The authors even found a discrepancy in sensitivity dependent on which sample storage medium was used, with Viral Transport Medium (VTM) yielding the highest proportion of positive samples when compared to duplicate samples stored in RNA*later*™ Stabilisation Solution (Thermo Fisher Scientific) (Osborne *et al.*, 2011). This is particularly concerning as it is widely accepted that RNA*later*™ is an appropriate sample storage medium and is reported to have been used in most bat studies screening for viruses (Dominguez *et al.*, 2007; Pfefferle *et al.*, 2009; Drexler *et al.*, 2010; August *et al.*, 2012; Shirato *et al.*, 2012; Tsuda *et al.*, 2012; Annan *et al.*, 2013; Corman *et al.*, 2013, 2014a; Fischer *et al.*, 2016). Additionally, the type of sample material screened was shown to influence detection rates, with faecal pellets commonly reported to be more likely to produce positive results (Woo *et al.*, 2005b; Tong *et al.*, 2009; Osborne *et al.*, 2011; Wacharapluesadee *et al.*, 2015; Anthony *et al.*, 2017b). This finding is further supported by the low detection rates of studies using non-faecal material for CoV screening purposes (Mühldorfer *et al.*, 2011; Maganga *et al.*, 2014b). It was also found that screening with conserved consensus primers such as those used in the Pan-CoV PCR assay was less sensitive than sequence-specific primers, supporting the use of lineage-specific primers for increased surveillance sensitivity (Osborne *et al.*, 2011).

As highlighted above, reduced sensitivity may be due to a number of factors that should be strongly considered when undertaking surveillance studies. However, primer sequences likely play the most important role. From the literature there are at least eight sets of published primer sequences with accompanying PCR protocols that have been used to screen bat samples with the aim of broadly detecting members of the CoV family in more than one published study, indicated in Table 5.1 (Woo *et al.*, 2005b; Chu *et al.*, 2006; de Souza Luna *et al.*, 2007; Tong *et al.*, 2009; Quan *et al.*, 2010; Watanabe *et al.*, 2010; Falcón *et al.*, 2011). Except for the primer set published by Quan *et al.* (2010) that targets a more downstream region, these primer sets all target similar conserved regions of the RdRp with most overlapping with the target region of the Pan-CoV PCR assay.

Table 5.1 Published primer sets cited in more than one study for the purpose of broadly detecting members of the coronavirus family. The original publication in which primers were listed as well as other studies that cite using these same primers are detailed in the table below. The target region with bat coronavirus HKU10 (Genbank ID NC_018871) as a reference sequence is indicated along with the sequence length obtained for phylogenetic analysis.

First publication	Use of published primers cited in other studies	Sequence length excluding primers	Target region according to the full genome sequence of bat coronvirus HKU10 (NC_018871)
(Stephensen *et al.*, 1999)	(Tang *et al.*, 2006; Cui *et al.*, 2007; Carrington *et al.*, 2008)	970 bp	14027 – 15036
(Woo *et al.*, 2005b)	(Dominguez *et al.*, 2007; Lau *et al.*, 2007, 2010, 2012b, 2013; Osborne *et al.*, 2011; Shirato *et al.*, 2012; Lelli *et al.*, 2013; Razanajatovo *et al.*, 2015; Góes *et al.*, 2016)	393 bp	14423 – 14862
(de Souza Luna *et al.*, 2007)	(Gloza-Rausch *et al.*, 2008; Pfefferle *et al.*, 2009; Drexler *et al.*, 2010, 2011; August *et al.*, 2012; Annan *et al.*, 2013; Ithete *et al.*, 2013; Corman *et al.*, 2013, 2014a, 2015; Ge *et al.*, 2013; Maganga *et al.*, 2014b; Meyer *et al.*, 2014; De Benedictis *et al.*, 2014; Goffard *et al.*, 2015; Fischer *et al.*, 2016)	395 bp	14417 – 14871
(Quan *et al.*, 2010)	(Anthony *et al.*, 2013, 2017a; Memish *et al.*, 2013)	284 bp	17581 – 17907
(Watanabe *et al.*, 2010)	(Memish *et al.*, 2013; Wacharapluesadee *et al.*, 2015; Anthony *et al.*, 2017a)	393 bp	14423 – 14862
(Tong *et al.*, 2009)	(Wacharapluesadee *et al.*, 2013; Tao *et al.*, 2017)	152 bp	14420 – 14619
(Chu *et al.*, 2006)	(Shirato *et al.*, 2012; Suzuki *et al.*, 2014)	389 bp	14423 – 14857
(Falcón *et al.*, 2011)	(He *et al.*, 2014a)	493 bp	14420 – 14952

Although conserved at the amino acid level this target region appears to be more variable at the nucleotide level, likely affecting primer binding across diverse CoVs. Figure 5.1 depicts an alignment of sample sequence from the RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) database highlighting the regions targeted by the Pan-CoV PCR assay's primers at the amino acid and nucleotide level.
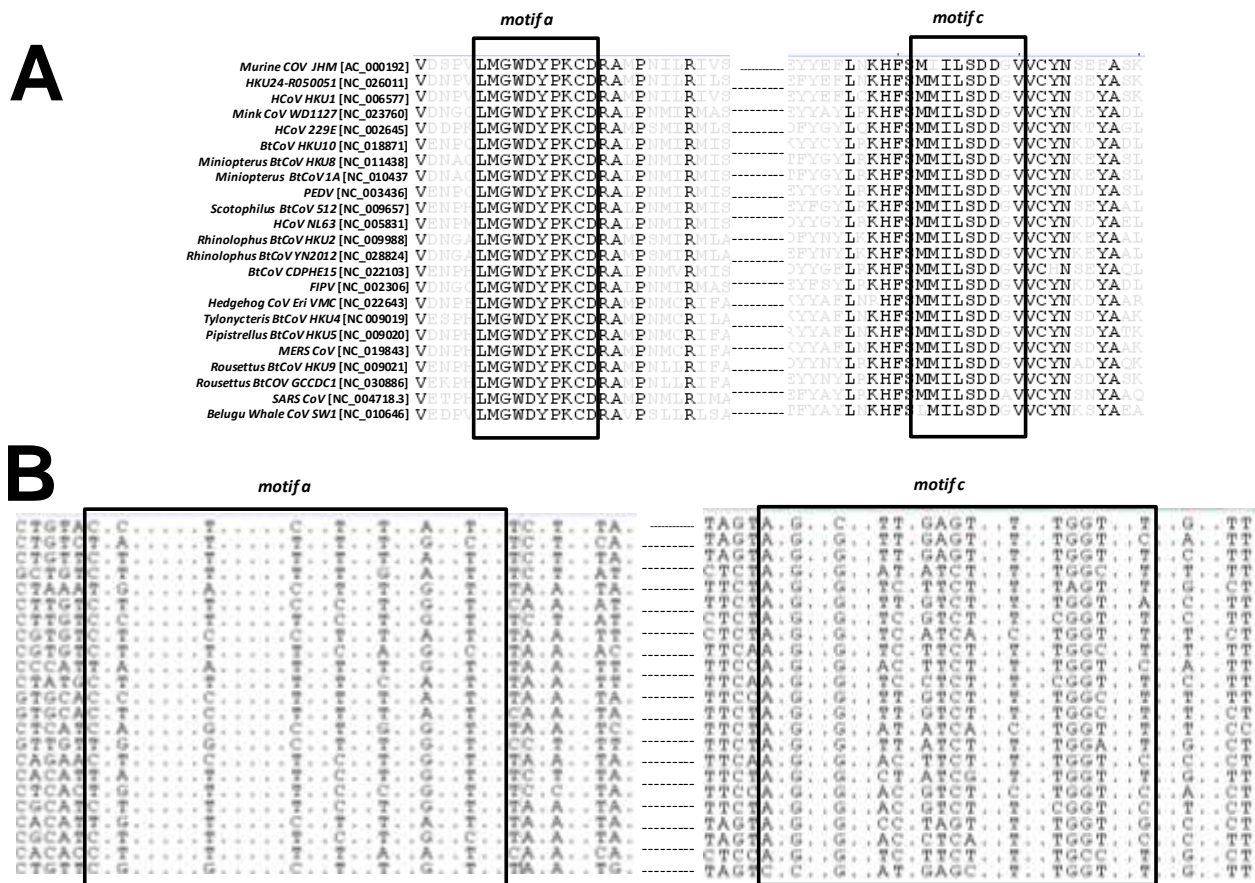
Figure 5.1 The regions of the coronavirus RNA dependent polymerase targeted by the Pan-CoV PCR assay's primers. The regions are highly conserved on the amino acid level (Figure A) but show more variability on the nucleotide level (Figure B). Coronavirus sequences in the alignment were obtained from the RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) database and the sequence alignment graphic was adapted from the sequence alignment created in Geneious R10 (Biomatters Inc., New Zealand).

Nucleotide variability at primer binding sites resulting in primer sequence mismatches reduces PCR amplification efficiency (Green *et al.*, 2015). The Pan-CoV PCR assay employs two different forward primers during each round of PCR meant to improve the broad range targeting of CoVs (de Souza Luna *et al.*, 2007). The use of multiple primers may increase the likelihood of primer heterodimer formation and may also weaken the robustness of the PCR reaction by competing for template which, coupled with a number of mismatches due to nucleotide diversity, may result in lower detection rates.

At high concentration of virus RNA, these mismatches are likely more tolerable in PCR reactions than at low concentrations. This might explain the detection of the sequence from sample 20150106CDK_NC1 during this study with the Pan-CoV PCR assay primers despite multiple mismatches with the forward primer sequences, including ones at the 3' end, while sample 20160906DC95_CP was not detected despite having minimal primer sequence mismatches as shown in Figure 5.2.
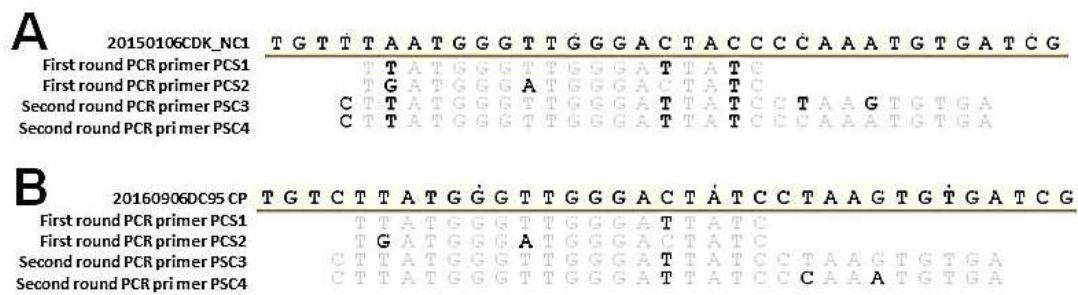
Figure 5.2 Primer binding mismatches in a sequence detected by the Pan-CoV PCR assay (Figure A) and primer mismatches in the primer binding region of a sequence not detected by the Pan-CoV PCR assay (Figure B). This diagram indicates the potential role that virus RNA concentrations in samples may play in the detection of CoV using broadly targeting primers. Mismatches may be more tolerable at high viral RNA concentrations than at lower concentrations. The sequence alignment graphic was adapted from the sequence alignment created in Geneious R10 (Biomatters Inc., New Zealand).

This is further supported by a CoV sequence from 20160411DC63_NN that was not detected by the Pan-CoV assay, despite being closely related to the sequence from 20150106CDK_NC1 that was detected by the Pan-CoV PCR assay. Both had a number of mismatches in the primer binding region. In an experiment not included in this thesis, this *N. nana*- derived CoV sequence identified using the Extended RdRp PCR assay was not detected at all by the Pan-CoV PCR assay across a 10x dilution series of quantified *in vitro* transcribed sample RNA starting with 1.6 million copies per µl. This further indicates that a single set of primer sequences, although targeting conserved amino acid regions, may not be optimal for the detection of all members of the CoV family. Primer binding against sequences obtained in this study could only be assessed for long sequences obtained with the Extended RdRp PCR assay against the Pan-CoV PCR forward primers due to the same reverse primer being used in both assays with resulting sequences containing matching primer sequences that do not reflect the true sequence variability in this region.

From published literature it is evident that there is little consensus regarding a standard approach to screening for CoV in bats, exemplified by the many different screening primers currently in use (Table 5.1) (Osborne *et al.*, 2011). With this a maturing field, it would be beneficial, for comparative purposes and for increased accuracy in reporting, for a standardised approach to be used. Towards the end of this study, *in vitro* transcribed positive control RNA from several different BtCoVs detected during this study was generated. Due to time constraints, assessing these positive controls with all published screening primers and protocols was beyond the scope of this study, but future investigations should prove insightful to formulate supported recommendations for the standardisation of current screening approaches in the growing BtCoV research field.

Despite not completing an extensive assessment of available screening PCR assays and primer sets, some suggestions and recommendations based on observations from this study can be made.

Optimisation of the Pan-CoV PCR assay for screening bat samples may result in increased sensitivity but with the observation that this assay's primers may be biased towards the detection of certain CoV sequences, care would need to be taken if optimising this assay with control material from a single BtCoV to not introduce detection biases towards a specific CoV. Ideally, optimisation to improve sensitivity should be conducted with a panel of diverse BtCoVs. However, this protocol still requires a nested PCR approach that is time consuming and only generates a short sequence requiring further amplification to extend the sequence for adequate phylogenetic analyses. Following a formal assessment, it is possible that other published screening PCR assays may prove more sensitive for the general detection of BtCoVs with some having the added benefit of not requiring a nested amplification protocol (Poon *et al.*, 2005; Watanabe *et al.*, 2010; Lau *et al.*, 2015). However, these assays still only generate a short sequence requiring extension to be of phylogenetic value.

Ultimately, based on these shortcomings, it is recommended that the Extended RdRp PCR assay with its lineage-specific primers be assessed for use as a screening PCR assay. Regular assessment and if necessary update of lineage-specific primers is recommended to maintain optimum detection of diverse BtCoVs as new BtCoV sequences are identified. This study is the first report where this assay was used as a screening PCR, rather than a confirmatory PCR to extend sample sequences for improved phylogenetic resolution. The use of this PCR assay as a screening PCR with three sets of lineage-specific primers during this study not only proved to increase the detection rate of positive samples with more diverse CoV sequences but it also generated a longer fragment for improved phylogenetic analysis after just a single non-nested PCR, essentially saving time and reagents. A recommendation as to a standardised CoV screening protocol however, can only be made following further investigations into the detection rates of different screening assays across a diverse panel of BtCoVs.

Based on observations made during this study it is suggested that for studies where only short sequences were generated using the Pan-CoV PCR assays, these sequences be extended using the Extended RdRp PCR assay for inclusion in more valuable phylogenetic analyses (Maganga *et al.*, 2014a; Goffard *et al.*, 2015; Wang *et al.*, 2017). Furthermore, where studies reported low detection rates and / or diversity, it may prove worthwhile to undertake repeat screening with multiple lineage-specific primer sets as demonstrated in this study. With only a subset of samples undergoing repeat screening with the Extended RdRp PCR in this study, it would be worthwhile to complete this in future investigations for an improved report of BtCoVs present in these samples.

140

From the findings in this study it can be recommended that investigators wanting to screen bat samples for CoVs thoroughly assess available screening assays. From the literature, as can be expected, collaborating research groups tend to use the same screening PCR assay protocols. With each protocol's continued use likely due to previous successes, it is important that protocols regularly be assessed to ensure that they still represent an optimal approach and that inter-laboratory variations or differences in research settings have not resulted in a reduced sensitivity that may otherwise go unnoticed (Osborne *et al.*, 2011).

Despite the initial use of a potentially suboptimal screening PCR assay during this study, the results obtained from its use together with those from the subset that underwent repeat screening with lineage-specific primers indicate the presence of a number of different BtCoV species in South African bats. With very few published studies on CoV surveillance in South African bats, these findings contribute significantly to the current description of BtCoV diversity in this region (Müller *et al.*, 2007; Geldenhuys *et al.*, 2013; Ithete, 2013; Ithete *et al.*, 2013; Corman *et al.*, 2014a).

5.2 General surveillance findings

With the use of the putative CoV species classification system established by Drexler *et al.* (2010), CoV sequences obtained during this study could be classified into nine different CoV species. With the exception of studies detailing the detection of a MERS-related betaCoV, NeoCoV, previous studies from South Africa have only reported on the detection of short CoV sequences, less than 400 bp in length, that did not meet the requirements for putative CoV species classification using the RGU criterion (Drexler *et al.*, 2010; Geldenhuys *et al.*, 2013; Ithete, 2013; Ithete *et al.*, 2013; Corman *et al.*, 2014a). With some CoV sequences obtained during this study showing high sequence identity to these short South African sequences, by the assumption that the high sequence identity extends beyond the available alignment, these South African sequences have been putatively classified based on the phylogenetic analyses of the longer sequences obtained during this study.

Based on this assumptive classification, CoV species detected previously in South African bats include the putatively novel unclassified *Neoromicia* BtCoV 1 and 2, *Miniopterus* BtCoV HKU8 and 1A, unclassified *Chaerephon* BtCoV KY22, and MERS-related CoV (Geldenhuys *et al.*, 2013; Ithete, 2013; Ithete *et al.*, 2013). This study additionally detected CoV sequences belonging to *Scotophilus* BtCoV 512, unclassified *Cardioderma* BtCoV KY43*,* and a putatively novel CoV species, unclassified *Miniopterus* BtCoV. The previously published CoV sequences from South African bats were obtained from four different bat species, namely *N. capensis*, *N. nana*, *M. natalensis*, and *M. condylurus*. During this study, CoV sequences were detected in an additional six bat species namely *C. pumilus, S. dinganii*, *R. capensis, R. clivosus*, *R. simulator*, and *P. hesperidus*. The findings from this study thus significantly contribute to the description of CoV

diversity and distribution in South Africa. Table 5.2 summarises the putative CoV species detected in previous South African studies based on phylogenetic analyses carried out in this study.

Table 5.2 Summary of bat CoV sequences with putative species classification detected in previous studies from South Africa.

Publication	Putative CoV species	Number detected	Bat species
Geldenhuys *et al.*	Unclassified *Neoromicia* BtCoV 1	1	*Neoromicia capensis*
	Miniopterus BtCoV 1A	1	*Miniopterus natalensis*
	Unclassified *Chaerephon* BtCoV KY22	1	*Mops condylurus*
Ithete 2013	Unclassified *Neoromicia* BtCoV 1	8	*Neoromicia capensis* *Neoromicia nana*
	Unclassified *Neoromicia* BtCoV 2	1	*Neoromicia capensis*
	Miniopterus BtCoV HKU8	2	*Miniopterus natalensis*
	Unclassified *Chaerephon* BtCoV KY22	1	*Mops condylurus*
	MERS-related CoV	1	*Neoromicia capensis*

To date, there are only 16 published sequences from South African bats. Except for one CoV sequence detected in a *N. capensis* bat that was shown phylogenetically to be a MERS-related betaCoV, these published sequences all belong to the alphaCoV genus. During this study, a total of 90 CoV sequences were obtained with 59 identified as alphaCoVs and 31 identified as betaCoVs. This greatly adds to the available CoV sequence dataset from South Africa for future investigations. With not all bat species sampled equally, it is likely that if these different bat species are sampled more extensively in future studies, additional BtCoVs will be identified as suggested by a recent study investigating global CoV diversity patterns (Anthony *et al.*, 2017b).

Points of interest regarding the BtCoVs detected during this study are highlighted in the following sections.

5.3 Alphacoronaviruses in South African bats

5.3.1 *Scotophilus* BtCoV 512-related coronaviruses

Based on phylogenetic analysis and the use of the RGU criteria, a sequence obtained from an *S. dinganii* bat captured at a wastewater treatment facility was shown to be closely related to *Scotophilus* BtCoV 512. This is the first report of a CoV sequence detected in South African *Scotophilus* sp. bats and from an assessment of sequences available on Genbank may represent the first report of a *Scotophilus* BtCoV 512 sequence from the African continent.

The RefSeq (www.ncbi.nlm.nih.gov/refseq/) sequence (Genbank ID NC_009657) for *Scotophilus* BtCoV 512 was obtained from an *S. kuhli* bat from China (Tang *et al.*, 2006). This is currently the only full genome sequence available for this CoV species. Genome characterisation and phylogenetic analysis have shown strong similarities with PEDV, with these viruses thought to likely share a common ancestor (Tang *et al.*, 2006; Huang *et al.*, 2013). Studies conducted in the

142

Philippines, Thailand, Taiwan, and Viet Nam have detected *Scotophilus* BtCoV 512-related sequences that showed high similarity to the sequence identified from study sample 20160411DC68_SD in *S. kuhli*, and *S. heathii* bats (Watanabe *et al.*, 2010; Wacharapluesadee *et al.*, 2015; Berto *et al.*, 2017). A recent study has also detected *Scotophilus* BtCoV 512-related sequences in *Rhinolophus monoceros*, *Kerivoula titania*, and *Miniopterus fuliginosus* bats from Taiwan indicating that although evidence of co-roosting bat species may have facilitated inter-species transmission in some instances, the host range of this CoV species may not be restricted to *Scotophilus* sp. bats (Chen *et al.*, 2016). The sequences from these studies are all ~400 bp in length hindering further phylogenetic analyses within this CoV species (Watanabe *et al.*, 2010; Wacharapluesadee *et al.*, 2015; Chen *et al.*, 2016; Berto *et al.*, 2017).

Despite the finding that other bat species might host *Scotophilus* BtCoV 512-related sequences, it appears from available studies that as the name suggests, *Scotophilus* BtCoV 512, is most commonly associated with *Scotophilus* sp. bats (Watanabe *et al.*, 2010; Wacharapluesadee *et al.*, 2015; Chen *et al.*, 2016; Berto *et al.*, 2017). This bat genus is distributed across Africa and southern Asia and with at least 15 different recognised species (www.iucnredlist.org); it is likely that with further investigations more variations of this CoV species will be detected.

From geographic distribution maps on the IUCN Red List website (www.iucnredlist.org), the distributions of *S. kuhli* and *S. heathii* overlap with each other but not with *S. dinganii* (Bates *et al.*, 2008a, 2008b; Griffin, 2008). This possibly indicates that either closely related CoVs are present in other bat species as indicated in the study by Chen et al (2016) with overlapping distributions to facilitate virus transmission, or that these *Scotophilus* BtCoV 512-related viruses perhaps co-evolved along with the speciation of the *Scotophilus* bat genus.

Further investigations into this CoV species are warranted considering the possibility that a cross-species transmission event of an ancestral CoV related to *Scotophilus* BtCoV 512 from bats to swine may be linked to the emergence of PEDV, a virus causing considerable economic damage in the USA and Asian swine industries (Tang *et al.*, 2006; Song and Park, 2012; Huang *et al.*, 2013; Lee, 2015). *Scotophilus* sp. bats are commonly reported to co-exist with humans and domestic animals; thus the potential exists for cross-species transmission events that could drive the emergence of a new zoonotic CoV in humans and / or domestic animals, such as swine (Bates *et al.*, 2008a, 2008b; Griffin, 2008). With little known about this bat CoV species, future investigations attempting full genome sequencing may prove insightful for understanding the evolution and ecology of these viruses.

5.3.2 *Miniopterus* BtCoV 1A and *Miniopterus* BtCoV HKU8-related coronaviruses

During this study alphaCoVs were detected in *M. natalensis* bats with phylogenetic analyses indicating these sequences to likely belong to the established *Alphacoronavirus* species, *Miniopterus* BtCoV 1A and HKU8 first identified in *Miniopterus* sp. bats from China (Chu *et al.*, 2006, 2008a). AlphaCoVs related to these two CoV species have been widely detected in *Miniopterus* sp. bats from countries such as China, Japan, Kenya, South Africa, Australia, and Bulgaria to name a few, indicating a wide geographic distribution associated with the geographic distribution of its main reservoir host, *Miniopterus* sp. bats; this strong association with *Miniopterus* sp. bats likely indicates the existence of an ancestral *Miniopterus* BtCoV that diversified over time as the *Miniopterus* family evolved and underwent speciation (Poon *et al.*, 2005; Chu *et al.*, 2006, 2008a; Shirato *et al.*, 2012; Tao *et al.*, 2012; Geldenhuys *et al.*, 2013; Ithete, 2013; He *et al.*, 2014a; Smith *et al.*, 2016). Together with the related CoV sequences previously published from South Africa, the *Miniopterus* sp.-derived CoV sequences detected in this study provides evidence of the circulation of *Miniopterus* BtCoV 1A and HKU8-related viruses in South African bats.

To date there is no evidence linking these *Miniopterus* sp.-derived CoVs to zoonotic spillover events in human or domestic animals signalling that these CoVs are likely harmless to the human population. However, it should be noted that *M. natalensis* bats have been reported to be coinfected with multiple CoVs that could facilitate future recombination events that might result in new bat CoVs with zoonotic potential (Chu *et al.*, 2008a). With their extensive geographic distribution and some species such as *M. schreibersii* being migratory in nature, these bats could facilitate rapid transmission of newly emerging viruses (Hutson *et al.*, 2008). Continued surveillance and monitoring of CoVs in *Miniopterus* sp. bats would therefore be beneficial to better understand how this CoV species has diversified over time across bat species and geographic regions.

5.3.3 Unclassified bat coronaviruses in South African bats

Two putative *Alphacoronavirus* species were detected in *N. capensis* and *N. nana* bats during this study and designated as unclassified *Neoromicia* BtCoV 1 and 2. Shorter sequences of these two putative Alphacoronavirus species were previously published in other studies from South Africa (Geldenhuys *et al.*, 2013; Ithete, 2013). The two putative CoV species show distinct phylogenetic clustering (Figures 4.5 and 4.6 Chapter 4). The sequences from unclassified *Neoromicia* BtCoV 1 showed closest relation to alphaCoV sequences from *Nyctalus* sp. bats from China and Bulgaria. *Nyctalus* sp. bats belong to the Nycteridae bat family while *Neoromicia* sp. bats belong to the Vespertilionidae family, likely explaining the phylogenetic indication that these sequences represent two different BtCoV species. The sequences from unclassified *Neoromicia* BtCoV 2

144

demonstrated two distinctly different phylogenetic clustering patterns based on whether the short or the longer partial RdRp sequence was analysed. The phylogenetic clustering based on the shorter sequence indicated these alphaCoVs to cluster with an alphaCoV sequence from a *Chaerephon pumilus* bat from Kenya near sequences related to HCoV-229E and HCoV-NL63. However, ML and Bayesian inference phylogenetic analyses of the longer sequences indicated these alphaCoVs to cluster outside the known *Alphacoronavirus* phylogeny. Further investigation of this CoV with attempted full genome sequences may prove beneficial for understanding the evolutionary history of these *Neoromicia*-derived CoV and how they fit into the *Alphacoronavirus* phylogeny. The unclassified *Neoromicia* BtCoV 1 sequences from *N. nana* bats demonstrated some phylogenetic divergence, likely reflecting coevolution of this alphaCoV with the evolutionary speciation of the *Neoromicia* bat genus.

Sequences detected in the putatively novel unclassified *Chaerephon* BtCoV KY22 species were detected in *C. pumilus* and *R. simulator* bats during this study with sequences previously detected in *M. condylurus* bats in South Africa showing strong relatedness to these sequences (Geldenhuys *et al.*, 2013; Ithete, 2013). The putative prototype sequence for this species was detected in a *C. pumilus* bat from Kenya (Tao *et al.*, 2012). That these sequences were detected in three different bat families may indicate that unlike the majority of other BtCoVs, this putative CoV species demonstrates low host species-specificity. This finding could indicate that the unclassified *Chaerephon* BtCoV KY22 either represents a new BtCoV that has not coevolved with a bat host species or that it has recently switched hosts.

Sequences obtained from *Rhinolophus* sp. bats during this study were shown by phylogenetic analyses and the RGU criteria to be related to the putative unclassified *Cardioderma* BtCoV KY43 species previously identified in *Cardioderma* sp. bats from Kenya.

One sequence obtained from a *M. natalensis* bat during this study was shown not to cluster with any other known BtCoV. Phylogenetic analyses and the RGU criteria indicate that this sequence may represent a novel BtCoV, further investigation is required for confirmation. This sequence was found to cluster between the two CoV species associated with *Miniopterus* sp. bats, *Miniopterus* BtCoV HKU8 and 1A and may therefore represent a new miniopterid BtCoV.

5.4 Betacoronaviruses in South African bats

During this study 31 betaCoV sequences were detected. Phylogenetic analyses and putative classification using the RGU criteria indicate that they all belong to the MERS-related CoV species (Drexler *et al.,* 2010). One other MERS-related CoV sequence has been published from South Africa and as previously mentioned, was termed NeoCoV based on its detection from an *N. capensis* bat (Ithete *et al.*, 2013; Corman *et al.*, 2014a).

Following the identification of MERS-CoV, a lineage C betaCoV, during an outbreak of respiratory disease in the Middle East, a number of studies reported the detection of closely related CoVs in bats. The virus found to be most closely related to MERS-CoV at the time of the outbreak was identified in a *Taphazous perforatus* bat trapped near the home of an infected patient but with only short sequences available, these findings were phylogenetically inconclusive as to whether or not MERS-CoV had evolved from bats (Li *et al.*, 2005; Reusken *et al.*, 2010; Zaki *et al.*, 2012; Anthony *et al.,* 2013; Memish *et al.*, 2013). Although the genome of MERS-CoV showed relatedness to *Tylonycteris* HKU4 and *Pipistrellus* HKU5 BtCoVs, the sequences differed significantly for it to be classified as a separate *Betacoronavirus* lineage C species (Woo *et al.,* 2007, 2009, 2012a; van Boheemen *et al.*, 2012; Adams *et al.*, 2016).

NeoCoV represents the genetically closest full genome lineage C betaCoV from a bat to MERS-CoV (Corman *et al.*, 2014a). However, due to significant differences in the S protein's subunit 1, NeoCoV does not represent the likely direct ancestor to this human pathogen, leaving the MERS-CoV progenitor still to be identified (Ithete *et al.*, 2013; Corman *et al.*, 2014a; Anthony *et al.*, 2017a). Despite not being the genetically closest relative to MERS-CoV, *Tylonycteris* BtCoV HKU4 has demonstrated an ability to utilise the same receptor as MERS-CoV further supporting bats as the ancestral host of MERS-CoV (Wang *et al.*, 2014). This finding indicates that likely there are other diverse MERS-related CoVs in other vespertilionid bats that perhaps have the same S1 subunit as observed in MERS-CoV. Alternatively, with the demonstration of MERS-related CoVs in hedgehogs, the MERS-CoV progenitor may be circulating in other small mammals not yet surveyed (Corman *et al.*, 2014b).

