AIX – MARSEILLE UNIVERSITE

FACULTE DE MEDECINE DE MARSEILLE

ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

THESE DE DOCTORAT

Présentée par

Mrs Denise MARSTON

Née le 26 Juin 1976 á Newton Abbot, Royaume-Uni

En vue de l'obtention du grade de *DOCTEUR d'AIX-MARSEILLE UNIVERSITE*

Mention : Pathologie Humaine

Spécialité : Maladies Infectieuses

Utilisation of next generation sequencing to characterise novel lyssaviruses, improve phylogenetic inferences and investigate cross species transmission events

Soutenue publiquement le 17 Novembre 2017

JURY :

Monsieur le Professeur Jan Felix DREXLER	Rapporteur
Monsieur le Docteur Nőel TORDO	Rapporteur
Monsieur le Professeur Anthony FOOKS	Co-directeur de thése
Madam le Docteur Lorraine MCELHINNEY	Co-directeur de thése
Monsieur le Professeur Xavier de LAMBALLERIE	Directeur de thése

UMR "Emergence des Pathologies Virales" (EPV : Aix-Marseille Univ – IRD 190 – Inserm 1207 – EHESP – IHU Méditerranée Infection), Marseille, France

Wildlife Zoonoses & Vector-Borne Diseases Research Group, Animal and Plant Health Agency, UK

Acknowledgements

As a more 'mature' student I fear this might be lengthy, as so many people have been influential in getting me to this stage!

I would like to begin by thanking my previous work colleagues at IAH (Compton), in particular **Professor Jim Kaufman** who quietly and regularly asked me to consider a PhD. My response was always the same: I love being in the lab, I have no desire to run a research group and be frustrated with being cooped up in an office all day. My early 20's 'black and white' opinions have since mellowed; and here I am, grateful that you had the confidence in me, even then, to complete a PhD. I have often thought of those conversations whilst writing this thesis, (indeed stuck in an office!) but with the opportunity to expand my hypotheses, design and complete experiments and tackle the highs and lows, all skills influenced by you. **Dr Sally Rogers**, **Fiona Gadd, Dr Iain Shaw, Dr Tim Powell** and **Dr Jansen Jacob**, your expertise and work ethics inspired me in my scientific 'salad days' and gave me the broad scientific foundation which has served me so well so far.

Dr Pete Kaiser and **Dr Phil Wakeley**, both exceptional scientists in their own fields, taken from our lives far too quickly. You influenced my early scientific careers in very different ways. Pete, you led by example, and had the patience of a Saint! Phil, when I first arrived at APHA you blew my mind with your molecular diagnostic experience.

To my APHA colleagues past and present, there have been some interesting characters over the years, but I can honestly say VI1 really is VIfun! I can't name everyone, but thank you all! **Emma Wise** and **Daisy Jennings** do get a special mention. You have both been awesome to manage and have helped in so many ways throughout my PhD. You are both studying for your own higher degree qualifications, I know you will both complete them and move onto further successes.

Special thanks go to my first manager and co-supervisor **Dr Lorraine McElhinney**. You have given me the encouragement and confidence to complete this PhD and become the scientist I am today. Your selflessness has not gone unnoticed, I am where I am today because of you.

Prof Tony Fooks, apart from Jim, you have been the biggest influence in my scientific career, you are both completely different, but both have shown unfaltering confidence in my abilities as a scientist. You have given me so many opportunities to develop, scientifically, but also as a manager, scientific paper writer and presenter and I will always be grateful.

Dr Ash Banyard, thank you for supporting me during this PhD, I am completely aware that at times what you needed me to do (my job) and what I was doing (my PhD) were not compatible! Your support is hugely appreciated.

Dr Richard Ellis, the sequencing guru! I am truly proud of our development of virus NGS capabilities. Thank you for your patience and responses to my many requests, this PhD wouldn't have been possible without you.

Dr Greg Moureau, you are the reason I started this PhD, I will never forget Tony returning from your defence and discussing the possibility of me registering for a manuscript-based thesis. Your thesis has been next to me these last few months, being remote from Marseille, it was the tangible link I had to the University and

to the end result. **Géraldine Piorkowski**, **Soline Buisine and Raphaelle Klitting**, thank you for helping me throughout with so many questions and with the French administration!

Prof. Xavier de Lamballerie, thank you for agreeing to be my supervisor! I am truly honoured to have had the opportunity to work in your laboratory. I have enjoyed my times spent in Marseille and hope our collaborations continue.

To the many collaborators I have the honour of working with, I thank you. Without you, the papers contained within this thesis and others besides, would not have been written. In particular, to **Dr Thomas Müller** and **Dr Conrad Freuling** at FLI, **Dr Kirstyn Brunker** and **Dr Katie Hampson** in Glasgow and **Dr Claude Sabeta** in RSA. The rabies field is uniquely collaborative, and supportive!

Dr Dan Horton, from the time you started your PhD at APHA to now, we have both spent time training and mentoring each other. Thank you for all your support, particularly on the host shift manuscript, I look forward to many future collaborative ventures with you.

To my amazing husband Dave and my gorgeous boys James and Tommy, you are my world. Thank you for supporting me through this journey without complaint. I could have started my PhD in Compton in my 20's like everyone else, but it was not the right time for me. Completing it with a young family was not without its challenges, but I am proud of what I have achieved and hope you are too.

Finally, thanks to my dear supportive friends and family, you have kept me going with plenty of wine and good times! Not many people have the chance to co-author a paper with a family member; **Dr Barry Atkinson**, article 23 – we need to sign and frame it for Mum and Dad!

Abstract

Viruses are microscopically small particles, unable to replicate without infecting other living organisms. Common to all three domains of life, viruses successfully thrive within eukaryotic, bacterial and archaeal hosts, ever present throughout the history of modern organisms. Zoonotic viruses have posed significant threat to humans, jumping from animal reservoirs into the human population. Rabies virus (RABV) is a classic zoonotic virus, with many animal host reservoirs, including the domestic dog (*Canis lupus familiaris*). *Rabies lyssavirus* is the prototype species in the *Lyssavirus* genus, which currently has 16 species. All lyssaviruses cause the disease rabies, which is contracted via a bite from an infected animal, and once symptoms appear, death is inevitable. Unlike many other zoonotic viruses (such as Ebola or influenza) transmission between people rarely occurs. With cheap effective vaccines available, rabies is an entirely preventable disease. Despite this, tens of thousands of people die every year. Rabies is considered by the WHO (World Health Organisation) as a neglected tropical disease, alongside HIV, malaria and tuberculosis. Within the past five years, a challenging target date of 2030 to eliminate human-dog-mediated rabies has been set. One of the many challenges will be preventing re-emergence of RABV in the dog population after elimination, due to multiple animal reservoirs including wolves, foxes, raccoons and bats. Pivotal to preventing re-emergence in the dog population is understanding how lyssaviruses cross species barriers and maintain in new host populations, termed 'host switching'. RABV circulates within specific host reservoirs, evolving and adapting; and although cross species transmission (CST) events occur, they are rarely followed by onward transmission. Unlike the majority of lyssaviruses, RABV has multiple animal host reservoirs. Why there is such a difference is unknown, hypothesised in this thesis to be due to viral heterogeneity.

By utilising HTS (high throughput sequencing), whole genome sequences can be generated from clinical or cultured viruses, and the depth of coverage generated using these processes can be interrogated to identify and quantify diversity within the sub-consensus viral population. At the beginning of this PhD, the technologies which were developed and refined for human and bacterial genomes were still in their infancy for viral genomics. One might be forgiven for thinking that obtaining a complete genome for a virus which is 12,000 bases long may be easier than for a bacterium; however there are a number of virus specific critical issues, including low levels of viral RNA in clinical samples. Before attempting an investigation of viral heterogeneity, developing sample preparation methodologies, followed by bioinformatics to analyse the 'big data' obtained from the sequencing machines were essential. Once HTS was established, the genome sequences obtained were analysed to investigate the origins of lyssaviruses, transmission of RABV within dogs in Tanzania, and asses a host shift event from dogs to foxes. High viral heterogeneity was seen in fox samples after the host shift than in other fox samples from different regions of Turkey, suggesting it

is important for a successful host shift. Complete genome sequences have provided fine-resolution for highly related viruses, to a level inconceivable with partial genome data. Furthermore, HTS enabled assembly of genome sequences from novel viruses.

The resilience and adaptation of viruses to ever changing environments, whether natural or due to human intervention, should never be underestimated. Improvements in sequencing capabilities have revolutionised our ability to obtain complete viral genome sequences and to investigate the sub-consensus population at an unprecedented level. These data will be pivotal to improve our understanding of how viruses evolve in new hosts and change over time, informing policy makers as eradication plans are implemented.

List of Figures

List of Tables:

Table of Contents

List of Abbreviations

ABLV: Australian bat lyssavirus

ARAV: Aravan virus

ANJV: Anjozorbe virus

APC: Affinity propagation clustering

BBLV: Bokeloh bat lyssavirus

BEAST: Bayesian Evolutionary Analysis Sampling Trees

BI: Bayesian Inferences

bp: base pairs

BWA: Burrow-Wheeler algorithm

CNS: central nervous system

DI: Defective Interfering

DUVV: Duvenhage lyssavirus

dRIT: direct Rabies Immunoglobulin Test

EBLV-1 and -2: European bat lyssavirus type -1 and -2

FAT: Florescent Antibody Test

FLI: Frederick Loeffler Institute

FMDV: Foot and Mouth Disease Virus

FTA: Flinders Technology Associates

G: Glycoprotein

GBLV: Gannoruwa bat lyssavirus

gDNA: genomic DNA

HCV: hepatitis C virus

HIV: human immunodeficiency virus

hnRT-PCR: hemi-nested Reverse Transcriptase Polymerase Chain Reaction

ICTV: International Committee for the Taxonomy of Viruses

IKOV: Ikoma virus

Indel: Insertion or deletion in one sequence relative to others

IRKV: Irkut virus

ISKV: Issyk-Kul virus

IVA: (Iterative Virus Assembler)

KHUV: Khujand virus

L: Large protein (or RNA-dependent-RNA polymerase)

LBV: Lagos bat virus

LLEBV: Lleida bat virus

M: Matrix Protein

MCC: Maximum Clade Credibility

ML: Maximum Likelihood

MOKV: Mokola virus

MRCA: Most recent common ancestor

N: Nucleoprotein

nts: nucleotides

OIE: World Health Organisation

ORF: open reading frame

P: Phosphoprotein

PHE: Public Health England

RABV: Rabies virus

RdRP: RNA dependent RNA polymerase

RER: rough endoplasmic reticulum

RNP: Ribonucleoprotein

rRNA: ribosomal RNA

RT-PCR: Reverse Transcription – Polymerase Chain Reaction

SNP: single nucleotide polymorphism

TCSN: tissue culture supernatant

TMRCA: Time to most recent common ancestor

UTR: Untranslated region

WGS: Whole genome sequencing

WCBV: West Caucasian bat virus

Chapter 1: Introduction

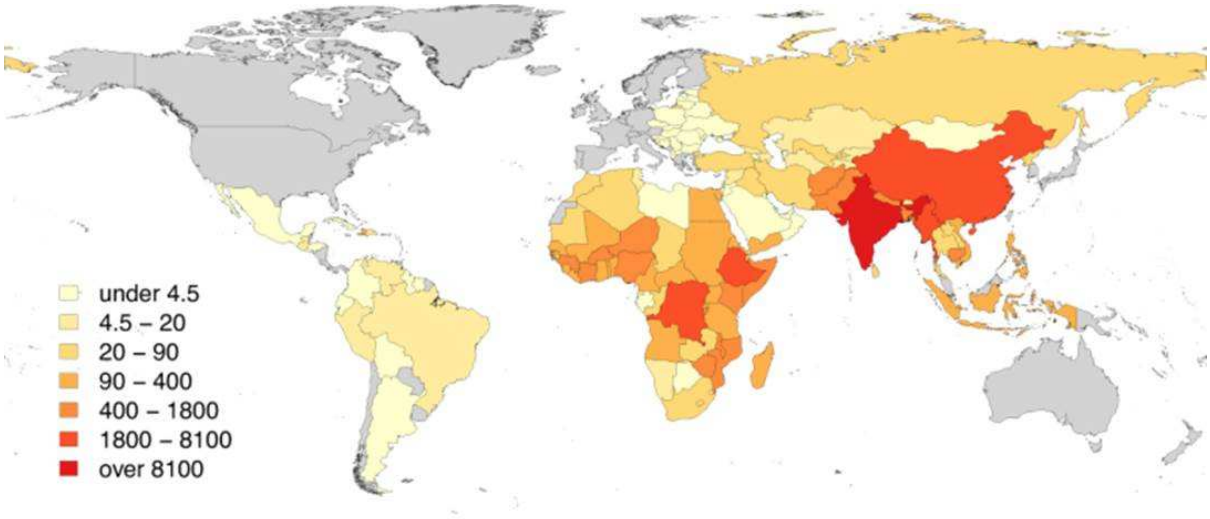
Chapter 1: Introduction

1.1 Lyssavirus Introduction

Lyssaviruses are negative sense non-segmented, single stranded RNA viruses, which are of significance to both human and animal health. All viruses within this genus are highly neurotropic and are capable of causing a fatal encephalitis, known as rabies. The prototypic lyssavirus is rabies virus (RABV), which is primarily transmitted through the bite of an infected dog and causes a higher burden of disease in both humans and animals than any other lyssavirus. Indeed, the term 'lyssa' is of Greek origin and is believed to refer to a 'raging dog' (Neville, 2004). Of the remaining lyssaviruses, all have been detected in bats with only two exceptions, Mokola virus and Ikoma lyssavirus. Despite the availability of safe and effective prophylactic and post-exposure tools, RABV remains endemic across much of the globe, particularly in Africa and the Indian sub-continent. Lyssaviruses are able to infect and cause disease in a wide host range, with all mammalian species being theoretically susceptible to infection (Rupprecht et al., 2017) . Susceptibility to infection varies with several factors hypothesised to contribute to the duration of the incubation period, progression of disease and overall outcome of infection including: site of wound, virus dose, infecting strain and immunocompetency of the host. Despite the availability of tools to control rabies infection, the presence of virus in sylvatic populations has made elimination difficult and eradication almost impossible. From a human perspective, there is effective prophylactic treatment which, if taken before symptoms appear, prevents progression of the disease, resulting in full recovery of the patient. However, in the majority of cases prophylaxis is not given due to a range of factors including a lack of awareness of the disease, misdiagnosis and a distrust of Western medicine. As a result, clinical disease develops, inevitably resulting in a fatal outcome. The most recent estimations of human mortality due to rabies is 59,000 (25,000-159,000) (Fooks et al., 2014; Hampson et al., 2015) (Figure 1), although there are recognised limitations with these estimations, mainly due to poor surveillance data (Taylor et al., 2017).

Figure 1: The distribution of the global burden of rabies: human rabies deaths.

Modified from Hampson et al., 2015



A target for global elimination of human and dog rabies by 2030, using a 'One Health' approach has been set by WHO (World Health Organisation) and PAHO (Pan American Health Organisation). Vaccination and sterilisation of free-roaming dogs is an essential part of the elimination program, which although has recorded successes in Latin America and small targeted regions in Africa and the Indian sub-continent, is significantly behind in the rest of the world. One major limitation is the amount of canine rabies vaccine available. The World Animal Health Organization (OIE) established a canine rabies vaccine bank in 2012, with multi-donor support to assist countries with mass-dog vaccination campaigns (OIE, 2013). It has been predicted that almost 1.3 billion doses (more than 260 million doses annually including a 10% wastage rate) are needed from 2015 to 2019 alone to efficiently tackle the issue of dog mediated rabies (Rupprecht et al., 2017).

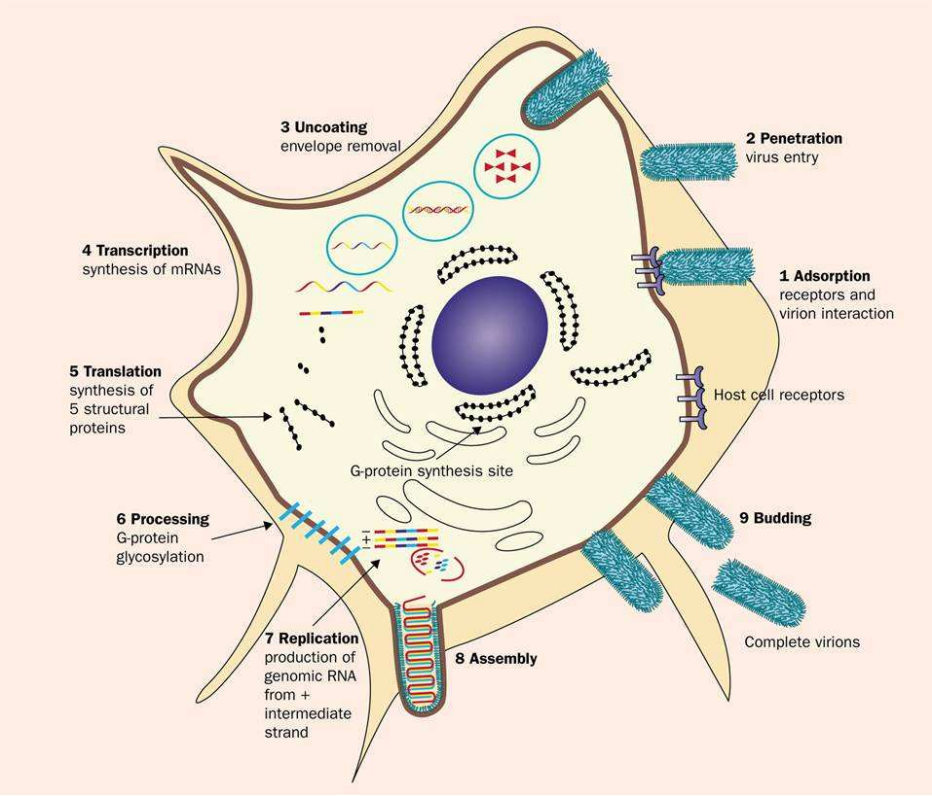
The success of Latin American countries in reducing dog-mediated rabies have provided an excellent foundation for others, including quantifying the economic advantages (Shwiff et al., 2013). There have been some early successes in reducing dog-mediated rabies on island nations. In particular, Sri Lanka is on target to declare freedom from dog-mediated rabies by 2020 (Lionel Harischandra et al., 2016), with the Philippines also making excellent progress (Ferguson et al., 2015). Island nations are inherently good locations to begin eradication programmes, with distinct boarders and minimal entry points.

1.2: Lyssavirus Life Cycle

Rabies is most commonly transmitted via infected saliva transfer from an animal bite. It is thought that the virus initially replicates in muscle cells at the site of the wound (Bellinger et al., 1983) and then is able to enter either motor or sensory nerves to travel by retrograde transport along axons to the central nervous system (CNS) (Lo et al., 2001; Mrak and Young, 1993). Virus particles are able to travel along axons in a retrograde fashion at a rate of 50-100mm/day (Tsiang et al., 1991). The mechanism of axonal transport may in part be due to interactions between the P protein and actin and microtubule networks including the microtubule motor molecule, dynein LC8 (Jacob et al., 2000; Raux et al., 2000). The virus is able to spread rapidly within the CNS due to the ubiquitous nature of the proposed receptors whilst movement in the periphery appears limited, which restricts the spread of virus in these areas (Ugolini, 2008). Once the virus reaches the brain it continues to replicate and spread through the tissue via cell-to-cell transmission or trans-synaptic spread. From the brain, the virus is able to travel along neurones to the salivary glands for onward transmission and also to many other areas of the body (Jackson et al., 1999). At the individual neuron level, replication can be divided into a number of different processes, see below and Figure 2.

Figure 2: Conceptual flow of lyssavirus life cycle in a generalised host cell.

Taken from (Rupprecht et al., 2002)



1.2.1: Adsorption

The lyssavirus glycoprotein is the main viral protein that enables virion attachment and entry into host cells. Whilst there are postulated to be as yet undefined receptors utilised by lyssaviruses for cell entry, several receptor molecules have been proposed to be involved in virus binding and entry (Lafon, 2005). Once the virus has entered the host it may either enter the nervous system directly or may initially replicate in the muscle tissue surrounding the locus of infection before entering the nervous system. The exact sites of virus replication prior to entering the CNS is yet to be defined, however, the suggestion of low level replication in the non-neuronal tissue following infection may provide some explanation for the long incubation periods, which have been reported (Lafon, 2005). It has been demonstrated *in vitro* that the virus is able to enter motor nerves via neuromuscular junctions (NMJs). As well as this, viral antigen has been shown to be present in proprioceptors, sensory spindles and stretch receptors *in vitro* (Lewis et al., 2000).

The virus is able to enter neural tissues via one/a combination of three proposed receptors which are present in all mammals: the neural cell adhesion molecule (NCAM) which is present in NMJs at post-synaptic membranes (Thoulouze et al., 1998), the nicotinic acetylcholine receptor (nAchR) which is involved in interneuronal signalling within both the peripheral and central nervous systems (Gastka et al., 1996) and p75NTR which is a neurotrophin receptor responsible for apoptosis, axonal elongation and synaptic transmission (Dechant and Barde, 2002). Each of these three proposed receptors are not only expressed in the CNS but can be expressed on cells in the dermis; NCAM and p75NTR are expressed from dermal fibroblasts associated with hair follicles (Botchkareva et al., 1999; Muller-Rover et al., 1998) whilst nAchR has been detected in the squamous epithelium (Arredondo et al., 2002). This suggests that they may be available around sites of infection as a route to the CNS for the virus; however the extent to which rabies and the lyssaviruses utilise each of these proposed receptors is undefined.

1.2.2: Penetration and Uncoating

Upon attachment to a permissive cell, the virus enters via fusion of the viral envelope with the plasma membrane of the cell (Wunner and Conzelmann, 2013). The virus is also able to enter cells through attachment and engulfment by coated pits (viropexis) or via internalisation of vesicles (pinocytosis), both of which can result in the internalisation of multiple virions simultaneously (Tsiang, 1993). Once internalised, the G protein initiates fusion of the viral envelope with the endosomal membrane to release the RNP into the endosomal vesicles (Wunner and Conzelmann, 2013). Lysosomes then fuse

with these vesicles to release the RNP into the cell cytoplasm (Gosztonyi, 1994) to initiate transcription (Finke and Conzelmann, 2005).

1.2.3: Primary Transcription

The lyssavirus genome consists of a strand of linear, negative sense, single stranded RNA, therefore virus replication requires transcription of mRNA. The L protein, as a RdRP, initiates transcription of the genomic RNA to produce leader RNA which is an exact complement of the 3' end of the genome RNA. It is phosphorylated at the 5' end but has no 3' poly(A) tail. Five complementary positive sense mRNAs are also transcribed which are capped at the 5' end and polyadenylated at the 3' end, each corresponding to one of the five viral genes (Banerjee, 1987). The transcripts are generated in consecutive mono-cistronic fashion with the sequential nature of transcription controlled by the transcriptional complex. This complex is able to recognise the start, stop and polyadenylation signals which flank each gene. A gradient effect of transcriptional efficiency has been observed from the 3' to 5' genes. This process allows structural proteins at the 3' end of the genome to be produced in higher concentrations than the 5' proximal genes (Tordo and Kouknetzoff, 1993).

1.2.4: Translation and Protein Processing

The viral mRNAs are translated using the host cell machinery. The N, M, P and L proteins are translated on free ribosomes within the cytoplasm whereas the mRNA encoding G is translated on membrane bound ribosomes (Tordo, 1996). Completion of the G processing is carried out co-translationally: firstly, core glycosylation occurs in the rough endoplasmic reticulum (RER) and this is followed by final processing of the carbohydrate side chains in the RER and Golgi body (Shakin-Eshleman et al., 1992). The now glycosylated transmembrane G protein is transported to the cell membrane, the glycosylated ectodomain protruding from the exterior of the cell (Tordo, 1996).

1.2.5: Replication

Although the exact mechanisms that drive a switch from transcription to replication remain unclear, at some stage following infection a switch occurs to drive the generation of a positive strand full length genome replicative intermediate. This replication is dependent on continued synthesis of viral N protein. Once a full length antigenome has been produced, this acts as the template for genome replication to produce nascent genome sense RNA. During the switch from primary transcription to

genome replication the activity of the viral polymerase changes from a transcriptive to a replicative mode. As the positive sense RNA is synthesised it is immediately associated with the newly generated N protein to form positive sense RNP and it is this which acts as the template for replication. Negative sense genome is then amplified; again with the negative sense RNA being immediately encapsidated by the N protein. Once the negative sense RNP has been amplified this acts as a template for secondary transcription which results in amplification of all viral proteins for subsequent virion assembly (Banerjee, 1987).