With the majority of reported lineage C betaCoVs to date detected in *Pipistrellus* sp. bats, there appears to be a strong association of these CoVs to vespertilionid bats (Woo *et al.*, 2007; Reusken *et al.*, 2010; Annan *et al.*, 2013; Anthony *et al.*, 2017a). The MERS-related CoV sequences identified during this study were all detected in *N. capensis* and *P. hesperidus* bats, further supporting this notion. The partial RdRp study sequences from *P. hesperidus* and *N. capensis* bats shared high sequence identities with NeoCoV. Phylogenetically, these sequences clustered into three groups within the MERS- related CoV species group. With the exception of one sequence, sequences from *P. hesperidus* were noted to cluster together with the recently reported MERS-related CoV sequence from a *P. hesperidus* bat in Uganda (Anthony *et al.*, 2017a). The remaining sequences from *N. capensis* bats and one *P. hesperidus* bat formed two separate sequence clusters, one containing only *N. capensis*-derived sequences from this study and one containing both *N. capensis* and a *P. hesperidus*-derived study sequences clustering with NeoCoV just outside the group of MERS-CoV sequences from humans and camels. This divergence of MERS-related CoVs hosted by closely related bat species suggests a high level of diversity amongst these viruses not previously reported.

The study sequence from a *P. hesperidus* bat that clustered with *N. capensis*-derived sequences does not fit the bat-species associated phylogenetic clustering observed for the other *P. hesperidus*-derived study sequences. It should be noted that the cyt *b* sequence for this sample, 20161014DC168_PH, could not be obtained despite repeated attempts. Sample material was limited and non-host DNA, likely from insects digested during feeding, was repeatedly amplified. With this sample collected in the middle of a series of *N. capensis* bats that have similar morphological traits making accurate in-field identification difficult, it is possible that this sequence was in fact obtained from an *N. capensis* bat. However, it also cannot be ruled out that this finding might indicate a cross-species transmission between co-roosting bat species as reported for other CoVs (Lau *et al.*, 2012a; Smith *et al.*, 2016).

The distribution of lineage C betaCoVs appears widespread with reports from bats in China, Mexico, Thailand, Netherlands, Romania, Ukraine, Italy, Ghana, Uganda, and in hedgehogs from Germany (Woo *et al.*, 2006, 2007; Reusken *et al.*, 2010; Annan *et al.*, 2013; Anthony *et al.*, 2013, 2017a; Wacharapluesadee *et al.*, 2013; Corman *et al.*, 2014b; De Benedictis *et al.*, 2014; Yang *et al.*, 2014). Together with the strong association with vespertilionid bats that have diverged into more than 400 different species, the diversity and distribution of these CoVs is likely much greater than previously reported.

With the many MERS-related CoV sequences detected during this study, it is likely that these viruses are circulating in South African bats with considerable prevalence rates. Along with the many non-identical MERS-related CoV sequences generated from this study that add significant value to furthering the understanding of the evolution of the MERS-related CoV species, continued surveillance of vespertilionid bats in this region is warranted as much remains unanswered about the evolutionary history of MERS-CoV and how it emerged. With camels identified as the source of MERS-CoV spillover to human populations, it is not yet known whether or not a recombination event in bats led to a spillover into camels or if perhaps camels, or another intermediate host, were coinfected with different CoVs leading to recombination and emergence of MERS-CoV (Reeves *et al.,* 2015; de Wit *et al.*, 2016; Funk *et al.*, 2016).

Full genome amplification of one *P. hesperidus*-derived and two *N. capensis*-derived study MERS-related CoV sequences was attempted to further investigate the diversity of these CoV. Although full genome amplification was not achieved, a number of full and partial gene sequences were obtained allowing additional phylogenetic analyses. These analyses showed these bat-derived MERS-related CoV sequences to consistently cluster with NeoCoV and the *P. hesperidus*-derived PDF 2180 sequence from Uganda (Genbank ID KX574227). For most regions phylogenetic analysis indicated an apparent genetic divergence between sequences from the two different bat species hosts; this was most noticeable for the M protein. Sample study sequences showed similar phylogenetic clustering between the S subunit 1 and 2 as NeoCoV indicating that these study

147

samples similarly do not represent the MERS-CoV progenitor and likely do not pose a direct threat to humans as demonstrated by the inability of the closely related *P. hesperidus*-derived betaCoV to infect human cell lines (Corman *et al.*, 2014a; Anthony *et al.*, 2017a).

Additionally, not only was genetic divergence observed at the host species level but also at the geographical level where for certain alignments, ORF1 and N protein, South African-derived sequences were seen to be more closely related to each other despite being obtained from different bat host species. These findings indicate the potential for a much greater diversity of MERS-related CoVs. With genetic divergences observed at the bat host species level, the potential for recombination events between co-roosting bats could result in recombination events leading to the emergence of new BtCoVs.

5.5 Coronavirus diversity and ecology in the *Neoromicia capensis* bat

To gain insight into the diversity and ecology of BtCoVs, species-specific surveillance was conducted on the bat species *N. capensis*. Overall 61 CoV positive *N. capensis* individuals were identified with ten further identified as being coinfected with two different BtCoVs.

5.5.1 Coronavirus diversity at the species level

CoVs belonging to three different CoV species were detected in *N. capensis* bats, namely unclassified *Neoromicia BtCoV 1* and *2*, and *MERS-related CoV*. Overall, 43, 2, and 26 unclassified *Neoromicia* BtCoV 1 and 2, and MERS-related CoV sequences were identified respectively. The number of *Neoromicia* BtCoV *2* detected appears considerably lower than the other CoVs detected in this bat species and may be reflective of suboptimal primers for the sensitive detection of this BtCoV.

A recent publication by Anthony et al (2017) estimated that each bat species likely hosts 2.7 different CoV species that have the highest probability of being detected for sample sizes with more than 110 bats per bat species. Here the screening of 196 *N. capensis* bats identified 3 different CoV species.

Although the unclassified *Neoromicia* BtCoV 1 sequences were well conserved on the amino acid level, greater variation was observed on the nucleotide level. Woo *et al.* (2006) reported that nucleotide differences could be observed when comparing sequences of the same CoV collected from bats at different locations implying that closely related CoV are more likely to be found among bats at the same location (Woo *et al.*, 2006). This is to be expected as CoV are continuously evolving and as bats move and settle in different areas these viruses will start to diverge from each other resulting in geographic clustering of CoV sequences. This can be seen by the apparent

clustering of unclassified *Neoromicia* BtCoV 1 sequences by bat trapping sites in similar geographic regions.

The alpha- and betaCoV sequence obtained from a bat captured at the CGR bat trapping site was found to consistently cluster with sequences from a different geographic location. This may indicate that this individual bat has not been associated with the colony at the CGC bat trapping site, on the same property as the CGR trapping site, for an extended period of time and likely represents a sequence from a bat that has recently moved from a site with CoV sequences more closely associated with those from the HWK bat trapping site in the Overberg.

As demonstrated by this finding, the sequence variations observed may prove useful in future studies that attempt to monitor potential inter-colony movements across large enough distances.

5.5.2 Coronavirus diversity at the individual bat level

Coinfection with unclassified *Neoromicia* BtCoV 1 and MERS-related CoV was identified in ten *N. capensis* bats. This coinfection of two different CoV species from two different genera has been reported in the literature previously but represents the first account of coinfection in bats from South Africa (Tang *et al.,* 2006; Reusken *et al.,* 2010).

The sample, 20140923LNR_NC7 that was shown to be carrying two unclassified *Neoromicia* BtCoV 1 strains present in bats from the same colony provided some additional insight into the diversity of CoV sequences in individual bats. As similarly noted in the literature, clonal sequencing data showed the existence of multiple strains, a quasispecies, of the same bat alphaCoV in this individual (Wacharapluesadee *et al.,* 2015). With this individual bat, 20140923LNR_NC7, harbouring both dominant strains present in other bats from the same colony, intra-species transmission within this colony is probable. Furthermore, comparison of sequences obtained from two different sampling events at this bat trapping site demonstrated the evolutionary change over time within this short partial fragment of the conserved RdRp.

As with bat trapping site LNR, sequences obtained from some other sites also indicated the presence of more than one variant. Comparison of sequences obtained from bat trapping sites that were sampled on different occasions showed evidence of evolutionary change over time at some sites but not at others, suggesting that factors driving evolutionary changes may differ between different bat colonies.

5.5.3 Predictors of CoV infection in *Neoromicia capensis* bats

Understanding what influences the ecology of viruses in their reservoir hosts is important for the development of preventative approaches and improving preparedness against future zoonotic epidemics. Although the BtCoVs detected during this study likely pose no immediate zoonotic

149

threat, insight gained into the ecology of these viruses increases the current knowledge and provides base line data for future studies. Binomial logistic regression was used to identify possible intrinsic and extrinsic predictors of unclassified *Neoromicia* BtCoV 1 (NeoBtCoV1) infection in *N. capensis* bats (Chapter 4). The best model, based on AIC analysis, identified the intrinsic variable, sex, and the extrinsic variable, Forest biome, as significant predictors of NeoBtCoV1 infection in *N. capensis* bats.

The sex variable was significantly negatively associated with male bats indicating that female bats were more likely to be infected with NeoBtCoV1. Overall, 35% of sampled female *N. capensis* bats were positive compared to 19% of sampled male *N. capensis* bats.

There is mixed evidence for sex as a predictor of CoV infection in bats. Reproductive status and age are commonly reported as significant predictors of CoV infection in bats (Gloza-Rausch *et al.*, 2008; Drexler *et al.*, 2011; August *et al.*, 2012; Annan *et al.*, 2013; Anthony *et al.*, 2017b). For example, studies that detected CoVs in *Nycteris, Myotis* and *Pipistrellus* bats found that both juveniles and lactating females were more likely to be CoV infected than adult or male bats (Gloza-Rausch *et al.*, 2008; Annan *et al.*, 2013). In females, the reproductive cycle and ecological trade-offs during reproduction result in weaker immune defences (Luis *et al.*, 2013). Juvenile bats frequently have a strong association with CoV infection (Gloza-Rausch *et al.*, 2008; August *et al.*, 2012; Annan *et al.*, 2013; Anthony *et al.*, 2017b) because juvenile bats comprise naïve populations that are susceptible to new infections mediated by vertical transmission between mothers and pups, as well as horizontal transmission between pups within the colony following the waning of maternal immunity (Drexler *et al.*, 2011; Luis *et al.*, 2013). The influence of birth pulses on virus shedding has been documented for other viruses, such as the lyssa- and filoviruses (George *et al.*, 2011; Amman *et al.*, 2012; Hayman, 2015; Plowright *et al.*, 2016). By contrast, other studies have reported neither sex, nor age, nor reproductive state to be significant predictors of CoV infection in bats (Reusken *et al.*, 2010; Seltmann *et al.*, 2017). A recent study reported that male bats had a stronger association with CoV infection than female bats (Anthony *et al.*, 2017b). Male bats may be more susceptible to CoV infection and infections from ectoparasites, bacteria, and other viruses, because sex steroids modulating host immunity have a greater negative trade-off on male bats than females (Klein, 2000).

In this study, reproductive status was not a significant predictor of infection and was not present in the top five logistic models. Further, only eight juvenile bats were sampled; hence age was not included as a factor in the logistic models. Nonetheless, five juveniles (62.5%) were positive, lending some support for the hypothesis that juveniles are more susceptible to CoV infection than adults.

The Forest biome variable was significantly positively associated with NeoBtCoV1 infection. Although the Forest biome might present some unique characters driving NeoBtCoV1 prevalence

in *N. capensis* bats, there is no obvious indication as to what these might be. *N. capensis* bats are widely distributed across most biomes of South Africa (Monadjem *et al.*, 2010). There is little evidence that their roosting behaviour and food sources differ significantly across the different biomes (Monadjem *et al.*, 2010; Schoeman and Jacobs, 2011). The Forest biome represents the smallest of the South African biomes and has a long history of human use with logging activities leading to habitat fragmentation and loss (Geldehuys and Mucina, 2006; Rutherford *et al.*, 2006)

A recent study investigating the effect of human activities on habitat: specifically fragmented, recovering, and actively logged forest regions, found no association between CoV infection in bats and habitat type (Seltmann *et al.*, 2017). A second recent study found the detection of CoV in bats to be positively associated with habitat regions with an animal-use interface associated with hunting, wild animal farms, markets, restaurants, animal trade, and handling by veterinarians or researchers (Anthony *et al.*, 2017b). Although the small sample size with a large proportion of CoV positive samples for this biome may account for some skewing of the results, the small sample size with all negative results did not render the Nama-Karoo biome a significant negative predictor of NeoBtCoV1 infection in *N. capensis* bats. Further investigations in different biomes will be required to elucidate this finding.

Altitude was negatively associated with NeoBtCoV1 positivity i.e. as the altitude increased bats were less likely to be CoV positive. A similar relationship between influenza infection and altitude was found in avian hosts (Magee *et al.*, 2014; Shekaili *et al.*, 2015). This association likely reflects habitat preferences of the host. As altitudes increase, environmental conditions become harsher and less favourable for habitation by bats leading to sparse host distribution that may be unfavourable for the maintenance of viruses (Turmelle and Olival, 2009).

The top five models identified by AIC had similar AICc values suggesting that other intrinsic – particularly body condition - and extrinsic variables may also influence the likelihood of CoV infection. Few studies reported significant negative relationships between body condition and CoV infection (Lau *et al.*, 2010, 2012a). However, a recent study found that low body weight was associated with astrovirus infection but not CoV infection (Seltmann *et al.*, 2017).

There are a number of caveats regarding these results. Due to the small number of juveniles present in the dataset, age was excluded as a variable for inclusion in the analyses. As previously discussed, age has been shown to be an important variable associated with CoV infection in bats (Gloza-Rausch *et al.*, 2008; August *et al.*, 2012; Annan *et al.*, 2013; Anthony *et al.*, 2017b). This variable should be investigated in future studies with an improved dataset. CoV infection dynamics may differ between different bat host species due to different population structures and might even differ between different bat CoVs. For example, infection with BtCoV HKU10 was associated with low body weight in *Hipposideros pomona* but not in *Rousettus leschenaulti*, and SARSr-Rh-BatCoV was associated with low body weight in *Rhinolophus sinicus*, but Rh-BatCoV HKU2 was

not (Lau *et al.,* 2010, 2012a). Moreover, differences in findings could be related to the underreporting of infected individuals due to the screening protocol employed; there were discrepancies in detections rates between different CoV screening PCR assays during this study. Logistic regression analyses were limited to the dataset representing the Pan-CoV PCR screening results for NeoBtCoV1 because of the small sample size for the Extended RdRp PCR. The analyses were further limited by a dataset that contained considerably more negative values than positive values, only 43 of 196 *N. capensis* bats were NeoBtCoV1 positive. Despite these limitations, there was strong statistical support for the accuracy and goodness of fit of the best model providing evidence for the influence of both intrinsic and extrinsic factors on the prevalence of NeoBtCoV1 in *N. capensis* bats.

The lack of studies on predicting variables of CoV infection indicates this to be a key area for future investigation. Future investigations with larger datasets using more sensitive screening approaches such as the Extended RdRp PCR assay with lineage-specific primers may elucidate the preliminary findings from this study. Of interest, would be to assess whether the influence of intrinsic and extrinsic factors as predictors of CoV infection among different bat host species are mediated by bat species' ecology, population structures, and / or roosting and habitat preferences. Additionally, longitudinal colony studies investigating CoV infection dynamics will assist in understanding how extrinsic and intrinsic factors might drive CoV shedding patterns.

5.6 The importance of host species confirmation by molecular means

As indicated in Chapter 4, for a number of samples the in-field host species identification based on morphological traits was not in agreement with the species as identified by molecular methods. Although experienced zoologists are probably unlikely to incorrectly identify different bat species in the field, this finding indicates the value of confirming species identity by molecular methods, particularly for bat species with similar morphological characteristics. From the literature species confirmation by molecular methods appears not to be standard practice, or at least not the reporting thereof, with a number of studies not reporting the confirmation of bat species by molecular methods, indicating the potential for incorrect conclusions (Li *et al.,* 2005; Chu *et al.,* 2006; Gloza-Rausch *et al.,* 2008; Reusken *et al.,* 2010; Tsuda *et al.,* 2012).

N. capensis bats were most commonly misidentified as *P. hesperidus* likely due to their high morphological similarity. In fact these two species were up until recently thought to belong to the same species (Kearney *et al.,* 2002; Monadjem *et al.,* 2010). An inexperienced fieldworker could therefore quite easily fail to make the distinction between the two species. This is the likely reason for the consistent misidentification of *N. capensis* bats during a fieldtrip across sites in the Eastern Cape. With some sites sampled repeatedly, the potential exists for identification to become

nonchalant with species identifications of species with highly similar morphology traits to be made based on previous site visits. It is also possible that not all species were truly misidentified in the field but rather that errors occurred during data entry. Species identification by morphological traits and molecular methods are important tools that should be used to improve the accuracy of reported data.

5.7 The unsuccessful longitudinal investigation of coronavirus diversity and ecology within a bat colony

As indicated in Chapter 4, despite the identification of a bat colony found to have CoV infected bats, subsequent testing of monthly samples failed to generate positive results. A number of possible reasons for the failure of this study objective are described here.

The longitudinal sampling of this bat colony involved the passive collection of pooled faecal pellets from underneath the bat boxes at regular intervals in a similar manner as that described in other studies (Drexler *et al.*, 2011). Training was given to the farm owner with instructions to layout plastic sheeting below the bat boxes late evening and to collect pools of bat faecal pellets first thing the next morning. Since the collection of pooled samples was not conducted by our research group it cannot be ruled out that the collection protocol was not followed correctly. The bat boxes were located on the side of a building exposed to dry, arid conditions. If faecal pellets were collected midmorning, exposure to high temperatures, low humidity, and the sun's UV rays may have destroyed the viral RNA (Duan *et al.*, 2003; Geller *et al.*, 2012).

A second point to consider is the changing seasonal patterns observed in the region. Higher temperatures coupled with recent drought conditions resulted in a marked drop in bat numbers in the area likely linked to a decrease in insect activity related to weather conditions (personal communication with farm owner Mr Laubscher). Although little information is available on how weather patterns or climate changes affect bats there is some evidence that drought episodes can result in reduced population numbers due to insufficient food supplies to support reproduction (Amorim *et al.*, 2015). During the first year, the bats temporarily migrated during winter, returning end of August when insect activity increased. With worsening drought conditions, it was noted that when the bats moved away during the second year's winter period that the returning population was considerably smaller. It is therefore possible that with reduced numbers, the CoV infection previously detected has not been maintained or has shifted with the migrated bats. This however does not explain why no CoV infection was detected during the first few months immediately after identifying CoV positive individual in the colony.

A third consideration is the effect of pooling samples. The *N. capensis* colony initially identified was large in number. With little known about the ecology of CoV infection within a colony it is difficult to know if sufficient pools of an adequate number of faecal pellets were collected. Our approach was

153

to collect ten pooled samples containing of five pellets each. However, with hundreds of pellets potentially available for collection, it is possible that an insufficient number of positive pellets were collected for CoV detection in a pooled sample.

Furthermore, the number of faecal pellets was chosen at random based on previous publications (Drexler *et al.*, 2011). A concern was that perhaps CoV positive faecal pellets with low viral loads were going undetected in pooled samples. For a number of pooled samples, pools were teased apart and individual pellets screened; no positive results were obtained using the conventional PCR assay and additionally a real-time PCR assay with higher sensitivity also failed to detect any CoV positive results.

With the failure to detect CoV in longitudinally collected pooled faecal samples, this aspect of the study had to be halted. Progress has been made with the installation of bat boxes at a number of locations nearby which in future may be used for longitudinal bat colony studies.

5.8 Concluding remarks

This study aimed to describe the existing CoV diversity within South African bat populations as well as factors that might influence bat-CoV ecology and diversity. Four hypotheses were assessed and found to hold true.

It was demonstrated that different BtCoVs can be found in different bat species of Southern Africa. Overall nine different CoV species were detected from nine different bat species across different regions of South Africa. This study detected additional support for the circulation of MERS-related CoVs in South African bats. Divergent variants were detected in two different vespertilionid species, *N. capensis* and *P. hesperidus*, further supporting the notion that vespertilionid bats likely represent the ancestral reservoir host for these viruses.

A species-specific surveillance of *N. capensis* bats revealed CoV diversity at both the species and individual bat level. At the species level, three different CoV species were detected that demonstrated genetic variability across different geographic regions. Within an individual bat, the presence of multiple strains of the same BtCoV was demonstrated. Furthermore, coinfection with two different CoV species was identified in ten *N. capensis* bats.

Host and environmental factors were found to influence CoV ecology with sex, altitude and biome found to be possible predictors for the presence of NeoBtCoV1 in *N. capensis* bats. Based on binomial regression results, female *N. capensis* bats and bats trapped low altitudes within the Forest biome had the highest likelihood of being NeoBtCoV1 positive.

In addition to confirming the study hypotheses, this study generated longer sequences than previously reported from South Africa that allowed for the putative classification of detected

BtCoVs at the species level. Based on phylogenetic analyses, assumptive classification of previously identified BtCoVs from South Africa, for which only short sequences were available, was also possible.

With discrepancies observed during this study between different CoV PCR assays and an extensive literature search showing little consensus on how to conduct CoV surveillance of bat populations, this study has highlighted the need for a standardised approach that not only improves the detection of diverse CoVs but generates sufficient sequence lengths for adequate phylogenetic resolution. For these purposes, this study suggests the use of the Extended RdRp PCR assay with a panel of lineage-specific primers to be used directly as a screening PCR assay; however further assessment and validation are required before a formal recommendation can be made.

South Africa currently has a high level of biodiversity but with continual population expansion, accompanied by habitat loss, scenarios observed in other regions of the world where encroachment onto wildlife populations coupled with a loss of biodiversity led to the emergence of devastating disease are not to be unexpected here. It is therefore important to continue the monitoring and surveillance of our bat populations to detect potentially novel EIDs and to better understand the ecology of host-parasite relationship between bats and their associated viruses.

References

Abascal, F., Zardoya, R. and Posada, D. (**2005**) 'ProtTest: selection of best-fit models of protein evolution', *Bioinformatics*, 21(9), pp. 2104–2105.

Abdul-Rasool, S. and Fielding, B. C. (**2010**) 'Understanding human coronavirus HCoV-NL63.', *Open Virol J.* 4(1), pp. 76–84.

Adams, M. J. and Carstens, E. B. (**2012**) 'Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2012).', *Arch Virol.*, 157(7), pp. 1411–22.

Adams, M. J., Lefkowitz, E. J., King, A. M. Q., Harrach, B., Harrison, R. L., Knowles, N. J., *et al.* (**2016**) 'Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2016).', *Arch Virol.,* 161(10), pp. 2921–49.

Addgene (**no date**) *Creating bacterial glycerol stocks for long-term storage of plasmids*, *Basic Molecular Biology Protocols*. Available at: https://www.addgene.org/protocols/create-glycerol-stock/ (Accessed: 1 April 2014).

Aguiire, A., Tabor, G. M. and Ostfeld, R. S. (**2012**) 'Conservation Medicine: Ontogeny of an emerging discipline', in Aguirre, A. A., Ostfeld, R. S., and Daszak, P. (eds) *New Directions in Conservation Medicine: Applied Cases of Ecological Health.* New York: Oxford University Press, pp. 3–16.

Alexander, D. J. (2006) 'Summary of avian influenza activity in Europe, Asia, Africa, and Australasia, 2002–2006.', *Avian Dis.*, 51(s1), pp. 161-166.

Alexander, K. A. and Appel, M. J. G. (**1994**) 'African wild dogs (*Lycaon pictus*) endangered by a canine distemper epizootic among domestic dogs near the Masai Mara National Reserve, Kenya.', *J Wildl Dis.*, 30(4), pp. 481–485.

Almeida, J. D. and Tyrrell, D. A. J. (**1967**) 'The morphology of three previously uncharacterized human respiratory viruses that grow in organ culture.', *J Gen Virol.*, 1(2), pp. 175–178.

Altizer, S., Bartel, R. and Han, B. A. (**2011**) 'Animal migration and infectious disease risk.', *Science*, 331(6015), pp. 296–302.

Altizer, S., Harvell, D. and Friedle, E. (**2003**) 'Rapid evolutionary dynamics and disease threats to biodiversity.', *Trends Ecol Evol.*, 18(11), pp. 589–596.

Amman, B. R., Carroll, S. A., Reed, Z. D., Sealy, T. K., Balinandi, S., Swanepoel, R., *et al.* (**2012**) 'Seasonal pulses of Marburg virus circulation in juvenile Rousettus aegyptiacus bats coincide with periods of increased risk of human infection', *PLoS Pathog.*, 8(10), p. e1002877.

Amorim, F., Mata, V. A., Beja, P. and Rebelo, H. (**2015**) 'Effects of a drought episode on the reproductive success of European free-tailed bats (Tadarida teniotis).', *Mamm Biol.*, 80(3), pp. 228–236.

Annan, A., Baldwin, H. J., Corman, V. M., Klose, S. M., Owusu, M., Nkrumah, E. E., *et al.* (**2013**) 'Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe.', *Emerg Infect Dis.*, 19(3), pp. 456–9.

Anon (**1994**) 'Addressing emerging infectious disease threats: a prevention strategy for the United States. Executive summary.', *MMWR Recomm Rep.*, 43(RR-5), pp. 1–18.

Anon (**2017**) *Understanding evolution*, *University of California Museum of Paleontology*. Available at: http://evolution.berkeley.edu/ (Accessed: 1 April 2017).

Anthony, S. J., Gilardi, K., Menachery, V. D., Goldstein, T., Ssebide, B., Mbabazi, R., *et al.* (2017a) 'Further evidence for bats as the evolutionary source of Middle East respiratory syndrome coronavirus.', *MBio.*, 8(2), pp. e00373-17.

Anthony, S. J., Johnson, C. K., Greig, D. J., Kramer, S., Che, X., Wells, H., *et al.* (2017b) 'Global patterns in coronavirus diversity.', *Virus Evol.*, 3(1), pp. 1814–20.

Anthony, S. J., Ojeda-Flores, R., Rico-Chávez, O., Navarrete-Macias, I., Zambrana-Torrelio, C. M., Rostal, M. K., *et al.* (2013) 'Coronaviruses in bats from Mexico.', *J Gen Virol.*, 94(Pt 5), pp. 1028–38.

Arai, S., Nguyen, S. T., Boldgiv, B., Fukui, D., Araki, K., Dang, C. N., *et al.* (2013) 'Novel bat-borne hantavirus, Vietnam', *Emerg Infect Dis.*, 19(7), pp. 1159–61.

Asano, K. M., Hora, A. S., Scheffer, K. C., Fahl, W. O., Iamamoto, K., Mori, E., *et al.* (2016) 'Alphacoronavirus in urban Molossidae and Phyllostomidae bats, Brazil.', *Virol J.*, 13(1), p. 110.

Ashford, R. W. (1997) 'What it takes to be a reservoir host.', *Belg J Zool.*, 127(Suppl1), pp. 85–90.

Ashford, R. W. (2003) 'When is a reservoir not a reservoir?', *Emerg Infect Dis.*, 9(11), pp. 1495–6.

August, T. A., Mathews, F. and Nunn, M. A. (2012) 'Alphacoronavirus detected in bats in the United Kingdom.', *Vector Borne Zoonotic Dis.*, 12(6), pp. 530–3.

Azhar, E. I., El-Kafrawy, S. A., Farraj, S. A., Hassan, A. M., Al-Saeed, M. S., Hashem, A. M., *et al.* (2014) 'Evidence for camel-to-human transmission of MERS coronavirus.', *N Engl J Med.*, 370(26), p. 2499-505.

Badrane, H. and Tordo, N. (2001) 'Host switching in lyssavirus history from the Chiroptera to the Carnivora orders.', *J Virol.*, 75(17), pp. 8096–104.

Baker, D. G. (1998) 'Natural pathogens of laboratory mice, rats, and rabbits and their effects on research.', *Clin Microbiol Rev.*, 11(2), pp. 231–66.

Baker, K. S., Todd, S., Marsh, G. A., Crameri, G., Barr, J., Kamins, A. O., *et al.* (2013) 'Novel, potentially zoonotic paramyxoviruses from the African straw-colored fruit bat Eidolon helvum.', *J Virol.*, 87(3), pp. 1348–58.

Baker, S. C. (2008) 'Coronaviruses: Molecular biology.', in Mahy, B. W. J. and van Regenmortel, M. H. V. (eds) *Encyclopedia of Virology*, pp. 554–562.

Barr, J., Smith, C., Smith, I., De Jong, C., Todd, S., Melville, D., *et al.* (2015) 'Isolation of multiple novel paramyxoviruses from pteropid bat urine.', *J Gen Virol.*, 96(Pt1), pp. 24-9.

Bastos, A. D., Nair, D., Taylor, P. J., Brettschneider, H., Kirsten, F., Mostert, E., *et al.* (2011) 'Genetic monitoring detects an overlooked cryptic species and reveals the diversity and distribution of three invasive Rattus congeners in South Africa.', *BMC Genet.*, 12(1), p. 26.

Bates, P., Csorba, G., Molur, S. and Srinivasulu, C. (2008a) *Scotophilus heathii*, *The IUCN Red List of Threatened Species*. Available at: http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T20067A9142155.en. (Accessed 20 July 2017).

Bates, P., Kingston, T., Francis, C., Rosell-Ambal, G., Heaney, L., Gonzales, J.-C., *et al.* (2008b) *Scotophilus kuhli*, *The IUCN Red List of Threatened Species*. Available at: http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T20068A9142479.en. (Accessed 20 July 2017).

Beach, J. R. and Schalm, O. W. (1936) 'A filterable virus, distinct from that of laryngotracheitis, the cause of a respiratory disease of chicks.', *Poult Sci.*, 15(3), pp. 199–206.

Bean, W. J., Kawaoka, Y., Wood, J. M., Pearson, J. E. and Webster, R. G. (**1985**) 'Characterization of virulent and avirulent A/chicken/Pennsylvania/83 influenza A viruses: potential role of defective interfering RNAs in nature.', *J Virol.*, 54(1), pp. 151–60.

Beaudette, F. and Hudson, C. (**1937**) 'Cultivation of the virus of infectious ., *J Am Vet Med Assoc.*, 90, pp. 51–58.

Beckert, B. and Masquida, B. (**2011**) 'Synthesis of RNA by in vitro transcription.', in Nielsen, H. (ed.) *RNA: Methods and Protocols.* Springer Science, pp. 29–41.

Beerenwinkel, N., Günthard, H. F., Roth, V. and Metzner, K. J. (**2012**) 'Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data.', *Front Microbiol.*vol 3, p. 329.

Behjati, S. and Tarpey, P. S. (**2013**) 'What is next generation sequencing?', *Arch Dis Child Educ Pract Ed.*, 98(6), pp. 236–8.

Belouzard, S., Millet, J. K., Licitra, B. N. and Whittaker, G. R. (**2012**) 'Mechanisms of coronavirus cell entry mediated by the viral spike protein.', *Viruses*, 4(6), pp. 1011–1033.

De Benedictis, P., Marciano, S., Scaravelli, D., Priori, P., Zecchin, B., Capua, I., *et al.* (**2014**) 'Alpha and lineage C betacoronavirus infections in Italian bats.', *Virus Genes.*, 48(2), pp. 366–71.

Bengis, R. G., Kock, R. A. and Fischer, J. (**2002**) 'Infectious animal diseases: the wildlife/livestock interface.', *Rev Sci Tech.*, 21(1), pp. 53–65.

Berger, A., Drosten, C., Doerr, H. W., Stürmer, M. and Preiser, W. (**2004**) 'Severe acute respiratory syndrome (SARS)-paradigm of an emerging viral infection.', *J Clin Virol.*, 29(1), pp. 13–22.

Berkelman, R. L., Bryan, R. T., Osterholm, M. T., LeDuc, J. W. and Hughes, J. M. (**1994**) 'Infectious disease surveillance: a crumbling foundation.', *Science.*, 264(5157), pp. 368–371.

Bermingham, A., Chand, M. A., Brown, C. S., Aarons, E., Tong, C., Langrish, C., *et al.* (**2012**) 'Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012.', *Euro Surveill.*, 17(40), p. 20290.

Berry, D. M., Cruickshank, J. G., Chu, H. P. and Wells, R. J. (**1964**) 'The structure of infectious bronchitis virus.', *Virology*, 23, pp. 403–7.

Berto, A., Anh, P. H., Carrique-Mas, J. J., Simmonds, P., Van Cuong, N., Tue, N. T., *et al.* (**2017**) 'Detection of potentially novel paramyxovirus and coronavirus viral RNA in bats and rats in the Mekong Delta region of southern Viet Nam.', *Zoonoses Public Health.*, [Epub ahead of print].

Binder, S., Levitt, A. M. and Hughes, J. M. (**1999**) 'Preventing emerging infectious diseases as we enter the 21st century: CDC's strategy.', *Public Health Rep.*, 114(2), pp. 130–4.

van Boheemen, S., de Graaf, M., Lauber, C., Bestebroer, T. M., Raj, V. S., Zaki, A. M., *et al.* (**2012**) 'Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans.', *MBio*, 3(6), pp. 1–9.

Booth, C. M., Matukas, L. M., Tomlinson, G. A., Rachlis, A. R., Rose, D. B., Dwosh, H. A., *et al.* (**2003**) 'Clinical features and short-term outcomes of 144 patients with SARS in the greater Toronto area.', *JAMA*, 289(21), pp. 2801–9.

Boshoff, C. I., Bastos, A. D. S., Gerber, L. J., and Vosloo, W. (2007) 'Genetic characterisation of African swine fever viruses from outbreaks in southern Africa (1973–1999).', *Vet Microbiol.*, 121(1-2), pp.45-55.

Breiman, R. F., Evans, M. R., Preiser, W., Maguire, J., Schnur, A., Li, A., *et al.* (**2003**) 'Role of China in the quest to define and control severe acute respiratory syndrome.', *Emerg Infect Dis.*, 9(9), pp. 1037–41.

Brierley, L., Vonhof, M. J., Olival, K. J., Daszak, P. and Jones, K. E. (**2015**) 'Quantifying global drivers of zoonotic bat viruses: A process-based perspective.', *Am Nat.*, 187(2), pp. E53–E64.

Brody, J. R. and Kern, S. E. (**2004**) 'Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis.', *Biotechniques.*, 36(2), pp. 214–6.

Brosset, A. (**1990**) 'The migrations of Pipistrellus nathusii in france possible implication on the spreading of rabies.', *Mammalia.*, 54(2), pp. 207–212.

Brown, I. H. (2010) 'Summary of avian influenza activity in Europe, Asia, and Africa, 2006–2009.', *Avian Dis.*, 54(s1), pp.187-193.

Brunet-Rossinni, A. K. and S., W. (**2009**) 'Methods for age estimation and the study of senescence in bats.', in Kunz, T. H. and Parsons, S. (eds) *Ecological and behavioral methods for the study of bats*. 2nd edn. Baltimore: John Hopkins Press.

Burke, D. S. (**1998**) 'The evolvability of emerging viruses.', in Nelson, A. M. and Horsburgh, C. R. (eds) *Pathology of emerging infections 2*. Washington, D.C: American Society for Microbiology, pp. 1–12.

Bushnell, L. D. and Brandly, C. A. (**1933**) 'Laryngotracheitis in Chicks.', *Poul Sci.*, 12(1), pp. 55–60.

Calcagno, V. and de Mazancourt, C. (**2010**) 'glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models.', *J Stat Softw.*, 34(12).

Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V and Schountz, T. (**2006**) 'Bats: important reservoir hosts of emerging viruses.', *Clin Microbiol Rev.*, 19(3), pp. 531–45.

Carrington, C. V. F., Foster, J. E., Zhu, H. C., Zhang, J. X., Smith, G. J. D., Thompson, N., *et al. (**2008**) 'Detection and phylogenetic analysis of group 1 coronaviruses in South American bats.', *Emerg Infect Dis.*, 14(12), pp. 1890–3.

Carstens, E. B. (**2010**) 'Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2009).', *Arch Virol.*, 155(1), pp. 133–46.

Cavanagh, D. (**2005**) 'Coronaviruses in poultry and other birds.', *Avian Pathol.*, 34(6), pp. 439–448.

Cavanagh, D. and Britton, P. (**2008**) 'Coronaviruses: General Features.', in Mahy, B. W. J. and van Regenmortel, M. H. V. (eds) *Encyclopedia of Virology*. Elsevier, pp. 549–554.

Chan, J. F. W., Lau, S. K. P., To, K. K. W., Cheng, V. C. C., Woo, P. C. Y. and Yuen, K.-Y. (**2015**) 'Middle East respiratory syndrome coronavirus: Another zoonotic betacoronavirus causing SARS-like disease.', *Clin Microbiol Rev.*, 28(2), pp. 465–522.

Charron, D. F. (**2012**) 'Ecosystem approaches to health for a global sustainability agenda.', *Ecohealth*, 9(3), pp. 256–266.

Chen, L., Liu, B., Yang, J. and Jin, Q. (**2014**) 'DBatVir: the database of bat-associated viruses.', *Database (Oxford).*, p. bau021.

Chen, Y.-N., Phuong, V. N., Chen, H. C., Chou, C.-H., Cheng, H.-C. and Wu, C.-H. (**2016**) 'Detection of the severe acute respiratory syndrome-related coronavirus and alphacoronavirus in the bat population of Taiwan.', *Zoonoses Public Health.* 63(8), pp. 608-615.

Cheng, V. C. C., Lau, S. K. P., Woo, P. C. Y. and Yuen, K. Y. (**2007**) 'Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection.', *Clin Microbiol Rev.*, 20(4), pp. 660–94.

Cho, S. Y., Kang, J.-M., Ha, Y. E., Park, G. E., Lee, J. Y., Ko, J.-H., *et al. (**2016**) 'MERS-CoV outbreak following a single patient exposure in an emergency room in South Korea: an epidemiological outbreak study.', *Lancet.*, 388(10048), pp. 994–1001.

Chomel, B. B. (**1998**) 'New emerging zoonoses: a challenge and an opportunity for the veterinary profession.', *Comp Immunol Microbiol Infect Dis.*, 21(1), pp. 1–14.

Chowell, G., Abdirizak, F., Lee, S., Lee, J., Jung, E., Nishiura, H., *et al.* (**2015**) 'Transmission characteristics of MERS and SARS in the healthcare setting: a comparative study.', *BMC Med.*, 13(1), p. 210.

Chu, D. K. W., Peiris, J. S. M., Chen, H., Guan, Y. and Poon, L. L. M. (**2008a**) 'Genomic characterizations of bat coronaviruses (1A, 1B and HKU8) and evidence for co-infections in Miniopterus bats.', *J Gen Virol.*, 89(Pt 5), pp. 1282–7.

Chu, D. K. W., Poon, L. L. M., Chan, K. H., Chen, H., Guan, Y., Yuen, K. Y., *et al.* (**2006**) 'Coronaviruses in bent-winged bats (Miniopterus spp.).', *J Gen Virol.*, 87(Pt 9), pp. 2461–6.

Chu, D. K. W., Poon, L. L. M., Guan, Y. and Peiris, J. S. M. (**2008b**) 'Novel astroviruses in insectivorous bats.', *J Virol.*, 82(18), pp. 9107–14.

Chua, K. B. (**2003**) 'Nipah virus outbreak in Malaysia.', *J Clin Virol.,* 26(3), pp. 265–75.

Chua, K. B. (**2010**) 'Risk factors, prevention and communication strategy during Nipah virus outbreak in Malaysia.', *Malays J Pathol.*, 32(2), pp. 75–80.

Chua, K. B., Wang, L.-F., Lam, S. K., Crameri, G., Yu, M., Wise, T., *et al.* (**2001**) 'Tioman virus, a novel paramyxovirus isolated from fruit bats in Malaysia.', *Virology.*, 283(2), pp. 215–229.

Cleaveland, S., Laurenson, M. K. and Taylor, L. H. (**2001**) 'Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence.', *Philos Trans R Soc Lond B Biol Sci.*, 356(1411), pp. 991–9.

Cohen, M. L. (**2000**) 'Changing patterns of infectious disease.', *Nature.*, 406(6797), pp. 762–767.

Cook, J. K. A., Jackwood, M. and Jones, R. C. (**2012**) 'The long view: 40 years of infectious bronchitis research.', *Avian Pathol.*, 41(3), pp. 239–250.

Cook, R. A., Karesh, W. B. and Osofsky, S. (**2004**) *The Manhattan principles on 'One World, One Health'*, *One World One Health.* Available at: http://www.oneworldonehealth.org/. (Accessed: 12 May 2014).

Corman, V. M., Baldwin, H. J., Fumie Tateno, A., Melim Zerbinati, R., Annan, A., Owusu, M., *et al.* (**2015**) 'Evidence for an ancestral association of human coronavirus 229E with bats.', *J Virol.*, 89(23), pp. 11858-70.

Corman, V. M., Ithete, N. L., Richards, L. R., Schoeman, M. C., Preiser, W., Drosten, C., *et al.* (**2014a**) 'Rooting the phylogenetic tree of Middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat.', *J Virol.*, 88(19), pp. 11297–303.

Corman, V. M., Kallies, R., Philipps, H., Göpner, G., Müller, A., Eckerle, I., *et al.* (**2014b**) 'Characterization of a novel betacoronavirus related to Middle East respiratory syndrome coronavirus in European hedgehogs.', *J Virol.*, 88(1), pp. 717–724.

Corman, V. M., Müller, M. A., Costabel, U., Timm, J., Binger, T., Meyer, B., *et al.* (**2012**) 'Assays for laboratory confirmation of novel human coronavirus (hCoV-EMC) infections.', *Euro Surveil.,* 17(49), pp. 1–9.

Corman, V. M., Rasche, A., Diallo, T. D., Cottontail, V. M., Stöcker, A., Souza, B. F. de C. D., *et al.* (**2013**) 'Highly diversified coronaviruses in neotropical bats.', *J Gen Virol.*, 94(Pt 9), pp. 1984–94.

Crameri, G., Todd, S., Grimley, S., McEachern, J. A, Marsh, G. A, Smith, C., *et al.* (**2009**) 'Establishment, immortalisation and characterisation of pteropid bat cell lines.', *PloS One.*, 4(12), p. e8266.

Crick, J., Tignor, G. H. and Moreno, K. (**1982**) 'A new isolate of Lagos bat virus from the Republic of South Africa.', *Trans R Soc Tropl Med Hyg.*, 76(2), pp. 211–3.

Cui, J., Han, N., Streicker, D., Li, G., Tang, X., Shi, Z., *et al.* (**2007**) 'Evolutionary relationships between bat coronaviruses and their hosts.', *Emerg Infect Dis.*, 13(10), pp. 1526–32.

Cunningham, A. A., Daszak, P. and Wood, J. L. N. (**2017**) 'One Health, emerging infectious diseases and wildlife: two decades of progress?', *Philos Trans R Soc B Biol Sci.*, 372(1725), p. 20160167.

Dacheaux, L., Cervantes-Gonzales, M., Guigon, G., Thiberge, J., Vandenbogaert, M., Maufrais, C., *et al.* **(2014) '**A preliminary study of viral metagenomics of French bat species in contact with humans: identification of new mammalian viruses.', *PLoS One.*, 9(1), p. e87194.

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

Daszak, P. (**2000**) 'Emerging infectious diseases of wildlife-- Threats to biodiversity and human health.', *Science*, 287(5452), pp. 443–449.

Daszak, P., Cunningham, A. A., and Hyatt, A. D. (**2001**) 'Anthropogenic environmental change and the emergence of infectious diseases in wildlife.', *Acta Trop.*, 78(2), pp. 103–16.

Denison, M. R., Graham, R. L., Donaldson, E. F., Eckerle, L. D. and Baric, R. S. (**2011**) 'Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity.', *RNA Biol.*, 8(2), pp. 270–9.

Dobson, A. and Foufopoulos, J. (**2001**) 'Emerging infectious pathogens of wildlife.', *Philos Trans R Soc Lond B Biol Sci.*, 356(1411), pp. 1001–12.

Dobson, A. P. (**2005**) 'Virology. What links bats to emerging infectious diseases?', *Science.*, 310(5748), pp. 628–9.

Dobson, G. (**1875**) 'Conspectus of the suborders, families and genera of Chiroptera arranged according to their natural affinities.', *Ann Mag Nat Hist* 16(95), pp. 345–357.

Domingo, E. and Holland, J. J. (**1997**) 'RNA virus mutations and fitness for survival.', *Annu Rev Microbiol.*, 51, pp. 151–78.

Dominguez, S. R., O'Shea, T. J., Oko, L. M. and Holmes, K. V (**2007**) 'Detection of group 1 coronaviruses in bats in North America.', *Emerg Infect Dis.*, 13(9), pp. 1295–300.

Doyle, L. P. and Hutchings, L. M. (**1946**) 'A transmissible gastroenteritis in pigs.', *J Am Vet Med Assoc.*, 108, pp. 257–9.

Drexler, J. F., Corman, V. M. and Drosten, C. (**2014**) 'Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS.', *Antiviral Res.*, 101, pp. 45–56.

Drexler, J. F., Corman, V. M., Müller, M. A., Maganga, G. D., Vallo, P., Binger, T., *et al.* (**2012**) 'Bats host major mammalian paramyxoviruses.', *Nat Commun.*, 3, p. 796.

Drexler, J. F., Corman, V. M., Wegner, T., Tateno, A. F., Zerbinati, R. M., Gloza-Rausch, F., *et al.* (**2011**) 'Amplification of emerging viruses in a bat colony.', *Emerg Infect Dis.*, 17(3), pp. 449–56.

Drexler, J. F., Gloza-Rausch, F., Glende, J., Corman, V. M., Muth, D., Goettsche, M., *et al.* (**2010**) 'Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences.', *J Virol.*, 84(21), pp. 11336–49.

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., *et al.* (2003) 'Identification of a novel coronavirus in patients with severe acute respiratory syndrome.', *N Engl J Med.*, 348(20), pp. 1967–76.

Duan, S.-M., Zhao, X.-S., Wen, R.-F., Huang, J.-J., Pi, G.-H., Zhang, S.-X., *et al.* (2003) 'Stability of SARS coronavirus in human specimens and environment and its sensitivity to heating and UV irradiation.', *Biomed Environ Sci.*, 16(3), pp. 246–55.

Dzingirai, V., Bukachi, S., Leach, M., Mangwanya, L., Scoones, I. and Wilkinson, A. (2017) 'Structural drivers of vulnerability to zoonotic disease in Africa.', *Philos Trans R Soc Lond B Biol Sci.*, 372(1725), pii: 20160169.

Eckerle, I., Ehlen, L., Kallies, R., Wollny, R., Corman, V. M., Cottontail, V. M., *et al.* (2014) 'Bat airway epithelial cells: a novel tool for the study of zoonotic viruses.', *PloS One.*, 9(1), p. e84679. .

Eick, G. N., Jacobs, D. S. and Matthee, C. A. (2005) 'A nuclear DNA phylogenetic perspective on the evolution of echolocation and historical biogeography of extant bats (chiroptera).', *Mol Biol Evol.*, 22(9), pp. 1869–86.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T. and Arshanapalli, B. G. (2014) 'Appendix E: Model Selection Criterion: AIC and BIC.', in *The Basics of Financial Econometrics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 399–403.

Falcón, A., Vázquez-Morón, S., Casas, I., Aznar, C., Ruiz, G., Pozo, F., *et al.* (2011) 'Detection of alpha- and betacoronaviruses in multiple Iberian bat species.', *Arch Virol.*, 156(10), pp. 1883–1890.

Farell, E. M. and Alexandre, G. (2012) 'Bovine serum albumin further enhances the effects of organic solvents on increased yield of polymerase chain reaction of GC-rich templates.', *BMC Res Notes.*, 5(1), p. 257.

Fehr, A. R. and Perlman, S. (2015) 'Coronaviruses: an overview of their replication and pathogenesis.', *Methods Mol Biol.*, 1282, pp. 1–23.

Fenner, F. (1976) 'The classification and nomenclature of viruses - Summary of results of meetings of the International Committee on Taxonomy of Viruses in Madrid, September 1975.', *J Gen Virol.*, 31, pp. 463–470.

Field, H., Young, P., Yob, J. M., Mills, J., Hall, L. and Mackenzie, J. (2001) 'The natural history of Hendra and Nipah viruses.', *Microbes Infect.,* 3(4), pp. 307–14.

Fischer, K., Zeus, V., Kwasnitschka, L., Kerth, G., Haase, M., Groschup, M. H., *et al.* (2016) 'Insectivorous bats carry host specific astroviruses and coronaviruses across different regions in Germany.', *Infect Genet Evol.*, 37, pp. 108–16.

Fisher, R. (2016) *Next generation sequencing demonstrates minor variant HIV drug resistance mutations*. Stellenbosch University.

Folmer, O., Black, M., Hoeh, W., Lutz, R. and Vrijenhoek, R. (1994) 'DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.', *Mol Mar Biol Biotechnol.*, 3(5), pp. 294–9.

FAO, Food and Agriculture Organisation of the United Nations (2011) 'Investigating the role of bats in emerging zoonoses: Balancing ecology, conservation and public health interests.', in Newman, S. H. *et al.* (eds) *FAO Animal Production and Health Manual No.12*. Rome: Food and Agriculture Organisation of the United Nations.

Fouchier, R. A. M., Hartwig, N. G., Bestebroer, T. M., Niemeyer, B., de Jong, J. C., Simon, J. H., *et al.* (**2004**) 'A previously undescribed coronavirus associated with respiratory disease in humans.', *Proc Natl Acad Sci U S A.*, 101(16), pp. 6212–6.

Fouchier, R. A. M., Kuiken, T., Schutten, M., van Amerongen, G., van Doornum, G. J. J., van den Hoogen, B. G., *et al.* (**2003**) 'Aetiology: Koch's postulates fulfilled for SARS virus.', *Nature.*, 423(6937), p. 240.

Fragaszy, E. and Hayward, A. (**2014**) 'Emerging respiratory infections: influenza, MERS-CoV, and extensively drug-resistant tuberculosis.', *Lancet Respir Med.*, 2(12), pp. 970–2.

Francis, C. M. (**1989**) 'A comparison of mist nets and two designs of harp traps for capturing bats.', *J Mammal.*, 70(4), pp. 865–870.

Funk, A. L., Goutard, F. L., Miguel, E., Bourgarel, M., Chevalier, V., Faye, B., *et al.* (**2016**) 'MERS-CoV at the animal–human interface: inputs on exposure pathways from an expert-opinion elicitation.', *Front Vet Sci.*, 3, p. 88.

Ge, X.-Y., Li, J.-L., Yang, X.-L., Chmura, A. a, Zhu, G., Epstein, J. H., *et al.* (**2013**) 'Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor.', *Nature.*, 503(7477), pp. 535–8.

Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y., *et al.* (2012) 'Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China.', *J Virol.*, 86(8), pp. 4620-30.

Ge, X.-Y., Wang, N., Zhang, W., Hu, B., Li, B., Zhang, Y.-Z., *et al.* (**2016**) 'Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft.', *Virol Sin.*, 31(1), pp. 31–40.

Geldehuys, C. and Mucina, L. (**2006**) 'Towards a new national forest classification for South Africa.', in Ghazanfar, S. and Beentje, H. (eds) *Taxonomy and ecology of African plants, their conservation and sustainable use.* Kew Publishing, pp. 111–129.

Geldenhuys, M., Weyer, J., Nel, L. H. and Markotter, W. (**2013**) 'Coronaviruses in South African bats.', *Vector Borne Zoonotic Dis.*, 13(7), pp. 516–9.

Geller, C., Varbanov, M. and Duval, R. E. (**2012**) 'Human coronaviruses: insights into environmental resistance and its influence on the development of new antiseptic strategies.', *Viruses.*, 4(11), pp. 3044–68.

George, D. B., Webb, C. T., Farnsworth, M. L., O'Shea, T. J., Bowen, R. A., Smith, D. L., *et al.* (**2011**) 'Host and viral ecology determine bat rabies seasonality and maintenance.', *Proc Natl Acad Sci U S A.*, 108(25), pp. 10208–13.

Germi, R., Lupo, J., Semenova, T., Larrat, S., Magnat, N., Grossi, L., *et al.* (**2012**) 'Comparison of commercial extraction systems and PCR assays for quantification of Epstein-Barr virus DNA load in whole blood.', *J Clin Microbiol.*, 50(4), pp. 1384–9.

Gledhill, A. W. and Andrewes, C. H. (**1951**) 'A hepatitis virus of mice.', *Br J Exp Pathol.*, 32(6), pp. 559–68.

Gloza-Rausch, F., Ipsen, A., Seebens, A., Göttsche, M., Panning, M., Drexler, J. F., *et al.* (**2008**) 'Detection and prevalence patterns of group I coronaviruses in bats, northern Germany.', *Emerg Infect Dis.*, 14(4), pp. 626–31.

Góes, L. G. B., de Almeida Campos, A. C., de Carvalho, C., Ambar, G., Queiroz, L. H., Cruz-Neto, A. P., *et al.* (**2016**) 'Genetic diversity of bats coronaviruses in the Atlantic Forest hotspot biome, Brazil.', *Infect Genet Evol.,* 44, pp. 510-3.

Góes, L. G. B., Ruvalcaba, S. G., Campos, A. A., Queiroz, L. H., de Carvalho, C., Jerez, J. A., *et al.* (**2013**) 'Novel bat coronaviruses, Brazil and Mexico.', *Emerg Infect Dis.*, 19(10), pp. 1711–3.

Goffard, A., Demanche, C., Arthur, L., Pin?on, C., Michaux, J. and Dubuisson, J. (**2015**) 'Alphacoronaviruses detected in French bats are phylogeographically linked to coronaviruses of European bats.', *Viruses.*, 7(12), pp. 6279–6290.

Goller, K. V., Fyumagwa, R. D., Nikolin, V., East, M. L., Kilewo, M., Speck, S., *et al.* (**2010**) 'Fatal canine distemper infection in a pack of African wild dogs in the Serengeti ecosystem, Tanzania.', *Vet Microbiol.*, 146(3–4), pp. 245–252.

González, J. M., Gomez-Puertas, P., Cavanagh, D., Gorbalenya, A. E. and Enjuanes, L. (**2003**) 'A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae.', *Arch Virol.*, 148(11), pp. 2207–35.

Green, S. J., Venkatramanan, R., Naqib, A., Tseng, C.-P. and Stombaugh, J. (**2015**) 'Deconstructing the polymerase chain reaction: Understanding and correcting bias associated with primer degeneracies and primer-template mismatches.', *PloS One.*, 10(5), p. e0128122.

Griffin, M. (**2008**) *Scotophilus dinganii*, *The IUCN Red List of Threatened Species*. Available at: http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T20066A9141665.en. (Accessed: 20 July 2017).

Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., *et al.* (**2003**) 'Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China.', *Science.*, 302(5643), pp. 276–8.

Gubler, D. J. (**2001**) 'Human arbovirus infections worldwide.', *Ann N Y Acad Sci.*, 951, pp. 13–24.

Haarsma, A.-J. (**2008**) *Manual for assessment of reproductive status, age and health in European Vespertilionid bats*. V1. Hillegom, Holland: Electronic Release.

Hamre, D. and Procknow, J. J. (**1966**) 'A new virus isolated from the human respiratory tract.', *Proc Soc Exp Biol Med.,* 121(1), pp. 190–3.

Han, B. A., Schmidt, J. P., Bowden, S. E. and Drake, J. M. (**2015a**) 'Rodent reservoirs of future zoonotic diseases.', *Proc Natl Acad Sci U S A.*, 112(22), pp. 7039–44.

Han, H., Wen, H., Zhou, C., Chen, F., Luo, L., Liu, J.-W., *et al.* (**2015b**) 'Bats as reservoirs of severe emerging infectious diseases.', *Virus Res.*, 205, pp. 1–6.

Hawkey, P. M. (**2003**) 'Severe acute respiratory syndrome (SARS): breath-taking progress.', *J Med Microbiol.*, 52(8), pp. 609–613.

Haydon, D. T., Cleaveland, S., Taylor, L. H. and Laurenson, M. K. (**2002**) 'Identifying reservoirs of infection: A conceptual and practical challenge.', *Emerg Infect Dis.*, 8(12), pp. 1468–73.

Hayman, D. T. S. (**2015**) 'Biannual birth pulses allow filoviruses to persist in bat populations.', *Proc R Soc B Biol Sci.*, 282(1803), pp. 20142591–20142591.

He, B., Li, Z., Yang, F., Zheng, J., Feng, Y., Gui, H., *et al.* (2013) 'Virome profiling of bats from Myanmar by metagenomic analysis of tissue samples reveals more novel mammalian viruses.', *PLoS One.*, 8(4), p. e61950.

He, B., Zhang, Y., Xu, L., Yang, W., Yang, F., Feng, Y. Y., *et al.* (**2014a**) 'Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China.', *J Virol.*, 88(12), pp. 7070–82.

He, X., Korytář, T., Zhu, Y., Pikula, J., Bandouchova, H., Zukal, J., *et al.* (**2014b**) 'Establishment of Myotis myotis cell lines - Model for investigation of host-pathogen interaction in a natural host for emerging viruses.', *PloS One.*, 9(10), p. e109795.

van Helden, P. D., van Helden, L. S. and Hoal, E. G. (**2013**) 'One world, one health. Humans, animals and the environment are inextricably linked--a fact that needs to be remembered and exploited in our modern approach to health.', *EMBO Rep.*, 14(6), pp. 497–501.

Hijawi, B., Abdallat, M., Sayaydeh, A., Alqasrawi, S., Haddadin, A., Jaarour, N., *et al.* (**2013**) 'Novel coronavirus infections in Jordan, April 2012: Epidemiological findings from a retrospective investigation.', *East Mediterr Health J*, 19 Suppl 1, pp. S12-8.

van der Hoek, L., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J. M., Wolthers, K. C., *et al.* (**2004**) 'Identification of a new human coronavirus.', *Nat Med.*, 10(4), pp. 368–73.

Hogue, B. G. and Machamer, C. E. (**2008**) 'Coronavirus structural proteins and assembly.', in Perlman, S., Gallagher, T. M., and Snijder, E. J. (eds) *Nidoviruses.* Washington, D.C: ASM Press.

Holmes, E. C., Dudas, G., Rambaut, A. and Andersen, K. G. (**2016**) 'The evolution of Ebola virus: Insights from the 2013–2016 epidemic.', *Nature.*, 538(7624), pp. 193–200.

Holmes, E. C. and Zhang, Y. (**2015**) 'The evolution and emergence of hantaviruses.', *Curr Opin Virol.*,10, pp. 27–33.

Hsu, L.-Y., Lee, C.-C., Green, J. A., Ang, B., Paton, N. I., Lee, L., *et al.* (**2003**) 'Severe acute respiratory syndrome (SARS) in Singapore: Clinical features of index patient and initial contacts.', *Emerg Infect Dis.*, 9(6), pp. 713–7.

Hu, B., Ge, X., Wang, L.-F. and Shi, Z. (**2015**) 'Bat origin of human coronaviruses.', *Virol J.*, 12(1), p. 221.

Huang, Y.-W., Dickerman, A. W., Piñeyro, P., Li, L., Fang, L., Kiehne, R., *et al.* (**2013**) 'Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the United States.', M*Bio.*, 4(5), pp. e00737-13.

Huelsenbeck, J. P. and Ronquist, F. (**2001**) 'MRBAYES: Bayesian inference of phylogenetic trees.', *Bioinformatics.*, 17(8), pp. 754–5.

Hutcheon, J. M. and Kirsch, J. A. W. (**2006**) 'A moveable face: deconstructing the Microchiroptera and a new classification of extant bats.', *Acta Chiropterol.*, 8(1), pp. 1–10.