1.2.6: Assembly

Assembly begins with the formation of the RNP during replication. M then associates with the RNP complex in the cell cytoplasm and down-regulates RNA transcription while up-regulating replication (Finke and Conzelmann, 2003). M initiates tight coiling of the RNP complex and localises the coiled RNP-M at the cell membrane. At the membrane of the cell the M protein bridges the RNP and the cytoplasmic tales of G and as the components assemble the polymerase activity of L is halted by M within the very tightly coiled RNP and virus budding is initiated (Wunner and Conzelmann, 2013).

1.2.7: Budding

Once the RNP-M complex has associated with the transmembrane G proteins at the cell membrane, the mature virions bud through the cell membrane, acquiring their lipid bilayer envelope during the process (Wunner and Conzelmann, 2013). The interaction of the G cytoplasmic tail with the M protein within the virion is thought to hold the virus in its classic bullet-like morphology It is possible that the RNP-M complex is able to bud though the cell membrane in areas where no viral G trimers are expressed and in this case the viral particles produced are non-infectious (Mebatsion et al., 1996a). It has also been observed that virions may bud through the endoplasmic reticulum or Golgi body within the cells (Lahaye et al., 2009) into the lumen of vesicles produced by these membranes. In this case it is possible that these vesicles would be secreted from the infected cells by the normal secretory pathway (Wunner and Conzelmann, 2013).

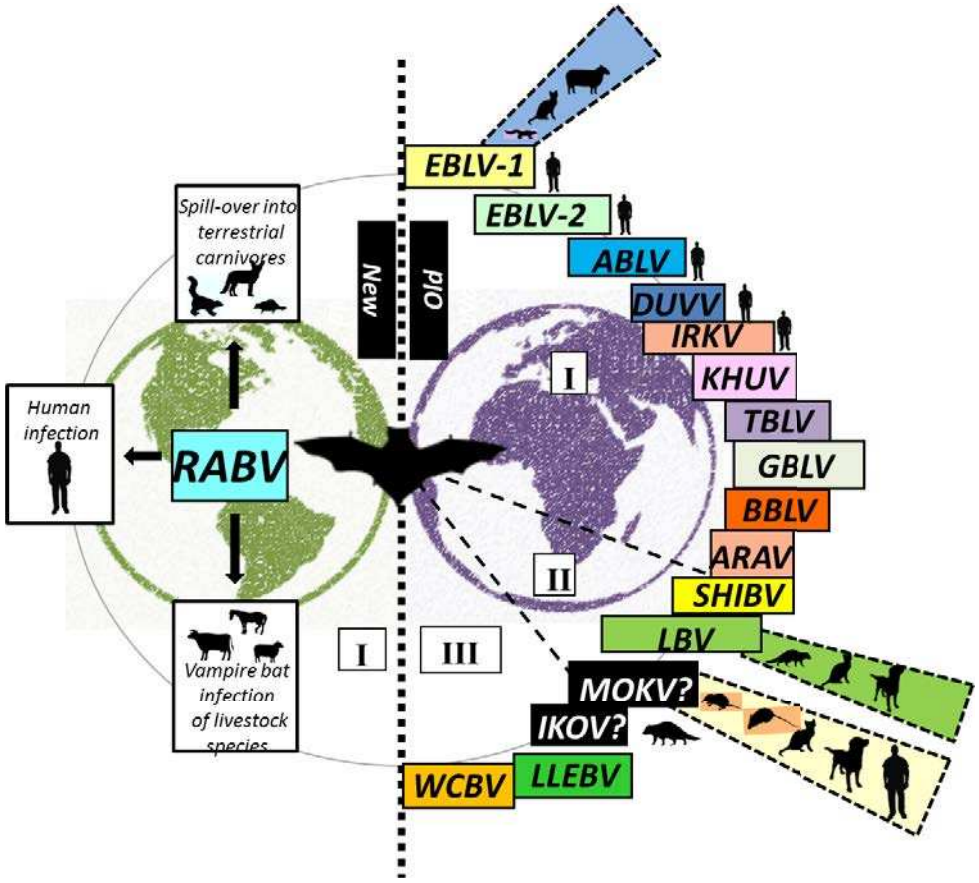
1.3: Lyssavirus Epidemiology

RABV is distributed globally with only a few, principally island nations, being free of disease (Fooks et al., 2014). Concerted vaccination campaigns with domestic animals across both the USA and Western

Europe have now eliminated the burden of rabies from domestic carnivore populations. In the USA, sylvatic rabies remains a constant threat to human populations with racoons, skunks and foxes acting as reservoirs for the classic rabies viruses but with bats causing increasingly more human infections (Blanton et al., 2011). Extensive wildlife vaccination campaigns have managed to eliminate sylvatic rabies from Western Europe. However, globally the existence of lyssaviruses circulating in Chiropteran reservoirs remains and so a complete eradication of rabies remains unrealistic (Rupprecht et al., 2017; Rupprecht et al., 2008). Within the USA, bat species maintain RABV whilst, in contrast across the Old World, classical RABV appears to be completely absent from bat populations, instead, all non-RABV lyssaviruses persist (apart from MOKV and IKOV) (Figure 3). These populations of bats are very difficult to access, and across the EU, are protected by law (Freuling et al., 2012) making eradication of lyssaviruses an unfeasible target.

Figure 3: Global distribution of lyssaviruses in bats.

All lyssaviruses have been isolated from bats apart from MOKV and IKOV. Spill over events of lyssaviruses are indicated (terrestrial RABV distribution is not described here). Adapted from figure by Dr Ashley Banyard.



1.3.1: Lyssaviruses in Europe

Alongside RABV in terrestrial wildlife populations four other lyssavirus species have been identified: *European Bat type-1 Lyssavirus* (EBLV-1), *European Bat type-2 Lyssavirus* (EBLV-2) *Bokeloh Bat Lyssavirus* (BBLV) and *Lleida lyssavirus* (LLEBV)

EBLV-1 is genetically classified into two lineages; EBLV-1a is found throughout northern Europe and EBLV-1b has been detected in southern Germany, France and Spain. Both EBLV-1 lineages are associated with infection of *Eptesicus fuscus*, the Serotine bat. Infection of other bat species such as the Meridional Serotine (*E. isabellinus*) may suggest the circulation of a third lineage of EBLV-1 (Banyard et al., 2011; Vazquez Moron et al., 2011). EBLV-1 cross species transmission (CST) events

have occurred at a low frequency, although independent terrestrial transmission pathways have not been established (Dacheux et al., 2009; Muller et al., 2004; Tjornehoj et al., 2006).

In contrast, EBLV-2 is predominantly associated with infection of *Myotis daubentonii*, the Daubenton's bat, across much of northern Europe and no cross species transmission events have been reported in terrestrial species although two fatal human cases have occurred (Fooks et al., 2003; Lumio et al., 1986).

The first isolate of BBLV was obtained from a Natterer's bat (*Myotis natterei*) in Germany in November 2009 (Freuling et al., 2011). Since then further detections have been reported including another in Germany (Freuling et al., 2013) and France (Picard-Meyer et al., 2013).

LLEBV was identified in Spain in a common bent-wing bat (*Miniopterus schreibersi*). Initial sequencing of the N gene suggests considerable genetic divergence from the current European lyssaviruses (Arechiga Ceballos et al., 2013).

1.3.2: Lyssaviruses in Eurasia

A further six Eurasian lyssaviruses have been described, predominantly from single virus isolates from different bat species. Aravan virus (ARAV) was isolated from a lesser-mouse eared bat (*Myotis blythii*) in Kyrgyzstan (Botvinkin et al., 2003), Khujand virus (KHUV) isolated from a whiskered bat (*Myotis mystacinus)* in Tajikistan (Kuzmin et al., 2008), Irkut virus (IRKV) isolated from a Greater Tube-nosed bat (*Murina leucogaster*) in Eastern Siberia (Botvinkin et al., 2003) and West Caucasian Bat virus (WCBV) isolated from a common bent-wing bat (*Miniopterus schreibersi*) in the Caucasus mountains in Russia (Botvinkin et al., 2003). More recently, Gannoruwa bat lyssavirus (GBLV) was isolated from four Indian flying foxes (*Pteropus medius*) in Sri Lanka (Gunawardena et al., 2016), discussed later in this thesis (4.2.7 and 4.2.8). The most recently identified lyssavirus (still yet to be formally classified) is Taiwan bat lyssavirus (TWBLV) identified on two separate occasions in Taiwan in Japanese house bats (*Pipistrellus abramus*). Phylogenetic classification of these Eurasian isolates has shown them to be distinct species within the *Lyssavirus* genus with WCBV being the most genetically divergent (Horton et al., 2010). Of these, only Irkut virus has been isolated on more than one occasion (Liu et al., 2013b) and has been associated with human fatality (Belikov et al., 2009).

1.3.3: Lyssaviruses in Australasia

Australian bat lyssavirus (ABLV) has been isolated from five different bat species since its initial isolation in 1996 (McCall et al., 2000) and has also been implicated in three human fatalities (Allworth et al., 1996; Hanna et al., 2000; Weir et al., 2013). There have also been two confirmed cases of infection in horses with ABLV in Queensland (Symons, 2013), which have raised concerns about a theoretical yet possible risk of transmission of ABLV infection from horses to humans.

1.3.4: Lyssaviruses in Africa

Across Africa a number of bat lyssaviruses have been detected. The most frequently reported is Lagos bat virus (LBV) which has been isolated from multiple bat species distributed across sub-Saharan Africa (Hayman et al., 2008). There are four defined lineages of LBV; lineage A consists of Kenyan (2007), Senegalese (1985) and a French (1999) isolates, lineage B contains only of the original Nigerian (1956) isolate, lineage C contains isolates from Zimbabwe, South Africa and the Central African Republic and lineage D contains a single isolate identified in an Egyptian fruit bat in Kenya (Kuzmin et al., 2010; Markotter et al., 2008). It has been suggested that Lagos bat virus isolates could be subdivided into separate species based on criteria adopted by the ICTV whereby all isolates belonging to the same species have identities of greater than 80-82%. LBV isolates can be more divergent than this, within the variable glycoprotein, across each lineage and as such may ultimately be re-classified.

Duvenhage virus (DUVV) was initially isolated from a fatal human infection following a bat bite and was subsequently isolated from a number of bat species, mainly Nycteris species in South Africa and Zimbabwe. There have been two further cases of DUVV infection in humans, both of African origin (Paweska et al., 2006). Only a single isolation of Shimoni bat virus (SHIBV) has been made in 2009 from a striped leaf-nosed bat (*Hipposideros vittatus*) in Kenya (Kuzmin et al., 2010).

Mokola virus (MOKV) is an African lyssavirus, which is only one of two lyssavirus species (*Mokola lyssavirus* and *Ikoma lyssavirus*) that has not been detected within bat species. MOKV has been isolated from a range of mammals including a shrew, domestic cats and a dog. Most of the isolates have been from South Africa and neighbouring Zimbabwe but there have also been sporadic isolations from Nigeria, Cameroon, Central African Republic and Ethiopia. Interestingly, sera from LBV positive bats have been shown to neutralise MOKV making an assessment of the range of species infected with MOKV difficult and as a result bats cannot be ruled out as a reservoir for MOKV (Kgaladi et al., 2013). There have been two recorded human cases of MOKV infection (Familusi and Moore, 1972; Familusi et al., 1972).

The most recent African lyssavirus to be identified is Ikoma lyssavirus (IKOV) which is discussed later (4.2.1-4). IKOV was isolated from an African civet and is the second lyssavirus to have been isolated only from a terrestrial carnivore species although its origin within Chiropteran species is assumed from the circumstances within which it was detected (Marston et al., 2012b). IKOV is genetically divergent from all the other lyssaviruses (Marston et al., 2012a).

1.3.5: Host reservoirs and cross species transmission

Due to the direct mode of transmission (usually through saliva via a bite), lyssaviruses evolve within host conspecifics. Bats are the most common host reservoir for lyssaviruses, with all but MOKV and IKOV being identified in bats (Figure 3). In the New World species of insectivorous, frugivorous and hematophagous bats all represent important reservoirs of RABV. In contrast, RABV has never been detected in Old World bat populations despite being endemic in terrestrial mammals (Figures 3 and 4). Instead, both insectivorous and frugivorous bat species of across the Old World appear to act as reservoirs for the non-RABV lyssaviruses. Furthermore, RABV is the only lyssavirus which has over time switched hosts multiple times, all other lyssaviruses have evolved closely with a specific host reservoir with limited spill overs documented.

Figure 4: Major host reservoirs of RABV across the globe.

1.3.5.1: Article 1: In review with Current Opinions in Virology:

The lyssavirus host-specificity conundrum – rabies virus – the exception not the rule.

Summary:

Lyssaviruses are a diverse range of viruses which all cause the disease rabies. Of the 14 recognised species, only rabies viruses (RABVs) have multiple host reservoirs. Although lyssaviruses are capable of infecting all mammals, onward transmission in a new host reservoir requires adaptation of the virus, in a number of stages with both host and virus factors determining the outcome. How RABV has historically and contemporarily successfully crossed species barriers, resulting in the global multi-host pathogen it is today, when other lyssaviruses remain in single host reservoirs is explored in this review.

Although many bat species globally act as natural reservoir host for lyssaviruses, two bat species *Eptesicus* and *Myotis* spp in the Americas and Europe are identified for further investigation because in the Americas RABV circulates in both bat species, whereas in Europe, *Eptesicus* spp are reservoir hosts for EBLV-1 and *Myotis* spp are host for EBLV-2 and BBLV. Historical co-evolution of lyssaviruses with specific bat species over time is the most probable explanation for these observations, but our interest is to understand why RABV is the only lyssavirus that has successfully crossed species barriers multiple times to become the multi-host pathogen it is today.

We summarise a number of manuscripts investigating RABV host shifts concluding that no positive selection is occurring and expand on the previously suggested hypothesis that pre-adaptation may have occurred. We suggest that pre-adaptation in a RABV ancestor of the nonvolant lineage obtained a preadaptation that continues to the extant viruses observed today, resulting in increased host shifts without requiring positive selection. Furthermore, we suggest a pre-adaptation in an ancient RABV, ABLV and GBLV ancestor which increased host shifts with the *Chiroptera* order. What the pre-adaptations might be is discussed, including increasing virulence, cell replication, or increasing the viral heterogeneity to circumvent the host shift bottleneck.

The lyssavirus host-specificity conundrum – rabies virus – the exception not the rule.

Denise A. Marston[a,b], Ashley C. Banyard[a], Lorraine M. McElhinney[a,c], Conrad M. Freuling[d], Stefan Finke[d], Xavier de Lamballerie[b], Thomas Müller[d], Anthony R. Fooks[a,c]

[a]Wildlife Zoonoses & Vector-Borne Diseases Research Group, Animal and Plant Health Agency, New Haw, Addlestone, Surrey, United Kingdom

[b]UMR "Émergence des Pathologies Virales" (EPV: Aix-Marseille Univ – IRD 190 – Inserm 1207 – EHESP – IHU Méditerranée Infection), Marseille, France

[c]Institute of Infection and Global Health, University of Liverpool, United Kingdom

[d]Friedrich-Loeffler-Institute, (FLI), Institute of Molecular Virology and Cell Biology, Greifswald-Insel Riems, Germany

*Corresponding Author: Email: Tony.Fooks@apha.gsi.gov.uk

Total word count: max. 2000-2700 (not including abstract or references) **2258**

Abstract (100-200 words) **162**

Lyssaviruses are a diverse range of viruses which all cause the disease rabies. Of the 16 recognized species, only rabies viruses (RABV) have multiple host reservoirs. Although lyssaviruses are capable of infecting all mammals, onward transmission in a new host population requires adaptation of the virus, in a number of stages with both host and virus factors determining the outcome. Due to an absence of recorded non-RABV host shifts, RABV data is extrapolated to draw conclusions for all lyssaviruses. In this article, we have focused on evidence of host shifts in the same insectivorous bat reservoir species in North America (RABV) and Europe (EBLV-1, -2 and BBLV) are evaluated. How RABV has successfully crossed species barriers and established infectious cycles in new hosts to be the global multi-host pathogen it is today, whilst other lyssaviruses appear restricted in host species is explored in this review. It hypothesized that RABV is the exception, rather than the rule, in this fascinating genus of viruses.

Introduction

The *lyssavirus* genus is a group of negative-strand RNA viruses, which among other features, are characterized by their ability to cause fatal encephalitis [1]. Multiple lineages of rabies virus (RABV), which represent the type species of the genus *(Rabies lyssavirus)* circulate globally in a wide range of mammalian hosts within the *Carnivora* and *Chiroptera* orders. All lyssaviruses appear to have evolved closely with specific hosts. However, unlike all other lyssaviruses, multiple independent transmission cycles of RABV have been established in a

broad range of volant and non-volant host reservoirs, where particular RABV lineages circulate within host species [2]. Lyssavirus cross-species transmission (CST) events occur readily, mostly without onward transmission, often referred to as ‘spill over’ events, recorded through surveillance initiatives [3]. In contrast, host shifts are considered to be rare, because maintenance in a new host population is dependent on many viral, host and environmental factors [4,5]. Analysis of RABV genome sequences has provided a comprehensive analysis of host shift events and neither relatedness of host species, nor positive selection within the virus genome were identified as important influences on RABV host shift events [2]. A broad scale analysis of virus CST events at the virus family level, indicated host shifts play an important role in viral evolution, although RABV sequences predominated the lyssavirus dataset in these analyses [6]. The low level of detection of spill-over events and sustained host shifts for each of the lyssaviruses, apart from RABV circulating in non–volant species, suggests co-evolution with respective host reservoirs, favoring maintenance transmission within the host population.

The co-evolution of lyssaviruses with individual bat species has defined this genus as we understand it today. However, the global distribution of lyssaviruses has perplexed researchers for decades. Despite the widespread distribution of hundreds of susceptible bat species, RABV has only been detected within ‘New World’ bats, whereas all other lyssaviruses with known bat reservoirs circulate in ‘Old World’ bats. In contrast to the previously proposed ‘out of Africa hypothesis’, analysis of bat lyssavirus genomes using a probabilistic phylogeographic approach, concluded that the most likely origin of lyssaviruses is

71 the Palearctic region, which spans across Europe, North Africa, Northern and
72 central Arabian Peninsula and Asia (North of Himalayas) [7]. Regardless of the
73 geographic origins of lyssaviruses, accurately determining the timescales of
74 lyssavirus evolution is also challenging. It is well recognized that estimating
75 timescales for virus evolution, using a limited set of recently sampled viruses,
76 cannot provide accurate inferences for viruses under purifying selection [8,9]. To
77 address this shortcoming, recent analysis of hantaviruses using phylogenetic
78 analyses from both the viruses and their reservoir hosts provided 'fixed points' to
79 anchor the viral phylogenies which may indicate a more realistic time scale of
80 evolution [10]. Utilizing bat phylogenetic data to interpret lyssavirus phylogenies
81 could significantly improve calibration of the lyssavirus ancestry.
82

83 **Lyssavirus host restriction**

84 The co-evolution of all lyssavirus species, with the exception of RABV, to a
85 single or highly restricted host species is compelling. For each lyssavirus species
86 an element of geographical restriction, alongside an apparent host restriction, is
87 observed [11]. Comprehensive surveillance schemes exist for bat lyssaviruses
88 across Europe, yet the detection of EBLV-1, EBLV-2 and BBLV in any bat
89 species other than the natural host reservoir is rare (BBLV and EBLV-1) or absent
90 (EBLV-2) (Figure). Incidences of bat-associated rabies (RABV) in the Americas
91 are more evident than in the rest of the world. In North America, over a thousand
92 RABV positive bats are reported each year. Moreover, hundreds of RABV spill
93 over events are recorded, including multiple non-volant [3] and volant species [12]
94 (Figure).

Why related bat species between the Americas and the Old World (for example *Myotis* and *Eptesicus* species) evolved with different lyssaviruses is an interesting question. The restricted host reservoir observed for EBLV-1 (*Eptesicus* spp) and BBLV and EBLV-2 (*Myotis* spp), despite these bat species co-roosting and feeding with other bat species, directly contrasts to RABV circulating in a large range of bat species including *Eptesicus* and *Myotis* spp in the Americas (Figure). One explanation is that a bat morphologically similar to extant bats, Onychonycteris, which existed in the territory of Gondwana could have harboured an ancient RABV which compartmentalized in numerous bat species across the Americas and later switching to skunks, raccoons and other bat species [13]. Indeed, phylogenetic analysis of *Myotis* species clearly indicates that although *Myotis* species from America and Europe are morphologically similar, phylogenetically they are unrelated [14], strengthening the hypothesis that lyssaviruses have evolved within bats, and bat evolution and distribution has determined the division of lyssaviruses between the New and Old World.

Exploring the relationship between lyssaviruses and their host is paramount to unravelling the complexities of host switching. All lyssaviruses cause fatal encephalitis (rabies), therefore the balance between virulence and transmission is critical. The trade-off between virulence and successful transmission is perhaps the critical factor driving host specificity, without this, the host succumbs to rabies before opportunity to transmit the virus. From a population perspective, maintenance has been modelled for a number of bat species and indicated that long incubation periods associated with infection within individual bats was an important factor for long term maintenance in populations [15,16].

Host interactions with lyssaviruses – intra-host transmission

Lyssaviruses are most commonly transmitted via infected saliva transfer from an animal bite. The virus may initially replicate in muscle cells at the site of the wound before entering either motor or sensory nerves to travel by retrograde transport along axons to the central nervous system (CNS). Improved replication at site of infection could increase the success of the virus entering the CNS, although triggering the immune system at this stage would more likely result in clearance of the virus [17]. For lyssaviruses that have co-evolved with a single host reservoir, fine-tuning of replication at different sites could maximize the chance of CNS invasion and spread with limited immune responses being triggered within the host. Although some host interactions involved in lyssavirus replication have been described, virtually all studies have utilized laboratory animal models that must be considered poor surrogates for natural infection. Certainly, the limited knowledge of bat immunology remains a major obstacle to understanding the host factors that contribute to clearance following exposure, or the development of a productive infection with onward transmission. Occasionally, the development of clinical lyssavirus infections has been observed in bats following long periods of rehabilitation in captivity [18]. Similarly long incubation periods have been observed in other species, including human infection.

Where studies have been attempted, innate immune evasion by the phosphoprotein (P) is conserved across lyssaviruses by inhibition of IFN –α-dependent STAT1/2 signaling [19] and also through interaction with I-kappa B kinase epsilon which was shown for an attenuated RABV [20]. Also, differences in

145 virus uptake and subsequent infection of the CNS through long distance axonal
146 transport could contribute to differences in incubation periods seen and different
147 outcomes of infection. Despite p75NTR, NCAM and nAchR being accepted as
148 RABV receptors [21] and virus binding and co-internalization in primary dorsal
149 root ganglion neurons demonstrated by live imaging [22], it is still unknown which
150 repertoire of neuronal receptors are used by RABV in vivo, or whether this is
151 identical to the receptor usage of other lyssaviruses. Further, it is likely that other
152 receptor molecules exist which enable virus uptake at different sites. Notably, all
153 of the above listed receptors appear dispensable for RABV infection and p75NTR
154 may serve as a receptor for only some lyssaviruses, as interaction of glycoprotein
155 with p75NTR was shown for RABV and EBLV-2, but not for EBLV-1 or ABLV [23].
156 Whether this is due to a virus species-specific receptor usage or adaptation to
157 host specific variations of p75NTR is unknown. After entry via receptor mediated
158 endocytosis, RABV is transported inside of vesicles [22,24] and it is likely that the
159 mode of the intra-neuronal long-distance transport is predetermined by receptor
160 usage. As kinetic data about retrograde axonal transport are only available for
161 RABV [22,25] it is unknown whether other lyssaviruses differ in axonal transport
162 efficiencies. After virus replication, differential intra-neuronal transport may also
163 play a role in host adaptation. Whether mechanisms involved in the release of
164 lyssaviruses from infected neurons are influenced by host specific factors also still
165 requires investigation.

166 Overall, although some detailed knowledge about the molecular
167 mechanisms behind RABV replication and host interactions has been reported,
168 their contribution to a host species dependent replication and role in reservoir

maintenance is uncertain. Only by investigating other lyssaviruses can we conclude which factors are likely to increase host-shift events.

Lyssavirus inter-host transmission

Lyssaviruses are transmitted directly to humans (and other mammalian species) from host reservoirs with no intermediate reservoir, defined as 'type I zoonotic transmission' [26]. Often, the definition of 'reservoir' requires that the infectious agent in question be non-pathogenic to the reservoir host species. However, lack of pathogenicity does not necessarily equate to the maintenance of an infectious agent in a given population, and rabies is one such example [27]. There has been much debate regarding the use of the term 'reservoir' for pathogens, mostly focusing on human disease [27,28]. Throughout this review the term 'host reservoir' is used, to describe specific mammalian species identified as maintaining lyssaviruses as independent cycles, detailed information regarding lyssavirus reservoirs and the implications for human transmission are explored in box 1.