Hutson, A. M., Aulagnier, S., Benda, P., Karataş, A., Palmeirim, J. and Paunović, M. (**2008**) *Miniopterus schreibersii*, *The IUCN Red List of Threatened Species*. Available at: http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T13561A4160556.en. (Accessed: 12 July 2017).

Huynh, J., Li, S., Yount, B., Smith, A., Sturges, L., Olsen, J. C., *et al.* (**2012**) 'Evidence supporting a zoonotic origin of human coronavirus strain NL63.', *J Virol.*, 86(23), pp. 12816–25.

ICTV, International Committee on Taxonomy of Viruses. (**2005**) *Virus taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses*. Edited by C. Fauquet *et a*l. Academic Press.

ICTV, International Committee on Taxonomy of Viruses. (**2012**) *Virus taxonomy: classification and nomenclature of viruses: Ninth report of the International Committee on Taxonomy of Viruses*. Edited by A. M. Q. King *et al.* Elsevier Academic Press.

IOM, Committee on Emerging Microbial Threats to Health. (**1992**) *Emerging infections: Microbial threats to health in the United States*, *Sciences-New York.* Edited by J. Lederberg, R. E. Shope, and S. C. Oaks. Washington, D.C: National Academy Press.

IOM, Committee on Emerging Microbial Threats to Health in the 21st Century. (**2003**) *Microbial Threats to Health*. Edited by M. Smolinski, M. Hamburg, and J. Lederberg. Washington, D.C.: National Academies Press.

Irwin, D. M., Kocher, T. D. and Wilson, A. C. (**1991**) 'Evolution of the cytochrome b gene of mammals.', *J Mol Evol.*, 32(2), pp. 128–44.

Ithete, N. (**2013**) *Investigation of small mammal-borne viruses with zoonotic potential in South Africa*. Stellenbosch University.

Ithete, N. L., Stoffberg, S., Corman, V. M., Cottontail, V. M., Richards, L. R., Schoeman, M. C., *et al.* (**2013**) 'Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa.', *Emerg Infect Dis.*, 19(10), pp. 1697–9.

IUCN, International Union for Conservation of Nature. (**2012**) *The IUCN Red List of Threatened species.* Version 20. Edited by IUCN. Available at: http://www.iucnredlist.org. (Accessed: 02 October 2014).

Jacobs, D., Cotterill, F. P. D. and Taylor, P. J. (**2008**) *Neoromicia capensis*, *The IUCN Red List of Threatened Species.* Available at: http://dx.doi.org/10.2305/IUCN.UK.2014-3.RLTS.T44918A67358046.en. (Acessed: 04 March 2014).

Jacobsen, K. H., Aguirre, A. A., Bailey, C. L., Baranova, A. V., Crooks, A. T., Croitoru, A., *et al.* (**2016**) 'Lessons from the Ebola outbreak: Action items for emerging infectious disease preparedness and response.', *Ecohealth.*, 13(1), pp. 200–212.

Jansen van Vuren, P., Wiley, M., Palacios, G., Storm, N., McCulloch, S., Markotter, W., *et al.* (**2016**) 'Isolation of a novel fusogenic orthoreovirus from Eucampsipoda africana bat flies in South Africa.', *Viruses.*, 8(3), p. 65.

Jansen van Vuren, P., Wiley, M. R., Palacios, G., Storm, N., Markotter, W., Birkhead, M., *et al.* (**2017**) 'Isolation of a novel orthobunyavirus from bat flies (Eucampsipoda africana).', *J Gen Virol.*, 98(5), pp. 935–945.

Johnson, C. K., Hitchins, P. L., Evans, T. S., Goldstein, T., Thomas, K., Clements, A., Joly, D. O., *et al.* (2015) ' Spillover and pandemic properties of zoonotic viruses with high host plasticity.', *Sci Rep.,* 5, p. 14830.

Jones, K. E., Patel, N. G., Levy, M. a, Storeygard, A., Balk, D., Gittleman, J. L., *et al.* (**2008**) 'Global trends in emerging infectious diseases.', *Nature.*, 451(7181), pp. 990–3.

Jones, B. A., Grace, D., Kock, R., Alonso, S., Rushton, J., Said, M. Y., et al. (2013). ' Zoonosis emergence linked to agricultural intensification and environmental change.', *Natl Proc Acad Sci U S A.*, 110(21), pp. 8399-8404.

Jordan, J. and Mirick, G. S. (**1951**) 'Hepatitis in mice, of presumed viral origin; a preliminary report.', *Bull Johns Hopkins Hosp.*, 89(4), pp. 328–31.

Kan, B., Wang, M., Jing, H., Xu, H., Jiang, X., Yan, M., *et al.* (**2005**) 'Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms.', *J Virol.*, 79(18), pp. 11892–11900.

Kandeil, A., Shehata, M. M., El Shesheny, R., Gomaa, M. R., Ali, M. A. and Kayali, G. (**2016**) 'Complete genome sequence of Middle East respiratory syndrome coronavirus isolated from a dromedary camel in Egypt.', *Genome Announc.*, 4(2), pp. e00309-16.

Kawaoka, Y. and Webster, R. G. (**1988**) 'Molecular mechanism of acquisition of virulence in influenza virus in nature.', *Microb Pathog.*, 5(5), pp. 311–318.

Kearney, T. C., Volleth, M., Contrafatto, G. and Taylor, P. J. (2002) 'Systematic implications of chromosome GTG-band and bacula morphology for Southern African Eptesicus and Pipistrellus and several other species of Vespertilioninae (Chiroptera: Vespertilionidae).', *Acta Chiropterol.*, 4(1), pp. 55-76.

Keesing, F., Belden, L. K., Daszak, P., Dobson, A., Harvell, C. D., Holt, R. D., *et al.* (2010) 'Impacts of biodiversity on the emergence and transmission of infectious diseases.', *Nature.*, 468(7324), pp. 647–52.

Kim, H. K., Yoon, S.-W. S.-W., Kim, D.-J. D.-J., Koo, B.-S. B.-S., Noh, J. Y., Kim, J. H., *et al.* (2016) 'Detection of severe acute respiratory syndrome-like, Middle East respiratory syndrome-like bat coronaviruses and group H rotavirus in faeces of Korean bats.', *Transbound Emerg Dis.*, 63(4), pp. 365–72.

Kim, Y., Han, M.-S., Kim, J., Kwon, A. and Lee, K.-A. (2014) 'Evaluation of three automated nucleic acid extraction systems for identification of respiratory viruses in clinical specimens by multiplex real-time PCR.', *BioMed Res Int.*, 2014, p. 430650.

Kipar, A. and Meli, M. L. (2014) 'Feline infectious peritonitis.', *Vet Pathol.*, 51(2), pp. 505–526.

Klein, S. L. (2000) 'The effects of hormones on sex differences in infection: from genes to behavior.', *Neurosci Biobehav Rev.*, 24(6), pp. 627–38.

Konca, C., Korukluoglu, G., Tekin, M., Almis, H., Bucak, İ. H., Uygun, H., *et al.* (2017) 'The first infant death associated with human coronavirus NL63 infection.', *Pediatr Infect Dis J.*, 36(2), pp. 231–233.

Koopmans, M. (2013) 'Surveillance strategy for early detection of unusual infectious disease events.', *Curr Opin Virol.*, 3(2), pp. 185–91.

Kreader, C. A. (1996) 'Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein.', *Appl Environl Microbiol.*, 62(3), pp. 1102–1106.

Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., *et al.* (2003) 'A novel coronavirus associated with severe acute respiratory syndrome.', *N Engl J Med.*, 348(20), pp. 1953–66.

Kumar, S., Stecher, G. and Tamura, K. (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets.', *Mol Biol Evol.*, 33(7), pp. 1870–4.

Kunz, T. H. and Anthony, E. L. P. (1982) 'Age estimation and post-natal growth in the bat Myotis lucifugus.', *Journal of Mammalogy*. The Oxford University Press, 63(1), pp. 23–32.

Kunz, T. H., Hodgkison, R. and Weise, C. D. (2009) 'Methods of capturing and handling bats.', in Kunz, T. and Parsons, S. (eds) *Ecological and behavioral methods for the study of bats*. 2nd edition. Baltimore: The Johns Hopkins University Press, pp. 3–35.

Kunz, T. H. and Lumsden, L. (2003) 'Ecology of cavity and foliage roosting bats.', in Kunz, Thomas, H. and Fenton, M. B. (eds) *Bat Ecology*. Chicago: University of Chicago Press, pp. 3–89.

Kunz, T. and Peirson, E. (1994) 'Bats of the world: An introduction.', in Nowak, R. (ed.) *Walker's Bats of the World*. John Hopkins Press, pp. 1–46.

Kunz, T., Tideman, C. and Richards, G. (1996) 'Capturing mammals: Small volant mammals.', in Wilson, D. *et al.* (eds) *Measuring and monitoring biological diversity: standard methods for mammals*. Smithsonian Institute Press.

Kuzmin, I. V, Bozick, B., Guagliardo, S. A., Kunkel, R., Shak, J. R., Tong, S., *et al.* (2011) 'Bats, emerging infectious diseases, and the rabies paradigm revisited.', *Emerg Health Threats J.*, 4, p. 7159.

Lacroix, A., Duong, V., Hul, V., San, S., Davun, H., Omaliss, K., *et al.* (2017) 'Genetic diversity of coronaviruses in bats in Lao PDR and Cambodia.', *Infect Genet Evol.*, 48, pp. 10–18.

Lam, S. K. and Chua, K. B. (2002) 'Nipah virus encephalitis outbreak in Malaysia.', *Clin Infect Dis.,* 34 Suppl 2(Suppl 2), pp. S48-51.

Larsen, R. J. (2007) *Mist net interaction, sampling effort, and species of bats captured on Montserrat, British West Indies*. South Dakota State University.

Lau, S. K. P., Ahmed, S. S., Yeung, H. C., Li, K. S. M., Fan, R. Y. Y., Cheng, T. Y. C., *et al.* (**2016a**) 'Identification and interspecies transmission of a novel bocaparvovirus among different bat species in China.', *J Gen Virol.*, 97(12), pp. 3345–3358.

Lau, S. K. P. and Chan, J. F. W. (**2015**) 'Coronaviruses: emerging and re-emerging pathogens in humans and animals.', *Virol J.*, 12(1), p. 209.

Lau, S. K. P., Li, K. S. M., Huang, Y., Shek, C.-T., Tse, H., Wang, M., *et al.* (**2010**) 'Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events.', *J Virol.*, 84(6), pp. 2808–19.

Lau, S. K. P., Li, K. S. M., Tsang, A. K. L., Lam, C. S. F., Ahmed, S., Chen, H., *et al.* (**2013**) 'Genetic characterization of betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of Pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus.', *J Virol.*, 87(15), pp. 8638–50.

Lau, S. K. P., Li, K. S. M., Tsang, A. K. L., Shek, C.-T., Wang, M., Choi, G. K. Y., *et al.* (**2012a**) 'Recent transmission of a novel alphacoronavirus, bat coronavirus HKU10, from Leschenault's rousettes to pomona leaf-nosed bats: first evidence of interspecies transmission of coronavirus between bats of different suborders.', *J Virol.*, 86(21), pp. 11906–18.

Lau, S. K. P., Wernery, R., Wong, E. Y. M., Joseph, S., Tsang, A. K. L., Patteril, N. A. G., *et al.* (**2016b**) 'Polyphyletic origin of MERS coronaviruses and isolation of a novel clade A strain from dromedary camels in the United Arab Emirates.', *Emerg Microbes Infect.*, 5(12), p. e128.

Lau, S. K. P., Woo, P. C. Y., Li, K. S. M., Huang, Y., Tsoi, H., Wong, B. H. L., *et al.* (**2005**) 'Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats.', *Proc Natl Acad Sci U S A.*, 102(39), pp. 14040–5.

Lau, S. K. P., Woo, P. C. Y., Li, K. S. M., Huang, Y., Wang, M., Lam, C. S. F., *et al.* (**2007**) 'Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome.', *Virology.*, 367(2), pp. 428–39.

Lau, S. K. P., Woo, P. C. Y., Yip, C. C. Y., Fan, R. Y. Y., Huang, Y., Wang, M., *et al.* (**2012b**) 'Isolation and characterization of a novel Betacoronavirus subgroup A coronavirus, rabbit coronavirus HKU14, from domestic rabbits.', *J Virol.*, 86(10), pp. 5481–96.

Lau, S. S. K. P., Feng, Y., Chen, H., Luk, H. K. H., Yang, W.-H., Li, K. S. M., *et al.* (**2015**) 'Severe acute respiratory syndrome (SARS) coronavirus ORF8 protein is acquired from SARS-related coronavirus from greater horseshoe bats through recombination.', *J Virol.*, 89(20), pp. 10532–47.

Laurenson, K., Sillero-Zubiri, C., Thompson, H., Shiferaw, F., Thirgood, S. and Malcolm, J. (**1998**) 'Disease as a threat to endangered species: Ethiopian wolves, domestic dogs and canine pathogens.', *Anim Conserv.*, 1(4), pp. 273–280.

Le, S. Q. and Gascuel, O. (**2008**) 'An improved general amino acid replacement matrix.', *Mol Biol Evol.*, 25(7), pp. 1307–1320.

Lederberg, J. (**1993**) 'Emerging infections: microbial threats to health.', *Trends Microbiol.*, 1(2), pp. 43–4.

Lee, C. (**2015**) 'Porcine epidemic diarrhea virus: An emerging and re-emerging epizootic swine virus.', *Virol J.*, 12, p. 193.

Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G. M., *et al.* (**2003**) 'A major outbreak of severe acute respiratory syndrome in Hong Kong.', *N Engl J Med.*, 348(20), pp. 1986–1994.

Lelli, D., Papetti, A., Sabelli, C., Rosti, E., Moreno, A. and Boniotti, M. B. (**2013**) 'Detection of coronaviruses in bats of various species in Italy.', *Viruses.*, 5(11), pp. 2679–89.

Leroy, E. M., Epelboin, A., Mondonge, V., Pourrut, X., Gonzalez, J.-P., Muyembe-Tamfum, J.-J., *et al.* (**2009**) 'Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007.', *Vector Borne Zoonotic Dis.*, 9(6), pp. 723–728.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., *et al.* (**2005**) 'Bats are natural reservoirs of SARS-like coronaviruses.', *Science.,* 310(5748), pp. 676–9.

Licitra, B., Duhamel, G. and Whittaker, G. (**2014**) 'Canine enteric coronaviruses: Emerging viral pathogens with distinct recombinant spike proteins.', *Viruses.*, 6(8), pp. 3363–3376.

Lima, F. E. D. S., Campos, F. S., Kunert Filho, H. C., Batista, H. B. D. C. R., Carnielli, P., Cibulski, S. P., *et al.* (**2013**) 'Detection of alphacoronavirus in velvety free-tailed bats (Molossus molossus) and Brazilian free-tailed bats (Tadarida brasiliensis) from urban area of Southern Brazil.', *Virus Genes.*, 47(1), pp. 164–7.

Linder, H. P. and Hardy, C. R. (**2004**) 'Evolution of the species-rich Cape flora.', *Philos Trans R Soc Lond B Biol Sci.*, 359(1450), pp. 1623–32.

Longdon, B., Brockhurst, M. A., Russel, C. A., Welch, J. J. and Jiggins, F. M. (2014) ' The evolution and genetics of virus host shifts.', *PLoS Pathog.*, 10(11), p. e1004395.

Lourenço, E. C., Costa, L. M., Silva, R. M. and Esbérard, C. E. L. (**2010**) 'Bat diversity of Ilha da Marambaia, Southern Rio de Janeiro State, Brazil (Chiroptera, Mammalia).', *Braz J Biol.,* (3), pp. 511–9.

Ludwig, B., Kraus, F. B., Allwinn, R., Doerr, H. W. and Preiser, W. (**2003**) 'Viral Zoonoses – A threat under control?', *Intervirol.*, 46(2), pp. 71–78.

Luis, A. D., Hayman, D. T. S., O'Shea, T. J., Cryan, P. M., Gilbert, A. T., Pulliam, J. R. C., *et al.* (**2013**) 'A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special?', *Proc Biol Sci.*, 280(1756), p. 20122753.

Luis, A. D., O'Shea, T. J., Hayman, D. T. S., Wood, J. L. N., Cunningham, A. A., Gilbert, A. T., *et al.* (**2015**) 'Network analysis of host-virus communities in bats and rodents reveals determinants of cross-species transmission.', *Ecol Lett.*, 18(11), pp. 1153–1162.

Mackenzie, J. S. and Jeggo, M. (**2013**) 'Reservoirs and vectors of emerging viruses.', *Curr Opin Virol.*, 3(2), pp. 170–9.

Maganga, G. D., Bourgarel, M., Obame Nkoghe, J., N'Dilimabaka, N., Drosten, C., Paupy, C., *et al.* (**2014a**) 'Identification of an unclassified paramyxovirus in Coleura afra: a potential case of host specificity.', *PloS One.*, 9(12), p. e115588.

Maganga, G. D., Bourgarel, M., Vallo, P., Dallo, T. D., Ngoagouni, C., Drexler, J. F., *et al.* (**2014b**) 'Bat distribution size or shape as determinant of viral richness in African bats.', *PloS One.*, 9(6), p. e100172.

Magee, D., Beard, R., Suchard, M. A., Lemey, P. and Scotch, M. (**2014**) 'Combining phylogeography and spatial epidemiology to uncover predictors of H5N1 influenza A virus diffusion.', *Arch Virol.*, 160(1), pp. 215-24.

Mandl, J. N., Ahmed, R., Barreiro, L. B., Daszak, P., Epstein, J. H., Virgin, H. W., et al. (2015) 'Reservoir host immune responses to emerging zoonotic viruses.', *Cell.*, 160(0), pp. 20-35.

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

Markotter, W., Randles, J., Rupprecht, C. E., Sabeta, C. T., Taylor, P. J., Wandeler, A. I., *et al.* (**2006**) 'Lagos bat virus, South Africa.', *Emerg Infect Dis.*, 12(3), pp. 504–6.

Marques, T. A, Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A, Moretti, D. J., *et al.* (**2013**) 'Estimating animal population density using passive acoustics.', *Biol Rev Camb Philos Soc.*, 88(2), pp. 287–309.

Marra, M. A., Jones, S. J. M., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S. N., *et al.* (**2003**) 'The genome sequence of the SARS-associated coronavirus.', *Science.*, 300(5624), pp. 1399–1404.

Marston, D. A., Mcelhinney, L. M., Ellis, R. J., Horton, D. L., Wise, E. L., Leech, S. L., *et al.* (**2013**) 'Next generation sequencing of viral RNA genomes.', *BMC Genom.*, 14(1), p. 444.

Mau, B., Newton, M. A. and Larget, B. (**1999**) 'Bayesian phylogenetic inference via Markov chain Monte Carlo methods.', *Biometrics.*, 55(1), pp. 1–12.

May, R. M., Gupta, S. and McLean, A. R. (**2001**) 'Infectious disease dynamics: What characterizes a successful invader?', *Philos Trans R Soc Lond B Biol Sci.*, 356(1410), pp. 901–10.

McIntosh, K., Becker, W. B. and Chanock, R. M. (**1967**) 'Growth in suckling-mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease.', *Proc Nat Acad Sci U S A.*, 58(6), pp. 2268–73.

McMichael, A. J. (**2004**) 'Environmental and social influences on emerging infectious diseases: past, present and future.', *Philos Trans R Soc Lond B Biol Sci.*, 359(1447), pp. 1049–58.

McNamara, T. S., McLean, R. G., Saito, E. K., Wolfe, P. L., Gillin, C. M., Fischer, J. R., *et al.* (**2013**) 'Surveillance of wildlife diseases: Lessons from the West Nile virus outbreak.', *Microbiol Spectr.*, 1(1).

Memish, Z. A., Mishra, N., Olival, K. J., Fagbo, S. F., Kapoor, V., Epstein, J. H., *et al.* (**2013**) 'Middle East respiratory syndrome coronavirus in bats, Saudi Arabia.', *Emerg Infect Dis.*, 19(11), pp. 1819–23.

Menachery, V. D., Yount, B. L., Debbink, K., Agnihothram, S., Gralinski, L. E., Plante, J. A., *et al.* (**2015**) 'A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence.', *Nature Med.*, 21(12), pp. 1508-13.

Menachery, V. D., Yount, B. L., Sims, A. C., Debbink, K., Agnihothram, S. S., Gralinski, L. E., *et al.* (**2016**) 'SARS-like WIV1-CoV poised for human emergence.', *Proc Natl Acad Sci.*, p. 201517719.

Meng, F., Zhu, L., Huang, W., Irwin, D. M. and Zhang, S. (**2016**) 'Bats: Body mass index, forearm mass index, blood glucose levels and SLC2A2 genes for diabetes.', *Sci Rep.*, 6(1), p. 29960.

Meslin, F. X. (**1992**) 'Surveillance and control of emerging zoonoses.', *World Health Stat Q.,* 45(2–3), pp. 200–7.

Messenger, S. L., Rupprecht, C. E. and Smith, J. S. (**2003**) 'Bats, emerging virus infections, and the rabies paradigm.', in Kunz, T. H. and Brock Fenton, M. (eds) *Bat Ecology*. The University of Chicago Press, pp. 622–79.

Meyer, A. G., Spielman, S. J., Bedford, T. and Wilke, C. O. (**2015**) 'Time dependence of evolutionary metrics during the 2009 pandemic influenza virus outbreak.', *Virus Evol.*, 1(1), pii: vev006.

Meyer, B., Müller, M. A., Corman, V. M., Reusken, C. B. E. M., Ritz, D., Godeke, G.-J., *et al.* (**2014**) 'Antibodies against MERS coronavirus in dromedary camels, United Arab Emirates, 2003 and 2013.', *Emerg Infect Dis.*, 20(4), pp. 552–559.

Mickleburgh, S. P., Hutson, A. M. and Racey, P. A. (**2002**) 'A review of the global conservation status of bats.', *Oryx.*, 36(1), pp. 18–34.

Millar, B. C. and Moore, J. E. (**2006**) 'Emerging pathogens in infectious diseases: definitions, causes and trends.', *Rev Med Microbiol.*, 17(4), pp. 101–106.

Mitchell-Jones, A. (**2004**) 'Health and safety in bat work.', in Mitchell-Jones, A. J. and McLeish, A. P. (eds) *Bat Workers' Manual*. 3rd edn, pp. 23–58.

Miura, T. and Kitaoka, M. (**1977**) 'Viruses isolated from bats in Japan.', *Arch Virol.*, 53(4), pp. 281–6.

Monadjem, A., Goodman, S. M., Stanley, W. T. and Appleton, B. (**2013**) 'A cryptic new species of Miniopterus from south-eastern Africa based on molecular and morphological characters.', *Zootaxa.*, 3746(1–4), pp. 123–42.

Monadjem, A., Taylor, P. J., Cotterill, F. P. D. and Schoeman, M. C. (**2010**) *Bats of Southern and Central Africa*. Edited by M. Mossmer. Johannesburg: Wits University Press.

Moratelli, R. and Calisher, C. H. (**2015**) 'Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses?' *Mem Inst Oswaldo Cruz.*, 110(1), pp. 1–22.

Morse, S. S. (**1991**) 'Emerging viruses: defining the rules for viral traffic.', *Perspect Biol Med.*, 34(3), pp. 387–409.

Morse, S. S. (**1993**) 'Examining the origins of emerging viruses.', in Morse, S. S. (ed.) *Emerging viruses*. New York: Oxford University Press.

Morse, S. S. (**1995**) 'Factors in the emergence of infectious diseases.', *Emerg Infect Dis.*, 1(1), pp. 7–15.

Morse, S. S. (**1997**) 'The public health threat of emerging viral disease.', *J Nutr.*, 127(5 Suppl), p. 951S–957S.

Morse, S. S. (**2012**) 'Public health surveillance and infectious disease detection.', *Biosecur Bioterror.*, 10(1), pp. 6–16.

Morse, S. S. and Schluederberg, A. (**1990**) 'Emerging viruses: the evolution of viruses and viral diseases.', *J Infect Dis.*, 162(1), pp. 1–7.

Mucina, L., Rutherford, M. and Powrie, L. (**2006**) 'Vegetation atlas of South Africa, Lesotho and Swaziland.', in Mucina, L. and Rutherford, M. (eds) *The vegetation of South Africa, Lesotho and Swaziland:* Pretoria: SANBI, p. 748v–789.

Mühldorfer, K., Speck, S., Kurth, A., Lesnik, R., Freuling, C., Müller, T., *et al.* (**2011**) 'Diseases and causes of death in European bats: dynamics in disease susceptibility and infection rates.', *PloS One.*, 6(12), p. e29773.

Müller, M. A., Paweska, J. T., Leman, P. A., Drosten, C., Grywna, K., Kemp, A., *et al.* (**2007**) 'Coronavirus antibodies in African bat species.', *Emerg Infect Dis.*, 13(9), pp. 1367–70.

Nelson, J. B. (**1952**) 'Acute hepatitis associated with mouse leukemia. II. Etiology and host range of the causal agent in mice.', *J Exp Med.*, 95(5), pp. 303–12.

Nichol, S. T., Spiropoulou, C. F., Morzunov, S., Rollin, P. E., Ksiazek, T. G., Feldmann, H., *et al.* (**1993**) 'Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness.', *Science.*, 262(5135), pp. 914–7.

Ngeyen, T. T., Kwon, H. J., Kim, I. J., Hong, S. M., Seong, W. J., Jang, J. W., *et al.* (2013) 'Multiplex nested RT-PCR for detecting avian influenza virus, infectious bronchitis virus and Newcastle disease virus.', *J Virol Methods.*, 188(1-2), pp. 41-6.

O'Connell, J. (2002) 'The basics of RT-PCR - some practical consideration.', in O'Connell, J. (ed.) *Methods in molecular biology*. vol 193. Totowa, NJ: Humana Press Inc.

O'Shea, T. J., Cryan, P. M., Cunningham, A. A., Fooks, A. R., Hayman, D. T. S., Luis, A. D., *et al.* (**2014**) 'Bat flight and zoonotic viruses.', *Emerg Infect Dis.*, 20(5), pp. 741–5.

Olival, K., Epstein, J. H., Wang, L., Field, H. E. and Daszak, P. (**2012**) 'Are bats exceptional viral reservoirs?', in Aguirre, A. A., Ostfeld, R. S., and Daszak, P. (eds) *New Directions in Conservation Medicine: Applied Cases of Ecological Health.* New York: Oxford University Press, pp. 195–212.

Olival, K. J., Hosseini, P. R., Zambrana-Torrelio, C., Ross, N., Bogich, T. L., and Daszak, P. (2017) 'Host and viral traits predict zoonotic spillover from mammals.', *Nature.*, 546(7660), pp. 646-650.

Osborne, C., Cryan, P. M., O'Shea, T. J., Oko, L. M., Ndaluka, C., Calisher, C. H., *et al.* (2011) 'Alphacoronaviruses in New World bats: Prevalence, persistence, phylogeny, and potential for interaction with humans.', *PLoS One.*, 6(5), p. e19156.

Ostfeld, R. S. (**2009**) 'Biodiversity loss and the rise of zoonotic pathogens.', *Clin Microbiol Infect.*, 15 Suppl 1, pp. 40–3.

Oxford, J. S. (**2000**) 'Influenza A pandemics of the 20th century with special reference to 1918: virology, pathology and epidemiology.', *Rev Med Virol.*, 10(2), pp. 119-33.

Panigrahy, B., Naqi, S. A. and Hall, C. F. (**1973**) 'Isolation and characterization of viruses associated with transmissible enteritis (bluecomb) of turkeys.', *Avian Dis.*, 17(2), pp. 430–8.

Parnaby, H. (**1999**) 'An interim guide to identification of insectivorous bats of south-eastern Australia.', *Tech Rep Aust Mus.,* 8, pp. 1–33..

Parry, J. (**2003**) 'SARS may have peaked in Canada, Hong Kong, and Vietnam.', *BMJ.*, 326(7396), p. 947.

Pasternak, A. O., Spaan, W. J. M. and Snijder, E. J. (**2006**) 'Nidovirus transcription: how to make sense?', *J Gen Virol.*, 87(6), pp. 1403–1421.

Paton, N. I., Leo, Y. S., Zaki, S. R., Auchus, A. P., Lee, K. E., Ling, A. E., *et al.* (1999) 'Outbreak of Nipah-virus infection among abattoir workers in Singapore.', *Lancet.*, 354(9186), pp. 1253–1256.

Pavri, K. M., Hollinger, F. B., Singh, K. R. P. and Hollinger, F. B. (**1971**) 'Isolation of a new parainfluenza virus from a frugivorous bat, rousettus leschenaulti, collected at Poona, India.', *Am J Trop Med Hyg.*, 20(1), pp. 125–130.

Paweska, J. T., Storm, N., Grobbelaar, A. A., Markotter, W., Kemp, A. and Jansen van Vuren, P. (**2016**) 'Experimental inoculation of Egyptian fruit bats (Rousettus aegyptiacus) with Ebola virus.', *Viruses.*, 8(2), p. 29.

Peiris, J. and Poon, L. (**2009**) 'Coronaviruses and Toroviruses.', in Zuckerman, A. et al. (eds) *Principles and Practice of Clinical Virology*. 5th edn. John Wiley & Sons Ltd, pp. 511–531.

Penrith, M., Vosloo, W., Jori, F., Bastos, A. D. S. (2013) 'African swine fever eradication in Africa.', *Virus Res.*, 173(1), pp.228-246.

Perlman, S. and Netland, J. (**2009**) 'Coronaviruses post-SARS: Update on replication and pathogenesis.', *Nat Rev Microbiol.*, 7(6), pp. 439–450.

Pfefferle, S., Oppong, S., Drexler, J. F., Gloza-Rausch, F., Ipsen, A., Seebens, A., *et al.* (2009) 'Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana.', *Emerg Infect Dis.*, 15(9), pp. 1377–84.

Pike, B. L., Saylors, K. E., Fair, J. N., LeBreton, M., Tamoufe, U., Djoko, C. F., *et al.* (2010) 'The origin and prevention of pandemics.', *Clin Infect Dis.*, 50(12), pp. 1636–1640.

Plowright, R. K., Eby, P., Hudson, P. J., Smith, I. L., Westcott, D., Bryden, W. L., *et al.* (2014) 'Ecological dynamics of emerging bat virus spillover.', *Proc Biol Sci.*, 282(1798), pp. 20142124–20142124.

Plowright, R. K., Field, H. E., Smith, C., DivIjan, A., Palmer, C., Tabor, G., *et al.* (2008) 'Reproduction and nutritional stress are risk factors for Hendra virus infection in little red flying foxes (Pteropus scapulatus).', *Proc. Biol Sci.*, 275(1636), pp. 861–9.

Plowright, R. K., Peel, A. J., Streicker, D. G., Gilbert, A. T., McCallum, H., Wood, J., *et al.* (2016) 'Transmission or within-host dynamics driving pulses of zoonotic viruses in reservoir–host populations.', *PLoS Negl Trop Dis.*, 10(8), p. e0004796.

Poon, L. L. M., Chu, D. K. W., Chan, K. H., Wong, O. K., Ellis, T. M., Leung, Y. H. C., *et al.* (2005) 'Identification of a novel coronavirus in bats.', *J Virol.*, 79(4), pp. 2001–9.

Poutanen, S. M., Low, D. E., Henry, B., Finkelstein, S., Rose, D., Green, K., *et al.* (2003) 'Identification of severe acute respiratory syndrome in Canada.', *N Engl J Med.*, 348(20), pp. 1995–2005.

PREDICT Consortium (2014) '*Reducing pandemic risk, promoting global health*.', One Health Institute, Davis: University of California.

Preiser, W. (2012) 'SARS: A case study for factors driving disease emergence.', in Aguirre, A. A., Daszak, P., and Ostfeld, R. (eds) *New Directions in Conservation Medicine: Applied Cases of Ecological Health.* New York: Oxford University Press.

Priestnall, S. L., Mitchell, J. A., Walker, C. A., Erles, K. and Brownlie, J. (2014) 'New and emerging pathogens in canine infectious respiratory disease.', *Vet Pathol.*, 51(2), pp. 492–504.

Promkuntod, N. (2015) 'Dynamics of avian coronavirus circulation in commercial and non-commercial birds in Asia - a review.', *Vet Q.*, 36(1), pp. 1–24..

Quan, P., Firth, C., Street, C., Henriquez, J. A., Petrosov, A., Tashmukhamedova, A., *et al.* (2010) 'Identification of a severe acute respiratory syndrome coronavirus-like virus in a leaf-nosed bat in Nigeria.', *MBio.*, 1(4), pp. 1–9.

Racey, P. (2009) 'Reproductive assessment in bats.', in Kunz, T. (ed.) *Ecological and behavioral methods for the study of bats*. Smithsonian Institute Press.

Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H. W., Müller, M. a, Dijkman, R., *et al.* (2013) 'Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC.', *Nature.*, 495(7440), pp. 251–4.

Razanajatovo, N. H., Nomenjanahary, L. A., Wilkinson, D. A., Razafimanahaka, J. H., Goodman, S. M., Jenkins, R. K., *et al.* (2015) 'Detection of new genetic variants of betacoronaviruses in endemic frugivorous bats of Madagascar.', *Virol J.,* 12(1), p. 42.

Reagan, R. L. and Brueckner, A. L. (1952) 'Electron microscope studies of four strains of infectious bronchitis virus.', *Am J Vet Res.*, 13(48), pp. 417–8.

Reagan, R. L. and Hauser, J. E. (1948) 'Electron micrograph of the virus of infectious bronchitis of chickens.', *Cornell Vet.*, 38(2), p. 190.

Reed, K. D., Meece, J. K., Henkel, J. S. and Shukla, S. K. (2003) 'Birds, migration and emerging zoonoses: west nile virus, lyme disease, influenza A and enteropathogens.', *Clin Med Res.*, 1(1), pp. 5–12.

Reeves, T., Samy, A. M. and Peterson, A. T. (2015) 'MERS-CoV geography and ecology in the Middle East: analyses of reported camel exposures and a preliminary risk map.', *BMC Res Notes.*, 8(1), p. 801.

Ren, W., Li, W., Yu, M., Hao, P., Zhang, Y., Zhou, P., *et al.* (2006) 'Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis.', *J Gen Virol.*, 87(Pt 11), pp. 3355–9.

Reusken, C. B. E. M., Haagmans, B. L., Müller, M. A, Gutierrez, C., Godeke, G.-J., Meyer, B., *et al.* (2013) 'Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study.', *Lancet Infect Dis.*, 13(10), pp. 859–66.

Reusken, C. B. E. M., Lina, P. H. C., Pielaat, A., de Vries, A., Dam-Deisz, C., Adema, J., *et al.* (2010) 'Circulation of group 2 coronaviruses in a bat species common to urban areas in Western Europe.', *Vector Borne Zoonotic Dis.*, 10(8), pp. 785–91.

Reusken, C. B., Raj, V. S., Koopmans, M. P. and Haagmans, B. L. (2016) 'Cross host transmission in the emergence of MERS coronavirus.', *Curr Opin Virol.*, 16, pp. 55–62.

Reusken, C., Farag, E., Jonges, M., Godeke, G., El-Sayed, A., Pas, S., *et al.* (2014) 'Middle East respiratory syndrome coronavirus (MERS-CoV) RNA and neutralising antibodies in milk collected according to local customs from dromedary camels, Qatar, April 2014.', *Euro Surveill.*, 19(23).

Ritchie, A. E., Deshmukh, D. R., Larsen, C. T. and Pomeroy, B. S. (1973) 'Electron microscopy of coronavirus-like particles characteristic of turkey bluecomb disease.', *Avian Dis.*, 17(3), pp. 546–58.

Ronquist, F. and Huelsenbeck, J. P. (2003) 'MrBayes 3: Bayesian phylogenetic inference under mixed models.', *Bioinformatics.*, 19(12), pp. 1572–4.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., *et al.* (2012) 'MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space.', *Syst Biol.*, 61(3), pp. 539–542.

Rosenthal, S. R., Ostfeld, R. S., McGarvey, S. T., Lurie, M. N. and Smith, K. F. (2015) 'Redefining disease emergence to improve prioritization and macro-ecological analyses.', *One Health.*, 1, pp. 17–23.

Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., *et al.* (2003) 'Characterization of a novel coronavirus associated with severe acute respiratory syndrome.', *Science.*, 300(5624), pp. 1394–1399.

Rowlands, R. J., Michaud, V., Heath, L., Hutchings, G., Oura, C., Vosloo, W., *et al.* (2008) 'African swine fever isolate, Georgia, 2007.', Emerg Infect Dis., 14(12), pp. 1870-74.

Rutherford, M., Mucina, L. and Powrie, L. (2006) 'Biomes and bioregions of Southern Africa.', in Mucina, L. and Rutherford, M. . (eds) *The vegetation of South Africa, Lesotho and Swaziland*. Pretoria: SANBI, pp. 30–51.

Saif, L. J. (2010) 'Bovine respiratory coronavirus.', *Vet Clin North Am Food Anim Pract.*, 26(2), pp. 349–64.

Saitou, N. and Nei, M. (1987) 'The neighbor-joining method: a new method for reconstructing phylogenetic trees.', *Mol Biol Evol.*, 4(4), pp. 406–25.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) 'DNA sequencing with chain-terminating inhibitors.', *Proc Natl Acad Sci U S A.*, 74(12), pp. 5463–7.

Schalk, A. F. and Hawn, M. G. (1931) 'An apparently new respiratory disease of baby chicks.', *J Am Vet Med Assoc.*, 75, pp. 413–423.

Schoeman, M. C. and Jacobs, D. S. (2008) 'The relative influence of competition and prey defenses on the phenotypic structure of insectivorous bat ensembles in Southern Africa.', *PLoS One.*, 3(11), p. e3715.

Schoeman, M. C. and Jacobs, D. S. (2011) 'The relative influence of competition and prey defences on the trophic structure of animalivorous bat ensembles.', *Oecologia.*, 166(2), pp. 493–506.

Schrag, S. J. and Wiener, P. (1995) 'Emerging infectious disease: what are the relative roles of ecology and evolution?', *Trends Ecol Evol.*, 10(8), pp. 319–24.

Seltmann, A., Corman, V. M., Rasche, A., Drosten, C., Czirják, G. Á., Bernard, H., *et al.* (2017) 'Seasonal fluctuations of astrovirus, but not coronavirus shedding in bats inhabiting human-modified tropical forests.', *Ecohealth.*, 14(2), pp. 272–284.

Seto, W. H., Tsang, D., Yung, R. W. H., Ching, T. Y., Ng, T. K., Ho, M., *et al.* (2003) 'Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS).', *Lancet.*, 361(9368), pp. 1519–20.

Shekaili, T. Al, Clough, H., Ganapathy, K. and Baylis, M. (2015) 'Sero-surveillance and risk factors for avian influenza and Newcastle disease virus in backyard poultry in Oman.', *Prev Vet Med.*, 122(1–2), pp. 145–153.

Shirato, K., Maeda, K., Tsuda, S., Suzuki, K., Watanabe, S., Shimoda, H., *et al.* (2012) 'Detection of bat coronaviruses from Miniopterus fuliginosus in Japan.', *Virus Genes.*, 44(1), pp. 40–4.

Si, Y., de Boer, W. F. and Gong, P. (2013) 'Different environmental drivers of highly pathogenic avian Influenza H5N1 outbreaks in poultry and wild birds.', *PLoS One.*, 8(1), p. e53362.

Sikes, R. S. and Gannon, W. L. (2011) 'Guidelines of the American Society of Mammalogists for the use of wild mammals in research.', *J Mammal.*, 92(1), pp. 235–253.

Silander, J. A. (2001) 'Temperate Forests.', in Simon Levin (ed.) *Encyclopedia of Biodiversity*. Violume 5. Acadmic Press.

Simmons, N. B. (2005) 'Order Chiroptera.', in Wilson, D. E. and Reeder, D. M. (eds) *Mammal Species of the World: A Taxonomic and Geographic Reference*. 3rd edn. Baltimore: The Johns Hopkins University Press, pp. 312–529.

Singh, R. K., Dhama, K., Malik, Y. S., Ramakrishnan, M. A., Karthik, K., Khandia, R., *et al.* (2017) 'Ebola virus – epidemiology, diagnosis, and control: threat to humans, lessons learnt, and preparedness plans – an update on its 40 year's journey.', *Vet Q.*, 37(1), pp. 98–135.

Sleator, R. D. (2013) 'A beginner's guide to phylogenetics.', *Microb Ecol.*, 66(1), pp. 1–4.

Smith, C. S., de Jong, C. E., Meers, J., Henning, J., Wang, L.-F. and Field, H. E. (2016) 'Coronavirus infection and diversity in bats in the Australasian region.', *Ecohealth.*, 13(1), pp. 72–82.

Song, D. and Park, B. (2012) 'Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines.', *Virus Genes.*, 44(2), pp. 167–175.

Soto, S. M. (2009) 'Human migration and infectious diseases.', *Clin Microbiol Infect.,* 15 Suppl 1, pp. 26–8.

de Souza Luna, L. K., Heiser, V., Regamey, N., Panning, M., Drexler, J. F., Mulangu, S., *et al.* (2007) 'Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray.', *J Clin Microbiol.*, 45(3), pp. 1049–52.

Stephensen, C. B., Casebolt, D. B. and Gangopadhyay, N. N. (1999) 'Phylogenetic analysis of a highly conserved region of the polymerase gene from 11 coronaviruses and development of a consensus polymerase chain reaction assay.', *Virus Res.*, 60(2), pp. 181–9.

Suzuki, J., Sato, R., Kobayashi, T., Aoi, T. and Harasawa, R. (2014) 'Group B betacoronavirus in rhinolophid bats, Japan.', *J Vet Med Sci.*, 76(9), pp. 1267–9.

Tamura, K. and Nei, M. (1993) 'Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.', *Mol Biol Evol.*, 10(3), pp. 512–26.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) 'MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.', *Mol Biol Evol.*, 30(12), pp. 2725–9.

Tan, S. C. and Yiap, B. C. (**2009**) 'DNA, RNA, and protein extraction: The past and the present.', *J Biomed Biotechnol.*, 2009, pp. 1–10.

Tang, X. C., Zhang, J. X., Zhang, S. Y., Wang, P., Fan, X. H., Li, L. F., et al. (**2006**) 'Prevalence and genetic diversity of coronaviruses in bats from China.', *J Virol.*, 80(15), pp. 7481–90.

Tao, Y., Shi, M., Chommanard, C., Queen, K., Zhang, J., Markotter, W., et al. (**2017**) 'Surveillance of bat coronaviruses in Kenya identifies relatives of human coronaviruses NL63 and 229E and their recombination history.', *J Virol.*, 91(5), pp. e01953-16.

Tao, Y., Tang, K., Shi, M., Conrardy, C., Li, K. S. M., Lau, S. K. P., et al. (**2012**) 'Genomic characterization of seven distinct bat coronaviruses in Kenya.', *Virus Res.*, 167(1), pp. 67–73.

Taylor, L. H., Latham, S. M. and Woolhouse, M. E. (**2001**) 'Risk factors for human disease emergence.', *Philos Trans R Soc Lond B Biol Sci.*, 356(1411), pp. 983–9.

Tekes, G. and Thiel, H.-J. (**2016**) 'Feline coronaviruses: Pathogenesis of feline infectious peritonitis.', *Adv Virus Res.*, volume 96 pp. 193–218.

Tong, S., Chern, S.-W. W., Li, Y., Pallansch, M. A. and Anderson, L. J. (**2008**) 'Sensitive and broadly reactive reverse transcription-PCR assays to detect novel paramyxoviruses.', *J Clin Microbiol.*, 46(8), pp. 2652–8.

Tong, S., Conrardy, C., Ruone, S., Kuzmin, I. V, Guo, X., Tao, Y., et al. (**2009**) 'Detection of novel SARS-like and other coronaviruses in bats from Kenya.', *Emerg Infect Dis.*, 15(3), pp. 482–5.

Tong, S., Li, Y., Rivailler, P., Conrardy, C., Castillo, D. A. A., Chen, L.-M., et al. (**2012**) 'A distinct lineage of influenza A virus from bats.', *Proc Natl Acad Sci U S A.*, 109(11), pp. 4269–74.

Trombetta, H., Faggion, H. Z., Leotte, J., Nogueira, M. B., Vidal, L. R. R. and Raboni, S. M. (**2016**) 'Human coronavirus and severe acute respiratory infection in Southern Brazil.', *Pathog Glob Health.*, 110(3), pp. 113–118.

Tsuda, S., Watanabe, S., Masangkay, J. S., Mizutani, T., Alviola, P., Ueda, N., et al. (**2012**) 'Genomic and serological detection of bat coronavirus from bats in the Philippines.', *Arch Virol.*, 157(12), pp. 2349–55.

Tu, C., Crameri, G., Kong, X., Chen, J., Sun, Y., Yu, M., et al. (**2004**) 'Antibodies to SARS coronavirus in civets', *Emerg Infect Dis.*, 10(12), pp. 2244–2248.

Tumlin, J. T., Pomeroy, B. S. and Lindorfer, R. K. (**1957**) 'Bluecomb disease of turkeys. IV. Demonstration of a filterable agent.', *J Am Vet Med Assoc.*, 130(8), pp. 360–5.

Turmelle, A. S. and Olival, K. J. (**2009**) 'Correlates of viral richness in bats (order Chiroptera).', *Ecohealth.*, 6(4), pp. 522–39.

Tyrrell, D. A., Alexander, D. J., Almeida, J. D., Cunningham, C. H., Easterday, B. C., Garwes, D. J., et al. (**1978**) 'Coronaviridae: second report.', *Intervirology.*, 10(6), pp. 321–8.

Tyrrell, D. A., Almeida, J. D., Cunningham, C. H., Dowdle, W. R., Hofstad, M. S., McIntosh, K., et al. (**1975**) 'Coronaviridae.', *Intervirology.*, 5(1–2), pp. 76–82.

Tyrrell, D. A. and Bynoe, M. L. (**1965**) 'Cultivation of a novel type of common-cold virus in organ cultures.', *Br Med J.*, 1(5448), pp. 1467–70.

Tyrrell, D. A., Bynoe, M. L. and Hoorn, B. (**1968**) 'Cultivation of difficult viruses from patients with common colds.', *Brit Med J.*, 1(5592), pp. 606–10.

Vijaykrishna, D., Smith, G. J. D., Zhang, J. X., Peiris, J. S. M., Chen, H. and Guan, Y. (**2007**) 'Evolutionary insights into the ecology of coronaviruses.', *J Virol.*, 81(8), pp. 4012–20.

Vijgen, L., Keyaerts, E., Lemey, P., Maes, P., Van Reeth, K., Nauwynck, H., *et al.* (2006) 'Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43.', *J Virol.*, 80(14), pp. 7270–4.

Vijgen, L., Keyaerts, E., Moes, E., Thoelen, I., Wollants, E., Lemey, P., *et al.* (2005) 'Complete genomic sequence of human coronavirus OC43: Molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event.', *J Virol.*, 79(3), pp. 1595–1604.

Vu, H. T., Leitmeyer, K. C., Le, D. H., Miller, M. J., Nguyen, Q. H., Uyeki, T. M., *et al.* (2004) 'Clinical description of a completed outbreak of SARS in Vietnam, February-May 2003.', *Emerg Infect Dis.*, 10(2), pp. 334–8.

Wacharapluesadee, S., Duengkae, P., Rodpan, A., Kaewpom, T., Maneeorn, P., Kanchanasaka, B., *et al.* (2015) 'Diversity of coronavirus in bats from Eastern Thailand.', *Virol J.*, 12(1), p. 57.

Wacharapluesadee, S., Sintunawa, C., Kaewpom, T., Khongnomnan, K., Olival, K. J., Epstein, J. H., *et al.* (2013) 'Group C betacoronavirus in bat guano fertilizer, Thailand.', *Emerg Infect Dis.*, 19(8), pp. 1349–51.

Waggoner, J. J., Abeynayake, J., Sahoo, M. K., Gresh, L., Tellez, Y., Gonzales, K., *et al.* (2013) 'Single-reaction, multiplex, real-time rt-PCR for the detection, quantitation, and serotyping of dengue viruses.', PLoS Negl Trop Dis., 7(4), p. e2116.

Wang, L.-F. and Crameri, G. (2014) 'Emerging zoonotic viral diseases.', *Rev. Sci. Tech.,* 33(2), pp. 569–581.

Wang, L. F. and Eaton, B. T. (2007) 'Bats, civets and the emergence of SARS.', *Curr Top Microbiol Immunol.*, 315, pp. 325–44.

Wang, L., Fu, S., Cao, Y., Zhang, H., Feng, Y., Yang, W., *et al.* (2017) 'Discovery and genetic analysis of novel coronaviruses in least horseshoe bats in southwestern China.', *Emerg Microbes Infect.*, 6(3), p. e14.

Wang, M. and Hu, Z. (2013) 'Bats as animal reservoirs for the SARS coronavirus: Hypothesis proved after 10 years of virus hunting.', *Virol Sin.*, 28(6), pp. 315–7..

Wang, Q., Qi, J., Yuan, Y., Xuan, Y., Han, P., Wan, Y., *et al.* (2014) 'Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26.', *Cell Host Microbe.*, 16(3), pp. 328–337.

Wang, R., and Taubenberger, J. K. (2010) 'Methods for molecular surveillance of influenza.', *Expert Rev Anti Infect Ther.*, 8(5), pp. 517-527.

Waruhiu, C., Ommeh, S., Obanda, V., Agwanda, B., Gakuya, F., Ge, X.-Y., *et al.* (2017) 'Molecular detection of viruses in Kenyan bats and discovery of novel astroviruses, caliciviruses and rotaviruses.', *Virol Sin.*, 32(2), pp. 101-114.

Watanabe, S., Masangkay, J. S., Nagata, N., Morikawa, S., Mizutani, T., Fukushi, S., *et al.* (2010) 'Bat coronaviruses and experimental infection of bats, the Philippines.', *Emerg Infect Dis.,* 16(8), pp. 1217–23.

Weaver, R. F. (2008) 'Molecular tools for studying genes and gene activity.', in *Molecular Biology*. New York: McGraw-Hill, pp. 83–125.

Weiss, S., Witkowski, P. T., Auste, B., Nowak, K., Weber, N., Fahr, J., *et al.* (2012) 'Hantavirus in bat, Sierra Leone.', *Emerg Infect Dis.*, 18(1), pp. 159–161.

Wernery, U., Rasoul, Ih. El, Wong, E. Y., Joseph, M., Chen, Y., Jose, S., *et al.* (2015) 'A phylogenetically distinct Middle East respiratory syndrome coronavirus detected in a dromedary calf from a closed dairy herd in Dubai with rising seroprevalence with age.', *Emerg Microbes Infect.*, 4(12), p. e74.

Whelan, S. and Goldman, N. (2001) 'A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach.', *Mol Biol Evol.*, 18(5), pp. 691–9.

WHO, World Health Organisation. (**2004**) *Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003.* Available at: http://www.who.int/csr/sars/country/%0Atable2004_04_21/en/. (Accessed: 06 July 2017).

WHO, World Health Organisation. (**2015**) *Middle East respiratory syndrome coronavirus (MERS-CoV) – Republic of Korea*, *Disease outbreak news.* Available at: http://www.who.int/csr/don/25-october-2015-mers-korea/en/. (Accessed: 06 July 2017).

WHO, World Health Organisation. (**2017**) *Middle East respiratory syndrome coronavirus (MERS-CoV) – Saudi Arabia.* Available at: http://www.who.int/emergencies/mers-cov/en/. (Accessed: 06 July 2017).

Wilcox, B. and Kueffer, C. (**2008**) 'Transdisciplinarity in EcoHealth: Status and future prospects.', *Ecohealth.*, 5(1), pp. 1–3.

de Wit, E., van Doremalen, N., Falzarano, D. and Munster, V. J. (**2016**) 'SARS and MERS: recent insights into emerging coronaviruses.', *Nat Rev Microbiol.*, 14(8), pp. 523–534.

Witkowski, P. T., Drexler, J. F., Kallies, R., Ličková, M., Bokorová, S., Mananga, G. D., *et al.* (2016**)** 'Phylogenetic analysis of a newfound bat-borne hantavirus supports a laurasiatherian host association for ancestral mammalian hantaviruses.', *Infect, Genet Evol.*, 41, pp. 113–119.

Wong, S., Lau, S., Woo, P. and Yuen, K. (**2007**) 'Bats as a continuing source of emerging infections in humans.', *Rev Med Virol.*, 17(2), pp. 67–91.

Woo, P. C. Y., Lau, S. K. P., Li, K. S. M., Poon, R. W. S., Wong, B. H. L., Tsoi, H., *et al.* (2006**)** 'Molecular diversity of coronaviruses in bats.', *Virology.*, 351(1), pp. 180–7.

Woo, P., Huang, Y., Lau, S. K. P., Tsoi, H. and Yuen, K. (**2005a**) 'In silico analysis of ORF1ab in coronavirus HKU1 genome reveals a unique putative cleavage site of coronavirus HKU1 3C-like protease.', *Microbiol Immunol.*, 49(10), pp. 899–908.

Woo, P., Lau, S. K., Li, K. S., Tsang, A. K. and Yuen, K.-Y. (**2012a**) 'Genetic relatedness of the novel human group C betacoronavirus to Tylonycteris bat coronavirus HKU4 and Pipistrellus bat coronavirus HKU5.', *Emerg Microbes Infect.*, 1(11), p. e35.

Woo, P., Lau, S. K. P., Chu, C., Chan, K., Tsoi, H., Huang, Y., *et al.* (2005b**)** 'Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia.', *J Virol.*, 79(2), pp. 884–95.

Woo, P., Lau, S. K. P., Huang, Y. and Yuen, K.-Y. (**2009**) 'Coronavirus diversity, phylogeny and interspecies jumping.', *Exp Biol Med.,* 234(10), pp. 1117–27.

Woo, P., Lau, S., Lam, C., Lau, C., Tsang, A., Lau, J., *et al.* (2012b**)** 'Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus.', *J Virol.*, 86(7), pp. 3995–4008.

Woo, P., Wang, M., Lau, S. K. P., Xu, H., Poon, R. W. S., Guo, R., *et al.* (2007**)** 'Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features.', *J Virol.*, 81(4), pp. 1574–85.

Wood, J. L. N., Leach, M., Waldman, L., Macgregor, H., Fooks, A. R., Jones, K. E., *et al.* (2012**)** 'A framework for the study of zoonotic disease emergence and its drivers: spillover of bat pathogens as a case study.', *Philos Trans R Soc Lond B Biol Sci.*, 367(1604), pp. 2881–92.

Woolhouse, M. E. J. (**2001**) 'Population biology of multihost pathogens.', *Science.*, 292(5519), pp. 1109–1112..

Woolhouse, M. E. J. (**2002**) 'Population biology of emerging and re-emerging pathogens.', *Trends Microbiol.*, 10(10 Suppl), pp. S3-7.

Woolhouse, M. E. J. and Dye, C. (**2001**) 'Preface.', *Philos Trans R Soc B Biol Sci.*, 356(1411), pp. 981–982.

Woolhouse, M. E. J. and Gowtage-Sequeria, S. (**2005**) 'Host range and emerging and reemerging pathogens.', *Emerg Infect Dis.*, 11(12), pp. 1842–7.

Woolhouse, M. E. J., Haydon, D. T. and Antia, R. (**2005**) 'Emerging pathogens: the epidemiology and evolution of species jumps.', *Trends Ecol Evol.*, 20(5), pp. 238–244.

Woźniakowski, G., Frączyk, M., Niemczuk, K., Pejsak, Z. (2016) 'Selected aspects related to epidemiology, pathogenesis, immunity, and control of African swine fever.', *J Vet Res.*, 60(2), pp. 119-125.

Wu, Y., Wu, Y., Tefsen, B., Shi, Y. and Gao, G. F. (**2014**) 'Bat-derived influenza-like viruses H17N10 and H18N11.', *Trends Microbiol.*, 22(4), pp. 183–191.

Wu, Z., Yang, L., Ren, X., He, G., Zhang, J., Yang, J., *et al.* (**2016**) 'Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases.', *ISME J.*, 10(3), pp. 609–620.

Xu, L., Wu, J., He, B., Qin, S., Xia, L., Qin, M., *et al.* (**2015**) 'Novel hantavirus identified in black-bearded tomb bats, China.', *Infect Genet Evol.*, 31, pp. 158-60.

Yadav, P., Sarkale, P., Patil, D., Shete, A., Kokate, P., Kumar, V., *et al.* (**2016**) 'Isolation of Tioman virus from Pteropus giganteus bat in North-East region of India.', *Infect Genet Evol.,* 45, pp. 224–229.

Yang, J. Y., Brooks, S., Meyer, J. A., Blakesley, R. R., Zelazny, A. M., Segre, J. A., *et al.* (**2013a**) 'Pan-PCR, a computational method for designing bacterium-typing assays based on whole-genome sequence data.', *J Clin Microbiol.*, 51(3), pp. 752–758.

Yang, L., Wu, Z., Ren, X., Yang, F., He, G., Zhang, J., *et al.* (**2013b**) 'Novel SARS-like betacoronaviruses in bats, China, 2011.', *Emerg Infect Dis.*, 19(6), pp. 989–91.

Yang, L., Wu, Z., Ren, X., Yang, F., Zhang, J., He, G., *et al.* (**2014**) 'MERS-related betacoronavirus in Vespertilio superans bats, China.', *Emerg Infect Dis.*, 20(7).

Yang, S. and Rothman, R. E. (**2004**) 'PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings.', *Lancet Infect Dis.*, 4(6), pp. 337–348.

Yang, X.-L., Hu, B., Wang, B., Wang, M.-N., Zhang, Q., Zhang, W., *et al.* (**2015**) 'Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus.', *J Virol.*, 90(6), pp. 3253–6.

Yang, Y., Zhang, L., Geng, H., Deng, Y., Huang, B., Guo, Y., *et al.* (**2013c**) 'The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists.', *Protein Cell.*, 4(12), pp. 951–61.

Yang, Z. and Rannala, B. (**1997**) 'Bayesian phylogenetic inference using DNA sequences: A Markov Chain Monte Carlo Method.', *Mol Biol Evol.*, 14(7), pp. 717–24.

Yob, J. M., Field, H., Rashdi, A. M., Morrissy, C., van der Heide, B., Rota, P., *et al.* (**2001**) 'Nipah virus infection in bats (order Chiroptera) in peninsular Malaysia.', *Emerg Infect Dis.*, 7(3), pp. 439–41.

Yuan, J., Hon, C.-C., Li, Y., Wang, D., Xu, G., Zhang, H., *et al.* (**2010**) 'Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans.', *J Gen Virol.*, 91(Pt 4), pp. 1058–62.

Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. and Fouchier, R. A. M. (**2012**) 'Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia.', *N Engl J Med.*, 367(19), pp. 1814–20.

Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K. A., Fang, X., *et al.* (**2013**) 'Comparative analysis of bat genomes provides insight into the evolution of flight and immunity.', *Science.,* 339(6118), pp. 456–60.

Zhao, Z., Zhang, F., Xu, M., Huang, K., Zhong, W., Cai, W., *et al.* (**2003**) 'Description and clinical treatment of an early outbreak of severe acute respiratory syndrome (SARS) in Guangzhou, PR China.', *J Med Microbiol.*, 52(8), pp. 715–720.

Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, Xie, Z. H., Chan, K. H., *et al.* (**2003**) 'Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003.', *Lancet.*, 362(9393), pp. 1353–8.

Zhou, M. Y. and Gomez-Sanchez, C. E. (**2000**) 'Universal TA cloning.', *Cur Issues Mol Biol.*, 2(1), pp. 1–7.

APPENDIX A: Ethics

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot • your knowledge partner

2 March 2012

Prof W Preiser
Department of Pathology
Faculty of Health Science
Tygerberg Campus
Stellenbosch University

E-MAILED and **MAILED**

Dear Prof Preiser
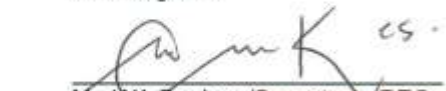
Application for Ethical Clearance:
Catching and non-invasive sampling of bats
Ref: SU-ACUM12-00001

Your application for ethics clearance has been approved by the Research Ethics Committee: Animal Care and Use (REC: ACU). Please note that this clearance is only valid for a period of twelve months. Ethics clearance of protocols spanning more than one year must be renewed annually through submission of a progress report, up to a maximum of three years.