Characterised lyssavirus host shifts are extremely rare, the majority being observed between non-volant host species [29,30]. RABV host shifts in rabies endemic areas are likely to go unrecognized, unless adequate surveillance is coupled with virus characterization. Where host shifts are detected, authorities respond with a plethora of interventions, including vaccination of susceptible animals, movement restrictions and increased surveillance, to successfully control the disease [29,31]. When interventions are not employed, or are not successful, the virus can become established in the new host, observed in Turkey, when a canine RABV lineage successfully established within the naïve

194 red fox population [30]. Indeed, RABV host shifts to wildlife reservoirs for which
195 limited control options exist, e.g. alien American raccoons (*Procyon lotor*) in
196 Europe, Golden jackals (*Canis aureus*) or mongoose (*Herpestes* spp), would
197 cause a tremendous setback and represent a potential challenge [32].
198
199 Pre-adaptation to host shift events
200 It has become apparent that RABV host shifts are often not accompanied
201 by obvious adaptive mutations in the virus genomes, as even sophisticated
202 phylogenetic analyses have not identified definitive positive selection events
203 correlating with host shifts [2,29,31,33]. That 'pre-adaptation' has resulted in
204 successful host shifts has been previously suggested [4,31]. We expand on this
205 concept and propose that pre-adaptation is not something that occurs before
206 each CST, rather that pre-adaptation has been acquired by rabies viruses at the
207 virus species level, which has increased the success of onward transmission after
208 a CST. The pre-adaptation event is most likely to have occurred in an ancestor to
209 RABV circulating in bats prior to critical host shifts into terrestrial mammals
210 (Figure). Within the Chiropteran Order, apart from RABV, ABLV is the only other
211 lyssavirus species known to circulate in more than one host reservoir at the family
212 level (*Pteropus* spp - fruit bats and *Saccolaimus* spp - insectivorous bats) [34].
213 However, there are a number of lyssavirus species which have single or low
214 numbers of virus isolations, including GBLV, which has a recent common
215 ancestor with ABLV, isolated from *Pteropus* spp in Sri Lanka [35]. If lyssavirus
216 surveillance in bat species within the Indian sub-continent were implemented,
217 detection of GBLV in other bat species, including insectivorous bats, might be
218 predicted [36,37].

Both intra and inter-host transmissions are enhanced by altering the infected individual's behavior, even spill over infected individuals display altered behavior. As a neurotropic virus, lyssaviruses replicate in the brain, seemingly targeting certain regions that affect behaviour, further details in box 2.

Concluding comments

Previous suggestions of 'pre-adaptation' have focused on molecular changes immediately preceding a host shift event [4]. However, we suggest that pre-adaptation is established within the *rabies lyssavirus* species. Perhaps pre-adaptation represents a changed function in the viral machinery, for example, in the phosphoprotein or glycoprotein, or replication efficiency? The controversial topic of recombination in lyssaviruses was proposed, to explain the RABV host shift from bats to raccoons in the Americas, although the requirement for a nonvolant species (skunk) RABV glycoprotein does not provide explanation for how the skunk variant came into existence [38]. Without experimental evidence for recombination in lyssaviruses, it is difficult to understand how a virus with such an acute disease progression can co-infect the same animal. Furthermore, how can recombination occur when the RNP protects the genome for the majority of the life cycle? Alternatively, utilising virus sub-population heterogeneity (quasispecies) within the host as observed during poliovirus infection [39], and after the bottleneck which occurs during a CST event, as observed in arboviruses during mosquito transmission are interesting possibilities for lyssavirus adaptation [40]. Furthermore, perhaps the switch from *Chiroptera* to *Carnivora* species provided an opportunity for RABV to increase virulence, because transmission

between non-volant species, such as dogs, is more frequent with higher infectious doses.

The strong association of all lyssavirus species with individual bat species, apart from RABV which has multiple host reservoirs across mammalian Orders is a fascinating conundrum of virology. Rather than focusing our attention solely on RABV, we should instead investigate the bat lyssaviruses host specificity determinants. If canine RABV is successfully eradicated by 2030, addressing the risk of other non-volant RABV reservoirs such as jackals, wolves and foxes is the next strategic hurdle. Understanding the differences between lyssaviruses that use non-volant and volant host reservoirs will become key to implementing control strategies to prevent another host shift back into the susceptible canine population.

Acknowledgements

The authors wish to thank Dr. Noel Tordo (Institut Pasteur, France), Dr. Adrian Vos (IDT, Germany) and Dr. Charles Rupprecht (Wistar Institute, USA) for constructive and informative discussions in the initial concepts of rabies virus evolution, pathogenesis and host shift events. This work was supported by the UK Department for Environment, Food and Rural Affairs (Defra), Scottish Government and Welsh Government by grant SE0431.

References

1. Fooks AR, Banyard AC, Horton DL, Johnson N, McElhinney LM, Jackson AC: **Current status of rabies and prospects for elimination**. *Lancet* 2014, **384**:1389-1399.
2. Troupin C, Dacheux L, Tanguy M, Sabeta C, Blanc H, Bouchier C, Vignuzzi M, Duchene S, Holmes EC, Bourhy H: **Large-Scale Phylogenomic Analysis**

Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts. *PLOS Pathogens* 2016, **12**:e1006041.
3. Wallace RM, Gilbert A, Slate D, Chipman R, Singh A, Cassie W, Blanton JD: **Right place, wrong species: a 20-year review of rabies virus cross species transmission among terrestrial mammals in the United States**. *PLoS One* 2014, **9**:e107539.
4. Mollentze N, Biek R, Streicker DG: **The role of viral evolution in rabies host shifts and emergence**. *Curr Opin Virol* 2014, **8**:68-72.
5. Badrane H, Tordo N: **Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders**. *J Virol* 2001, **75**:8096-8104.
6. Geoghegan JL, Duchene S: **Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families**. 2017, **13**:e1006215.
7. Hayman DT, Fooks AR, Marston DA, Garcia RJ: **The Global Phylogeography of Lyssaviruses - Challenging the 'Out of Africa' Hypothesis**. *PLoS Negl Trop Dis* 2016, **10**:e0005266.
8. Wertheim JO, Kosakovsky Pond SL: **Purifying selection can obscure the ancient age of viral lineages**. *Mol Biol Evol* 2011, **28**:3355-3365.
9. Taylor DJ, Ballinger MJ, Zhan JJ, Hanzly LE, Bruenn JA: **Evidence that ebolaviruses and cuevaviruses have been diverging from marburgviruses since the Miocene**. *PeerJ* 2014, **2**:e556.
10. Castel G, Tordo N, Plyusnin A: **Estimation of main diversification time-points of hantaviruses using phylogenetic analyses of complete genomes**. *Virus Res* 2017, **233**:60-69.
11. Banyard AC, Hayman DT, Johnson N, McElhinney L, Fooks AR: **Bats and Lyssaviruses**. In *Research Advances in Rabies*. Edited by Jackson AC: Elsevier/Academic Press; 2011:239-289. vol 79.]
12. Velasco-Villa A, Mauldin MR, Shi M, Escobar LE, Gallardo-Romero NF, Damon I, Olson VA, Streicker DG, Emerson G: **The history of rabies in the Western Hemisphere**. *Antiviral Res* 2017.
13. Rupprecht C, Kuzmin I, Meslin F: **Lyssaviruses and rabies: current conundrums, concerns, contradictions and controversies**. *F1000Res* 2017, **6**:184.
14. Ruedi M, Mayer F: **Molecular systematics of bats of the genus Myotis (Vespertilionidae) suggests deterministic ecomorphological convergences**. *Mol Phylogenet Evol* 2001, **21**:436-448.
15. George DB, Webb CT, Farnsworth ML, O'Shea TJ, Bowen RA, Smith DL, Stanley TR, Ellison LE, Rupprecht CE: **Host and viral ecology determine bat rabies seasonality and maintenance**. *Proc Natl Acad Sci U S A* 2011, **108**:10208-10213.
16. Blackwood JC, Streicker DG, Altizer S, Rohani P: **Resolving the roles of immunity, pathogenesis, and immigration for rabies persistence in vampire bats**. *Proc Natl Acad Sci U S A* 2013, **110**:20837-20842.
17. Scott TP, Nel LH: **Subversion of the Immune Response by Rabies Virus**. *Viruses* 2016, **8**.
18. Pajermo K, Harkess G, Goddard T, Marston D, McElhinney L, Johnson N, Fooks AR: **Isolation of European bat lyssavirus type 2 (EBLV-2) in a Daubenton's bat in the UK with a minimum incubation period of 9 months.** . *Rabies Bulletin Europe* 2008, **32**:6-8.

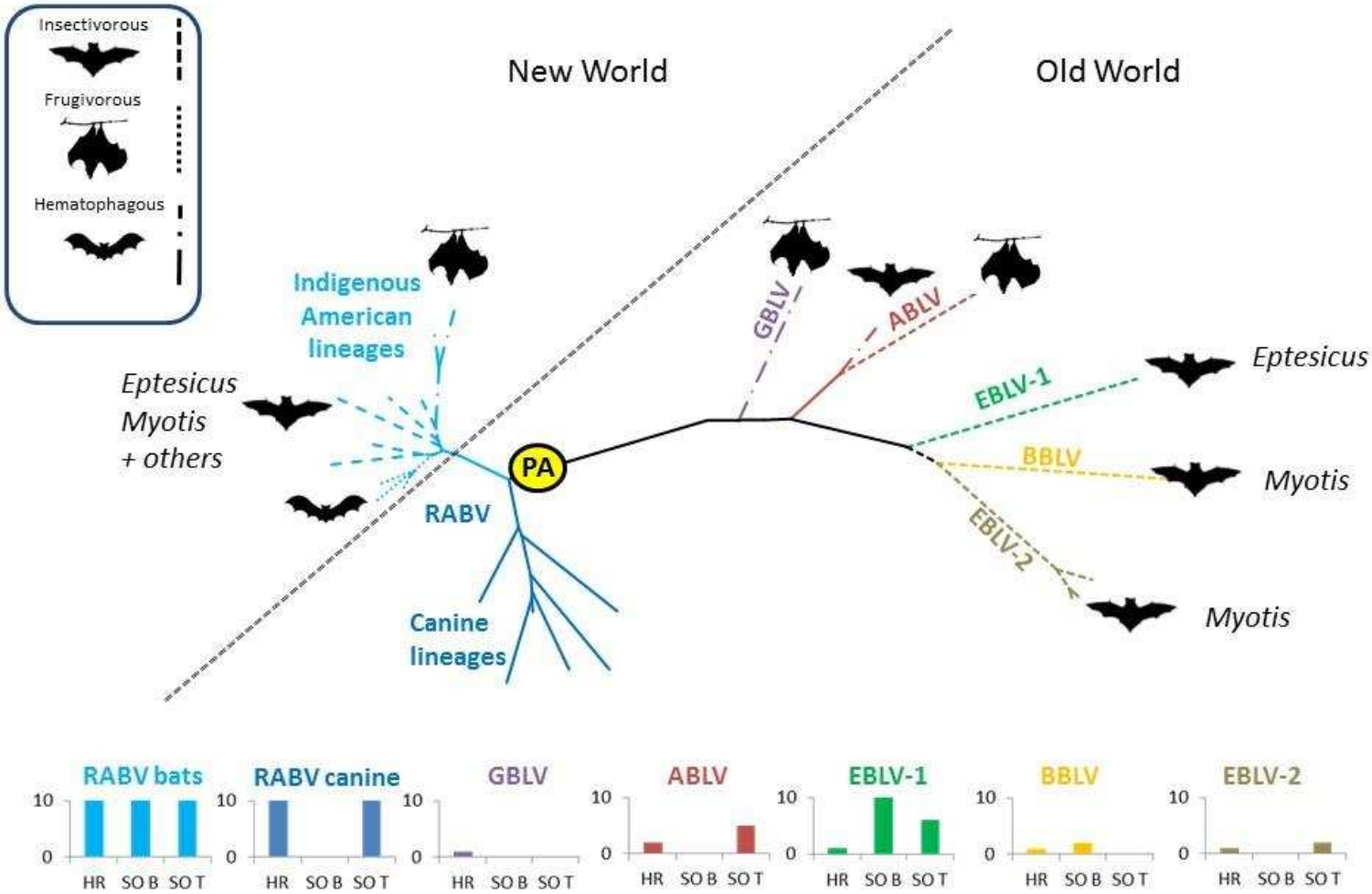
19. Wiltzer L, Larrous F, Oksayan S, Ito N, Marsh GA, Wang LF, Blondel D, Bourhy H, Jans DA, Moseley GW: **Conservation of a unique mechanism of immune evasion across the Lyssavirus genus**. *J Virol* 2012, **86**:10194-10199.
20. Brzozka K, Finke S, Conzelmann KK: **Identification of the rabies virus alpha/beta interferon antagonist: phosphoprotein P interferes with phosphorylation of interferon regulatory factor 3**. *J Virol* 2005, **79**:7673-7681.
21. Lafon M: **Rabies virus receptors**. *Journal of Neurovirology* 2005, **11**:82-87.
22. Gluska S, Zahavi EE, Chein M, Gradus T, Bauer A, Finke S, Perlson E: **Rabies Virus Hijacks and accelerates the p75NTR retrograde axonal transport machinery**. *PLoS Pathog* 2014, **10**:e1004348.
23. Tuffereau C, Desmezieres E, Benejean J, Jallet C, Flamand A, Tordo N, Perrin P: **Interaction of lyssaviruses with the low-affinity nerve-growth factor receptor p75NTR**. *J Gen Virol* 2001, **82**:2861-2867.
24. Klingen Y, Conzelmann KK, Finke S: **Double-labeled rabies virus: live tracking of enveloped virus transport**. *J Virol* 2008, **82**:237-245.
25. Bauer A, Nolden T, Schroter J, Romer-Oberdorfer A, Gluska S, Perlson E, Finke S: **Anterograde glycoprotein-dependent transport of newly generated rabies virus in dorsal root ganglion neurons**. *J Virol* 2014, **88**:14172-14183.
26. Cui JA, Chen F, Fan S: **Effect of Intermediate Hosts on Emerging Zoonoses**. *Vector Borne Zoonotic Dis* 2017.
27. Haydon DT, Cleaveland S, Taylor LH, Laurenson MK: **Identifying reservoirs of infection: a conceptual and practical challenge**. *Emerg Infect Dis* 2002, **8**:1468-1473.
28. Ashford RW: **When is a reservoir not a reservoir?** *Emerg Infect Dis* 2003, **9**:1495-1496.
29. Borucki MK, Chen-Harris H, Lao V, Vanier G, Wadford DA, Messenger S, Allen JE: **Ultra-deep sequencing of intra-host rabies virus populations during cross-species transmission**. *PLoS Negl Trop Dis* 2013, **7**:e2555.
30. Johnson N, Black C, Smith J, Un H, McElhinney LM, Aylan O, Fooks AR: **Rabies emergence among foxes in Turkey**. *J Wildl Dis* 2003, **39**:262-270.
31. Kuzmin IV, Shi M, Orciari LA, Yager PA, Velasco-Villa A, Kuzmina NA, Streicker DG, Bergman DL, Rupprecht CE: **Molecular Inferences Suggest Multiple Host Shifts of Rabies Viruses from Bats to Mesocarnivores in Arizona during 2001-2009**. *PLoS Pathog* 2012, **8**:e1002786.
32. Müller T, Freuling CM, Wysocki P, Roumiantzeff M, Freney J, Mettenleiter TC, Vos A: **Terrestrial rabies control in the European Union: Historical achievements and challenges ahead**. *Veterinary Journal* 2014.
33. Hanke D, Freuling CM, Fischer S, Hueffer K, Hundertmark K, Nadin-Davis S, Marston D, Fooks AR, Botner A, Mettenleiter TC, et al.: **Spatio-temporal Analysis of the Genetic Diversity of Arctic Rabies Viruses and Their Reservoir Hosts in Greenland**. *PLoS Negl Trop Dis* 2016, **10**:e0004779.
34. Weir DL, Annand EJ, Reid PA, Broder CC: **Recent observations on Australian bat lyssavirus tropism and viral entry**. *Viruses* 2014, **6**:909-926.

35. Gunawardena PS, Marston DA, Ellis RJ, Wise EL, Karawita AC, Breed AC, McElhinney LM, Johnson N, Banyard AC, Fooks AR: **Lyssavirus in Indian Flying Foxes, Sri Lanka**. *Emerg Infect Dis* 2016, **22**:1456-1459.
36. Mani RS, Dovih DP, Ashwini MA, Chattopadhyay B, Harsha PK, Garg KM, Sudarshan S, Puttaswamaiah R, Ramakrishnan U, Madhusudana SN: **Serological Evidence of Lyssavirus Infection among Bats in Nagaland, a North-Eastern State in India**. *Epidemiol Infect* 2017, **145**:1635-1641.
37. Pal SR, Arora B, Chhuttani PN, Broor S, Choudhury S, Joshi RM, Ray SD: **Rabies virus infection of a flying fox bat, Pteropus policephalus in Chandigarh, Northern India**. *Trop Geogr Med* 1980, **32**:265-267.
38. Ding NZ, Xu DS, Sun YY, He HB, He CQ: **A permanent host shift of rabies virus from Chiroptera to Carnivora associated with recombination**. *Sci Rep* 2017, **7**:289.
39. Xiao Y, Dolan PT, Goldstein EF, Li M, Farkov M, Brodsky L, Andino R: **Poliovirus intrahost evolution is required to overcome tissue-specific innate immune responses**. *Nat Commun* 2017, **8**:375.
40. Stapleford KA, Coffey LL, Lay S, Borderia AV, Duong V, Isakov O, Rozen-Gagnon K, Arias-Goeta C, Blanc H, Beaucourt S, et al.: **Emergence and transmission of arbovirus evolutionary intermediates with epidemic potential**. *Cell Host Microbe* 2014, **15**:706-716.

Figure Legends

Figure: Host restriction and reported spill over events for lyssaviruses.

Unrooted, phylogenetic tree of six lyssavirus species with bat reservoir host indicated. Graphs indicate approximate relative numbers, where a value of 10 = 10 or more, of recorded host reservoirs (HR), spill overs in bat (SOB) and spill overs in terrestrial mammals (SOT) for each lyssavirus (and for RABV the bat and canine lineages). The RAC/SK lineage of RABV is not included for clarity. RABV, rabies virus; GBLV, Gannoruwa bat lyssavirus; ABLV, Australian bat lyssavirus; EBLV-1, EBLV-2, European bat lyssavirus type 1 and 2; BBLV Bokeloh bat lyssavirus. PA, suggested pre-adaptation stage.



HR = host reservoir, SO B = Spill over bat, SO T = Spill over terrestrial

406

1.4: Lyssavirus Classification

Lyssaviruses are classified within the Family *Rhabdoviridae*, Order *Mononegavirales*. All viruses within this Order contain linear, negative-sense, single-stranded ribonucleic acid (RNA) genomes. As such, for replication to occur, first viral mRNAs are transcribed, and following the accumulation of viral proteins within an infected cell, the anti-genome sense RNA is generated as a full length positive sense replicative intermediate which acts as a template for replication to produce the negative sense genome (Whelan et al., 2004). Transcription and replication occurs in the cytoplasm of infected cells for most members of the *Mononegavirales* with the exception of bornaviruses, nyaviruses and the nucleorhabdoviruses where transcription and replication occurs in the host cell nucleus (Easton and Pringle, 2012). The majority of viruses classified within the *Mononegavirales* encode 5-10 genes in their genomes, although some have evolved to generate more through maximisation of coding potential (Pringle and Easton, 1997). Genetic reassortment is poorly characterised for most members of the *Mononegavirales*, most likely because the RNA genome is almost always very tightly encapsidated thus limiting the potential for hybridisation of an alternative template (Mebatsion et al., 1996b). So far there has been two documented cases of genetic reassortment in *Mononegavirales*; one between two deletion mutants of Respiratory Syncytial Virus (RSV), which were able to produce just one reassorted virus despite being under optimal laboratory controlled conditions for reassortment (Spann et al., 2003) and another case, which detected two potential recombination events in a panel of 44 rabies viruses, both within the polymerase protein, although independent sequencing of one of the samples identified as having a recombination has shown the sequence on Genbank is not correct and highlights the issue with using sequences deposited by other researchers (Liu et al., 2011).

The following article is a review on Rhabdoviruses

1.4.1: Article 2: In review with eLS:

Rhabdoviruses

Summary:

Within the *Mononegavirales* Order, different members of the *Rhabdoviridae* are able to infect a wide range of hosts including mammals, invertebrates, plants and fish, although only lyssaviruses and vesiculoviruses have been reported as the cause of clinical disease in animals and humans. In the last few years, the number of rhabdoviruses identified has increased substantially and there are now 18 genera and only one unassigned virus (Moussa virus) in this family.

This review summarises the latest information regarding rhabdoviruses, including representative genome organisation, whilst describing the basic viral properties of this fascinating group of viruses.

Version 4 a0027748

Rhabdoviruses

Daisy Jennings,
Denise A Marston,
Ashley C Banyard,
Anthony R Fooks,

AU:1

AU:2

Based in part on the previous versions of this eLS article 'Rhabdoviruses' (2003, 2006, 2013).

Advanced article

Article Contents

Online posting date: 15th December 2017

The family Rhabdoviridae is classified within the order *Mononegavirales* and is made up of 18 different genera and one other virus awaiting classification into this family. Generally, Rhabdovirus virions contain a nonsegmented negative-sense single-stranded ribonucleic acid (RNA) genome that, in its most basic form, encodes five proteins: the nucleocapsid protein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the RNA-dependent RNA polymerase protein (L). All Rhabdoviruses have a distinctive bullet-shaped morphology; however, two recently recognised genera have been shown to contain viruses with bipartite genomes. This diverse group of over 165 viruses can infect a range of species including mammals, reptiles, birds, fish, insects and plants. Several of these viruses cause notable economic losses to both the agriculture and aquaculture sector. However, only the lyssaviruses and the vesiculoviruses are able to infect both animals and humans to cause clinical disease, with rabies virus being the most important human pathogen.

Classification

The order *Mononegavirales*

The International Committee on Taxonomy of Viruses (ICTV) 2016 Release (http://www.ictvonline.org/virusTaxonomy.asp) included the order *Mononegavirales* expanding to encompass 69 novel species and application of non-Latinised binomial species names throughout the entire order (Amarasinghe *et al.*, 2017). The order *Mononegavirales* contains eight virus families, including the *Bornaviridae*, the *Filoviridae*, the *Paramyxoviridae* and the Rhabdoviridae families. Generally, all viruses within these families contain linear, nonsegmented and negative-sense single-stranded ribonucleic acid (RNA) genomes. These genomes act as templates for two distinct processes: (1) transcription of viral messenger ribonucleic acids (mRNAs) and (2) transcription of a full-length positive-sense copy of the virus genome, termed the antigenome, which in turn acts as the template for replication to produce negative-sense progeny genomes (Whelan *et al.*, 2004). Neither the negative-sense genome nor the positive-sense antigenome is considered infectious as they require viral proteins to drive transcription of viral messenger RNA and initiate the replicative cycle. The minimal replicative unit for all rhabdoviruses appears to be the genomic ribonucleic acid (gRNA) tightly wound as a helical structure of 30–70 nm encapsidated with N protein and in contact with P and L. This complex of proteins with nucleic acid is termed the ribonucleoprotein (RNP) complex. The generation of this complex *in vitro* is the basis for reverse genetics techniques whereby the virus is recovered following transfection of the deoxyribonucleic acid (DNA). Virus release occurs through budding from the plasma membrane, whereas virus replication and transcription are generally cytoplasmic for members of the order *Mononegavirales*, with the exception of the bornaviruses, the nucleorhabdoviruses and the dichorhaviruses, which perform both replication and transcription of their genetic material in the nucleus. Other features shared by the members of the order *Mononegavirales* include a helical nucleocapsid, the apparent inability to undergo genetic reassortment and a highly conserved gene order – with the N, P and M genes, always at or near the 3′ terminus, and the L gene always the furthest from the genome promoter. This is most likely an evolutionary adaptation that controls protein expression levels. **See also:** **Bornaviruses**; **Filoviruses**; **RNA Virus Genomes**

a0001011.pub3
a0001022
a0001488.pub2

Generally within this family of viruses, genomes encode 5–10 genes, the position of which is conserved to control gene expression. Although the majority of rhabdovirus genomes contain five genes, some species have evolved to contain extra genes apparently following gene duplication or the expansion of intergenic regions rather than through uptake and genomic incorporation of extraneous genetic information (Pringle and Easton, 1997).Continued genome characterisation and functional assessment of some of these extra genes may highlight mechanisms that have driven such genome expansion (Marriott and Easton, 1999; Wertz *et al.*, 1998) and demonstrate the role, if any, of smaller

eLS subject area: Virology

How to cite:
Jennings, Daisy; Marston, Denise A; Banyard, Ashley C; and Fooks, Anthony R (December 2017) Rhabdoviruses. In: eLS. John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0001085.pub4

AU:3

genes compared to a genome with fewer large ORFs. Certainly, other viruses express several genes from one ORF to maximise coding capacity, so a more fragmented genetic approach may link to reduced impact. Current understanding of the genome organisation of different rhabdoviruses is detailed in **Figure 1**. **See also**: **Genome Evolution: Overview**; **Viruses: Genomes and Genomics**

a0001810

a0001092.pub2

The family Rhabdoviridae

The distinctive bullet-shaped morphology of rhabdovirus virions distinguishes the Rhabdoviridae from other mono- or bipartite, negative-sense RNA viruses (**Figure 2a** and **b**). Despite the relatively large size of the rhabdoviruses, the genomes appear to have a limited coding capacity (11–16 kb) (Dietzgen *et al*., 2017). Within this virus family there are 18 different genera (**Table 1**). These include those that infect mammals, birds, reptiles, fish, plants and arthropods (**Figure 2**). An increasing number of rhabdoviruses are being isolated from humans displaying clinical disease; however, clear causal relationships are yet to be established. Importantly, only the lyssaviruses, ephemeroviruses and vesiculoviruses have been confirmed as being able to infect both animals and humans to cause clinical disease with the lyssavirus genus including one of the most notable viral infections known to man, rabies virus (RABV). Alongside viruses of human and veterinary significance, rhabdoviruses of plants and fish are responsible for considerable financial loss within both agricultural and aquaculture sectors.

Advancements in high-throughput technologies have highlighted the scale of the diversity within the family Rhabdoviridae. This has led to the development of new criteria by which to categorise rhabdoviruses and five novel genera to include unassigned and newly recognised rhabdoviruses (Amarasinghe *et al*., 2017).

The genus *Vesiculovirus*

The genus *Vesiculovirus* includes viruses that can infect a wide range of animals, some of which have serious economic impact. There are currently 16 viruses officially assigned to this genus based on their serological and genetic relationships. The most widely known viruses in this genus are the vesicular stomatitis viruses (VSVs), the type species of which is vesicular stomatitis Indiana virus (VSIV). VSVs can cause vesicular disease in horses, cattle and pigs (VSIV, vesicular stomatitis New Jersey virus, vesicular stomatitis Alagoas virus). Humans are also susceptible to some vesiculoviruses with symptoms ranging from mild febrile illness (VSV) to fatal encephalitis (Chandipura virus (CHPV)) (Rodriguez and Pauszek, 2012). The genomic organisation is similar to other members of the Rhabdoviridae, with a genome of approximately 11.2 kb consisting of five genes encoding N, P, M, G and L (**Figure 1**). **See also**: **Vesicular Stomatitis Virus**

a0001089.pub2

The genus *Lyssavirus*

The genus *Lyssavirus* includes one of the most important viruses known to man, the rabies RABV, which is the type species of this genus, which also includes 13 other defined species differentiated according to their genomic sequence. These are *Aravan lyssavirus, Australian bat lyssavirus*, *Bokeloh bat lyssavirus*, *Duvenhage lyssavirus*, *European bat -1 lyssavirus, European bat -2 lyssavirus, Ikoma lyssavirus. Irkut lyssavirus, Khujand lyssavirus, Lagos bat lyssavirus, Mokola lyssavirus*, *Shimoni lyssavirus* and *West Caucasian bat lyssavirus.* Two further isolates, *Gannoruwa bat lyssavirus* (Gunawardena *et al*., 2016) and *Lleida bat lyssavirus* (Ceballos *et al*., 2013), are awaiting classification. The discovery of novel lyssaviruses further highlights the diversity within this genus and illustrates the need for surveillance and screening for novel viruses. **See also**: ***Rabies Virus***

a0000427

The RABV genome is single-stranded negative-sense RNA and is approximately 12 kb in size although variations in length have been recorded between lyssaviruses species (Marston *et al*., 2007). This RNA contains the information necessary to produce the five different viral proteins (N, P, M, G and L).

With the exceptions of *Mokola lyssavirus* and *Ikoma lyssavirus* (Fooks *et al*., 2014), all of the classified species have been isolated from bats, and all are considered capable of causing fatal encephalitis in humans. Although RABV has a global distribution in multiple hosts, all other members of the genus appear, to a certain extent, to be host restricted.

AU:4

The genera *Ephemerovirus* and *Tibrovirus*

Virus species of the *Ephemerovirus* and *Tibrovirus* genera are primarily hosted by cattle (Dietzgen *et al*., 2017) and are transmitted by mosquitoes and biting midges. The genus *Ephemerovirus* currently consists of eight species including its type species *Bovine fever ephemerovirus*. The genus *Tibrovirus* consists of six species, with the type species being *Tibrogargan tibrovirus*. Animals infected with an Ephemerovirus or Tibrovirus generally appear asymptomatic. Exceptions to this are bovine fever ephemerovirus (BEFV), Kotonkan virus (KOTV) (genus *Ephemerovirus*) and Bas-Congo virus (BASV) (genus *Tibrovirus*), all of which can cause a debilitating febrile illness in cattle (Blasdell *et al*., 2012). In 2009, BASV was associated with an acute haemorrhagic fever outbreak in central Africa; however, a clear causal affiliation of BASV as the source of the outbreak was unable to be established (Grard *et al*., 2012).

Genetically, ephemeroviruses have larger genomes than those seen for other rhabdoviruses, with the genomes of Adelaide River virus (ARV) and BEFV being 14.6 and 14.9 kb, respectively. This increase in length is due to the inclusion of an extra gene, G_{ns}, alongside the standard N, G, M, P and L proteins expressed by other members of the Rhabdoviridae. There is some sequence homology evident between G and G_{ns} within these virus families and also with the G protein of other rhabdoviruses; however, further investigation is needed in order to determine G_{ns} as a structural protein or the result of gene duplication or recombination. Additional expressed coding sequences in the G–L intergenic gap are also evident in each of these viruses, although the function of this region remains unclear. From the point of view of development of a protective antibody response, the G protein has been partially characterised as containing four epitopes thought important for neutralisation (Walker and Kongsuwan, 1999). As seen with other rhabdoviruses, purified G proteins or subunit vaccines expressing the G protein alone are able to generate a protective humoral response in cattle. In contrast, the role of G_{ns} remains unclear although it does

not induce neutralising antibodies and therefore plays no role in B-cell-specific adaptive immune response, which can be correlated with protective immunity. Genetically, Ephemeroviruses are most closely related to tibroviruses; however, the genome organisation of tibroviruses is distinct. Tibroviruses do not contain a G_{ns} as occurs in ephemeroviruses but feature three additional genes encoding ORFs (U1, U2 and U3) that lie within an independent transcriptional unit preceding and following the G gene. U1 and U2 encode small proteins of unknown function, and U3 encodes a small viroporin-like protein (Dietzgen *et al.*, 2017).

The genus *Tupavirus*

The genus *Tupavirus* comprises three virus species – *Klamath tupavirus, Tupaia tupavirus* and the type species: *Durham tupavirus*. These species are characterised by their extensive host range having been isolated from both birds and mammals, notably an American coot (*Fulica americana)* (Durham virus – DURV), tree shrew (*Tupaia belangeri*) (Tupaia virus – TUPV) and a Montane vole (*Microtus montanus*) (Klamath virus – KLAV). Neutralising antibodies to KLAV have been detected in several deer species, American bison (*bison bison*) and humans; and *in vitro* host range analysis of DURV has demonstrated quail, monkey, cattle and bat cells to be permissive of infection and replication of DURV.

In addition to the five genes encoding the canonical rhabdovirus structural proteins (N, P, M, G and L), tupaviruses contain an additional long alternative ORF in the P gene (Px) and an ORF encoding a small hydrophobic protein (U1) between the M and G genes. Amino acid sequence identity of the N, L and G proteins illustrates KLAV to be sufficiently distinct from TUPV and DURV owing to an additional ORF encoding a small protein (U2) between genes G and L (**Figure 1**).

The genera *Ledantevirus, Sripuvirus* and *Hapavirus*

Ledantevirus, Sripuvirus and *Hapavirus* are three of the five novel Rhabdovirus genera included in the ICTV 2016 release (Amarasinghe *et al.*, 2017).

The Ledantevirus genus is composed of 14 new virus species, which is the largest volume of novel species assigned to any novel rhabdovirus genus. Ledanteviruses have been isolated from and/or detected in arthropods and mammals such as cattle, rodents and wild boar. Furthermore, there is a strong ecological association of ledanteviruses with bats, suggesting a spillover from their natural reservoir. Le Dantec virus (LDV) type species, *Le Dantec ledantevirus*, has additionally been isolated and/or detected in two unrelated human cases who were both hospitalised with acute febrile illness. Despite this, LDV was not determined to be the source of illness for either case (Blasdell *et al.*, 2015).

Ledanteviruses have similar genome organisations to that of vesiculoviruses and comprise five genes encoding the canonical rhabdovirus structural proteins, concise intergenic regions and a small additional ORF between the G and L in only three of the viruses. Ledanteviruses can be further subdivided into three subgroups based on full genome sequence comparison and genome organisation. Viruses in subgroup A contain an additional ORF between the G and L genes and those in subgroup B contain a substantially longer P gene than that found in subgroups A and C (Blasdell *et al.*, 2015).

The genus *Sripuvirus* accommodates five virus species isolated from either sandflies or lizards. Phylogenetically, they cluster closest with tupaviruses and are between 11.0 and 11.5 kb in size. Sripuviruses feature multiple additional ORFs, including a consecutive ORF, named Mx, within the M gene and another, Gx, which resides within an alternative reading frame within the G gene. Mx encodes small basic proteins and an initiation codon that interestingly overlaps the M genes termination codon, and Gx encodes small double-membrane-spanning proteins that are thought to be expressed by 'leaky' ribosomal scanning.

The *Hapavirus* genus contains 15 virus species – three of which were previously unassigned including the type species, *Flanders hapavirus*, and 12 newly recognised virus species. Most hapaviruses have been isolated from Culicine mosquitoes and are capable of infecting a diverse host range including birds, reptiles and mammals. Hapavirus genomes are large and complex, ranging from 12 403 to 15 721 nts in length. Phylogenetically, they are linked to ephemeroviruses and tibroviruses. Indeed, one hapavirus species, *ngaingan hapavirus*, has an additional G_{ns} gene within its genome, similar to that seen within the ephemeroviruses. All hapaviruses, apart from *marco hapavirus*, contain an ORF encoding a viroporin-like protein between the G and L genes and at least one ORF between the P and M genes. Despite these similarities, hapavirus genomes additionally feature multiple accessory genes that differ in numbers and location between each hapavirus species. It is these characteristics that create a genetically and ecologically distinct and diverse clade.

Arthropod-specific rhabdoviruses

The genera *Almendraviruses, Curioviruses* and *Sigmaviruses*

The novel genera *Almendravirus* and *Curiovirus* were published in the 2016 ICTV release (Amarasinghe *et al.*, 2017) and, as Sigmaviruses, comprise arthropod-specific viruses. Whilst the majority of known rhabdoviruses are arthropod-vectored (**Table 1**), studying arthropod-specific rhabdoviruses may provide insight into how and why many rhabdoviruses developed virulence towards vertebrates.

The *Almendravirus* genus was created to include five new species isolated from mosquitoes in the Americas. These species are able to replicate within mosquito cell cultures *in vitro*; however, they appear unable to replicate in mammalian cells or sucking mice, suggesting that they may be mosquito-specific. All almendravirus genomes are of similar size (~10 900–11 900 nt) and, in addition to the five canonical rhabdovirus structural protein genes, feature a long ORF between the G and L genes encoding a class 1A viroporin-like protein. Furthermore, viruses from the species *balsa almendravirus* contain an alternative long ORF (Px) near the start of the P gene, which is distinct from other almendravirus species.

The four *Curiovirus* species have been isolated from biting midges (*Culicoides* sp.), sandflies (*Lutzomyia* sp.) and mosquitoes (*Coquillettidia albicosta)*. As almendraviruses, curiovirus genomes encompass the additional ORF between the G

and L genes and also at least one further ORF between genes M and G, which makes their genome slightly larger in length (12 600–13 600 nts) in comparison to almendraviruses.

The genus *Sigmavirus* contains seven arthropod-specific rhabdovirus species. The type species, *drosophila melanogaster sigmavirus*, was discovered in 1937 and is considered the most widely studied species within the genus. Sigmavirus-infected hosts confer CO_2 sensitivity, which induces paralysis followed by death. Virus transmission occurs vertically through host eggs and sperm, with evidence of horizontal transmission yet to be observed. This has created an extreme host fidelity over ecological time (Longdon *et al*., 2010) and a significantly lower mutation rate than the estimated average for negative-strand RNA viruses. The genome organisation of sigmaviruses is typical of rhabdoviruses. The genome of Sigma virus of *Drosophila melanogaster* (Dme1SV) is ~12 kb and encodes six protein-coding genes – five which encode structural proteins N, P, M, G and L plus an additional X gene (also referred to as PP_3) whose function remains unclear; however, it may play a role in innate antiviral responses (Piontkivska *et al*., 2016).

Fish-adapted rhabdoviruses

The genera *Novirhabdovirus, Sprivivirus* and *Perhabdovirus*

There are currently three genera that encompass rhabdoviruses of aquatic hosts – *Novirhabdovirus, Sprivivirus* and *Perhabdovirus.* The *novirhabdovirus* genus contains four virus species including its type species *salmonid novirhabdovirus*, and the genera *sprivivirus* and *perhabdovirus* comprise two and three species, respectively, including their type species *carp sprivivirus* and *perch perhabdovirus*. All virus species within these genera infect finfish. Norvirhabdoviruses and perhabdoviruses have a diverse host range, whereas spriviviruses predominantly infect freshwater fish within the order *Cypriniformes* such as the common carp. Owing to their cold-blooded hosts, the temperature optima of these rhabdoviruses range from 15 to 25 °C. Transmission does not require a vector and occurs both horizontally as waterborne viruses and vertically as egg-associated viruses.

Rhabdoviruses of aquatic hosts have genomes of approximately 11 000 nt. The genus *Novirhabdovirus* was established as a new genus in the Rhabdoviridae family in 2009 owing to the presence of a nonvirion (NV) gene in members of this genus that is not present in other rhabdoviruses (**Figure 1**). Its function is unclear but may prevent apoptosis of host cells during early infection (Ammayappan and Vakharia, 2011). The morphological characteristics, genome organisation and sequence of spriviviruses and perhabdoviruses are most similar to that of mammalian vesiculoviruses.

Rhabdovirus infections in fish have a long history of severe disease impacts in freshwater and marine fishfarms, net-pen aquaculture and conservation fish hatcheries (Dietzgen *et al*., 2017). Fish rhabdoviruses cause 3 of the 10 globally reportable aquatic diseases: spring viremia of carp, infectious haematopoietic necrosis and viral haemorrhagic septicemia. Viruses from the species *Piscine novirhabdovirus* (VHSV) have been isolated in fish species from North America, Asia and Europe and have had a significant impact on trout farming in Europe. Viruses from the species *Salmonid novirhabdovirus* (IHNV), which originated in North America, have also been detected throughout Europe (Gomez-Casado *et al*., 2011). These viruses can cause infectious swelling of the kidneys and degeneration of the liver in their host species. A virus from the species *Snakehead novirhabdovirus* (SHRV) was first isolated in Thailand and causes necrotic ulcerations in infected fish (Johnson *et al*., 1999), whereas *hirame rhabdovirus* (HIRRV), first isolated in Japan, causes gonad congestion, haemorrhage of skeletal muscle and fins as well as an accumulation of fluid in the peritoneum (Sun *et al*., 2010).

Plant-adapted Rhabdoviruses

The genera *Cytorhabdovirus* and *Nucleorhabdovirus*

The *Cytorhabdovirus* genus currently has 11 defined species based on host range and vector specificity. This includes the type species, *lettuce necrotic yellows cytorhabdovirus*, which was first identified in lettuce. Lettuce necrotic yellows virus (LNYV) causes a flattened appearance in mature leaves with dark green spotting and necrosis (Martin *et al*., 2012). These rhabdoviruses are transmitted by aphids, plant hoppers or leaf hoppers, and while a large variety of plants are susceptible to this genus, each of the viruses normally has a restricted host range. As with other members of the Rhabdoviridae, the genome encodes for 5 proteins: N, P, M, G and L, as well as 1–5 accessory proteins and an additional gene encoding viral movement proteins located between P and M (Redinbaugh and Hogenhout, 2005; **Figure 1**). Replication occurs in the cytoplasm, with no evidence of nuclear involvement.

The *Nucleorhabdovirus* genus includes 10 different viruses, the type species of which is potato yellow dwarf nucleorhabdovirus (Bandyopadhyay *et al*., 2010). The genome organisation for potato yellow dwarf virus (PYDV) is similar to that of the other rhabdoviruses, with five genes (N, P, M, G and L) expressed as well as an additional putative movement protein gene expressed by the plant rhabdoviruses located between P and M; a further gene of unknown function located between N and P has been identified (Martin *et al*., 2012). The viruses in this genus are different from other rhabdoviruses in that replication occurs in the nucleus. The mechanism by which these viruses locate to the nucleus is unclear, although it appears that the inner nuclear membrane is required for particle maturation as enveloped viruses have been observed to accumulate within the perinuclear space.

The genera *Dichorhavirus* and *Varicosavirus*

The *Dichorhavirus* genus currently consists of two species: the type species *Orchid fleck dichorhavirus* and *Coffee ringspot dichorhavirus*. Both are transmitted by false spider mites (*Brevipalpus* spp.) and induce local chlorotic or necrotic spot symptoms in susceptible plant host species. The trade of orchid plants is likely responsible for the global identification of orchid fleck

virus (OFV), whereas Coffee ringspot virus (CoRSV) is yet to be identified outside of Brazil and Costa Rica (Rodrigues *et al.*, 2002). The *Varicosavirus* genus consists of a single species, *Lettuce big-vein associated varicosavirus* (LBVaV), which is distributed globally and transmitted by a soil-inhabiting cytrid fungus (*Olpidium virulentus*).

Dichorhaviruses and varicosaviruses feature the similar genome structure, transcriptional mechanism and gene order associated with rhabdoviruses; however, they are the first to be assigned to this family with segmented, bipartite genomes and nonenveloped particles. This classification has been made by the ICTV on the basis that from an evolutionary perspective, segmentation and particle morphology are relatively flexible characteristics when phylogenetically, these viruses are clearly related to the Rhabdoviridae with a similar gene order. The rod-shaped virons of OFV and CoRSV virons are approximately 40×100–110 nm in size and LBVaV typically 18×320–360 nm. The two negative-sense ssRNA segments that construct OFV and CoRSV genomes are composed of (1) RNA1 (~6.4 kb) encoding genes 3′-N-P-P3-M-G-5′ and (2) RNA2 (~6.1 kb) encoding the L polymerase. The LBVaV genome is of similar size but slightly differs in organisation: (1) RNA1 (6.7 kb) encoding L polymerase and an additional small ORF of unknown function and (2) RNA2 (6.1 kb) encoding an N protein homologue (coat protein (CP)), followed by four ORFs. The functions of these ORFs are currently unknown; however, they are thought to be equivalent to the plant rhabdoviral P, P3, M and G proteins (Dietzgen *et al.*, 2017).

Unassigned rhabdoviruses

The family Rhabdoviridae currently contains a single virus yet to be assigned to a genus – Moussa virus (MOUV). MOUV has the distinctive morphology and classical genome size (11.5 kb) and organisation of rhabdoviruses (**Figure 1**); however, sequence comparisons and phylogenetic analyses show no association of MOUV with a recognised rhabdovirus species or genus (Phenix-Lan *et al.*, 2010).

Structure

Morphology

Typically, rhabdoviruses appear as bullet-shaped particles, and this feature of rhabdovirus virions is a distinctive characteristic unique to the members of this diverse family. Across the family, virion size can vary significantly, especially where elongated virions are seen although generally particles display a diameter of 45–100 nm, while virion length can vary from 100 to 430 nm. The virion M_r is 300–1000×10^6, and virions have a CsCl (caesium chloride) buoyant density of approximately 1.19–1.20 g^{-1} cm^{-3} (1.17–1.19 g^{-1} cm^{-3} in sucrose). Some variation in dimensions may reflect the presence of virions containing truncated genomes termed defective interfering (DI) particles that are often seen as shorter particles within a preparation. These DI particles play an unknown role in the replicative life of viruses although their truncated genomes have been suggested to compete for virion and host cell proteins, which in turn can downregulate replication of full-length genomes.

Most rhabdovirus virions are protected by a lipoprotein envelope that is acquired from the membrane of the host cell by budding. Within this envelope the glycoprotein sits as a dense array of 5–10 nm spikes (peplomers) consisting of G trimers that enable docking to receptors on host cells and ultimately resulting in infection of the cell (Coll and INIA, 1995). Rhabdovirus virions are sensitive to heat (inactivation at 56 °C), UV, X-irradiation and exposure to lipid solvents.

Nucleic acid

The relative molecular mass (Mr) of the genome constitutes 1–2% of the virion by weight. With the exception of dichorhaviruses and varicosaviruses, the rhabdovirus genome consists of non-segmented, linear and negative-sense single-stranded RNA. Within a viral population, up to 5% of genome species are positive-stranded antigenomes; hairpin RNA forms are also observed (ICTV, 2017). At the genome termini, promoter regions exist, and there is often a high degree of terminal complementarity between these promoters suggesting similar roles. Indeed, these regions act as docking sites for the virus polymerase for the initiation of transcription and replication. **See also: RNA Virus Genomes** a0001488.pub2

Proteins

The virus genome encodes both structural and nonstructural proteins, and these make up approximately 65–75% of the particle weight (ICTVdB Management, 2006; **Table 2**).

Lipids and carbohydrates

Rhabdovirus virions are generally composed of 15–25% lipids by weight (although their composition is dependent on the host cell membrane). Of the total lipids present, phospholipids normally represent approximately 50–60% and sterols and glycolipids represent approximately 35–40%. The lipids present are derived from the host cell membranes (ICTVdB Management, 2006).

Carbohydrates have been detected in virus particles, constituting about 3%, and they are present as *N*-linked glycan chains on the G protein and also as glycolipids (ICTVdB Management, 2006). **See also: Glycolipids: Distribution and Biological Function; Lipids** a0001427.pub3 a0000711.pub3

Replication

The replication cycle is almost universal for all rhabdoviruses. Once the virus has entered the cell and uncoated, the nucleocapsid is released and transcription can occur. The viral genes are transcribed, translation of the viral proteins then occurs, and this is followed by cytoplasmic viral genome replication, encapsidation and budding of the virion from the cell membranes. Nucleorhabdoviruses and dichorhaviruses differ slightly due to the environment of plant cells and the establishment of replication factories in the nucleus, as opposed to the cytoplasm.

Attachment and adsorption

Adsorption involves the interaction of host cell surface receptors with the rhabdovirus G protein. Once the virus has bound to a

cellular receptor, it then generally enters the cell by fusion of the viral envelope with the cellular membrane (Albertini *et al.*, 2012). The identity of the specific receptors varies with the virus strain and can even involve numerous different receptors (as seen with RABV) which can utilise the nicotinic acetylcholine receptor, neural cell adhesion molecule, p75NTR, carbohydrates, phospholipids and gangliosides (Lafon, 2005). The virus can also enter the cell through coated pits and uncoated vesicles (viropexis or pinocytosis), which can lead to the incorporation of several virions at once (Tsiang *et al.*, 1983). Once the virus has been internalised, fusion of the endosomal membrane and the viral envelope is mediated by the virus glycoprotein (G), and the RNP is released in endosomal vesicles, the endosome then fuses with a lysosome and the enzymes allow the ribonucleocapsid to discharge into the cytoplasm (Wuner, 2007). The ribonucleocapsid represents the minimal replicative unit for these viruses and as such contains all of the components necessary to activate viral transcription and replication.