Applicants are reminded that they are expected to comply with accepted standards for the use of animals in research and teaching as reflected in the South African National Standards 10386: 2008. The SANS 10386: 2008 document is available on the Division for Research Development's website www.sun.ac.za/research .

Please feel free to contact me if any additional information is needed.

Kind regards

Mr. WA Beukes (Secretary: REC: ACU)

APPROVED
0 2 MAR 2012

DIE UNIVERSITEIT VAN STELLENBOSCH
0 2 MAR 2012
THE UNIVERSITY OF STELLENBOSCH

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot • your knowledge partner

Protocol Approval

Date: 15-Jun-2016

PI Name: Ithete, Ndapewa L

Protocol #: SU-ACUD16-00008

Title: Investigation of novel bat-borne viruses with zoonotic potential in South Africa

Dear Ndapewa Ithete, the Notification, was reviewed on 06-Jun-2016 by the Research Ethics Committee: Animal Care and Use via committee review procedures and was approved. Please note that this clearance is only valid for a period of twelve months. Ethics clearance of protocols spanning more than one year must be renewed annually through submission of a progress report, up to a maximum of three years.

Applicants are reminded that they are expected to comply with accepted standards for the use of animals in research and teaching as reflected in the South African National Standards 10386: 2008. The SANS 10386: 2008 document is available on the Division for Research Developments website www.sun.ac.za/research.

As provided for in the Veterinary and Para-Veterinary Professions Act, 1982. It is the principal investigator's responsibility to ensure that all study participants are registered with or have been authorised by the South African Veterinary Council (SAVC) to perform the procedures on animals, or will be performing the procedures under the direct and continuous supervision of a SAVC-registered veterinary professional or SAVC-registered para-veterinary professional, who are acting within the scope of practice for their profession.

Please remember to use your protocol number, SU-ACUD16-00008 on any documents or correspondence with the REC: ACU concerning your research protocol.

Please note that the REC: ACU has the prerogative and authority to ask further questions, seek additional information, require further modifications or monitor the conduct of your research.

Any event not consistent with routine expected outcomes that results in any unexpected animal welfare issue (death, disease, or prolonged distress) or human health risks (zoonotic disease or exposure, injuries) must be reported to the committee, by creating an Adverse Event submission within the system.

We wish you the best as you conduct your research.

If you have any questions or need further help, please contact the REC: ACU secretariat at or .

Sincerely,

REC: ACU Secretariat

Research Ethics Committee: Animal Care and Use

APPENDIX B: Primers used for extended betacoronavirus genome amplification

Table 1 Primers used to generate additional partial betacoronavirus genome sequences.

Primer pair	PCR Protocol	Forward primer 5' → 3'	Reverse primer 5' → 3'
08Fa|10R	SuperFI	modified: TTGAAACTGTTGTGGGTCAAT	Corman *et al.* (2013)
10Fn|11R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
10Fn|B_R1	SuperFI	Corman *et al.* (2013)	Dr T Suliman: CGAAGACTAGCCCTGTTGAAC
12Fn|12R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
12Fn|13R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
12Fn|14R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
12Fn|B_R1	SuperFI	Corman *et al.* (2013)	Dr T Suliman: CGAAGACTAGCCCTGTTGAAC
14F|17Ra	SuperFI	Corman *et al.* (2013)	modified: AGGGCGCAACTTATCCATAC
14F|14R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
14F|15Ra	SuperFI	Corman *et al.* (2013)	modified: ACTCAGCCAACAACACTACA
15F|16Fn	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
16Fn|17Ra	SuperFI	Corman *et al.* (2013)	modified: AGGGCGCAACTTATCCATAC
19F|19R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
22Fn|22R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
22Fn|23R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
33Fn|35R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
37Fn|39R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
37Fn|40R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
41Fn|44R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
52Fn|52R	Corman	Corman *et al.* (2013)	Corman *et al.* (2013)
53Fa|F_R1	SuperFI	modified: GAAGAGTTTAATGTCACAGCAGA	Dr T Suliman: AAGAATAGCACCCAGTTGCC
58Fn|58R	Corman	Corman *et al.* (2013)	Corman *et al.* (2013)
59F|61R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
59Fn|59R	Corman	Corman *et al.* (2013)	Corman *et al.* (2013)
60Fn|60R	Corman	Corman *et al.* (2013)	Corman *et al.* (2013)
61Fn|64Ra	SuperFI	Corman *et al.* (2013)	modified: CCTACAAATACTGCGACGAC
61Fn|G_R	SuperFI	Corman *et al.* (2013)	T Suliman: AACAGGAACCAGTTGTACAG
63F|66R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
63Fa|65R	SuperFI	modified: ATGGCTGTGGAACTCGTTT	Corman *et al.* (2013)
63Fa|67R	SuperFI	modified: ATGGCTGTGGAACTCGTTT	Corman *et al.* (2013)
67F|69Ra	SuperFI	Corman *et al.* (2013)	modified: GCTCAAGAAGCTCCAACCAT
67F|67R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
68Fn|68Ra	SuperFI	Corman *et al.* (2013)	modified: CCCTCTTCATGGACCCAAAT
70F|70R	SuperFI	Corman *et al.* (2013)	Corman *et al.* (2013)
70Fn|70R	Corman	Corman *et al.* (2013)	Corman *et al.* (2013)
G_F|G_R1	SuperFI	Dr T Suliman: AGTATAAGCGTGAAGTTTTT	Dr T Suliman: AAAGCAGCTACATAACCGCC
50F|52R	Corman	Corman *et al.* (2013)	Corman *et al.* (2013)
S1892|S52511	SuperFI	N Cronje: CAGGCACGGGTGTTTTTC	N Cronje: AACTGAGTCATCTTGGCGCA

APPENDIX C: Lists and tables relating to data collected

List 1: species and their corresponding abbreviated species code

Species code	Species name	Species code	Species name
AL	*Artibeus lituratus*	MN	*Minipterus natalensis*
C	*Caerephon sp.*	MNI	*Myotis nigricans*
CB	*Cynopterus brachyotis*	MR	*Myotis ricketti*
CC	*Cardioderma cor*	MRU	*Molossus rufus*
CP	*Chaerephon pumilus*	MS	*Miniopterus schreibersii*
CPE	*Carollia perspicillata*	MT	*Myotis tricolor*
CPL	*Cynomops planirostris*	N	*Nycteris sp.*
CS	*Cynopterus sphinx*	NC	*Neoromicia capensis*
CY	*Cynopterus sp.*	NL	*Nyctalus leisleri*
DM	*Dobsonia moluccensis*	NN	*Neoromicia nana*
DR	*Desmondus rotundus*	NT	*Nycteris thebaica*
EG	*Eumops glaucinus*	NV	*Nyctalus velutinus*
EH	*Eidolon helvum*	PH	*Pipistrellus hesperidus*
EL	*Epimorphorus labiatus*	PJ	*Ptenochirus jagori*
ES	*Eptesicus serotinus*	PN	*Pipistrellus nathusii*
GS	*Glossophaga soricina*	RA	*Rhinolophus affinis*
H	*Hipposideros sp.*	RAU	*Rhinonycteris aurantius*
HCO	*Hipposideros commersoni*	RB	*Rhinolophus blasii*
HP	*Hipposideros pratti*	RC	*Rhinolophus capensis*
HR	*Hipposideros ruber*	RCL	*Rhinolophus clivosus*
HV	*Hipposideros vittatus*	RCO	*Rhinolophus cornutus*
KL	*Kerivoula lanosa*	RD	*Rhinolophus darlingi*
LSP	*Laephotus botswanae*	RDNT	*Rhinolophus denti*
M	*Myotis sp.*	RF	*Rhinolophus ferrumequinum*
MB	*Myotis bocagii*	RH	*Rhinolophus hipposideros*
MC	*Mops condylurus*	RL	*Rousettus leschenaulti*
MD	*Myotis davidii*	RM	*Rhinolophus monoceros*
MDA	*Myotis daubentonii*	RSI	*Rhinolophus simulator*
MF	*Miniopterus fraterculus*	RSW	*Rhinolophus swinnyi*
MF	*Myotis fimbriatus*	SD	*Scotophilus dinganii*
MFU	*Miniopterus fuscus*	SL	*Sturnira lilium*
MI	*Miniopterus inflatus*	TA	*Tadarida aegyptiaca*
ML	*Murina leucogaster*	TAF	*Triaenops afer*
MM	*Mops midas*	TB	*Tadarida brasiliensis*
MM	*Myotis macropus*		

Table 1 Bat trapping sites. All bat trapping sites, with abbreviated site codes, where samples were collected for general and species-specific surveillance arms respectively, are indicated below. The site where samples for the longitudinal surveillance of a *N. capensis* colony were collected is indicated with an asterisk and is displayed in bold font. The geographic data, biome and altitude, are also detailed for each site.

Province	Site (site code)	GPS co-ordinates		Biome	Altitude (m)	Surveillance type		Sampling Dates (month-year)
		Latitude	Longitude			General	Species-Specific	
Eastern Cape	Aardvark Backpackers, Addo (ABA)	-33.5350	25.6955	Albany Thicket	57		X	Oct-14, July-15
	Arena Resort. East London (AEL)	-32.8839	28.0612	Albany Thicket	83	X	X	Oct-14
	Arminel Hotel. Hogsback (AHHB)	-32.5919	26.9332	Grassland	1227	X	X	Jan-14
	Babuhle Cottage. Hogsback (BCHB)	-32.8157	26.9153	Grassland	1229		X	Jan-14
	Hogsback Inn (HBI)	-32.5956	26.9403	Grassland	1202	X	X	Oct-14
	Phillip's Tunnel. Hankey (PTH)	-33.8211	24.8474	Albany Thicket	48	X		Oct-14, July-15
	Sandile's Cave (SPF)	-32.7166	27.2898	Grassland	1216	X		Jan-14
	Sandile's Rest (SRS)	-32.6714	27.2894	Grassland	830	X	X	Jan-14
	Sleepy Hollow. Maitland (SHM)	-33.9568	25.3132	Albany Thicket	167	X		Oct-14, July-15
	Table Farm. Grahamstown (TFG)	-33.2853	26.4276	Savanna	559	X		Jan-14, Oct-14, July-15
KwaZulu-Natal	Albert Falls (AF)	-29.4452	30.4301	Savanna	668	X		Oct-16
	Babanango Exploratory Mine 2 (BVLEM2)	-28.2852	31.0137	Savanna	938 - 992	X		Aug-14
	Babanango Exploratory Mine 1 (BVLEM1)	-28.2871	31.0129	Savanna	938 - 992	X		Aug-14
	Babanango Main Mine (BVLMM)	-28.2867	31.0133	Savanna	938 - 992	X		Aug-14
	Babango Valley (BVL)	-28.2867	31.0133	Savanna	938 - 992	X		Sept-14, May-15, Dec-15, Oct-16
	Buffelsdrift (BDF)	-29.7567	30.6791	Savanna	613	X		
	Dlinza Forest Nature Reserve (DFR)	-28.8959	31.4527	Savanna	516	X		Apr-14
	Doornhoek Mine (DHM)	-29.6000	30.5200	Savanna	597	X		May-16
	Ebenezer Farm (ENR)	-28.7317	31.5429	Savanna	212	X		Apr-14
	Entumeni Nature Reserve (ENR)	-28.8861	31.3758	Savanna	657	X		Apr-14
	Hilton Train Tunnel (HTT)	-29.5497	30.2958	Grassland	1118	X		Jan-16
	Inkunzi Londge. Babanango (ILB)	-28.5617	31.2404	Savanna	865	X	X	Sept-14
	Kwela Lodge (KL)	-29.4932	30.3611	Savanna	902	X	X	Oct-16
	Lotheni Nature Reserve (LNR)	-29.4375	29.5150	Grassland	1568		X	Sept-14, Jan-16
	Mooiplaas Gold Mine (MPG)	-28.5582	31.1653	Savanna	865	X		Sept-14, May-15, Dec-15, Oct-16
	Spionkop Lodge (SKL)	-28.6950	29.5355	Grassland	1140	X		Oct-16
	Umbilo WWTW (DC2)	-29.8455	30.8919	Coastal Belt	202	X		Mar-15, Apr-16, Sept-16
	Umdoni Forest (UF)	-30.3941	30.6808	Coastal Belt	65	X		Mar-15
	Verulum WWTW (DC1)	-29.6439	31.0636	Coastal Belt	20	X		Mar-15

185

	Zululand Inn (IZ)	-28.8878	31.4693	Savanna	534	X		Apr-14
Limpopo	Annette's Farm (LAF)	-23.0375	30.4301	Savanna	1038	X		Sept-16
	Kim's Farm (LKF)	-23.0222	29.7989	Savanna	1053	X		Sept-16
	Peter Taylor's House (LPT)	-23.0326	29.9296	Savanna	999	X		Sept-16
	Royal Macadamia (LRM)	-23.0553	30.1495	Savanna	780	X		Sept-16
Northern Cape	a La Fugue Guesthouse (LFU)	-28.4402	21.2945	Nama-Karoo	800	X	X	Aug-15
	Blinkklip Grotte (BKP)	-28.3001	23.1156	Savanna	1342	X		Nov-11, Aug-15
	Chapman Safaris Game Lodge (CSGL)	-27.4746	23.3856	Savanna	1354		X	Nov-15
	Hopefield Farm 1 (HFP1)	-28.6188	23.3242	Savanna	1033 - 1496	X		Aug-15
	Hopefield Farm 2 (HFP2)	-28.6305	23.3397	Savanna	1033 - 1496	X		Aug-15
	Hopefield Farm. Postmasburg (HFP)	-27.4746	23.3856	Savanna	1033 - 1496	X	X	Nov-14, Aug-15
Western Cape	Steenkampskraal Mine (SKK)	-30.9750	18.6343	Succulent Karoo	411	X		Sept-15, Mar-16
	Cloeteskraal Farm. Velddrif * (CCK)	-32.8732	18.2236	Fynbos	17	X	X	Jan-15
	DeKelders Cave 1 (CDK1)	-34.5556	19.3642	Fynbos	10 - 11	X		Jan-15, Sept-15
	DeKelders Cave 2 (CDK2)	-34.5500	19.3710	Fynbos	10 - 11	X		Jan-15, Sept-15
	Drie Kuilen Nature Reserve (CDK)	-33.5815	20.0312	Fynbos	1019		X	Jan-15, Sept-15
	Forest Edge. Knysna (FEK)	33.9294	22.9386	Forest	250	X	X	Jan-14, Mar-16
	Gecko Rock Cottage (CGC)	-33.5184	20.1310	Fynbos	892 - 912	X	X	Jan-15. Sept-15
	Gecko Rock Main House (CGR)	-33.5184	20.1188	Fynbos	892 - 912		X	Jan-15
	Haarwegskloof Nature Reserve (HWK)	-34.3383	20.3261	Fynbos	231	X	X	Sept-15
	Knysna Millwood Mines (KMM)	-33.8900	22.9910	Forest	539	X		Sept-15
	May's Lane. Greyton (GML)	-34.0548	19.6054	Fynbos	6		X	Mar-16
	Silver Ranch. Plettenberg Bay (SRP)	-34.0146	23.4060	Fynbos	224		X	Nov-15
	Total					45	20	

Table 2 Weather data collected for all sampling events at each bat trapping site where sample collection took place.

Province	Sites	Weather Station	GPS co-ordinates	Month	Month's maximum temperature (°C)	Month's minimum temperature (°C)	Total rainfall (mm)
Eastern Cape	Aardvark Backpackers[1]	[0055447A7] Addo	-33.4420; 25.7480	Oct 2014	39.6	4.6	25.4
				July 2015	26.9	0.2	108.4
	Arena Resort[2]	[0059541 5] East London	-33.0220; 27.8190	Oct 2014	30.3	10.0	94.6
	Arminel Hotel[2]	[0079811A0] Dohne	-32.5270; 27.4600	Jan 2014	38.1	9.8	52.2
	Babuhle Cottage[1]	[0079811A0] Dohne	-32.5270; 27.4600	Jan 2014	38.1	9.8	52.2
	Hogsback Inn[2]	[0079811A0] Dohne	-32.5270; 27.4600	Oct 2014	33.3	2.9	58.2
	Phillip's Tunnel	[0033556 5] Patensie	-33.7650; 24.8230	Oct 2014	38.2	5.9	25.0
				July 2015	27.7	2.5	122.6
	Sandile's Cave	[0079811A0] Dohne	-32.5270; 27.4600	Jan 2014	38.1	9.8	52.2
	Sandile's Rest[2]	[0079811A0] Dohne	-32.5270; 27.4600	Jan 2014	38.1	9.8	52.2
	Sleepy Hollow	[0033556 5] Patensie	-33.7650; 24.8230	Oct 2014	38.2	5.9	25.0
				July 2015	27.7	2.5	122.6
	Table Farm	[0056917 8]Grahamstown	-33.2900; 26.5020	Jan 2014	36.8	11.2	25.0
				Oct 2014	33.5	5.7	53.0
				July 2015	27.5	-1.7	0.0
KwaZulu-Natal	Albert Falls	[0239482 0] Cedara	-29.5410; 30.2650	Oct 2016	33.8	3.0	67.6
	Babanango Valley Sites	[0337382 5] Babanango	-28.3640; 31.2060	Aug 2014	31.7	3.3	9.4
				May 2015	31.7	7.1	29.6
				Dec 2015	42.0	13.3	79.2
				Sept 2014	38.2	2.8	23.6
				Oct 2016	36.1	3.6	86.8
	Buffelsdrift	[0239699 7] Oribi airport	-29.6470; 30.3990	March 2015	31.3	1.2	64.2
	Dlinza Forest Nature Reserve	[0304357 6] Mtunzini	-28.9470; 31.7070	April 2014	34.9	8.9	50.4
	Doornhoek Mine	[0239698 5] Pietermaritzburg	-29.6270; 30.4020	May 2016	29.7	3.3	21.0
	Ebenezer Farm	[0304357 6] Mtunzini	-28.9470; 31.7070	April 2014	34.9	8.9	50.4
	Entumeni Nature Reserve	[0304357 6] Mtunzini	-28.9470; 31.7070	April 2014	34.9	8.9	50.4
	Hilton Train Tunnel	[0239482 0] Cedara	-29.5410; 30.2650	Jan 2016	37.8	14.1	130.2
	Inkunzi Longe[2]	[0337382 5] Babanango	-28.3640; 31.2060	Sept 2014	38.2	2.8	23.6
	Kwela Lodge[2]	-29.5410; 30.2650	Oct 2016		33.8	3.0	67.6
	Lotheni Nature Reserve[1]	[0239482 0] Cedara	-29.5410; 30.2650	Sept 2014	34.5	1.9	42.4
				January 2016	37.8	14.1	130.2
	Mooiplaas Gold Mine	[0337382 5] Babanango	-28.3640; 31.2060	Sept 2014	38.2	2.8	23.6
				Oct 2016	36.1	3.6	86.8
				May 2015	31.7	7.1	29.6
				Dec 2015	42.0	13.3	79.2

187

	Site	Code/Town	Coordinates	Date			
	Spionkop Lodge	[0300454 3]Ladysmith	-28.5750; 29.7500	Oct 2016	40.9	2.2	52.0
	Umbilo WWTW	[0240837B7] Durban	-29.9560; 30.9730	Sept 2016	28.1	11.0	82.4
				April 2016	34.0	14.5	26.0
		[0241076 6] Virginia	-29.7720; 31.0550	March 2015	30.6	17.2	75.0
	Umdoni Forest	[0241076 6] Virginia	-29.7720; 31.0550	March 2015	30.6	17.2	75.0
	Verulum WWTW	[0241186 1]King Shaka airport	-29.6100; 31.1230	March 2015	32.0	14.4	117.2
	Zululand Inn	[0304357 6] Mtunzini	-28.9470; 31.7070	April 2014	34.9	8.9	50.4
Limpopo	Annette's Farm	[0723485A0] Levubu	-23.0940; 30.2860	Sept 2016	38.6	11.2	20.6
	Kim's Farm	[0722099 1] Mara	-23.1440; 29.5570	Sept 2016	37.1	5.7	0.40
	Peter Taylor's House	[0722099 1]Mara	-23.1440; 29.5570	Sept 2016	37.1	5.7	0.40
	Royal Macadamia	[0723485A0] Levubu	-23.0940; 30.2860	Sept 2016	38.6	11.2	20.6
Northern Cape	a La Fugue Guesthouse[2]	[0317475A8] Upington	-28.4110; 21.2640	Aug 2015	32.9	-3.8	9.0
	Blinkklip Grotte	[0360597B0] Vaalharts	-27.9570; 24.8400	Aug 2015	33.4	-5.5	0.0
				Nov 2014	34.3	5.8	136.4
	Chapman Safari Game Lodge[1]	[0393806 4] Kuruman	-27.4330; 23.4470	Nov 2014	32.0	-2.7	0.0
	Hopefield Farm Sites[2]	[0356880 4] Kathu	-27.6700; 23.0060	Nov 2014	35.8	6.3	15.4
				Aug 2015	31.3	-2.9	0.0
Western Cape	Steenkampskraal Mine	[0157843 6] Garies	-30.5610; 17.9870	March 2016	40.9	9.3	5.4
				Sept 2015	38.5	4.5	11.0
	Cloeteskraal Farm[2]	[0061268 6] Langebaanweg	-32.9780; 18.1640	Jan 2015	36.5	9.5	9.6
	De Kelders Caves	[0003020 4] Cape Agulhas	-34.8260; 20.0130	Jan 2015	27.1	15.1	14.4
				Sept 2015	31.0	10.1	47.2
	Drie Kuilen Nature Reserve[2]	[0023708A4] Robertson	-33.7940; 19.9010	Sept 2015	33.5	5.6	38.2
				Jan 2015[1]	39.5	10.7	10.8
	Forest Edge[2]	[0014123 3] Knysna	-34.0490; 23.0810	Jan 2014	35.5	13.6	104.8
				March 2016	35.1	11.6	68.6
	Gecko Rock Site[2]	[0023708A4]robertson	-33.7940; 19.9010	Sept 2015	33.5	5.6	38.2
				Jan 2015[1]	39.5	10.7	10.8
	May's Lane	[0007699A0] Tygerhoek	-34.1490; 19.9030	March 2016	42.3	10.8	40.8
	Haarwegskloof Nature Reserve[2]	[0003423A7] Overberg	-34.5510; 20.2520	Sept 2015	31.1	4.6	56.0
	Knysna Millwood Mines	[0014123 3] Knysna	-34.0490; 23.0810	Sept 2015	31.1	4.6	56.0
	Silver Ranch[1]	[0014545 4] Plettenbergbaai	-34.0890; 23.3250	Nov 2014	26.9	9.5	86.2

[1] Species-specific surveillance sampling site or event only [2] General surveillance and species-specific surveillance site

188

APPENDIX D: Tables relating to screening results

Table 1 The blastn results for all obtained sequences from the screening of general surveillance samples. The closest related sequence as ranked by percentage identity is indicated for sequences obtained using both the Pan-CoV PCR assay and Extended RdRp assay. Unless indicated, all sequences undergoing BLASTn were 395 nt in length.

| PCR | Sample code | Closest related sequence by blastn as ranked by identity | | | | | |
		Genus	Species	Country	Genbank no.	(%) coverage	(%) identity
Pan-CoV and Ext RdRp PCR	*SHM_MN1***	*Alpha*	*MIniopterus natalensis*	RSA	KF843851	100	99
	BVL_MN(DC253)	*Alpha*	*Miniopterus schreibersii*	Australia	EU834955	98	82
	BVL1_RCL2	*Alpha*	*Rhinolophus sp.*	Kenya	GU065418	99	89
	DHM_RSI(DC84)	*Alpha*	*Chaerephon sp.*	Kenya	HQ728486	100	98
	DC2_SD(DC68)	*Alpha*	*Scotophilus heathii*	Thailand	KJ020603	90	96
	LRM_MN(DC115)	*Alpha*	*Miniopterus natalensis*	RSA	KF843851	99	99
	*SKK_RC10***	*Alpha*	*Rhinolophus sp.*	Kenya	GU065417	96	88
	KMM_RCL4	*Alpha*	*Rhinolophus sp.*	Kenya	GU065417	96	89
	FEK_PH1	*Beta*	Dromedary camel	UAE	KT751244	100	94
	FEK_PH4	*Beta*	Dromedary camel	UAE	KT751244	100	94
Extended RdRp PCR only	*DC2_CP(DC95)*	*Alpha*	*Chaerephon sp.*	Kenya	HQ728486	100	96
	CDK1_MN1	*Alpha*	*Miniopterus natalensis*	Kenya	GU065410	100	91
	CDK1_MN2	*Alpha*	*Miniopterus natalensis*	Kenya	GU065411	100	96
	SKK_MN2	*Alpha*	*Miniopterus natalensis*	RSA	KF843851.	100	98
	DC2_NN1(DC63)*	*Alpha*	*Neoromicia capensis*	RSA	KF843858	43	94
	DC2_NN(DC93)	*Alpha*	*Neoromicia capensis*	RSA	KF843854	100	96
	KL_PH(DC1680	*Beta*	Dromedary camel	UAE	KT751244	100	94
	UF_PH(DC34)*	*Beta*	Dromedary camel	UAE	KT751244	100	91
	AEL_PH1	*Beta*	*Neoromicia capensis*	RSA	KC869678	100	98

* Length of sequence undergoing BLASTn was 816 nt **Not included in subset for repeat screening.

Table 2 Species-specific surveillance screening results for all *Neoromicia capensis* samples using both the Pan-CoV and Extended RdRp PCR assays by bat trapping site.

Province	Site	First trapping	Pan-CoV PCR No. positive | No. sampled	No. alpha|No. beta	Ext RdRp PCR No. positive | No. sampled	No. alpha|No. beta| No. coinfected	Second trapping	Pan-CoV PCR No. positive | No. sampled	No. alpha|No. beta	Ext RdRp PCR No. positive | No. sampled	No. alpha|No. beta| coinfected
Eastern Cape	ABA	14-Oct	5|11	5|0	3|3	3|0|0	15-Jul	1|7	0|1	2|7	0|2|0
	AEL	14-Oct	1|2	1|0	1|1	1|0|0	NA				
	AHHB	14-Jan	1|9	1|0	NA	NA	NA				
	BCHB	14-Jan	0|2	0|0	0|2	0|0|0	NA				
	HBI	14-Oct	0|11	0|0	1|2	0|1|0	NA				
	SRS	14-Jan	0|1	0|0	NA	Na	NA				
KwaZulu-Natal	LNR	14-Sep	7|19	7|0	4|7	4|0|0	16-Jan	4|10	4|0	5|6	5|0|0
	ILB	14-Sep	0|6	0|0	0|2	0|0|0	NA				
	KL	16-Oct	0|4	0|0	3|4	0|3|0	NA				
Northern Cape	HFP	14-Nov	0|20	0|0	7|13	1|6|0	15-Aug	0|13	0|0	NA	NA
	LFU	15-Aug	0|4	0|0	1|4	0|1|0	NA				
	CSGL	14-Nov	0|3	0|0	0|3	0|0|0	NA				
Western Cape	CDK	15-Jan	4|12	4|0	1|1	0|0|1	15-Sep	3|17	3|0	2|4	1|1|0
	FEK	14-Jan	1|1	1|0	3|4	0|0|3	16-Mar	3|4	3|0	NA	NA
	CGC	15-Jan	2|5	2|0	1|2	0|0|1	15-Sep	3|12	2|1	3|6	1|1|1
	CCK	15-Jan	3|5	3|0	3|3	1|1|1	NA				
	CGR	15-Jan	1|4	1|0	1|1	0|0|1	NA				
	GML	16-Mar	0|1	0|0	0|1	0|0|0	NA				
	HWK	15-Sep	6|8	6|0	6|8	4|0|2	NA				
	SRP	14-Nov	1|5	1|0	NA	NA	NA				

Table 3 Species confirmation by molecular methods. The table indicates the samples for which confirmation of species by molecular means was in disagreement with the in-field species allocated. The correct species along with the adjusted sample name is also indicated.