Transcription and replication

The presence of a negative-sense RNA genome means that mRNA must be transcribed in order to allow replication of the virus to occur. The RNA-dependent RNA polymerase (L protein), resident within the RNP, mediates transcription of the gRNA to produce the positive-stranded, capped and polyadenylated mRNAs (Albertini *et al.*, 2011). As the genome of most rhabdoviruses is not uncoated in the cytoplasm, the L protein recognises the viral RNA encapsulated within the N protein. The transcripts are produced in a consecutive monocistronic form. The generation of a transcriptional gradient whereby 3′ proximal genes are produced in abundance over 5′ genes is controlled by the transcription complex that recognises start, stop and polyadenylation transcription signals flanking each gene (Albertini *et al.*, 2011). This gradient allows major structural proteins such as N, which are required in relative abundance, to be expressed to a greater degree than 5′ proximal genes that are required in catalytic amounts. Following transcription, the nascent mRNAs are processed using the host cell translation machinery. The N, P, M and L proteins are expressed on free ribosomes in the cytoplasm, whereas the G-mRNA is translated on membrane-bound polysomes. The completion of the synthesis and glycosylation of G occurs in the endoplasmic reticulum and Golgi apparatus (Dietzgen, 2012).

The intracellular ratio of leader RNA to N protein is thought to regulate the switch from transcription to replication. When this switch is activated, replication of the viral genome begins (Dietzgen, 2012). The first step in viral replication is synthesis of full-length copies (positive-strand antigenomes) of the viral genome. When the switch to replication is triggered, RNA transcription becomes processive and ignores intergenic gene boundaries to produce a full-length antigenome-sense copy of the viral genome. These positive-strand replicative intermediate RNA species then serve as the template for synthesis of full-length negative-sense strands of the viral genome (CDC, 2011). **See also: RNA Plant and Animal Virus Replication**

a0001086

Assembly, Envelopment and Budding

During assembly, transcription and replication are inhibited by the downregulation of the viral-associated polymerase that occurs during infection. Downregulation is one of the roles of the M-protein, which causes a complete suppression of the transcriptase activity (Mebatsion *et al.*, 1999). The N–P–L complex encapsulates negative-stranded gRNA to form the RNP core, and the M-protein forms a capsule, or matrix, around the RNP. The RNP–M complex migrates to an area of the plasma membrane containing glycoprotein inserts, and the M-protein initiates coiling. Finally, the condensed M–RNP complex associates with the glycoprotein, and the complete virion buds from the plasma membrane (Mebatsion *et al.*, 1999).

Interestingly, vesiculoviruses and the ephemeroviruses replicate in the cytoplasm and bud from the plasma membrane. In contrast, the lyssaviruses bud predominantly from the internal membranes as do the plant rhabdoviruses, whereas the nucleorhabdoviruses bud from the inner nuclear membrane and are often seen to accumulate in the perinuclear space.

Pathobiology and Epidemiology of Some Clinically Important Rhabdoviruses

Pathobiology and epidemiology of vesiculoviruses

The vesiculoviruses contain a number of important pathogens such as VSIV, which causes disease in livestock and CHPV, which can cause disease in humans.

VSIV is one of the most well-characterised members of the rhabdovirus family, which can cause disease in both animals and humans. The virus causes vesicular lesions and febrile disease in livestock (horses, cattle and pigs). It is presumed that the virus is transmitted by blood-sucking arthropods, as VSIV and other vesiculoviruses have been isolated from mosquitoes, simulids and midges. Direct transmission of virus between infected animals is rare. It is rarely lethal but causes economic loss as a result of lameness, weight loss and interruption of lactation. VSIV is responsible for mild influenza-like febrile illness in the human population in rural areas of Central America. Significant levels of antibody are present in 25–90% of individuals in different communities. The virus is largely confined to the New World, although in the past it has appeared transiently in Africa and Europe before the implementation of effective veterinary control measures. In the United States, outbreaks occur in the summer months. **See also: Vesicular Stomatitis Virus**

a0001089.pub2

CHPV infects humans and can cause a mild febrile illness, as demonstrated in India in 1966, or sometimes a fatal encephalitis, as seen more recently with an outbreak in children from various parts of India (Rodriguez and Pauszek, 2012). The virus has been isolated directly from human serum in India and is thought to be transmitted by mosquitoes, ticks and sand flies.

Version 4 a0027748

Pathobiology and Epidemiology of Lyssaviruses

The type species of the lyssaviruses is *rabies lyssavirus*, which is one of the most lethal of all viruses. All lyssaviruses cause the disease rabies, which is a zoonotic disease endemic across much of the world. With a worldwide distribution, and a broad host range, the lyssaviruses have demonstrated a formidable versatility and adaptive capability. Where present, wildlife reservoirs of rabies include populations of foxes, jackals, raccoons, skunks and importantly bats. Rabid dogs remain the major source of RABV although where successful vaccination campaigns have eliminated terrestrial rabies (Western Europe), the presence of lyssaviruses in bat species maintains the threat to public health. Across the United States of America, the striped skunk and the raccoon continue to cycle RABV within populations. Most interestingly, the presence of lyssaviruses in bat populations continues to be problematic. It is of note that within the bat species of the New World, only classical RABV has been found with no other lyssaviruses being reported. In contrast, within the Old World, numerous other lyssaviruses have been reported in bat species with the apparent absence of classical RABV. The reasons for this, considering the historical distribution of classical RABV in terrestrial carnivore populations across Europe and the continued presence of rabies in fox and raccoon dog populations in Eastern Europe, are unclear. Importantly, within the United States of America, despite the elimination of terrestrial rabies from the domestic dog population, other terrestrial wildlife reservoirs remain, as does the threat of rabies from bat populations. Interestingly, a genetic analysis of viruses isolated from human cases of RABV has revealed the significance of bat strains of the virus, even where contact with bats could not be confirmed. **See also: Rabies: Virus and Disease**

a0000427.pub3

Transmission requires that RABV gain entry to the body through broken skin (a bite or scratch) or through the mucous membranes (eyes, nose and mouth) as the virus is unable to penetrate intact skin. Other routes of transmission are rare in humans, and the so-called cryptic infection usually involves an unnoticed bat bite (Messenger *et al*., 2002). Interestingly, infection has also been reported following the transplantation of biological tissues from a donor carrying the RABV (Dietzschold and Koprowski, 2004; Hellenbrand *et al*., 2005). The incubation period can vary dramatically (Charlton *et al*., 1997) and depends on a number of factors including the location of the initial bite wound and the quantity of virus introduced.

Initial clinical signs can include fever, chills, weakness, tiredness, sleeplessness, headache, anxiety, irritability and a lack of appetite. A sensation of numbness, tingling or itching can also sometimes be felt at the site of infection (Leung *et al*., 2007). The disease generally presents in one of the three forms: paralytic (dumb) rabies, encephalitic (furious) rabies or atypical presentation that includes a combination of both of these. Symptoms such as hypersexuality and hydrophobia may also occur in some patients. Death is normally due to circulatory failure. Human cases of rabies are rare in developed countries, where the establishment of domestic dog vaccination and removal of the threat from wildlife reservoirs has been achieved through extensive oral vaccination campaigns. Despite this, across developing countries, the annual death toll is estimated to exceed 55 000.

Control

As illustrated above, the rhabdoviruses are a diverse family of viruses infecting a broad range of hosts, and as such no single approach would suffice to control all of the viruses currently classified or awaiting classification. The following sections will aim to highlight the methods used to control and to protect against some of these viruses in a range of species.

Control of rhabdovirus infection in animals and humans

Infection of livestock with vesiculoviruses is not generally severe enough to warrant vaccination or further control methods. However, where severe disease is seen, control measures such as quarantine, vector removal and, where necessary, culling of infected animals are implemented.

Historically, control of RABV has also involved quarantine and culling of infected animals as the principal threat has been from rabid dogs. Preimmunisation of human populations has not been attempted as the threat of disease is generally rare and where knowledge and facilities allow, treatment following exposure prevents fatalities. Importantly, it is clear that simple measures can greatly reduce the incidence of dog and human rabies through responsible dog ownership, dog vaccination and informed reaction following human exposure including washing of the wound, administration of vaccine and, where necessary, postexposure rabies immunoglobulin (RIG). Where sylvatic rabies exists, oral vaccination campaigns have eliminated the virus from terrestrial carnivore populations through the use of oral baits and targeted campaigns of bait distribution in problem areas. Oral vaccines generally utilise either live attenuated vaccine strains or recombinant vaccinia viruses that express the rabies glycoprotein although other vector vehicles are under development. Such oral vaccination campaigns have successfully eliminated sylvatic rabies from Western Europe (Müller *et al*., 2015).

From a human health perspective, the availability of postexposure tools has long been an obstacle in the treatment of human cases following a suspected exposure and before the development of clinical disease. The development of rabies vaccines has a long history, beginning with the seminal works by Louis Pasteur more than 100 years ago (Vodopija and Clark, 1991). The development of tissue culture-generated vaccines for preimmunisation, and where needed as a postexposure tool, has revolutionised vaccine options for those exposed to virus. The current vaccines are safe and efficacious for pre- and postexposure prophylaxis and have none of the unwanted side effects associated with earlier nerve tissue-derived or duck embryo cell-derived vaccines. Alongside this, a shift has been made towards the development of alternative products to RIG, and monoclonal antibody cocktails have been evaluated as a suitable alternative. Such preparations have reduced production costs and therefore will enhance availability where needed. However, despite these advances in treatment options, once clinical disease has developed, there remains no reproducibly effective tool to prevent progression of disease and

a0000961
a0000490

ultimately death. **See also: Vaccination of Humans; Vaccines: Subunit**

Control of rhabdovirus infection in fish

Vaccination of large populations of fish for the commercial sector poses a number of problems including the delivery method of vaccination, as individual inoculation is not feasible for aquatic species. Furthermore, alternative control methods may have unwanted effects on other aquatic life impact on trade. Vaccination against VHSV and IHNV using live attenuated strains has been successful in a laboratory setting, but not in the field, with problems involving reversion to virulence. As a result, these vaccines are not suitable for commercial use. There are efforts to engineer live recombinant viruses using a reverse genetics approach. Inactivated viruses have also had some success but have not been produced in large scale. The most successful attempt has involved using DNA vaccines containing the G gene of either VHSV or IHNV; although used in North America, these are not currently used in Europe owing to concerns about the effects of DNA vaccination (Gomez-Casado *et al.*, 2011). Regardless, further tools are required to combat these economically important diseases. **See also: Vaccination of Animals**

a0000962

Control of rhabdovirus disease in plants

Plant rhabdoviruses are transmitted by a variety of arthropod species and infect a wide range of plant species. Importantly, numerous rhabdoviruses of plants remain uncharacterised and as such require further study to enable elucidation of potential control options. From the *nucleorhabdovirus* genus, the potato yellow dwarf virus is economically the most significant with infection of crops leading to huge economic losses. From the *cytorhabdovirus* genus, LNYV causes a serious disease of lettuce in Australia. In both cases, control has been achieved by a combination of vector removal, perennial weed eradication and introduction of virus-tolerant cultivars although members of these two genera require further study. **See also: Plant Virus Transmission by Insects; Virus Diseases of Potatoes; Viruses and Plant Disease**

a0000760.pub3
a0000763.pub2
a0000770.pub3

Rhabdoviruses as therapeutic and prophylactic agents

With the development of reverse genetics techniques, the linear genome and sequential gene expression of rhabdoviruses and other negative-strand RNA viruses have proven highly amenable to genetic engineering. Along with the existing knowledge that many rhabdoviruses contain more than the basic five proteins, it was very rapidly shown that they could support the insertion of foreign genes and were successfully used as vehicles to express foreign proteins.

No rhabdovirus has been utilised more than VSV for the development of recombinant tools for various purposes. Some examples of the use of VSV as a tool include the targeting of M-protein mutants of VSV to cancerous tissue and the inclusion of heterologous genes from other viruses as novel vaccines. In the former study, some specificity was evident as a consequence of the ability of the mutant virus to enhance interferon production in normal cells, while retaining the ability to induce apoptosis in tumour cells exhibiting a defective interferon response (Hastie and Grdzelishvili, 2012). In the latter case, recombinant VSV has been used prophylactically when engineered to express foreign viral proteins as novel vaccines, for example in the development of recombinant VSV-based vaccines directed against Ebola and Marburg viruses (Geisbert and Feldmann, 2011). Utilisation of rhabdoviruses as tools in these ways has been previously reviewed (Finke and Conzelmann, 2005).

With further development of molecular techniques, and further characterisation of rhabdoviruses, we may yet discover even more useful and versatile applications for this diverse family of viruses.

Glossary

Bacilliform Having a rod-like shape.

Gonad An organ that produces gametes, for example a testis or ovary.

Pleomorphic A variable appearance or morphology.

Sylvatic rabies A disease transmitted by wildlife, particularly foxes and wolves.

Vesicular disease A group of diseases resulting in vesicles on the skin of cloven-footed animals of major importance because of their high infectivity.

Viroporins Small, usually hydrophobic multifunctional viral proteins that are capable of modifying cellular membranes, rendering them more permeable and thus facilitating the exit of virions from the cell. Viroporins are not essential for viral replication but do enhance growth rates. Many viroporins also have effects on cellular metabolism and homeostasis mediated by interactions with host cell proteins.

References

Albertini AA, Ruigrok RW and Blondel D (2011) Rabies virus transcription and replication. *Advances in Virus Research* **79**: 1–22.

Albertini AA, Baquero E, Ferlin A and Gaudin Y (2012) Molecular and cellular aspects of rhabdovirus entry. *Viruses* **4** (1): 117–139.

Amarasinghe G, Bào Y, Basler C, *et al.* (2017) Taxonomy of the order Mononegavirales: update 2017. *Archives of Virology* **162**: 1–12.

Ammayappan A and Vakharia VN (2011) Nonvirion protein of novirhabdovirus suppresses apoptosis at the early stage of virus infection. *Journal of Virology* **85** (16): 8393–8402.

Bandyopadhyay A, Kopperud K, Anderson G, Martin K and Goodin M (2010) An integrated protein localization and interaction map for Potato yellow dwarf virus, type species of the genus Nucleorhabdovirus. *Virology* **402** (1): 61–71.

Blasdell KR, Voysey R, Bulach DM, *et al.* (2012) Malakal virus from Africa and Kimberley virus from Australia are geographic variants of a widely distributed ephemerovirus. *Virology* **433** (1): 236–244. DOI: 10.1016/j.virol.2012.08.008.

Blasdell K, Guzman H, Widen S, *et al.* (2015) Ledantevirus: a proposed new genus in the Rhabdoviridae has a strong ecological association with bats. *The American Journal of Tropical Medicine and Hygiene* **92** (2): 405–410.

CDC (2011) *The Rabies Virus*. http://www.cdc.gov/rabies/transmission/virus.html

Ceballos N, Morón S, Berciano JM, *et al.* (2013) Novel lyssavirus in bat, Spain. *Emerging Infectious Diseases* **19** (5): 793–795. DOI: 10.3201/eid1905.121071.

Charlton KM, Nadin-Davis S, Casey GA and Wandeler AI (1997) The long incubation period in rabies: delayed progression of infection in muscle at the site of exposure. *Acta Neuropathologica* **94** (1): 73–77.

Coll JM and INIA (1995) The glycoprotein G of rhabdoviruses. *Archives of Virology* **140** (5): 827–851.

Dietzgen RG (2012) Morphology, genome organization, transcription and replication of rhabdoviruses. In: Dietzgen RG and Kuzmin IV (eds) *Rhabdoviruses – Molecular Taxonomy, Evolution, Genomics, Ecology, Host–Vector Interactions, Cytopathology and Control*, pp. 5–11. Norfolk: Caister Academic Press.

Dietzgen RG, Kondo H, Goodin MM, Kurath G and Vasilakis N (2017) The family Rhabdoviridae: mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. *Virus Research* **227**: 158–170.

Dietzschold B and Koprowski H (2004) Rabies transmission from organ transplants in the USA. *Lancet* **364** (9435): 648–649.

Finke S and Conzelmann KK (2005) Recombinant rhabdoviruses: vectors for vaccine development and gene therapy. *Current Topics in Microbiology and Immunology* **292**: 165–200.

Geisbert TW and Feldmann H (2011) Recombinant Vesicular stomatitis virus-based vaccines against Ebola and Marburg virus infections. *Journal of Infectious Diseases* **204** (suppl. 3): S1075–S1081.

Gomez-Casado E, Estepa A and Coll JM (2011) A comparative review on European-farmed finfish RNA viruses and their vaccines. *Vaccine* **29** (15): 2657–2671.

Grard G, Fair JN, Lee D, *et al.* (2012) A novel Rhabdovirus associated with acute hemorrhagic fever in Central Africa. *PLoS Pathogens* **8** (9): e1002924. DOI: 10.1371/journal.ppat.1002924.

Gunawardena PS, Marston DA, Ellis RJ, *et al.* (2016) Lyssavirus in Indian flying foxes, Sri Lanka. *Emerging Infectious Diseases* **22** (8): 1456–1459. DOI: 10.3201/eid2208.151986.

Hastie E and Grdzelishvili VZ (2012) Vesicular stomatitis virus as a flexible platform for oncolytic virotherapy against cancer. *The Journal of General Virology* **93** (Pt 12): 2529–2545.

Hellenbrand W, Meyer C, Rasch G, Steffens I and Ammon A (2005) Cases of rabies in Germany following organ transplantation. *Eurosurveillance* **10** (2): E050224.6.

ICTVdB Management (2006) 01.062. Rhabdoviridae. In: Büchen Osmond C (ed) *ICTVdB The Universal Virus Database, version 3*. New York: Columbia University.

ICTV (2017) *Virus Taxonomy: The Classification and Nomenclature of Viruses – The Online (10th) Report of the ICTV*. Available: https://talk.ictvonline.org/ictv-reports/ictv_online_report/

Johnson MC, Maxwell JM, Loh PC and Leong JA (1999) Molecular characterization of the glycoproteins from two warm water rhabdoviruses: snakehead rhabdovirus (SHRV) and rhabdovirus of penaeid shrimp (RPS)/spring viremia of carp virus (SVCV). *Virus Research* **64** (2): 95–106.

Lafon M (2005) Rabies virus receptors. *Journal of Neurovirology* **11** (1): 82–87.

Leung AK, Davies HD and Hon KL (2007) Rabies: epidemiology, pathogenesis, and prophylaxis. *Advances in Therapy* **24** (6): 1340–1347.

Longdon B, Obbard D and Jiggins F (2010) Sigma viruses from three species of Drosophila form a major new clade in the rhabdovirus phylogeny. *Proceedings of the Royal Society B: Biological Sciences* **277**: 35–44. DOI: 10.1098/rspb.2009.1472.

Marston DA, McElhinney LM, Johnson N, *et al.* (2007) Comparative analysis of the full genome sequence of European bat Lyssavirus type 1 and type 2 with other Lyssaviruses and evidence for a conserved transcription termination and polyadenylation motif in the G-L 3′ non-translated region. *Journal of General Virology* **88** (Pt 4): 1302–1314.

Marston DA, Horton DL, Ngeleja C, *et al.* (2012) Ikoma Lyssavirus, highly divergent novel Lyssavirus in an African civet. *Emerging Infectious Diseases* **18** (4): 664–667.

Marriott AC and Easton AJ (1999) Reverse genetics of the Paramyxoviridae. *Advances in Virus Research* **53**: 321–340.

Martin KM, Dietzgen RG, Wang R and Goodin MM (2012) Lettuce necrotic yellows cytorhabdovirus protein localization and interaction map, and comparison with nucleorhabdoviruses. *Journal of General Virology* **93** (Pt 4): 906–914.

Mebatsion T, Weiland F and Conzelmann KK (1999) Matrix protein of rabies virus is responsible for the assembly and budding of bullet-shaped particles and interacts with the transmembrane spike glycoprotein G. *Journal of Virology* **73** (1): 242–250.

Mélade J, McCulloch S, Ramasindrazana B, *et al.* (2016) Serological evidence of lysaviruses among bats on Southwestern Indian Ocean Islands. *PLoS One* **11** (8): e0160553.

Messenger SL, Smith JS and Rupprecht CE (2002) Emerging epidemiology of bat-associated cryptic cases of rabies in humans in the United States. *Clinical Infectious Diseases* **35** (6): 738–747.

Müller TF, Schröder R, Wysocki P, Mettenleiter TC and Freuling CM (2015) Spatio-temporal use of oral rabies vaccines in fox rabies elimination programmes in Europe. *PLoS Neglected Tropical Diseases* **9** (8): e0003953.

Phenix-Lan Q, Junglen S, Tashmukhamedova A, *et al.* (2010) Moussa virus: a new member of the Rhabdoviridae family isolated from Culex decens mosquitoes in Côte d'Ivoire. *Virus Research* **147** (1): 17–24.

Piontkivska H, Matos L, Paul S, *et al.* (2016) Role of host-driven mutagenesis in determining genome evolution of sigma virus (DMelSV; Rhabdoviridae) in *Drosophila melanogaster*. *Genome Biology and Evolution* **8** (9): 2952–2963.

Pringle CR and Easton AJ (1997) Monopartite negative strand RNA genomes. *Seminars in Virology* **8**: 49–57.

Redinbaugh MG and Hogenhout SA (2005) Plant rhabdoviruses. *Current Topics in Microbiology and Immunology* **292**: 143–163.

Rodrigues JCV, Rodriguez CM, Moreira L, *et al.* (2002) Occurrence of Coffee ringspot virus, a Brevipalpus Miteborne Virus in Coffee in Costa Rica. *Plant Disease* **86** (5): 564.

Rodriguez LL and Pauszek SJ (2012) Genus vesiculovirus. In: Dietzgen RG and Kuzmin IV (eds) *Rhabdoviruses – Molecular Taxonomy, Evolution, Genomics, Ecology, Host–Vector Interactions, Cytopathology and Control*, pp. 23–35. Norfolk: Caister Academic Press.

Sun Y, Yue Z, Liu H, *et al.* (2010) Development and evaluation of a sensitive and quantitative assay for hirame rhabdovirus based on quantitative RT-PCR. *Journal of Virological Methods* **169** (2): 391–396.

Tsiang H, Derer M and Taxi J (1983) An *in vivo* and *in vitro* study of rabies virus infection of the rat superior cervical ganglia. *Archives of Virology* **76** (3): 231–243.

Vodopija I and Clark HF (1991) Human vaccination against rabies. In: Baer GM (ed) *The Natural History of Rabies*, 2nd edn. Boca Raton, FL: CRC Press.

AU:5

Walker PJ and Kongsuwan K (1999) Deduced structural model for animal rhabdovirus glycoproteins. *Journal of General Virology* **80** (Pt 5): 1211–1220.

Wertz GW, Perepelitsa VP and Ball LA (1998) Gene rearrangement attenuates expression and lethality of a nonsegmented negative strand RNA virus. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 3501–3506.

Whelan SP, Barr JN and Wertz GW (2004) Transcription and replication of nonsegmented negative-strand RNA viruses. *Current Topics in Microbiology and Immunology* **283**: 61–119.

Wuner H (2007) Rabies virus. In: Jackson AC and Wunner WH (eds) *Rabies*, 2nd edn, pp. 23–69. UK: Academic Press.

AU:6

Further Reading

Baer GM (1994) Rabies – an historical perspective. *Infectious Agents and Disease* **3**: 168–180.

Dietzgen RG and Kuzmin IV (2012) *Rhabdoviruses: molecular taxonomy, evolution, genomics, ecology, host–vector interactions, cytopathology and control*. Norfolk: Caister Academic Press.

Flint SJ, Enquist LW, Racaniello VR and Skalka AM (2015) *Principles of Virology*, 4th edn. Washington, DC: ASM Press.

Jackson AC and Wunner WH (2007) *Rabies*, 2nd edn. Amsterdam: Elsevier – Academic Press.

Version 4 a0027748

Article Title: Rhabdoviruses

Article ID: a0027748

Article DOI: 10.1002/9780470015902.a0001085.pub4

Version: 4

Previous version(s):
10.1002/9780470015902.a0001085.pub3

Article Type: Standard

Readership Level: Advanced article

Top level subject categories: Virology

Keywords: nonsegmented # negative-sense single-stranded RNA # Rhabdoviridae # genomes # bullet-shaped envelope # rabies virus

Key Concepts

- Only the lyssaviruses and the vesiculoviruses are able to infect both animals and humans to cause clinical disease.
- Rabies is the most notable human pathogen and is one of the most deadly infectious diseases known with a case-fatality rate approaching 100%.
- Virions are bullet-shaped and consist of an envelope, covered with peplomers surrounding a helically coiled cylindrical nucleocapsid.
- The genome of the majority of Rhabdoviruses consists of a single molecule of linear negative-sense single-stranded RNA, 13–16 kb in size.
- Dichorhaviruses and Varicosaviruses are newly described rhabdoviruses that lack an envelope and have bipartite genomes.
- The Rhabdoviridae are a diverse group of highly adaptable viruses with a broad host range.
- All Rhabdoviruses have non-Latinised binomial species names to indicate that genus the species belongs to and to facilitate differentiation between virus names and their taxonomic species name.

Author(s) and Affiliation(s):

Daisy Jennings,
Denise A Marston,
Ashley C Banyard,
Anthony R Fooks,

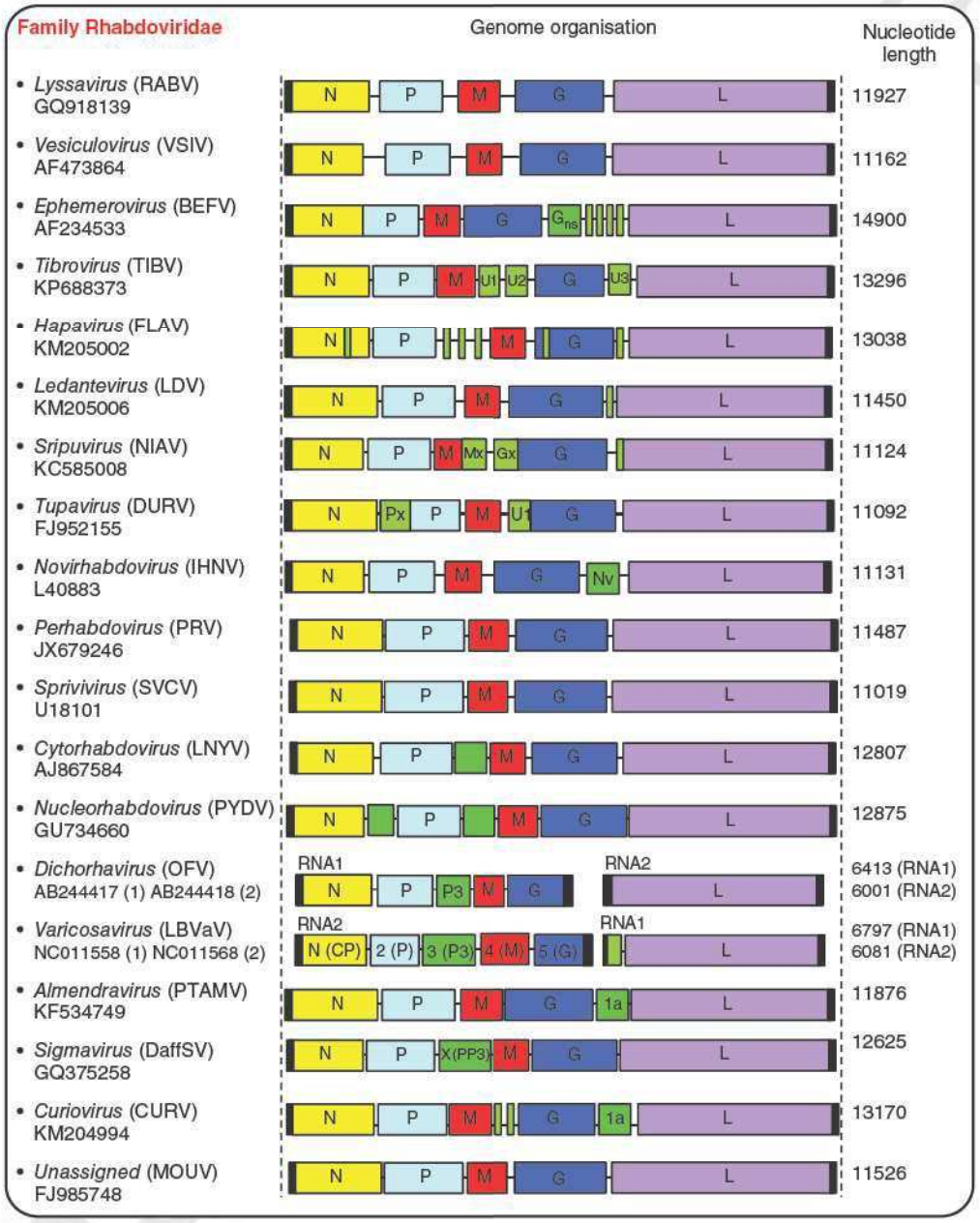
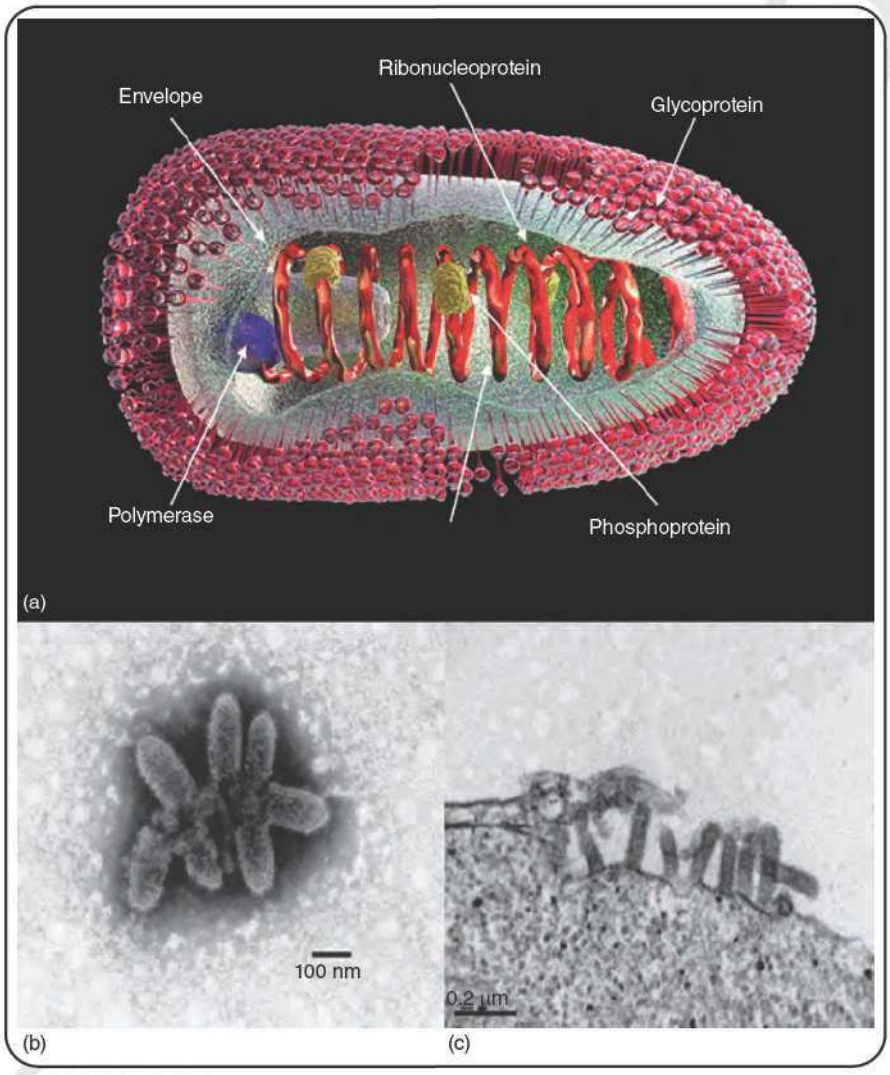


Figure 1 Genome organisation within the family Rhabdoviridae. Gene ORF sizes (the abbreviations for which are detailed in the text) are not shown to scale, and the total genome length has been standardised. Genome lengths and the accession number of an example virus are also stated next to genome schematics; extra genes are highlighted in green.



AU:7

Figure 2 (a) A pictorial representation of RABV, adapted from image designed by A R Fooks, APHA and produced by A Featherstone, Research Graphix, UK. (b) Electron micrograph 200 000× magnification (negative staining) of RABV (reproduced with kind permission from Bill Cooley, APHA). (c) Viral haemorrhagic septicaemia (VHS) budding from membranes of cells grown in culture. Reproduced from the EM unit, CEFAS. © UK Crown Copyright (2017). Reproduced with the permission of the Controller of Her Majesty's Stationary Office/Queen's Printer for Scotland and Animal and Plant Health Agency.

Table 1 Rhabdovirus taxonomy, host range and vector species

Genus	Number of species	Type species	Acronym	Main hosts	Vector
Almendravirus	5	*Puerto Almendras almendravirus*	PTAMV	Mosquitoes	n/a
Curiovirus	4	*Curionopolis curiovirus*	CURV	Mammals	Arthropods
Cytorhabdovirus	11	*Lettuce necrotic yellows cytorhabdovirus*	LNYV	Plants	Arthropods
Dichorhavirus	2	*Orchid fleck dichorhavirus*	OFV	Plants	Arthropods
Ephemerovirus	8	Bovine fever *ephemero*virus	BEFV	Cattle	Arthropods
Hapavirus	15	*Flanders hapavirus*	FLAV	Birds/reptiles/mammals	Arthropods
Ledantevirus	14	*Le Dantec ledantevirus*	LDV	Bats/man	Arthropods
Lyssavirus	14	Rabies lyssavirus	RABV	Bats/canines/man	n/a
Novirhabdovirus	4	*Salmonid novirhabdovirus*	IHNV	Fish	n/a
Nucleorhabdovirus	10	*Potato yellow dwarf nucleorhabdovirus*	PYDV	Plants	Arthropods
Perhabdovirus	3	*Perch perhabdovirus*	PRV	Fish	n/a
Sigmavirus	7	*Drosophila melanogaster sigmavirus*	DaffSV	Arthropods	n/a
Sprivivirus	2	*Carp sprivivirus*	SVCV	Fish	n/a
Sripuvirus	5	*Niakha sripuvirus*	NIAV	Reptiles	Arthropods
Tibrovirus	6	*Tibrogargan tibrovirus*	TIBV	Cattle	Arthropods
Tupavirus	3	*Durham tupavirus*	DURV	Birds/mammals	Unknown
Varicosavirus	1	*Lettuce big-vein associated varicosavirus*	LBVaV	Plants	Fungus
Vesiculovirus	16	*Indiana vesiculovirus*	VSIV	Cattle/horses/pigs	Arthropods
Unassigned	1	Moussa virus	MOUV	Mosquitoes	Unknown

Source: Adapted from Amarasinghe *et al.* (2017) and Dietzgen *et al.* (2017).

Table 2 The structural proteins of rhabdoviruses

Structural proteins	Identity across family[a] (%)	Range in gene length (nts)	Molecular weight range (kDa)	Principal role
Nucleoprotein (N)	28–57	1176–1419	95–116	Encapsidation of genome and antigenome RNA; protection of nucleic acid from host nucleases
Phosphoprotein (P)	28–41	714–1014	56–82	Interacts with N to form RNP; maintains N in soluble form in cytoplasm
Matrix protein (M)	26–42	480–762	39–72	Interacts with RNP and cytoplasmic tail of G; key structural component; interacts with host cell chaperones to facilitate budding
Glycoprotein (G)	25–47	1107–2463	92–197	Principal receptor-binding molecule; binds to host cell and facilitates virus entry
Large polymerase protein (L)	30–6259	5634–6435	459–531	Mainly enzymatic component of RNP responsible for enzymatic reactions that drive transcription and replication; binds to N and P to form the RNP; required in catalytic amounts

[a]The type species for each rhabdovirus genus was taken to calculate the amino acid identity across each ORF and to state the range of gene lengths.

1.4.2: ICTV Classification

The International committee of Taxonomy of Viruses (ICTV) has the task of developing, refining, and maintaining universal virus taxonomy. The goal of this undertaking is to categorize the multitude of known viruses into a single classification scheme that reflects their evolutionary relationships. The objectives are: (i) to develop an internationally agreed taxonomy for viruses; (ii) to develop internationally agreed names for virus taxa; (iii) to communicate taxonomic decisions to the international community of virologists; (iv) to maintain an Index of agreed names of virus taxa. The ICTV publish yearly reports which are available online.

Within the *Lyssavirus* genus, viruses are classified according to both genetic and antigenic data. Historically the viruses were divided into serotypes using monoclonal antibody typing. Serotypes were replaced by genotypes (but the groups remained the same, so that Serotype 1 became genotype 1 etc.) when genetic data was obtained from conserved regions of the genome (usually the nucleoprotein or polymerase genes). Subsequently, there was a shift in nomenclature from the definition of lyssavirus genotypes to that of lyssavirus species (ICTV, 2012). ICTV have updated the nomenclature for the entire Order of *Mononegavirales*, including lyssaviruses, to the standardised non-Latinized binomial species name format (Amarasinghe and Bao, 2017). The renaming, although slightly cumbersome, has resolved the inconsistencies in naming viruses and species within the genus (see table 1).

Table 1: List of all species in the genus Lyssavirus

Original nomenclature detailed alongside the current format. Virus names and abbreviations are included for completeness

Serotype	Genotype	Species	Updated species	Virus name	Virus Abbreviation	Approved by ICTV
1	1	*Rabies virus*	*Rabies lyssavirus*	rabies virus	RABV	Yes
2	2	*Lagos bat virus*	*Lagos bat lyssavirus*	Lagos bat virus	LBV	Yes
3	3	*Mokola virus*	*Mokola lyssavirus*	Mokola virus	MOKV	Yes
4	4	*Duvenhage virus*	*Duvenhage lyssavirus*	Duvenhage virus	DUVV	Yes
5	5	*European bat lyssavirus type 1*	*European bat 1 lyssavirus*	European bat lyssavirus type 1	EBLV-1	Yes
6	6	*European bat lyssavirus type 2*	*European bat 2 lyssavirus*	European bat lyssavirus type 2	EBLV-2	Yes

	7	*Australian bat lyssavirus*	*Australian bat lyssavirus*	Australian bat lyssavirus	ABLV	Yes
-	-	*Aravan virus*	*Aravan lyssavirus*	Aravan virus	ARAV	Yes
-	-	*Bokeloh bat lyssavirus*	*Bokeloh bat lyssavirus*	Bokeloh bat lyssavirus	BBLV	Yes
-	-	*Ikoma lyssavirus*	*Ikoma lyssavirus*	Ikoma lyssavirus	IKOV	Yes
-	-	*Irkut virus*	*Irkut lyssavirus*	Irkut virus	IRKV	Yes
-	-	*Khujand virus*	*Khujand lyssavirus*	Khujand virus	KHUV	Yes
-	-	*Shimoni bat virus*	*Shimoni bat lyssavirus*	Shimoni bat virus	SHIBV	Yes
-	-	*West Caucasian bat virus*	*West Caucasian bat lyssavirus*	West Caucasian bat virus	WCBV	Yes
-	-	*Gannaruwa bat lyssavirus*	*Gannaruwa bat lyssavirus*	Gannaruwa bat lyssavirus	GBLV	awaiting
-	-	*Lleida bat lyssavirus*	*Lleida bat lyssavirus*	Lleida virus	LLEBV	awaiting

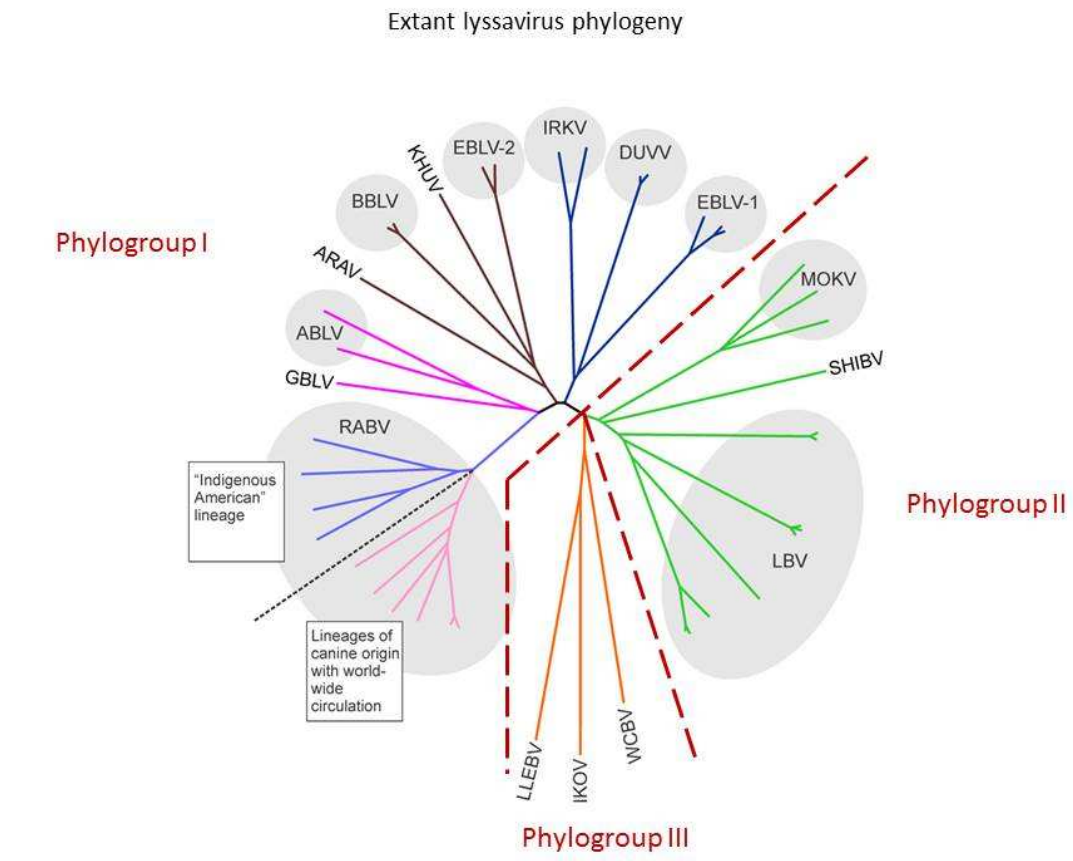
1.4.3: Lyssavirus Phylogroup Classification

In addition to the separation of genetically distinct viruses into species, these viruses can also be divided into phylogroups using genetic and antigenic data (Horton et al., 2010; ICTV, 2012; Nadin-Davis, 2007) (Figure 5). The differentiation is mainly based on phylogenetic analysis, but also the degree of cross protection afforded by rabies vaccines between each of the viruses characterised and cross protection of polyclonal antisera. Viruses within Phylogroup I are cross-neutralised by polyclonal sera raised against other phylogroup I viruses (although the level of cross reactivity does vary). Genetic data allows reliable quantitative comparison of lyssavirus divergence and sequence data are available for at least one isolate from all 14 virus species classified (and 2 of the 3 unclassified viruses) within the *Lyssavirus* genus.

There is a low degree of cross neutralisation between phylogroups I and II and also between both phylogroups and WBCV (Horton et al., 2010). *In vivo* vaccine challenge experiments have shown reduced or no efficacy of licensed rabies vaccines against viruses in phylogroup II (MOKV, LBV, SHIBV) (Badrane et al., 2001) and phylogroup III (WCBV) (Hanlon et al., 2005; Horton et al., 2014). There is also evidence for variable vaccine efficacy against some viruses in phylogroup I (Brookes et al., 2006; Brookes et al., 2005; Fekadu et al., 1988; Hanlon et al., 2005) which suggests a gradual loss of vaccine protection corresponding to antigenic distance from vaccine strains.

Figure 5: Unrooted phylogenetic tree of currently recognized and putative lyssaviruses

(neighbor-joining method, p-distances matrix).Phylogroups segregated by red dashed lines. Modified from (Rupprecht et al., 2017).



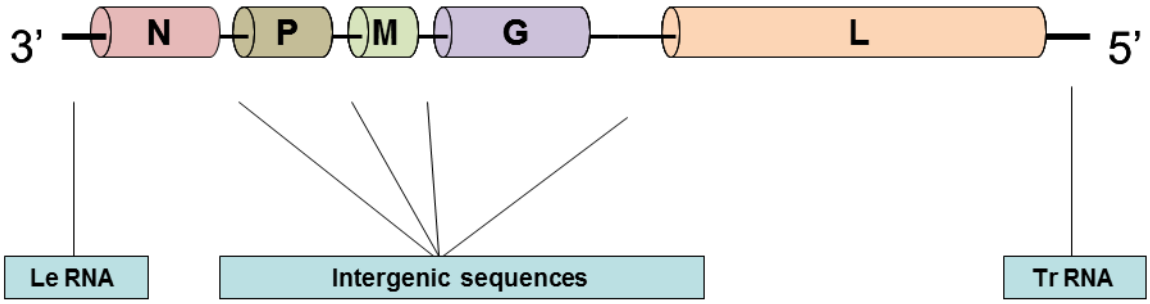
Rupprecht C, Kuzmin I and Meslin F 2017 [version 1; referees: 2 approved] F1000Research 2017, 6:184 (doi: 10.12688/f1000research.10416.1)

1.5: Lyssavirus Genome Organisation

All lyssaviruses are enveloped, deriving their envelope from the plasma membrane of the infected host cell upon budding; they exhibit the classic rhabdoviral bullet-like morphology. They generally measure ~75nm by 180nm (Warrell and Warrell, 2004). The genome organisation of members of the *Lyssavirus* genus is relatively simple, consisting of between 11,000 and 12,000 nucleotides which encode just five genes: N, P, M, G and L (Dietzschold et al., 2008). Across the *Rhabdovirus* family, the organisation of genes is highly conserved. The nucleoprotein (N) is positioned at the 3' end followed by the phosphoprotein (P) and matrix protein (M), the glycoprotein (G) and finally the large polymerase (L) at the 5' end. This conserved order of genes may be an adaptation to enable control

of protein expression (Hodges et al., 2012). The genome arrangement for lyssavirus is shown in figure 6.

Figure 6: Genomic organisation of lyssaviruses



1.5.1: The Nucleoprotein

The nucleoprotein (N) is the most well conserved of the lyssavirus proteins (Mannen et al., 1991). As a result the majority of diagnostic assays and phylogenetic studies target the N - gene. (Sacramento et al., 1991). In addition, ICTV demarcation criteria for lyssavirus species differentiation requires 80% or less nucleotide identity for the N-gene. The lyssavirus nucleoprotein is approximately 450 amino acids in length and one of its main functions is the encapsidation of the viral genome. The core of the virus genome; the ribonucleoprotein complex (RNP) consists of the genomic RNA tightly encapsidated by N in association with the phosphoprotein and large polymerase (Tordo, 1996) and is the minimal essential replication unit (Schnell et al., 2010). N also plays a role in the regulation of transcription and replication (Kissi et al., 1995). Accumulation of N in infected cells is thought to trigger a switch from transcription of the monocistronic messenger RNAs (mRNA) to transcription of the genome-length positive sense template RNA. It is this switch which induces replication of the negative sense RNA genome (Tordo et al., 1988).

The N protein has also been implicated in the induction of B cell proliferation and T helper cell (T_h) responses, both of which can induce long lasting humoral immunity to lyssaviruses.

1.5.2: The Phosphoprotein

The phosphoprotein (P) gene encodes a protein of between 297 and 303 amino acids (Tordo, 1996). It exists in two forms with a molecular weight of 38-41 kDa, depending on the degree of phosphorylation (Gupta et al., 2000). It has been reported to play a role in viral transcription and also pathogenicity (Tan et al., 2007). P is an essential component for the function of the large polymerase (L) protein (Marston et al., 2007). It is also able to initiate interaction with the cytoplasmic dynein light chain LC8 which may contribute to the axonal transport of RABV in infected neurones (Lo et al., 2001) and it may also play a role in down regulation of innate host immune responses post infection (Brzozka et al., 2006).

1.5.3: The Matrix protein

The matrix (M) is a small protein; 23-24 kDa which plays multiple roles within the lyssavirus replication cycle. It is the main structural component of the virus, being proposed to form a layer between the glycoprotein and the RNP core which holds the virion in its distinctive bullet-shaped form (Graham et al., 2008). It does this by interacting with the cytoplasmic tail of the glycoprotein at the cell membrane and promotes virus assembly and budding (Mebatsion et al., 1999). It has also been suggested that M may play a role in regulation of the balance between transcription and replication of the viral genome (Finke et al., 2003). Like P, M may be able to down regulate some host gene expression (Komarova et al., 2007) and may also be involved in the induction of apoptosis of infected cells (Kassis et al., 2004; Morimoto et al., 1999).

1.5.4: The Glycoprotein

The glycoprotein (G) is composed of 512 to 524 amino acids (Marston et al., 2007) and 65kDa in weight (Tordo, 1996). G is the only protein present on the virion surface, therefore plays an essential role in controlling virus and cell interactions (Rupprecht et al., 2002) including: determination of cell tropism (Wunner et al., 1984), fusion of the viral membrane and host cell endosome to release the nucleocapsid into the host cell cytoplasm (Gaudin et al., 1993) and formation of the virion envelope by its interaction with M (Mebatsion et al., 1999). The variation in the sequence of G between lyssavirus isolates is thought to play a role in the differing degrees of pathogenicity seen between different RABV strains or lyssavirus species (Faber et al., 2005; Morimoto et al., 2000; Yan et al., 2002).

The G protein is the main immunogenic component of the virion, thus the primary target for virus neutralising antibodies (VNAs) (Benmansour et al., 1991; Dietzschold et al., 1982). G is also able to stimulate cytotoxic T cells (T_c) (Macfarlan et al., 1986) and in combination with the induction of VNAs is the sole protein responsible for the induction of protective immunity against challenge (Wiktor, 1985).