Original sample ID	In-field species identification	Molecular species identification	Adjusted sample ID
20160112HTT_MN2	MN	MF	20160112HTT_MN2mf
20160112HTT_MN3	MN	MF	20160112HTT_MN3mf
20160112HTT_MN4	MN	MF	20160112HTT_MN4mf
20160112HTT_MN5	MN	MF	20160112HTT_MN5mf
20160112HTT_MN6	MN	MF	20160112HTT_MN6mf
20160112HTT_MN7	MN	MF	20160112HTT_MN7mf
20160112HTT_MN8	MN	MF	20160112HTT_MN8mf
20151208MPG_RSW2	RSW	RSI	20151208MPG_RSW2rsi
20151208MPG_RSW3	RSW	RSI	20151208MPG_RSW3rsi
20151208MPG_RSW5	RSW	RSI	20151208MPG_RSW5rsi
20151208MPG_RSW6	RSW	RSI	20151208MPG_RSW6rsi
20151208MPG_RSW7	RSW	RSI	20151208MPG_RSW7rsi
20140921MPG_RSW3	RSW	RSI	20140921MPG_RSW3rsi
20140921MPG_RSW4	RSW	RSI	20140921MPG_RSW4rsi
20140921MPG_RSW2	RSW	RSI	20140921MPG_RSW2rsi
20150921MPG_RSW5	RSW	RSI	20150921MPG_RSW5rsi
20150512MPG_RSW5	RSW	RSI	20150512MPG_RSW5rsi
20140923LNR_PH1	PH	NC	20140923LNR_PH1nc
20140923LNR_PH4	PH	NC	20140923LNR_PH4nc
20140923LNR_PH5	PH	NC	20140923LNR_PH5nc
20150819LFU_PH1	PH	NC	20150819LFU_PH1nc
20150815HFP_PH5	PH	NC	20150815HFP_PH5nc
20150815HFP_PH6	PH	NC	20150815HFP_PH6nc
20150815HFP_PH1	PH	NC	20150815HFP_PH1nc
20150815HFP_PH2	PH	NC	20150815HFP_PH2nc
20141107HFP_PH1	PH	NC	20141107HFP_PH1nc
20141107HFP_PH2	PH	NC	20141107HFP_PH2nc
20141108HFP_PH1	PH	NC	20141108HFP_PH1nc
20141108HFP_PH3	PH	NC	20141108HFP_PH3nc
20141022HBI_PH1	PH	NC	20141022HBI_PH1nc
20141022HBI_PH2	PH	NC	20141022HBI_PH2nc
20141022HBI_PH3	PH	NC	20141022HBI_PH3nc
20141022HBI_PH4	PH	NC	20141022HBI_PH4nc
20141022HBI_PH5	PH	NC	20141022HBI_PH5nc
20141022HBI_PH6	PH	NC	20141022HBI_PH6nc
20141022HBI_PH7	PH	NC	20141022HBI_PH7nc
20141022HBI_PH8	PH	NC	20141022HBI_PH8nc
20141022HBI_PH9	PH	NC	20141022HBI_PH9nc
20141022HBI_PH10	PH	NC	20141022HBI_PH10nc
20141022HBI_PH11	PH	NC	20141022HBI_PH11nc
20160304FEK_PH1	PH	NC	20160304FEK_PH1nc
20140115AHHB_PH1	PH	NC	20140115AHHB_PH1nc

20141021AEL_PH4	PH	NC	20141021AEL_PH4nc
20141021AEL_PH3	PH	NC	20141021AEL_PH3nc
20142027ABA_PH5	PH	NC	20142027ABA_PH5nc
20142027ABA_PH9	PH	NC	20142027ABA_PH9nc
20142027ABA_PH10	PH	NC	20142027ABA_PH10nc
20142027ABA_PH12	PH	NC	20142027ABA_PH12nc
20142027ABA_PH13	PH	NC	20142027ABA_PH13nc
20150720ABA_PH1	PH	NC	20150720ABA_PH1nc
20142027ABA_PH2	PH	NC	20142027ABA_PH2nc
20142027ABA_PH11	PH	NC	20142027ABA_PH11nc
20142027ABA_PH4	PH	NC	20142027ABA_PH4nc
20142027ABA_PH6	PH	NC	20142027ABA_PH6nc
20142027ABA_PH8	PH	NC	20142027ABA_PH8nc
20142027ABA_PH7	PH	NC	20142027ABA_PH7nc

Table 4 Fragments obtained from each betacoronavirus positive sample that underwent additional genome amplification. The fragment length, its corresponding region on the NeoCoV (Genbank ID KC869678) genome, its translated amino acid length, and the highest sequence identity match based on blastn results is indicated in the table below.

Sample	Frag.	Length	Region	Gene sequences identified (no. aa)	BLASTn highest sequence identity results
20150720ABA_NC1	1	3756 nt	2979 - 6735	partial ORF1a (1192)	93% NeoCoV
	2	1457 nt	21635 - 23091	partial S (399)	92% NeoCoV
	3	4491 nt	25312 - 29802	partial N (414)	99% NeoCoV
				ORF 3 (103)	93% NeoCoV
				ORF4b (246)	95% NeoCoV
				ORF 5 (224)	97% NeoCoV
				M (219)	98% NeoCoV
				partial S (64)	97% NeoCoV
				E (82)	98% Neo CoV
	4	5334 nt	13918 - 19251	partial ORF1ab (1773)	99% NeoCoV
	5	455 nt	8405 - 8859	partial ORF1a (151)	96% NeoCoV
	6	625 nt	23320 - 23944	partial S (208)	96% NeoCoV
20160303FEK_NC1	1	1575	17602 - 19176	partial ORF 1ab (525)	97% NeoCoV
	2	895	14813 - 15707	partial ORF 1ab (297)	95% NeoCoV
	3	2231	26089 - 28319	partial ORF 4b (116)	97% NeoCoV
				ORF 5 (224)	97% NeoCoV
				E (82)	96% NeoCoV
				partial M (160)	95% NeoCoV
201603040FEK NC1	1	3309	25310 - 28618	partial S (64)	94% NeoCoV
				ORF 3 (103)	92% NeoCoV
				ORF 4a (109)	95% NeoCoV
				ORF 4b (246)	96% NeoCoV
				ORF 5 (224)	97% NeoCoV
				M (219)	99% NeoCoV
				E (82)	96% NeoCOV
	2		23319 - 23951	partial S (210)	94% NeoCoV
	3		14834 - 15709	partial ORF 1ab (291)	95% NeoCoV
20160304FEK_N C2	1	630	3671 - 4300	partial ORF1a (210)	93 % NeoCoV
	2	1241	4492 - 5732	partial ORF1a (413)	95 % NeoCoV
	3	455	8401 - 8855	partial ORF1a (151)	95 % NeoCoV

4	3671	13927 - 17598	partial ORF1ab (1223)	95 % NeoCoV
5	630	23321 - 23950	partial S (210)	93 % NeoCoV
6	3008	25310 - 28317	partial S (64)	94 % NeoCoV
			ORF 3 (103)	96 % NeoCoV
			ORF 4a (109)	95 % NeoCoV
			ORF 4b (246)	96 % NeoCoV
			ORF 5 (224)	97 % NeoCoV
			E (82)	96 % NeoCoV
			partial M (159)	95 % NeoCoV

20140122FEK_PH1

1	725	3613 - 4337	partial ORF 1a (241)	90 % NeoCoV
2	1093	4492 - 5585	partial ORF 1a (364)	90 % NeoCoV
3	1076	5661 - 6736	partial ORF 1a (358)	94 % NeoCoV
4	636	7042 - 7678	partial ORF 1a (212)	90 % human MERS (KU851859)
5	1042	8391 - 9432	partial ORF 1a (346)	dromedary camel MERS 90 % (KX108943)
6	5340	13914 - 19257	partial ORF 1ab (1780)	93 % human MERS (KX154694)
7	895	21370 - 22261	partial S (265)	83 % PREDICT Pipistrellus hesperidus MERS (KX574227)
8	5321	24506 - 29833	partial S (333)	89 % NeoCoV
			ORF 3 (101)	82 % NeoCoV
			ORF 4a (109)	88 % NeoCoV
			ORF 4b (258)	81 % NeoCoV
			ORF 5 (225)	85 % NeoCoV
			E (82)	96 % NeoCoV
			M (219)	92 % dromedary camel MERS (KU740200)
			N (413)	90 % NeoCoV

20140122FEK_PH4

1	845	14808 - 16562	partial ORF1ab (281)	95 % dromedary camel MERS (KT751244)
2	681	21634 - 22311	partial S (226)	82 % PREDICT Pipistrellus hesperidus (KX574227) MERS
3	4418	24159 - 28579	partial S (447)	89 % NeoCoV
			ORF 3 (101)	83 % NeoCoV
			ORF 4a (109)	89 % NeoCoV
			ORF 4b (258)	81 % NeoCoV
			ORF 5 (225)	85 % NeoCoV
			E (82)	96 % NeoCoV
			M (219)	92 % dromedary came l MERS (KU740200)
4	1203	28627 - 29832	partial N (389)	90 % NeoCoV

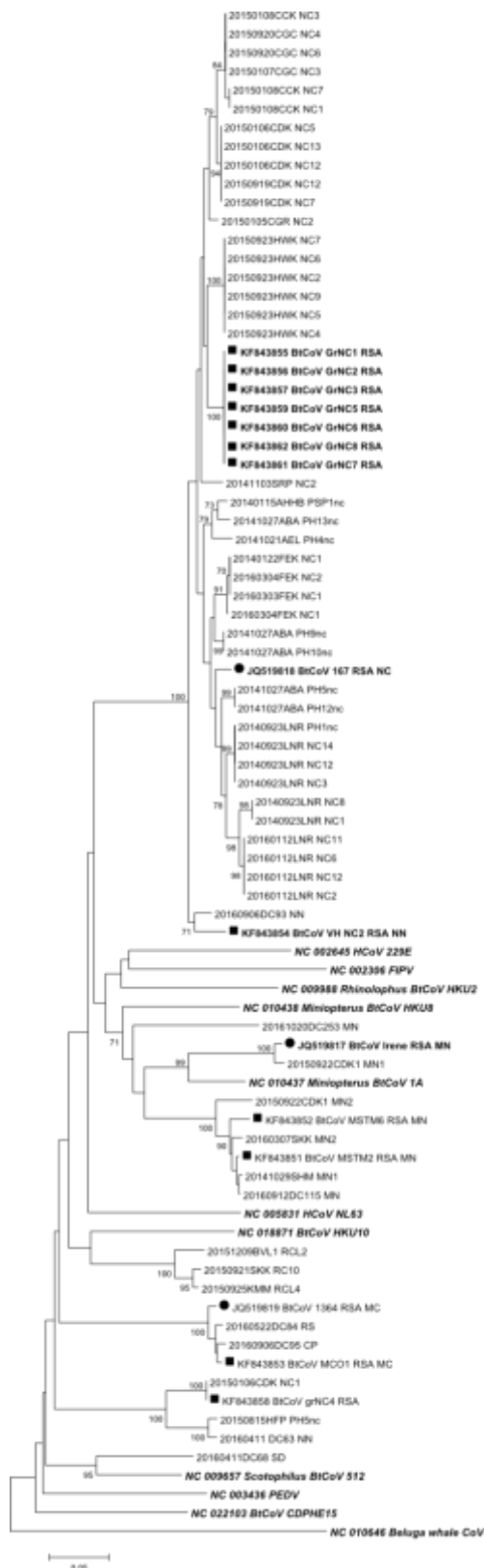## APPENDIX E: Additional phylogenetic trees

Figure 1 Phylogeny of all known bat-derived alphacoronaviruses from South Africa. Phylogeny was inferred by the Neighbour Joining Method with p-distances in MEGA7 and 1000 bootstrap (BS) replicates (Saitou and Nei, 1987; Kumar *et al.*, 2016). BS values less than 70% have been omitted. The alignment contained 275 nucleotide positions. Sequences from Ithete et al (2013) are indicated with ■ while sequences from Geldenhuys et al. (2013) are indicated with ●. ICTV-recognised CoV species and RefSeq sequences of interest are indicated in bold oblique font.

195
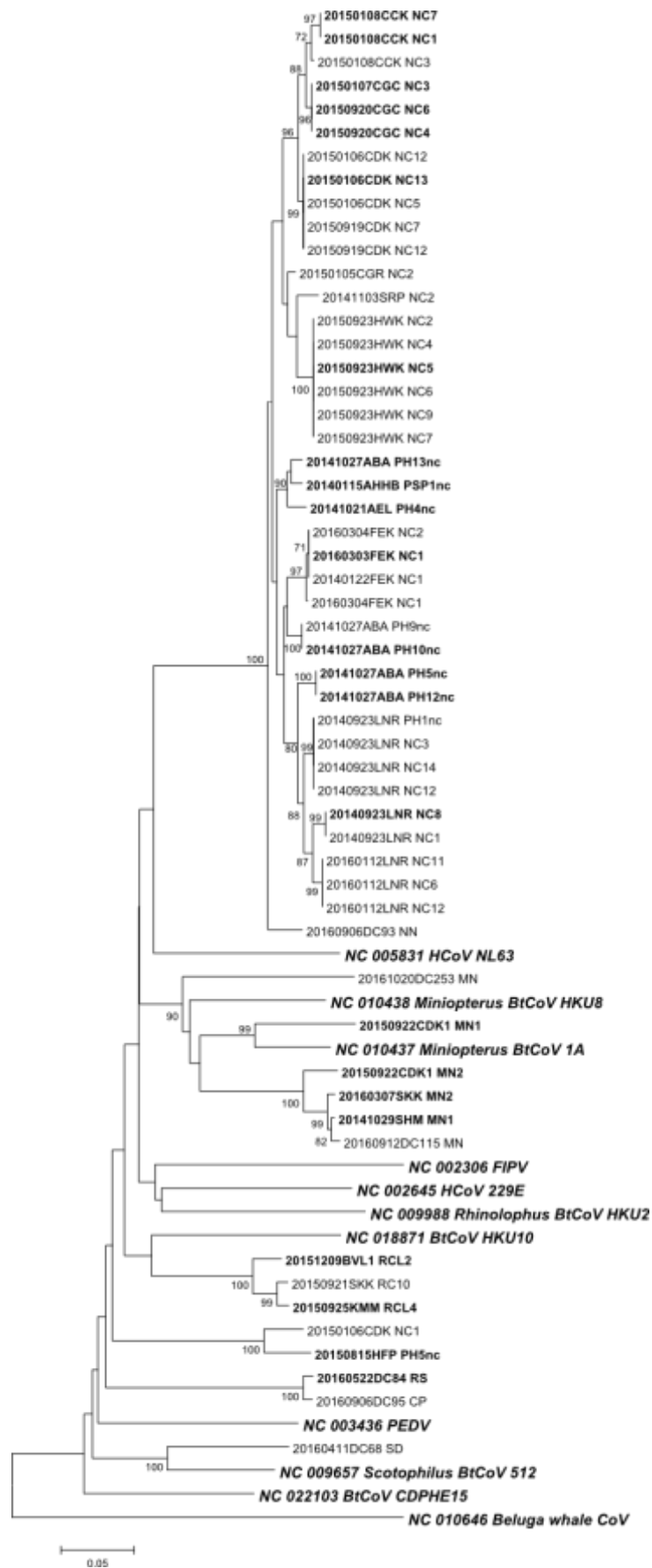
Figure 2 Phylogeny of all 395bp alphacoronavirus sequences obtained in this study. Phylogeny was inferred by the Neighbour Joining Method with p-distances in MEGA7 and 1000 bootstrap (BS) replicates (Saitou and Nei, 1987; Kumar *et al.*, 2016). BS values less than 70% have been omitted. The final alignment contained 395 nucleotide positions. Sample sequence names in bold font are only represented by a 395 bp partial RdRp sequence. ICTV-recognised CoV species and RefSeq sequences of interest are indicated in bold oblique font.
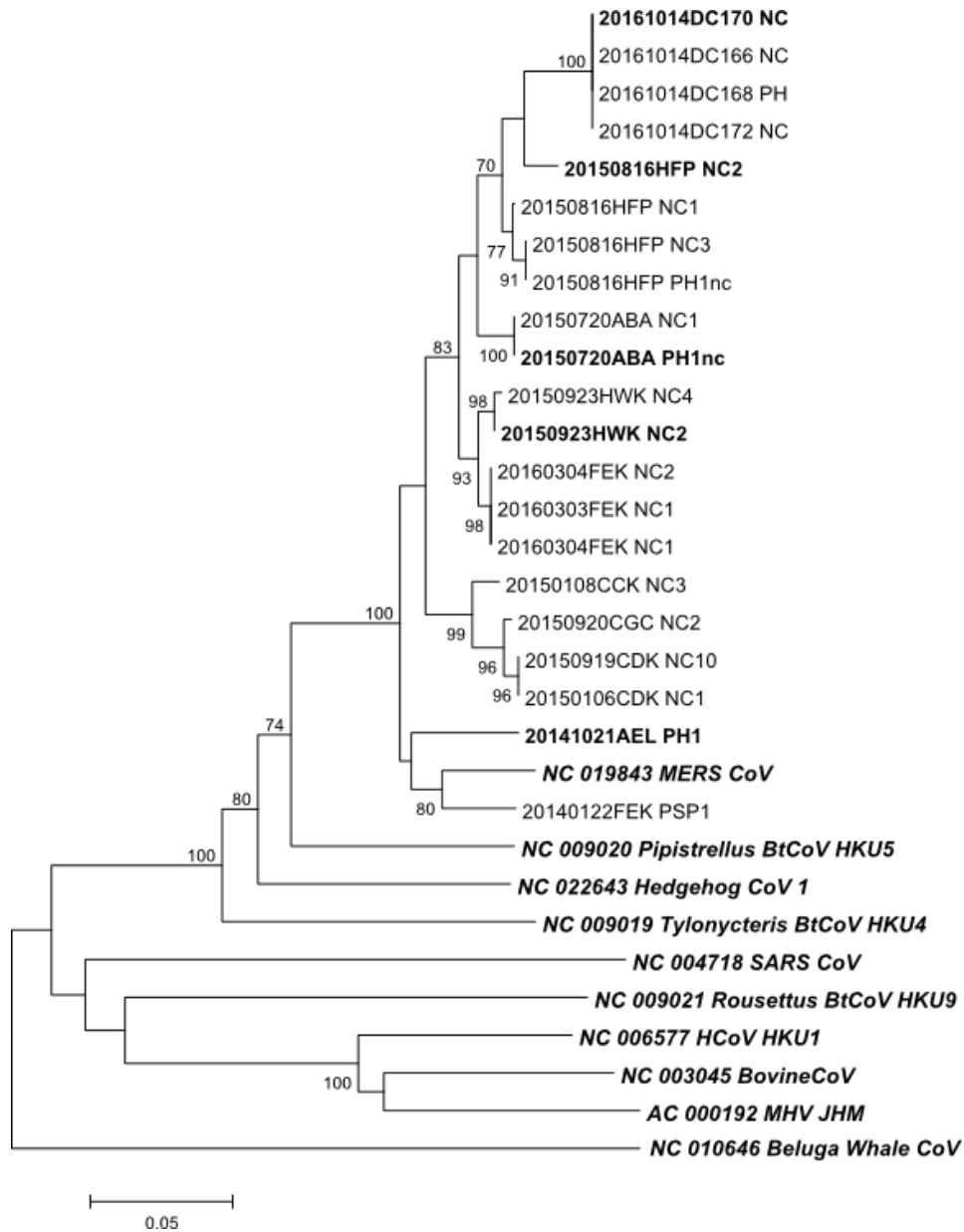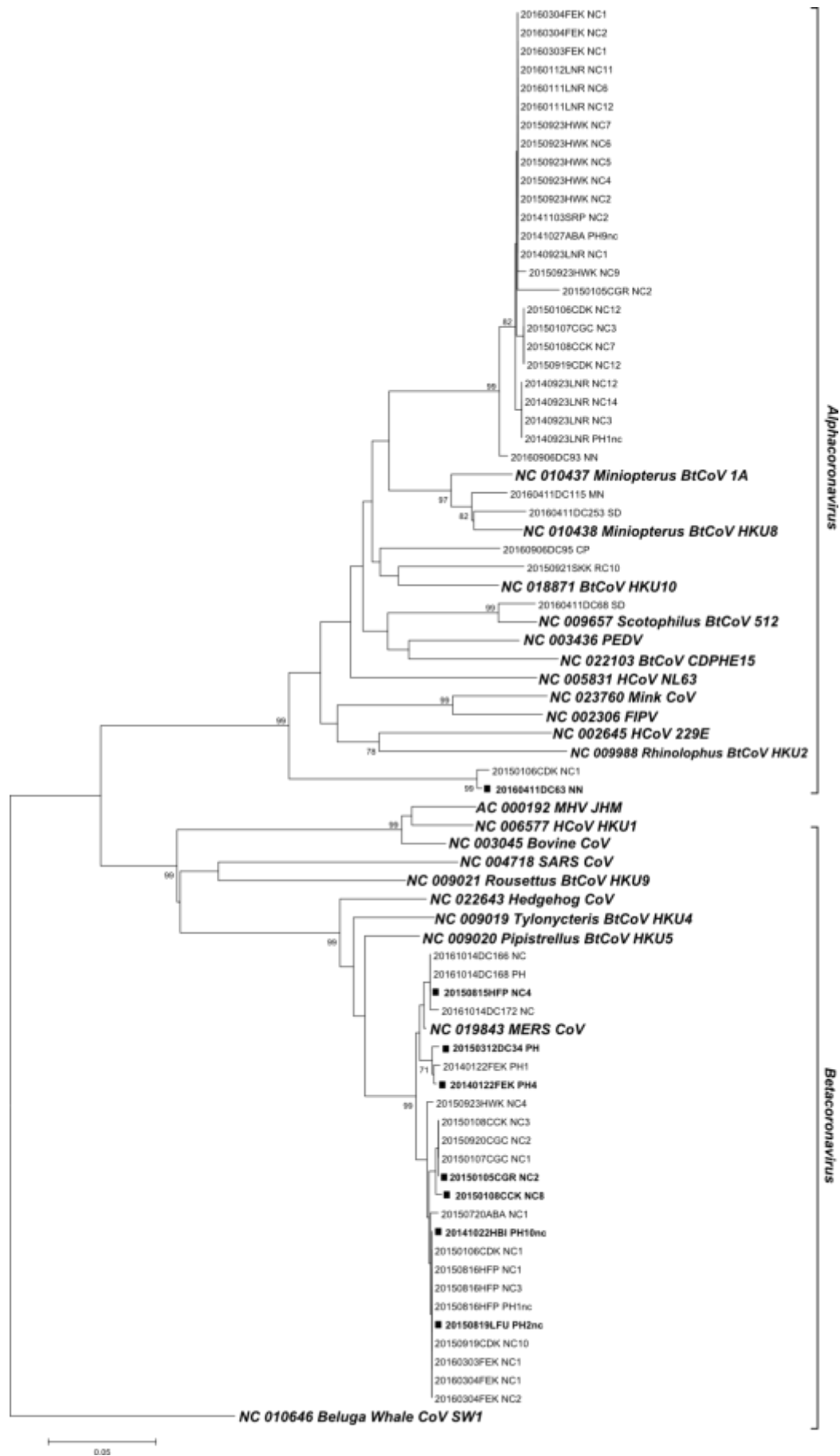
Figure 3 Phylogeny of all 395bp betacoronavirus sequences obtained in this study. Phylogeny was inferred by the Neighbour Joining Method with p-distances in MEGA7 and 1000 bootstrap (BS) replicates (Saitou and Nei, 1987; Kumar *et al.*, 2016). BS values less than 70% have been omitted. The alignment contained 395 nucleotide positions. Sample sequence names in bold font are only represented by a 395 bp partial RdRp sequence. ICTV-recognised CoV species and RefSeq sequences of interest are indicated in bold oblique font.

Figure 4 Phylogeny of all 272 aa partial RNA dependent RNA polymerase sequences obtained in this study. Phylogeny was inferred by the Neighbour Joining Method with p-distances in MEGA7 and 1000 bootstrap (BS) replicates (Saitou and Nei, 1987; Kumar *et al.*, 2016). BS values less than 70% have been omitted. The alignment contained 272 amino acid positions. Sample sequence names in bold font with ■ are only represented by a 272 aa partial RdRp sequence. ICTV-recognised CoV species and RefSeq sequences of interest are indicated in bold oblique font.

198
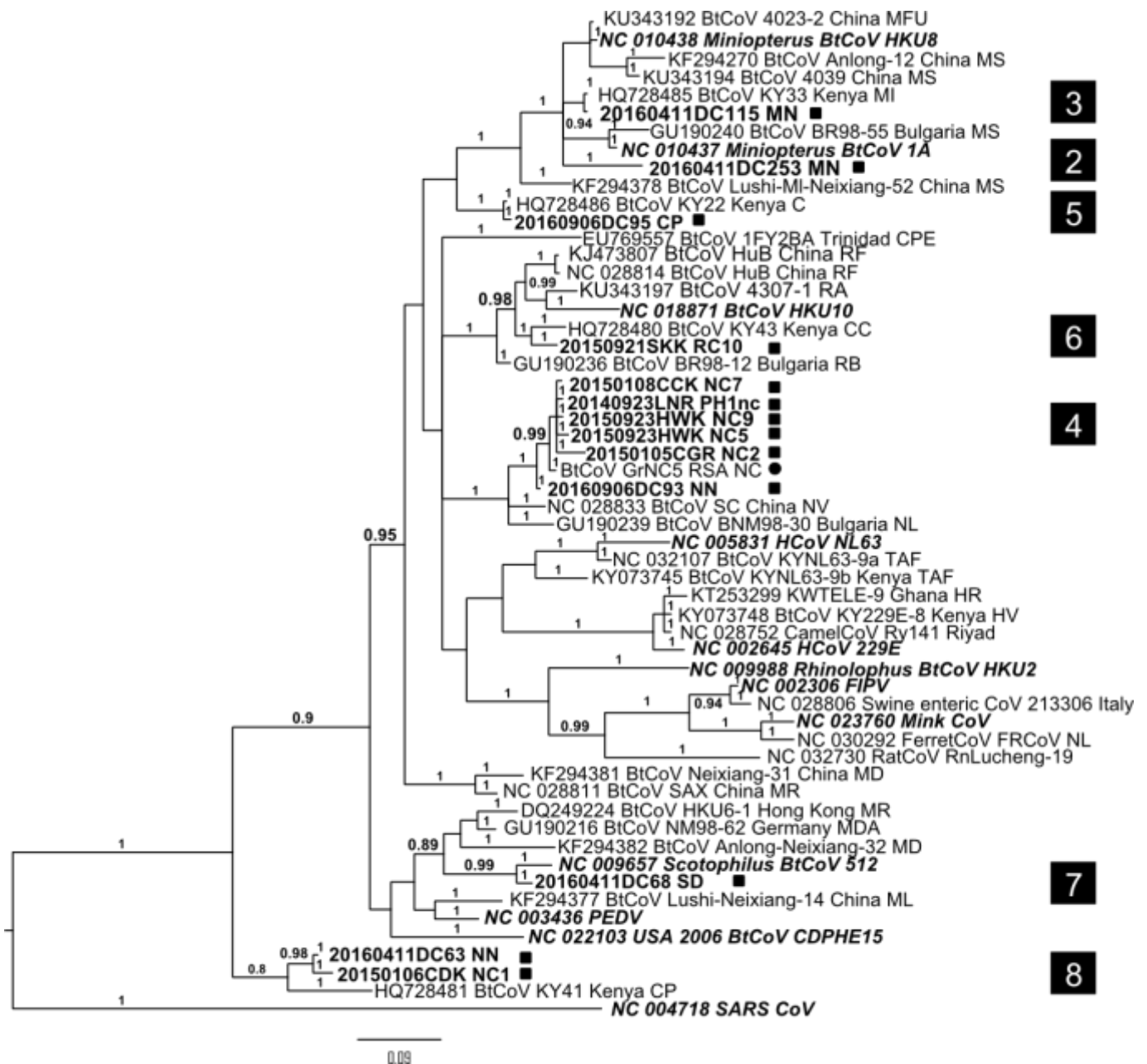
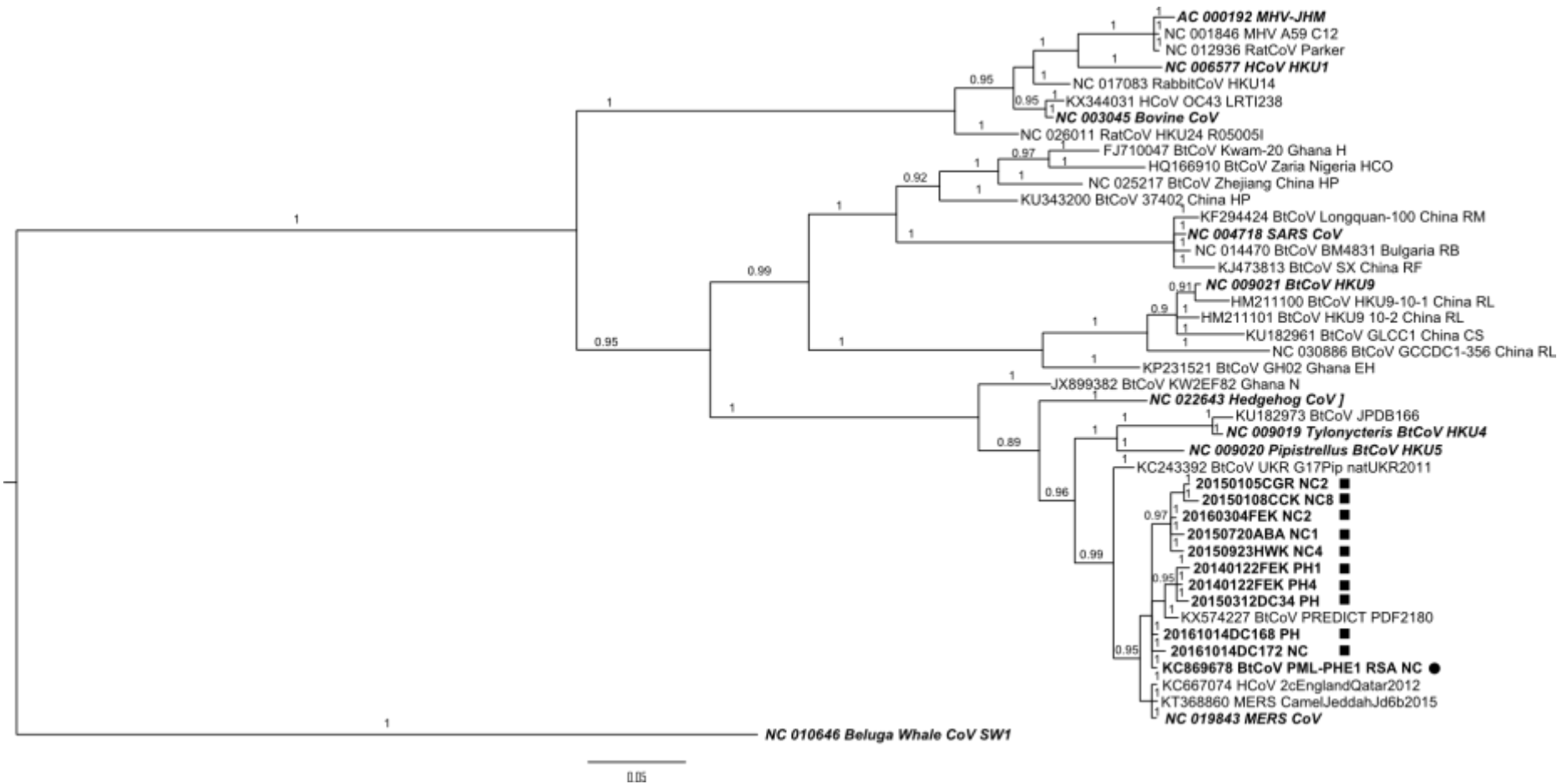**Figure 5 Phylogeny of alphacoronavirus sequences obtained during this study using Bayesian Inference.** Phylogenetic inference was determined by the Bayesian inference method in Mr Bayes 3.2.6 with two million generations and the initial 25% of sample trees were discarded. The Le Gascuel model with gamma distributed rate variation and invariant sites was used (LG+G+I). The tree was midpoint rooted with SARS-CoV (NC_004718). ICTV-recognised CoV species and RefSeq sequences of interest are indicated in bold oblique font. ■ = study sequence; ● = non-study South African sequence. Indicated numbering corresponds to that depicted in Figures 4.5 and 4.6.