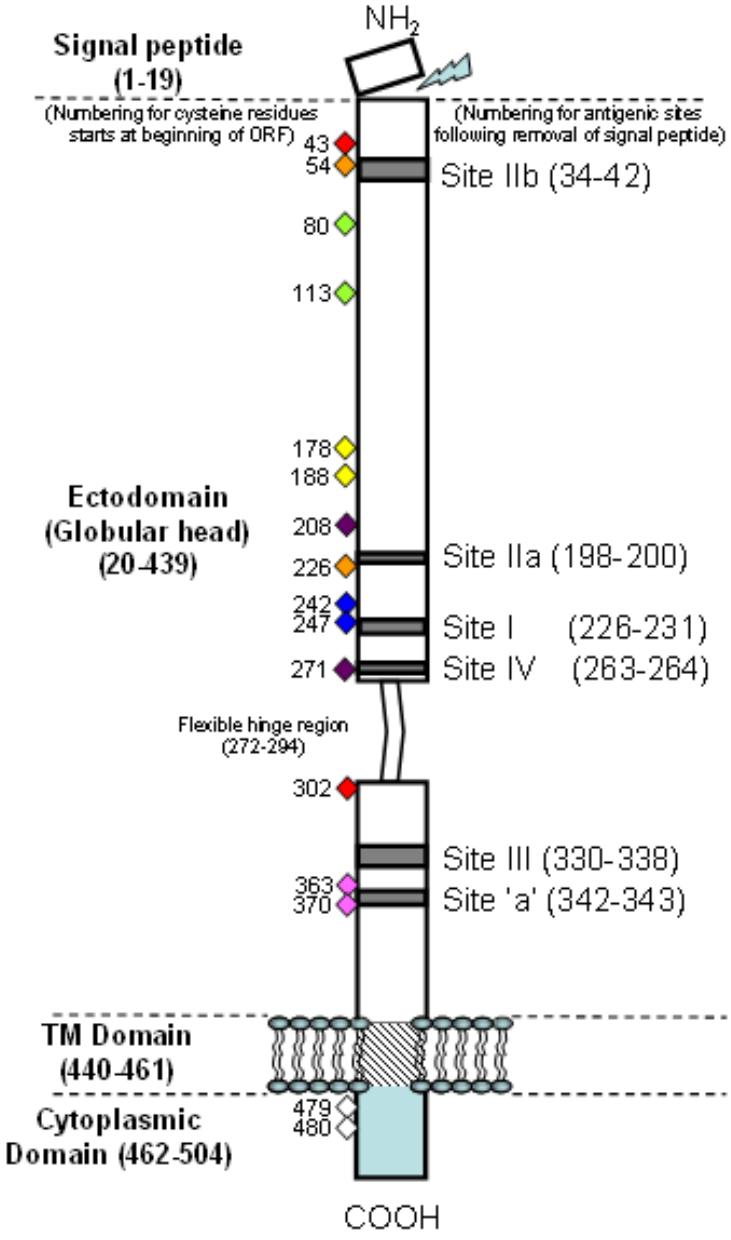
1.5.4.1: Antigenic Sites on the Glycoprotein

Present on the surface of the virion, the lyssavirus G protein is the sole viral component responsible for induction of a host antibody response and as such is the target of host neutralising antibodies (Gaudin et al., 1992). Furthermore all lyssavirus proteins contain 14 highly conserved cysteine residues alongside the antigenic domains that contribute to the structure of the protein (Figure 7).

Currently, defined antigenic domains within the G protein include four major antigenic sites and one minor antigenic site. Antigenic site I is a region containing both conformational and linear epitopes as it is delineated by the monoclonal antibody mAb 509-6 that recognises a conformational epitope as well as antibody CR57 that recognises a linear epitope (Benmansour et al., 1991). This site also appears to be the region recognised by the potent neutralising antibody, 62-7-13, a candidate antibody that has been proposed for inclusion in a monoclonal antibody based passive immunisation cocktail (Both et al., 2013; Muller et al., 2009). Antigenic site II has two domains, IIa and IIb, and is thought to be present as a discontinuous conformational epitope (Benmansour et al., 1991). Antigenic site II is proposed to be recognised by the neutralising antibody, E559 (Muller et al., 2009). Antigenic site III is continuous, though no site specific mAbs are able to bind unfolded protein which indicates that this epitope forms part of a loop on the protein surface and is this tertiary structure which enables recognition by antibodies or neuronal receptors (Benmansour et al., 1991).

Figure 7: Schematic representation of the lyssavirus G protein.

Defined antigenic sites and conserved cysteine residues are shown. C residues are coloured according to proposed linkages within the mature protein. TM: transmembrane; NH2: amino terminus; COOH: Carboxyl-terminus. Taken from (Evans et al., 2012).



The location of antigenic site IV is currently undefined. Studies have suggested that site IV is present at residues 263-264 so consists of only two amino acids and is continuous though it contains overlapping linear epitopes (Ni et al., 1995). It has also been suggested that site IV contains overlapping linear epitopes with key residues at 251 and 264 (Both, 2013) though another study suggested that site IV comprises only of residue 251 (Liu et al., 2013a). This suggests that a more conclusive investigation into the exact location of antigenic site IV may be required. For the purposes of this investigation, antigenic site IV has been defined as residues 263-264. Minor site 'a' is located in close proximity to site III but contains no overlapping epitopes and consists of only two amino acids (Benmansour et al., 1991).

1.5.5: The Large Polymerase

The polymerase (L) is the largest of the rhabdoviral proteins at around 2142 amino acids and it represents 54% of the viral genome (Koser et al., 2004). It is responsible for the viral transcription activities with four main functions: as an RNA dependent RNA polymerase (RdRP), mRNA methyltransferase, mRNA guanylyltransferase and a poly(A) synthetase (Wunner, 1991). The overall sequence of rhabdoviral L proteins is fairly homologous to other Mononegavirales including vesicular stomatitis virus (VSV) and Sendai virus however some regions show considerable variability, likely due to the multifunctional nature of L (Poch et al., 1990).

1.6: Virus Diversity

The publication of *The Origin of Species* in 1859, Darwin proposed natural selection as the mechanism of evolution which could explain the diversity of life. Since then, biologists have compared living species and populations to themselves and fossil to give insights into how populations have changed over time. The advent of DNA sequencing revolutionised this field, enabling populations to be compared at the genetic level, and phylogenies were utilised to resolve evolutionary relationships. Mircoorganisms, and in particular viruses, are ideal models to study evolution because they provide many advantages. Firstly, large viral populations can propagated in the laboratory, for example some lyssaviruses can reach titres up to 10^7 virus particles per mL when grown in optimum conditions in tissue culture. Secondly viral generation times are short, going through up to 5 generations in 24 hours, resulting in up to 100 generations in a 20 day serial passage experiment (Turner and Chao,

1998). Thirdly, viruses have high mutation rates in the region of 10^3-10^8 mutations per nucleotide per cell (m/n/c) (Sanjuan, 2010). Finally, small genome sizes enable researches to obtain complete genome sequences and manipulate the genome through site-directed mutagenesis to study the effects of specific mutations onto known genetic backgrounds.

It has long been observed that there is a link between genome length and viral mutation rate in DNA viruses (Drake, 1993). Although not so obvious, there is also evidence that mutation rates and genome size has co-evolved in RNA viruses (Sanjuan, 2012). RNA viruses have an average genome length of 10,000 nts, lyssaviruses have an approximate genome size of 12,000 nts. Much of the diversity observed within RNA genomes is a direct result of the lack of fidelity of the RNA polymerase encoded within the virus genome. The absence of the 3' exonuclease proofreading activity explains the higher mutation rates of RNA viruses compared to DNA viruses. Other factors affecting viral mutation rates include modes of replication, and host factors such as availability of dNTPs and mRNA-editing cytidine deaminases. The double stranded RNA dependent adenosine deaminase (ADAR) hypermutates adenosines to inosines for many viruses including rhabdoviruses (Carpenter et al., 2009)

The following article is a book review of the following book: Virus Evolution: Current Research and Future Directions.

1.6.1: Article 3 Submitted to Journal of Wildlife Diseases:

Book Review of 'Virus Evolution: Current Research and Future Directions'.

Summary:

A book review can be defined as an analysis of a book based on content style and merit (Peh and Ng, 2010). The Journal of Wildlife Diseases who commissioned the review provided a reference which proved useful as a basis (Scarnecchia, 2004). Writing a book review is an art in itself, a completely different style to scientific writing, which at first I found challenging. The opportunity to include personal observations, wit and creative literacy colour, was particularly daunting at the onset, yet ultimately rewarding. Finding relevant examples to use as a basis was not straightforward, as many book reviews are not retrievable using usual search engines.

The resulting review, attempts to capture the essence of the book, highlight areas of particular interest and provide context for potential readers. It was a welcome relief to writing this thesis and scientific manuscripts held within, whilst increasing my knowledge of the diversity that is virus evolution.

Journal of Wildlife Diseases, 54(1), 2018, pp. 000–000

Book Review

Edited by Charles E. Rupprecht
charles_rupprecht@yahoo.com

Book reviews express the opinions of the individual authors regarding the value of the book's content for Journal of Wildlife Diseases *readers. The reviews are subjective assessments and do not necessarily reflect the opinions of the editors, nor do they establish any official policy of the Wildlife Disease Association.*

Virus Evolution: Current Research and Future Directions. S. C. Weaver, M. Dennison, M. Roossinck and M. Vignuzzi (editors). Caister Academic Press, Poole, UK. 2016. 366 pp. Paperback ISBN: 978-1-910190-23-4, Ebook ISBN: 978-1-910190-24-1. £159, US $319.

Review by Denise A. Marston

?1 This is a highly informative book that summarizes recent advances in virus evolution in a comprehensive yet accessible manner. Traditionally monopolized by a few human pathogens such as HIV, Hepatitis C, and in uenza virus, the virus evolution field has recently expanded to include the viruses that cause emerging zoonoses such as severe acute respiratory syndrome, Middle East respiratory syndrome (MERS), dengue, and Ebola. Edited by world-renowned viral evolutionists, this book benefits from 12 chapters, all written by prominent scientists within the relevant subject matters.

The initial chapters are particularly commendable as introductory material; the first addresses why viruses have high mutation rates and how they can be quantified and then discusses the implications for disease emergence and control. The second chapter is an excellent introduction to viral informatics, including where to find related resources such as viral sequence databases and analytical software. Clearly, there is something for everyone in these chapters; although aimed at the novice, even the most experienced viral evolutionist can glean something.

Preconceived notions of the contents and focus of this book are not disappointed. For example, there is a good smattering of the old faithful bugs (i.e., HIV, In uenza A, and Ebola virus). However, there are some welcome surprises, such as Chikungunya virus and parvovirus, which help to keep the reader's attention. The paleovirology chapter (Chapter 10) is one of the more interesting, as viruses are much older than previously realized. Likely to become a firm favorite in future virus evolution books, here it is modestly introduced, providing examples across the seven main groups of viruses classified by the Baltimore system. This ?2 rapidly expanding field has capitalized on the overall expansion of high throughput sequencing technologies, utilizing host genomic data to identify endogenous viral elements, which ultimately are the closest we have to 'fossils' of contemporary viruses.

Whether readers of the *Journal of Wildlife Diseases* should invest in this book is a matter of context. One could argue that viruses are the most abundant and genetically diverse entities, present and ultimately 'parasitic' in all major domains of life, evolving over hundreds of millions of years to the enormous expanse observed today, and hence the content is applicable to every reader. Indeed, I would venture so far as to say that chapters within this book would inform and inspire wildlife disease professionals to better appreciate and understand the evolutionary strategies of their favorite viruses or perhaps those that affect their most-liked wildlife populations. Perhaps more importantly, reading this

book would broaden any wildlife virologist's knowledge due to the inclusion of plant viruses and bacteriophages. Inevitably, there are evolutionary models and some more-complex population genetic modelling that required greater focus and introspection, but these topics are introduced clearly and in manageable amounts, encouraging the reader to continue reading and rewarding their persistence with tangible evidence explaining why these models are so beneficial for consideration. For example, understanding that MERS has infected so many people—not because the R_0 is high but because multiple cross-species transmissions have occurred. Without exception, each chapter is written with accessibility to all readers in mind. More difficult concepts are concisely explained and the English is to a high standard (albeit American English!). Chapter 6, "Evolution of Viral Virulence: Empirical Studies," stretched the limit of my focus. Arguably, this is one of the more pivotal chapters, relating evolutionary theories to empirical data, fulfilling in brief to include both animal and plant viruses—yet it lost me (the reader) amongst the pages.

One distinct advantage of this book is the independence of each chapter from the others, bound together as a collection of individual chapters, which readers can dip in and out of ad hoc, rather than a book which must be read in its entirety for comprehension. The design of each chapter, consisting of an abstract, an introduction, and main content followed by concluding remarks, which in the main address the future directions, provides a certain familiarity and uniformity across the book without appearing repetitive. Apart from Chapter 11, "Population Genetic Modelling of Viruses," which did prewarn the reader of an indulgent focus on In uenza A virus, the breadth of viruses discussed is impressive. I couldn't help but enjoy finding a sprinkling of references to my favorite virus (rabies virus). At first, the inclusion of plant viruses seemed a little forced. The chapters which incorporated plant and animal viruses in the context of a common theme were more successful than chapters that focused entirely on only one or the other.

In general, this book is an easy read and is suitable for a wide audience with an interest in virus evolution. This book is a valuable resource for the up-to-date references alone. For the wildlife disease-focused animal virologists among us, I suggest reading the plant virus chapters when the mood strikes you. I found myself curiously interested in the difference between plant and animal viruses. For example, why would persistent plant viruses still go to the effort of encapsidating virions when they transmit vertically? I can visualize this book sitting next to *Fields Virology* (Knipe et al. 2013), quite confidently, in many students' and professors' offices alike.

LITERATURE CITED

Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B, editors. 2013. *Fields Virology*, 6th Ed. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania, 3,177 pp.

Denise A. Marston, Animal and Plant Health Agency, Wildlife Zoonoses and Vector-borne Diseases Research Group, Weybridge, Surrey, UK; AIX–Marseille Université –UMR "Emergence des Pathologies Virales," Marseille, France (Denise.Marston@apha.gsi.gov.uk).

1.7: Viral Heterogeneity

RNA virus populations exhibit a heterogeneous population of viruses within single individuals, often referred to as 'quasispecies' or 'mutant swarms' (Ciota et al., 2012) but perhaps more accurately termed 'viral heterogeneity' (Holmes and Moya, 2002). Viruses that contribute to a virus population may include those with single mutations, those with multiple mutations and those that have had complete deletions of stretches of the genome; DI (defective interfering) particles. Published data on viral heterogeneity is limited to a few target virus, mainly HIV, HCV and influenza virus. Viral heterogeneity has been investigated for other viruses: Dengue virus, arenaviruses and chikungunya viruses, although these areas are still in their infancy.

1.7.1: Role of heterogeneity in lyssaviruses

Viral heterogeneity, representing nucleotide sequence heterogeneity of the viral RNA, was originally proposed for rabies virus as a mechanism for viral evolution and adaptation to a new host (Benmansour et al., 1992). The heterogeneous population of a RABV street strain (European fox isolate) has been investigated previously (Kissi et al., 1999) using cloned PCR products covering 19% of the genome. This study identified sub-consensus sequences and concluded no host specific consensus level changes were observed, although the amount of heterogeneity observed varied depending on the host used to passage the virus. Furthermore, the fixed RABV strain CVS was passaged experimentally in BHK cells, resulting in selection of a dominant variant that differs genetically and phenotypically from the consensus sequence present in mice or neuroblastoma cell passaged CVS (Morimoto et al., 1998). Since the late 1990's until the development of Next Generation Sequencing (NGS) techniques, little progress in the area of viral heterogeneity was made.

1.8: Investigating viral heterogeneity using NGS

NGS techniques generate unprecedented cost-effective large-scale data, enabling both the generation of complete genomes (Marston et al., 2013; McGinnis et al., 2016; Montmayeur and Ng, 2017; Parker and Chen, 2017) and the analysis of viral sub-consensus populations (Eriksson et al., 2008; Poh et al., 2013; Raghwani et al., 2016; Wright et al., 2011). The inter-host dynamics of persistent viral infections where few virus particles are transmitted (Fischer et al., 2010) verses acute infections of equine influenza virus and norovirus, which have broad transmission bottlenecks have

been examined (Bull et al., 2012; Murcia et al., 2010). Detailed analysis of Foot and mouth disease virus (FMDV) traced minor variants within and between animal hosts using NGS data (Morelli et al., 2013). Without doubt there are issues with using NGS technologies to investigate virus heterogeneity including the issue of short reads and using a reference sequence to assemble these reads (Posada-Cespedes et al., 2016). Equally detailed in vivo studies with CHIV, reveals that although the consensus sequence remains unchanged, positively selected minority variants circulate as evolutionary intermediates before they rise to dominance (Stapleford et al., 2014). The field of viral heterogeneity analysis utilising NGS data is expanding, resulting in improving methods to extract the most relevant data (Seifert and Beerenwinkel, 2016). In addition, NGS has been used to identify DI particles in rabies virus (CVS) (Hoper et al., 2012). In the lyssavirus field, a number of studies have used NGS to obtain consensus whole genome sequences for phylogenetic analysis (Hanke et al., 2016; Hoper et al., 2015; Troupin et al., 2016). The application of NGS to study lyssavirus CSTs is becoming more commonplace, however, utilising the deep sequence data to investigate viral population dynamics is an area still to be explored (Borucki et al., 2013; Nadin-Davis et al., 2017).

Chapter 2:
Development of NGS technologies for sequencing whole viral genomes

Chapter 2: Development of NGS technologies for sequencing whole viral genomes

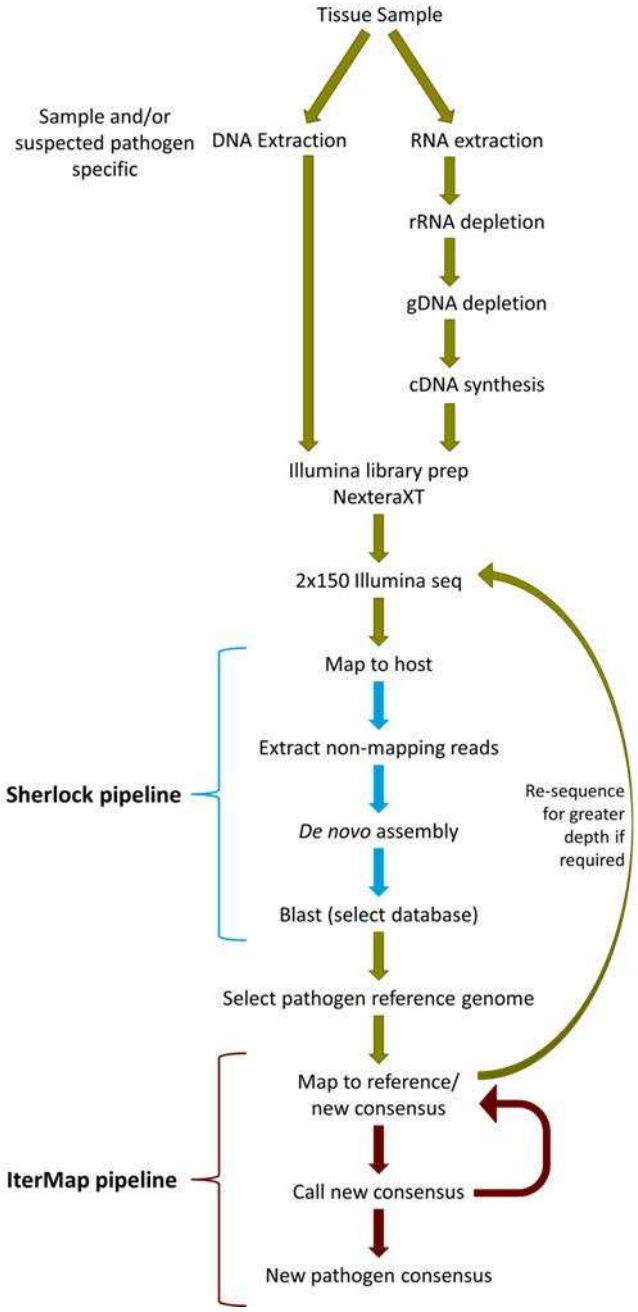
2.1: Introduction

For newly discovered novel viruses, a full length viral genome is required for downstream analysis and is required as part of the process to be recognised by ICTV. This was traditionally a challenging exercise using degenerate primers and 'walking' the length of the genome. Obtaining complete genomes from known viruses can be expensive and time consuming, requiring optimisation of primers for each virus and multiple PCRs to complete the genome. Furthermore, phylogenetic analyses using partial regions of the genome have limitations when the viruses being investigated were highly related, or when phenotypic changes were being mapped to an unknown region of the genome.

With the advent of Next Generation Sequencing (NGS) technologies, the ability to generate large amounts of sequence data has revolutionized the genomics field. In the virology field, genome walking and primer re-optimisation could potentially be made redundant. However, the promise of complete genomes being deduced from viruses was not as quick to realise as might have been envisaged. Most RNA viruses have relatively small genomes in comparison to other organisms and as such, would appear to be an obvious success story for the use of NGS technologies. However, due to the relatively low abundance of viral nucleic acid in relation to host nucleic acid, viruses have proved relatively difficult to sequence using NGS technologies. In addition to this main problem, other issues include: i) complementary ends on segmented viruses preventing short reads from mapping (see section 2.2.2 for details), ii) many viruses are present in the clinical material at negligible levels yet cannot be cultured to amplify the viral genome, and iii) some viruses are integrated into the host genome.

Over the past 5 years, NGS methodologies have been developing at an exponential rate, and the number of platforms available has also increased. The Roche 454 platform was universally considered the most reliable NGS platform, however was soon superseded by Illumina. Indeed, Roche has removed all support for 454 machines, including reagent production. At APHA, all work was initiated on the Roche 454, then moved across to Illumina by 2014. Figure 8 is a schematic of the entire NGS process developed at APHA from sample to consensus sequence using Illumina. In this Chapter, the workflow is discussed in detail and a number of publications where this workflow was used are presented. The chapter is broken down into 'Sample preparation' and 'Bioinformatics' for clarity, although all manuscripts describe the workflow in its entirety.

Figure 8: Schematic of the entire NGS process from sample to obtaining full length genome sequence.

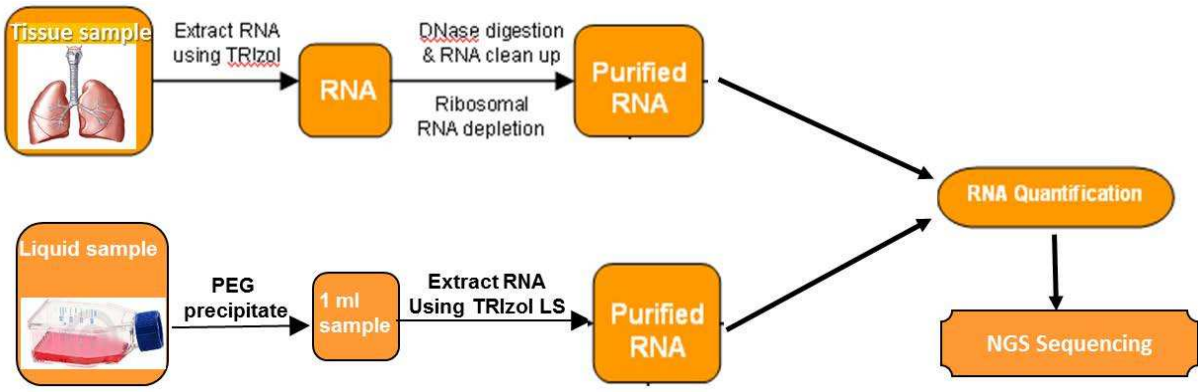


2.2: Development of sample preparation

2.2.1: Introduction

The principal issue encountered during this PhD was low viral loads in clinical samples. In order to overcome this I trialed different methods to increase the ratio of viral reads in comparison to host. The most successful method involved depleting the extracted RNA from host nucleic acid, namely gDNA using DNAse I and rRNA using a 5-phosphate-dependent Exonuclease. I showed that using both treatments on clinical tissue samples yielded the best results, but neither were critical (indeed rRNA depletion decreased the viral reads) when sequencing RNA extracted from tissue culture supernatant (TCSN). The final preparation methodology used throughout this thesis is described in Figure 9 and published in article 4. Samples are divided into two categories: 1) tissue samples (which includes any organ and cell culture pellets) and 2) liquid samples (which includes TCSN and saliva). TRIzol is used for both sample types, and Trizol LS is used for the liquid samples. The difference between the two TRIzol products is simply the concentration of phenol, with TRIzol LS containing an increased concentration to compensate for the volume of the liquid sample. For liquid samples, greater than 1 ml a PEG precipitation is performed to concentrate the virus down to a small enough volume for extraction. Downstream processing of the extracted RNA only occurs for the tissue sample where gDNA and rRNA is depleted (see above). However, the liquid samples do not require this downstream depletion due to the limited amount of host RNA present in the samples.