Figure 6 Phylogenetic analysis of betacoronavirus partial RdRp sequences obtained during this study using Bayesian inference. Phylogenetic inference was determined by the Bayesian inference method in Mr Bayes 3.2.6 with two million generations and 25% of sample trees were discarded. The Le Gascuel model with gamma distributed rate variation and invariant sites was used (LG+G+I). The tree was midpoint rooted with Beluga Whale CoV SW1 (NC_010646). ICTV-recognised CoV species and RefSeq sequences of interest are indicated in bold oblique font. ■ = study sequence ● = non-study South African sequence.
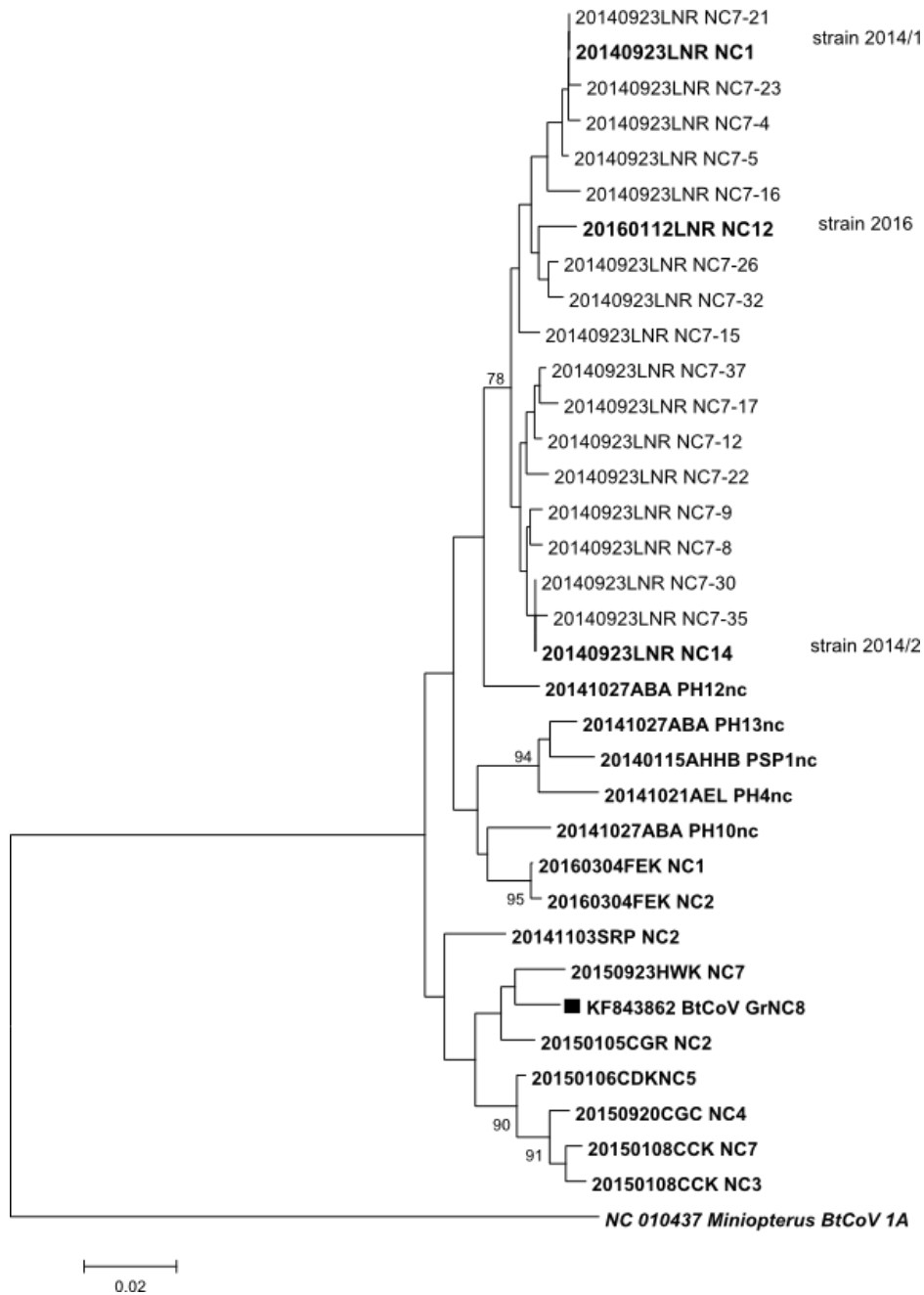
Figure 7 Phylogenetic analysis of non-identical clonal sequences from sample 20150923LNR_NC7 and non-identical unclassfied *Neoromicia* BtCoV 1 sequences from *Neoromicia capensis* bats. Phylogeny was inferred by the Neighbour Joining Method with p-distances in MEGA7 and 1000 bootstrap (BS) replicates (Saitou and Nei, 1987; Kumar *et al.*, 2016). BS values less than 70% have been omitted. The alignment contained 395 nt positions. Non-identical *N. capensis*-derived sequences are indicated in bold font. The three different strain sequences are indicated. The tree was midpoint rooted with *Miniopterus* BtCoV 1A (NC_010437).
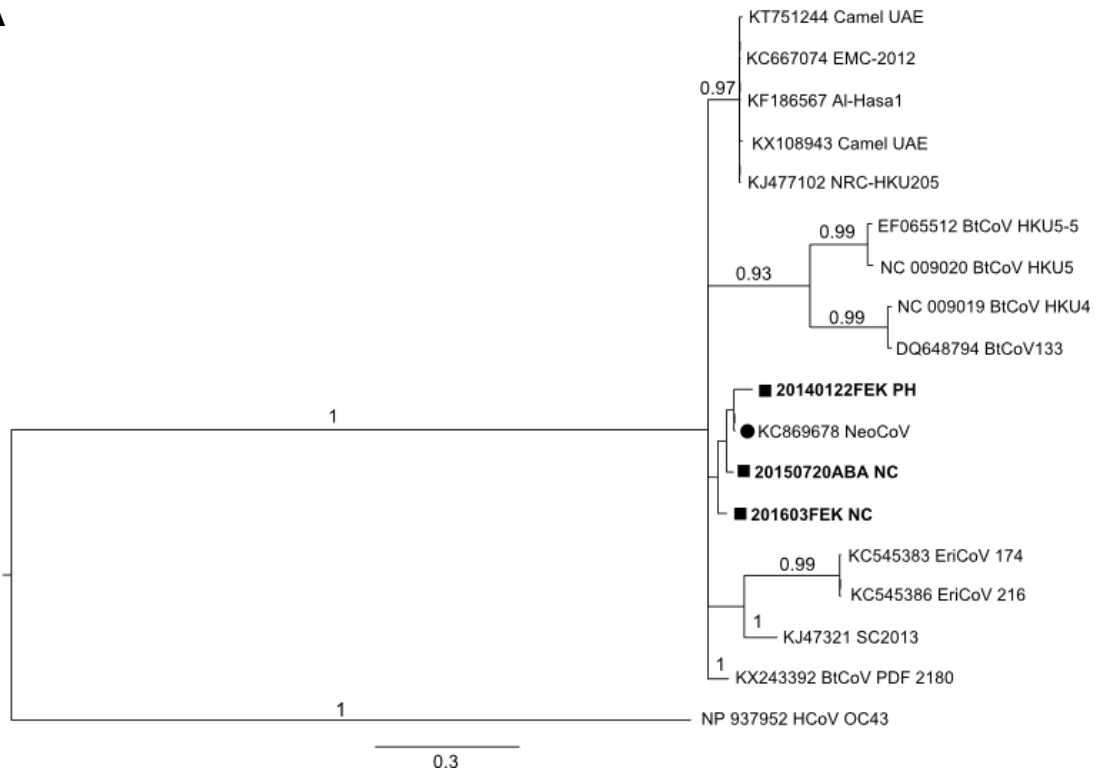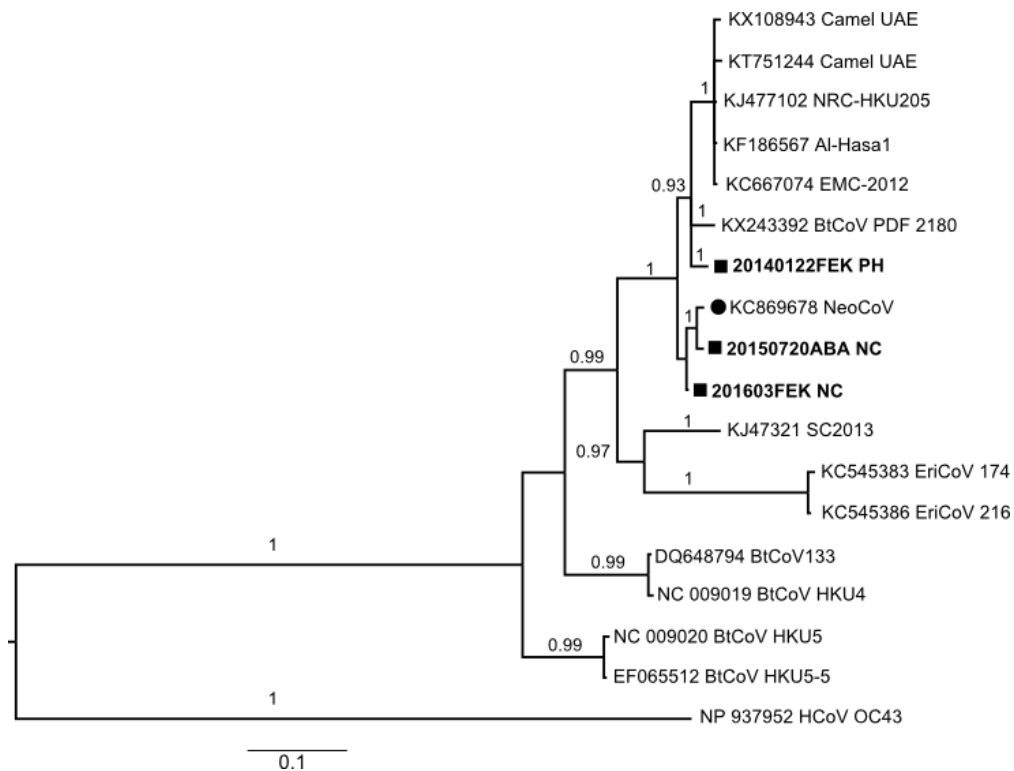
Figure 8 Bayesian phylogeny of the complete envelope (E) and membrane (M) genes of lineage C betacoronaviruses. Phylogenies were constructed with Mr Bayes with 2 million generations with the first 25% of sampled trees discarded. Posterior probabilities less than 0.90 have been omitted. Study sample sequences are indicated in bold with ■ and NeoCoV (Genbank ID KC869678) is indicated with ●. The tree was midpoint rooted with human coronavirus OC43 (hCoV OC43) (Genbank ID NP 937952). Figure A depicts the E gene phylogeny based on the Hasegawa-Kishino-Yano model with gamma distributed evolutionary rates (HKY + G). There were a total of 246 nucleotide positions in the final dataset. Figure B depicts the M gene phylogeny based on the Le Gascuel model with gamma distributed evolutionary rates (LG + G). There were a total of 218 amino acid positions in the final dataset.
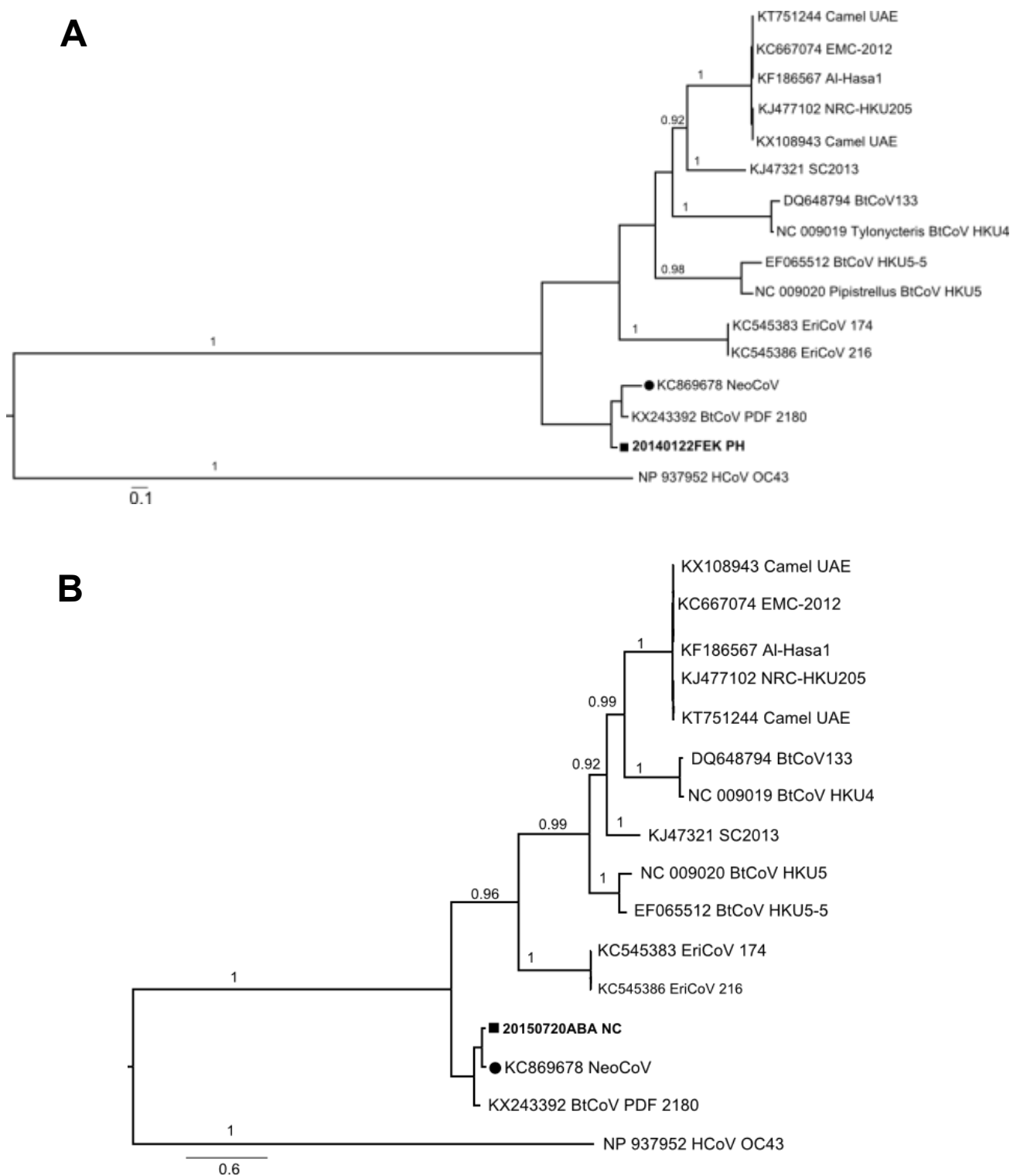
Figure 9 Bayesian phylogeny of partial Spike (S) subunit 1 sequences of lineage C betacoronaviruses.
The **Bayesian ph**ylogenies were constructed in Mr Bayes based on the Whelan AND Goldman model with gamma distributed evolutionary rates (WAG + G). Statistical support for each tree was measured with 2 million generations with the first 25% of sampled trees discarded, posterior probability values less than 0.90 have been omitted. Study sample sequences are indicated in bold with ■ and NeoCoV (Genbank ID KC869678) is indicated with ●. The tree was midpoint rooted with human coronavirus OC43 (hCoV OC43) (Genbank ID NP 937952). Figure A depicts the phylogeny of partial S1 sequences based on an alignment including the partial S1 sequence from study sample 20140122FEK_PH. There were a total of 315 amino acid positions in the final dataset. Figure B depicts the phylogeny of partial S1 sequences based on an alignment including the partial S1 sequence from study sample 20150720ABA_NC1. There were 415 amino acid positions in the final dataset.
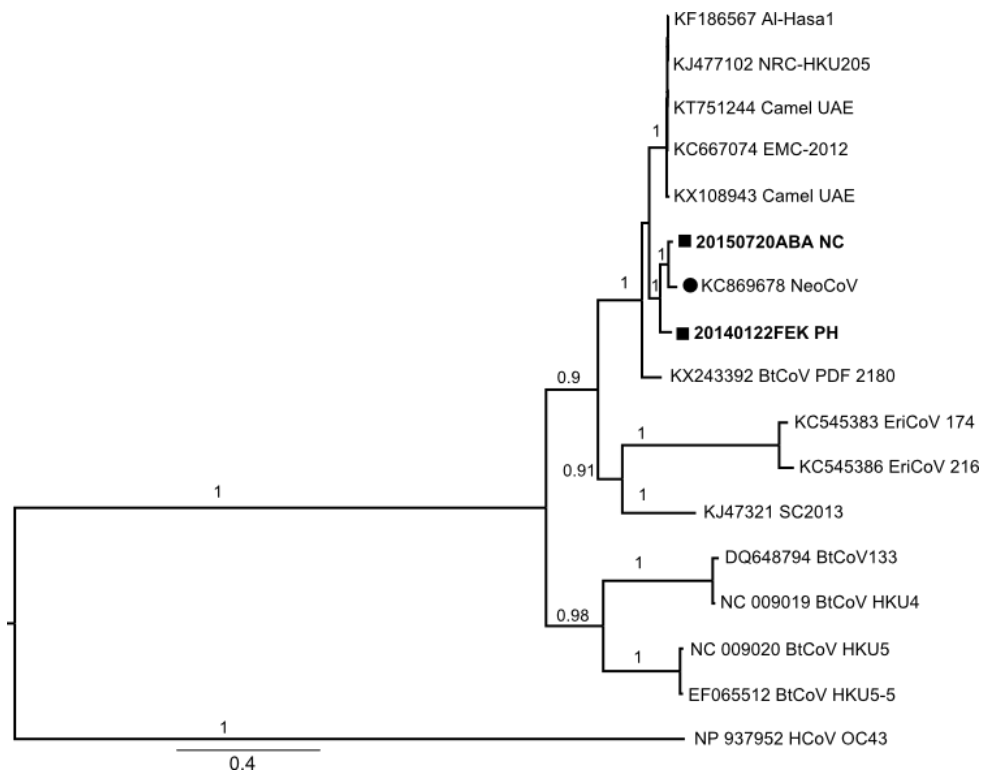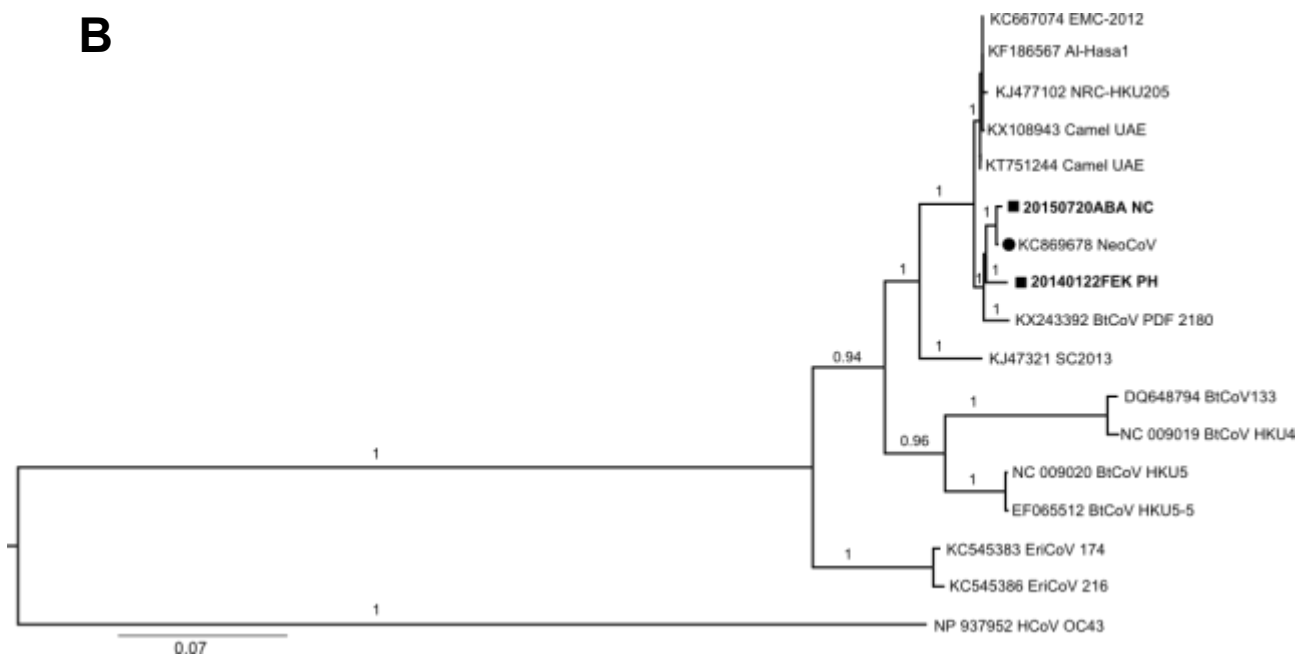
Figure 10 Bayesian phylogeny of partial ORF1a and ORF1b sequences of lineage C betacoronaviruses. The Bayesian phylogenies were constructed in Mr Bayes based on the Le Gascuel model with gamma distributed evolutionary rates (LG + G) and 2 million generations. The first 25% of sample trees were discarded, posterior probabilities less than 0.90 have been omitted. Study sample sequences are indicated in bold with ■ and NeoCoV (Genbank ID KC869678) is indicated with ●. The tree was midpoint rooted with human coronavirus OC43 (hCoV OC43) (Genbank ID NP 937952). Figure A depicts the phylogeny of partial ORF1A sequences. There were a total of 866 amino acid positions in the final dataset. Figure B depicts the phylogeny of partial ORF1B sequences. There were 1735 amino acid positions in the final dataset.
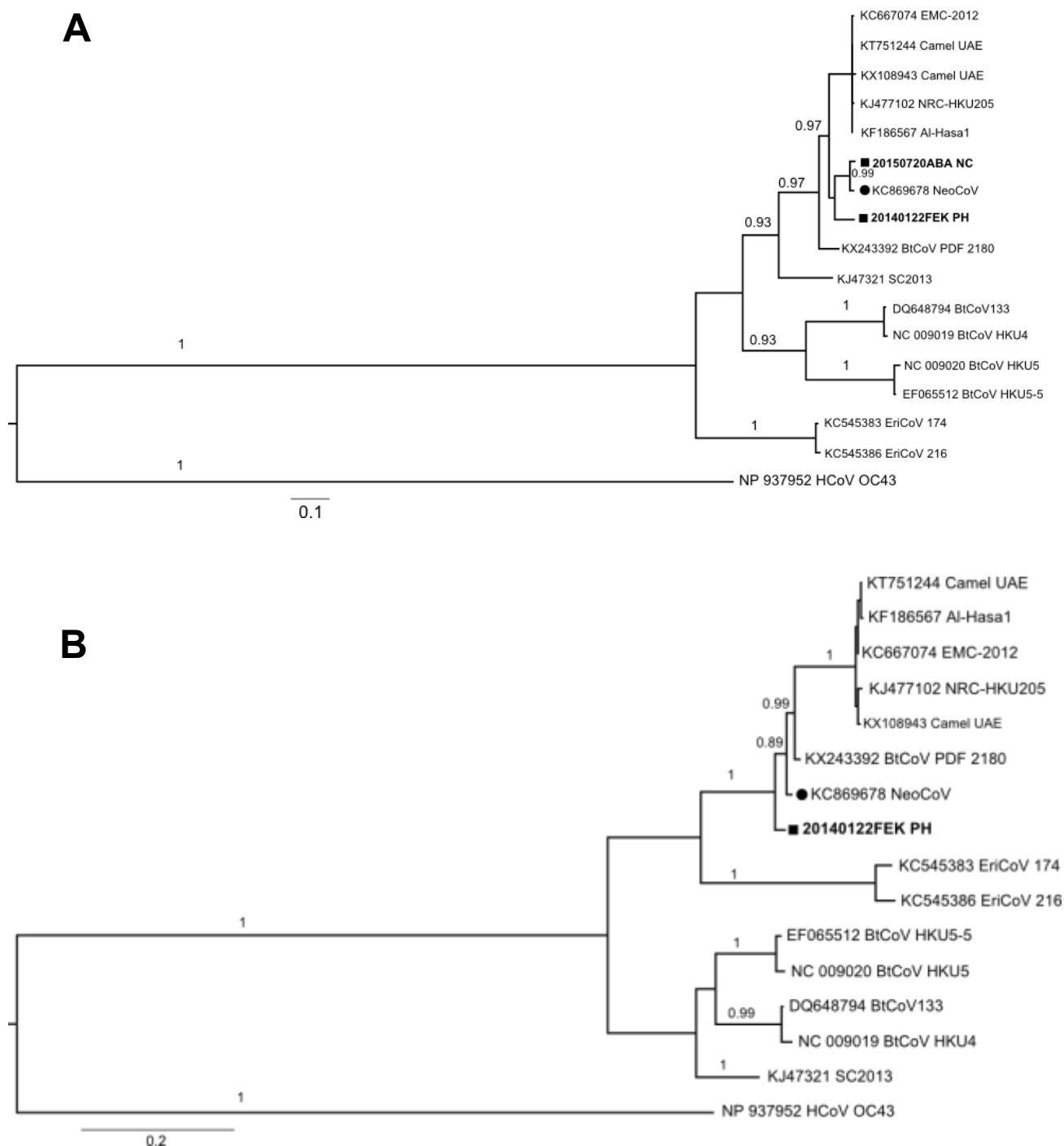
A

B

Figure 11 Bayesian phylogeny of the complete nucleocapsid (N) gene and partial Spike subunit 2 sequences of lineage C betacoronaviruses. The Bayesian phylogeny was constructed in Mr Bayes with two million generations and the first 25% of sampled trees discarded, posterior probability values less than 0.90 have been omitted. Study sample sequences are indicated in bold with ■ and NeoCoV (Genbank ID KC869678) is indicated with ●. The tree was midpoint rooted with human coronavirus OC43 (hCoV OC43) (Genbank ID NP 937952). Figure A depicts the phylogeny of the N protein based on the Le Gascuel model with gamma distributed evolutionary rates and empirical base frequencies (LG + G + F). Figure B depicts the phylogeny of partial S2 sequences based on the Le Gascuel model with gamma distributed evolutionary rates (LG + G).

APPENDIX F: Permission numbers for the use of copyrighted images

Image	Reference	Permission number
Figure 2.1	(Daszak *et al.*, 2001)	4155281269706
Figure 2.2	(Pike *et al.*, 2010)	4158820701942
Figure 2.4	(de Wit *et al.*, 2016)	4155831346829

APPENDIX G: Genbank accession numbers obtained for partial RNA dependent RNA polymerase sequences generated during this study

Sample ID	Genbank accession number
20140115AHHB_PSP1nc	MG310257
20140122FEK_NC1	MG310256
20140122FEK_PSP1	MG252861
20140122FEK_PSP4	MG310224
20140923LNR_NC1	MG205586
20140923LNR_NC12	MG205587
20140923LNR_NC14	MG205588
20140923LNR_NC3	MG193613
20140923LNR_NC8	MG310255
20140923LNR_PH1nc	MG252865
20141021AEL_PH1	MG310254
20141021AEL_PH4nc	MG310253
20141022HBI_PH10nc	MG310225
20141027ABA_PH10nc	MG310251
20141027ABA_PH12nc	MG310250
20141027ABA_PH13nc	MG310249
20141027ABA_PH5nc	MG310252
20141027ABA_PH9nc	MG193608
20141029SHM_MN1	MG310248
20141103SRP_NC2	MG205589
20150105CGR_NC2	MG205590, MG310226
20150106CDK_NC1	MG193615, MG205591
20150106CDK_NC12	MG252863
20150106CDK_NC13	MG310232
20150106CDK_NC5	MG310222
20150107CGC_NC1	MG310227
20150107CGC_NC3	MG193616
20150108CCK_NC1	MG310247
20150108CCK_NC3	MG252864, MG310235
20150108CCK_NC7	MG205592
20150108CCK_NC8	MG310228
20150312_DC34	MG310230
20150720ABA_NC1	MG205593
20150720ABA_PH1nc	MG310245
20150815HFP_NC4	MG205594
20150815HFP_PH5nc	MG310244
20150816HFP_NC1	MG205595
20150816HFP_NC2	MG310243
20150816HFP_NC3	MG205596
20150816HFP_PH1nc	MG205597
20150819LFU_PH2nc	MG310229

Sample ID	Genbank accession number
20150919CDK_NC10	MG252862
20150919CDK_NC12	MG205598
20150920CGC_NC2	MG252859
20150920CGC_NC4	MG310242
20150920CGC_NC6	MG310241
20150921SKK_RC10	MG193603
20150922CDK1_MN1	MG310240
20150922CDK1_MN2	MG310239
20150923HWK_NC2	MG205585, MG310238
20150923HWK_NC4	MG193604, MG252860
20150923HWK_NC5	MG205599
20150923HWK_NC6	MG193610
20150923HWK_NC7	MG252866
20150923HWK_NC9	MG193605
20150925KMM_RCL4	MG252867
20151209BVL1_RCL2	MG310237
20160111LNR_NC12	MG193614
20160111LNR_NC6	MG193612
20160112LNR_NC11	MG252868
20160303FEK_NC1	MG252869, MG252870
20160304FEK_NC1	MG252871, MG252872
20160304FEK_NC2	MG252873, MG252874
20160307SKK_MN2	MG310236
20160411_DC63	MG310231
20160411_DC68	MG193606
20160522_DC84	MG310234
20160906_DC93	MG252875
20160906_DC95	MG193607
20160921_DC115	MG193611
20161011_DC170	MG310233
20161014_DC_168	MG310223
20161014_DC166	MG252876
20161014_DC172	MG193617
20161020_DC253	MG193609