Figure 9: Schematic of the NGS sample processing pathway.

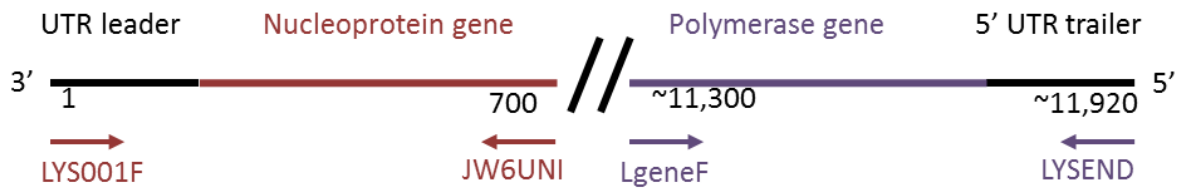


Samples prepared using these protocols can be sequenced on any NGS platform and I have sequenced RNA prepared in this way on Illumina, Roche 454 and MinIon platforms (see below).

2.2.2: Use of RT-PCR and Sanger Sequencing to confirm intergenic regions and genomic termini

Regardless of the platform used to obtain genomic sequences, obtaining reads that map to the genomic termini can be problematic. This issue is not that the reads are available, rather that the genomic termini of many viruses (including lyssaviruses) have complementarity. This complementarity at the genomic termini is important for successful replication, and therefore is a conserved feature. During the read assembly, the programs (de novo and mapping) cannot confidently position the reads at either the beginning or end of the genome sequence so they are left unassembled. I designed a quick RT-PCR method to obtain the genomic termini sequence (from known lyssaviruses using the conserved genomic termini sequence as a primer site for both ends, and designing complementary primers in the N and L gene (see Figure 10).

Figure 10: Schematic of primers used for amplification of genomic termini



By using these primers and sequencing the resulting amplicons the genomic termini were obtained. It must be noted that the sequence covered by the genomic termini primers cannot be deduced using this method. This is not an issue for sequences obtained from a known lyssavirus species, because the primer regions are 100% conserved. However, for a novel lyssavirus another method such as RACE or genomic ligation must be used to ensure the genomic termini sequence is correct (see 2.3.3.1 IKOV genomic sequence)

The following articles describe the optimized protocols developed.

2.2.3: Article 4 published in BMC genomics:

Next Generation Sequencing of viral RNA genomes

DOI: 10.1186/1471-2164-14-444

Summary: This manuscript combined the results of the work undertaken to optimize the sample preparation of viral tissue and cultured samples. The depleted samples were sequenced using Roche 454 platform which was the first NGS platform used at APHA. Since then the Illumina platform has been established, which has substantially improved our results, due to the increased number of reads obtained on this platform. The manuscript details a simple, robust methodology, without the use of ultra-centrifugation, filtration or viral enrichment protocols, to prepare RNA from diagnostic clinical tissue samples, cell monolayers and tissue culture supernatant, for subsequent sequencing on the Roche 454 platform. The most important message is that this protocol can be applied to all RNA viruses and both de novo assembly and mapping has been successfully undertaken on samples prepared using these methods.

Citations since publication: 69

BMC Genomics

METHODOLOGY ARTICLE Open Access

Next generation sequencing of viral RNA genomes

Denise A Marston[1,5], Lorraine M McElhinney[1,3*], Richard J Ellis[2], Daniel L Horton[1], Emma L Wise[1], Stacey L Leech[1], Dan David[4], Xavier de Lamballerie[5] and Anthony R Fooks[1,3]

Abstract

Background: With the advent of Next Generation Sequencing (NGS) technologies, the ability to generate large amounts of sequence data has revolutionized the genomics field. Most RNA viruses have relatively small genomes in comparison to other organisms and as such, would appear to be an obvious success story for the use of NGS technologies. However, due to the relatively low abundance of viral RNA in relation to host RNA, RNA viruses have proved relatively difficult to sequence using NGS technologies. Here we detail a simple, robust methodology, without the use of ultra-centrifugation, filtration or viral enrichment protocols, to prepare RNA from diagnostic clinical tissue samples, cell monolayers and tissue culture supernatant, for subsequent sequencing on the Roche 454 platform.

Results: As representative RNA viruses, full genome sequence was successfully obtained from known lyssaviruses belonging to recognized species and a novel lyssavirus species using these protocols and assembling the reads using *de novo* algorithms. Furthermore, genome sequences were generated from considerably less than 200 ng RNA, indicating that manufacturers' minimum template guidance is conservative. In addition to obtaining genome consensus sequence, a high proportion of SNPs (Single Nucleotide Polymorphisms) were identified in the majority of samples analyzed.

Conclusions: The approaches reported clearly facilitate successful full genome lyssavirus sequencing and can be universally applied to discovering and obtaining consensus genome sequences of RNA viruses from a variety of sources.

Keywords: Next generation sequencing, Pyrosequencing, Lyssavirus, Genome, RNA, Virus

Background

RNA viruses have small, simple genomes, which have a high level of diversity, due to the low-fidelity viral polymerase used for replication. Traditionally, due to their small genomes, 'genome-walking' was used to obtain a reference sequence, from which primer pairs can be designed for down-stream use on similar viruses [1-6]. However, this methodology can take a large amount of time, effort and expense, and be cumbersome, not least because of the variation within virus species, which results in the need to redesign primer pairs and frequent re-optimization of conditions. In addition to these optimization issues, every novel virus discovered will require 'genome walking'. Moreover, the introduction of PCR errors using Sanger-based sequencing is problematic, particularly when sequencing from cloned DNA. The use of high-fidelity enzymes and sequencing PCR products directly can overcome this issue in part, but errors occurring early in the amplification process will be sequenced incorrectly [7]. These problems can be avoided by utilizing Next Generation Sequencing (NGS), a high-throughput sequencing methodology which generates millions of sequences simultaneously from one sample [8]. Multiple platforms are available, the two commonly used are Roche 454 pyrosequencing (454 Life Science) and Illumina (Solexa) [9]. These platforms were initially developed and applied to mammalian and bacterial genomes, where Sanger methods were prohibitively expensive.

* Correspondence: Lorraine.McElhinney@ahvla.gsi.gov.uk
[1]Wildlife Zoonoses & Vector-Borne Diseases Research Group, Animal Health & Veterinary Laboratories Agency (AHVLA), New Haw, Addlestone, Surrey KT15 3NB, UK
[3]National Consortium for Zoonosis Research, Leahurst, Neston, Wirral, UK
Full list of author information is available at the end of the article

Because viral genome sequencing is achievable using the traditional PCR/Sanger methods, the virology field has had less necessity to embrace this new technology. The current NGS assays are optimized for large bacterial or mammalian genomes, where nucleic acid is in abundance and obtaining the hundreds of nanograms of nucleic acid required, is relatively simple. Due to the low percentage of viral RNA in relation to host RNA in any given sample, obtaining the quantity of viral RNA required by the protocols can be problematic.

The advantages of NGS technologies in virology are numerous, and its use is becoming more commonplace [10], particularly to detect and characterize pathogens without prior knowledge of their existence, or in association with the sequelae/disease outcome, in primary or cultured material, without the requirement of specific primers [11,12]. Often however, the virus is at such low abundance in these samples that only one or two reads can be obtained, or at best, a number of unlinked contiguous sequences (contigs), for which traditional Sanger protocols are employed to complete the genome [13,14]. Conversely, known positive samples have been 'deep sequenced' using NGS technologies, for a plethora of uses including determining viral heterogeneity, or the effect of the immune system or pharmaceutical drugs on viruses [15-19]. These deep sequencing methodologies employ the use of viral PCR amplicons to obtain a depth of coverage sufficient to detect variants which occur down to 0.1% frequency. Between these two approaches, a methodology to obtain full genome sequences directly from known positive clinical specimens or cultured material, without the use of amplicons, viral enrichment or virus concentration is still lacking. To this end, we describe a broadly applicable approach to obtaining full genome sequences from clinical or cultured samples, which we have successfully applied to lyssavirus infected samples.

Lyssaviruses (family *Rhabdoviridae*) have negative-sense, single-stranded RNA genomes approximately 12 kb long. Traditionally, partial gene sequence (usually the N-gene) was utilized for viral speciation and phylogenetic analysis. However, limitations on the information obtained from these partial sequences and a requirement to use full genome sequence as part of the ICTV (International Committee for Taxonomy of Viruses) criteria to propose new lyssavirus species, in combination with recent advances in high-throughput sequencing, have resulted in an increase of full genome sequences available on Genbank. From the submission of the first full length lyssavirus genome sequence of RABV (prototype) in 1988 [20] to the completion of representative genomes from each of the established twelve lyssavirus species (RABV, LBV, MOKV, DUVV, EBLV-1, EBLV-2 and ABLV, ARAV, KHUV, IRKV, WCBV and SHIBV) in 2010, the commonly used methodology for obtaining the genomes was relatively unchanged, utilizing reverse transcription then either cloning the cDNA directly, or cloning amplified PCR products, and Sanger sequencing [2,5]. Of the most recently described lyssavirus species, BBLV [21] and IKOV [22], only the IKOV characterization employed NGS technologies to obtain the full genome sequence, pioneering the way forward to obtaining genome sequences without the requirement to design multiple primers necessary for genome walking.

This study describes the optimization of protocols, required for a variety of starting material, describing a robust, simple, reliable methodology to obtain full genome sequences from original clinical material and cultured samples, both from the cell monolayer and the supernatant. Furthermore, we investigate the SNPs (Single Nucleotide Polymorphisms) observed in the sequences obtained.

Results

Optimization of the extraction protocol for tissue samples

Two different extraction methodologies were compared to investigate the suitability of the RNA obtained. Three brain tissue samples (RV2772, RV2627 and RV2516) were extracted, using the same amount of starting material, either by TRIzol® extraction followed by isopropanol precipitation, or RNeasy® plus mini kit (see Table 1 for isolate details and Table 2 for details of extraction). Genomic DNA depletion was carried out on-column, either as a separate step for the TRIzol® extracted RNA, or as part of the RNA extraction procedure, using the RNeasy® plus mini kit. The RNA concentration from the RNeasy® kit extracted samples were significantly lower than the TRIzol® extracted equivalents (Table 2). The required 200 ng RNA (10 ng/μl), was not achieved for the RNeasy® kit samples, or indeed for the RV2772 TRIzol® extracted sample. For these samples the maximum RNA (20 μl) was used with varying success.

The number of total reads and viral reads obtained for the RNeasy® kit samples were lower in comparison to the TRIzol® extracted RNA, most likely due to the difference in total RNA available for these samples. On the whole, viral RNA was not enriched by the RNeasy® kit, as the percentage of viral reads was less for RV2627 and RV2516, although for RV2772 there was a slight increase in viral specific reads. Without exception, none of the RNeasy® kit extracted sample reads were sufficient to obtain a single consensus sequence, due to the low number of viral reads obtained. Furthermore, *de novo* assembly on two of the three samples (RV2516 and RV2627) failed to align viral reads into contigs for further analysis resulting in only host contigs being identified (Table 2).

Analysis of depletion methodology

Regardless of the originating sample (brain tissue, cell monolayer, tissue culture supernatant) the concentration of the TRIzol® extracted RNA after gDNA depletion was

Table 1 Details of virus isolates analysed in this study

Virus	Original reference	Host species	Origin	Year of isolation	GenBank no.
Lyssavirus RABV					
RV2324	ISR-50	Dog	Israel	1950	KF154998
RV50	AZBAT 7453	Bat (*Eptesicus fuscus*)	US, Az	1975	JQ685956*
RV61	-	Human (canine strain)	UK (ex-India)	1987	KF154996
RV437	269	Raccoon Dog	Estonia	-	KF154997
RV2417	R16/08	Dog	UK (ex Sri Lanka)	2008	KF154999
RV2627	CASA 09/08	Cow (Canine strain)	Morocco	2009	KF155001
RV2772	Rab91/D10-867	Dog	Tanzania	2010	KF155002
RV2516	23	Cow (Canine strain)	Iraq	2010	KF155000
Lyssavirus EBLV-1					
RV20	RA552	Bat (*Eptesicus serotinus*)	Denmark	1986	KF155003
Lyssavirus EBLV-2					
RV1787	603-04[†]	Bat (*Myotis Daubentonii*)	UK (Staines)	2004	KF155004
Lyssavirus IKOV					
RV2508	snp0971[#]	African Civet	Tanzania	2011	JX193798

* Published after sequence was obtained prior to this publication.
[†] Sequence confirmed by Sanger sequencing from the bat brain material.
[#]Both brain and TCSN were sequenced. For analysis see Table 2.
- Data not available.

significantly less than the original extract RNA sample (Table 2). The largest reduction was for RV2772 where RNA at 1,833 ng/μl was depleted to 3.27 ng/μl (600-fold reduction) after removal of genomic DNA. Interestingly, this sample was part of a cohort of samples that were highly degraded upon receipt, therefore the majority of RNA had already been degraded. Otherwise a reduction of concentration between 3-fold and 100-fold was observed (Table 2). The subsequent depletion of rRNA resulted in a more conservative fold change of concentration between 30-fold (RV2417) and 2-fold (RV2772 and RV2508).

We investigated the requirement to deplete gDNA and rRNA in cultured viral samples after RNA extraction, since the amount of cellular material would be minimal in these supernatant preparations. Comparison at the RiboGreen stage determined that RV20 and RV1787 depleted samples are 45-fold and 10-fold less than the RV20 and RV1787 non-depleted RNA samples respectively. Indeed, for RV20 the total amount of RNA was too low to obtain 200 ng RNA for fragmentation. The virus titer of RV1787 and RV20 has been calculated previously [23,24] with RV1787 (EBLV-2) approximately 1 log lower than RV20, therefore the difference in the percentage of viral reads is likely to be a reflection of this. Despite the marked difference between the percentage of viral reads of RV20 and RV1787, the difference within samples regarding whether the RNA was depleted or not, is not so obvious. Indeed, for both samples, the RNA sequenced directly without depletion provided more viral-specific reads.

The success of the methodology for both tissue material and cell cultured material, particularly the ability to detect viral sequence after *de novo* assembly, is illustrated in Table 2. Apart from the column-extracted samples, all sequences obtained were sufficient to obtain viral specific reads either as a single contig, or a number of contigs, which subsequently can be identified by a BLAST search.

Depth of coverage analysis

The depth of viral read coverage of samples from which a single genome contig was obtained was investigated (Figure 1). Brain tissue sample read depths varied from RV50, total viral reads 293, maximum read depth 16, and average read depth of 8.6; to RV437, total viral reads 1,489, maximum read depth 103, and average read depth of 44 (Figure 1a, Tables 2 and 3). RV61 did have a higher number of viral reads, but from two runs, therefore not directly comparable. Read depths from cell cultured viral samples were higher, with average read depths of 31 (RV20 depleted), 125 (RV20) and 65 (RV2324 cell pellet) and maximum read depths of 69 (RV20 depleted), 228 (RV20) and 272 (RV2324 cell pellet) (Figure 1b, Table 3).

Viral heterogeneity analysis

Due to the processes involved, each read is obtained from a single cDNA strand, and therefore where multiple reads cover a region viral heterogeneity can be detected within the reads indicating the presence of a heterologous viral population. Although this methodology is not optimized for investigation of low level viral populations, which would require viral read depths of over 10,000, it is still possible to observe dominant or high level single nucleotide polymorphisms (SNPs). Even with this dataset, we were able to observe minimum SNPs (equating to one read with a SNP) at 1% of the population in cultured material, and 3% in brain tissue

Table 2 Details of brain and BHK cultured samples, relating to the extraction method used and RNA concentration during preparation stages resulting in number of reads obtained and outcome of obtaining full viral genome sequence

Virus	Sample origin	Extraction method	Concentration after RNA extraction (ng/μl)	Concentration after gDNA depletion (ng/μl)	Concentration RNA using Ribogreen[1] (ng/μl)	Total N° of reads	N° of vial reads	% viral specific reads	Genome as single contig	De Novo detection
Brain tissue samples[†]										
RV437	mouse	TRIzol®	5,404.4	410.77	121.77	282,370	1,489	0.53	Yes[a]	Yes
RV61	mouse	TRIzol®	3,362.8	ND	58.9	191,508[#]	1,810	0.95	Yes[c]	Yes
RV2417	Dog	TRIzol®	5,431.3	1799.21	57.78	96,014	1,488	1.55	Yes[b]	Yes
RV2508	African civet^	TRIzol®	6,537.4	66.55	28.18	139,841	626	0.45	Yes[a]	Yes
RV2627	Cow	TRIzol®	2,385.8	111.81	26.60	11,470	682	5.95	Yes	Yes
		RNeasy mini	N/A	4.99	0.95	239	2	0.84	No	No
RV2516	Cow	TRIzol®	1,891.2	45.84	11.09	154,068[#]	550	0.36	Yes[a]	Yes
		RNeasy mini	N/A	11.12	2.14	7,991	17	0.21	No	No
RV50	mouse	TRIzol®	2,186.2	18.76	7.64	76,568	293	0.38	Yes	Yes
RV2772	Dog	TRIzol®	1,833.9	3.27	1.50	305,110	391	0.13	Yes	Yes
		RNeasy mini	N/A	7.19	1.08	48,849	180	0.37	No	Yes
BHK cultured virus samples										
RV2324	cell pellet	TRIzol®	10,232.9	870.72	75.63	101,599	2,027	1.99	Yes	Yes
RV20 depleted	TCSN 100 ml	TRIzol® LS	3,732.6	2.8	0.52	2,537	1,054	41.55	Yes	Yes
RV20			Diluted to 1,000	ND	23.47	10,332	4,369	42.29	Yes	Yes
RV1787 depleted	TCSN 100 ml	TRIzol® LS	4,603	37.87	11.56	8,162	53	0.65	No	Yes
RV1787			Diluted to 1,000	ND	119.47	6,861	87	1.27	No	Yes
RV2508 cells	TCSN 250 μl	TRIzol® LS	129.44	ND	11.45	14,587[#]	223	1.529	No*	Yes
RV2508 no cells			127.05	ND	0.46	326	42	12.88	No*	Yes
RV2508 high titre			112.5	ND	1.31	1,132	12	1.06	No*	Yes

TCSN tissue culture supernatant; N/A – Not Applicable as DNAse treatment is incorporated in the RNA extraction protocol, therefore no concentration before gDNA depletion is available; ND - Not Done as gDNA was not depleted from these samples; [1]This column contains the RNA concentration after rRNA depletion for all samples, except RV20, RV1787 and RV2508 TCSN where no rRNA depletion was undertaken. [†]brain either from host, or from one passage in a mouse, *combining these reads resulted in a single genome contig; # combined results from two 454 runs of same library; ^ Stored in RNAlater; [a] 3′UTR not represented; [b] 5′UTR not represented; [c] 3′ and 5′ UTR not represented.

Brain tissue samples ordered by concentration after the depletion process. Where rRNA depletion was not undertaken (RV20, RV1787 and RV2508 TCSN), the concentration values in the Ribogreen column are directly comparable to the RNA extraction concentration.

samples (Table 3). In general, cell cultured samples had more SNPs than tissue samples, where some tissue samples (RV2417, RV2627, RV2516) had no detectable SNPs (Table 3). Of the 28 nucleotide substitutions observed, 22 (79%) were conservative (pyrimidine to pyrimidine or purine to purine). Substitutions were identified throughout the genome, although only one was identified outside of a coding region (RV20 at position 3143 in the M-G untranslated region). At the amino acid level, 8 (29%) were synonymous substitutions resulting in no amino acid change (Table 3). Two stop codon substitutions were identified in this dataset, both in the RV20 tissue culture supernatant samples, and both at a low level (2%) equating to a single read (Table 3). Furthermore, one of these (N-gene position 276) had two reads which differed from the consensus, one encoding a stop codon, resulting in a truncated Nucleoprotein transcript, the other a synonymous change (Tyr/Tyr^{92}) (Figure 2a).

RV61, an Arctic-like RABV, had one SNP detected in the 454 read data at genome position 3503 (Figure 2b). This SNP is located in the glycoprotein gene resulting in an amino acid substitution at position 70 Ile/Thr at

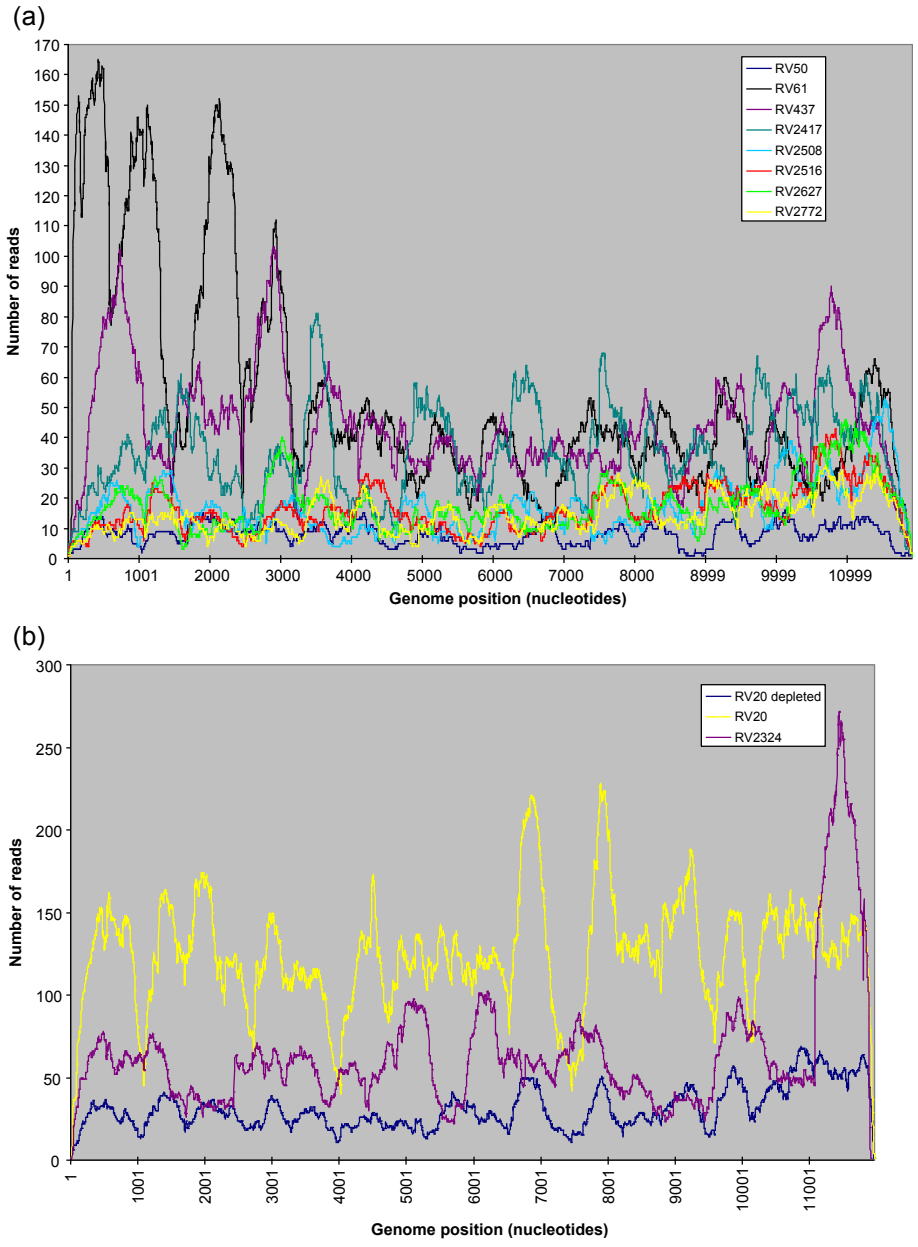


Figure 1 Depth of sequence obtained from (a) infected brain samples and (b) from cultured samples.

a frequency of 18%. Analysis of currently available lyssavirus glycoprotein sequences, indicates this residue is Isoleucine in the majority of lyssavirus species. RV61 glycoprotein sequence is available in Genbank, deposited separately by two independent laboratories. No original material is available for analysis and the published sequences are from passaged viruses. Interestingly, alignments of the published sequences and the sequence obtained in this study are 100% identical, apart from this residue, where JQ685981 had Ile^{70} and GU936881 had Thr^{70} (Figure 2b).

RV50, a US isolate (isolated from *Eptesicus fuscus*), was genome sequenced using Sanger sequencing as part of a larger dataset [6], providing a unique opportunity to analyze the same virus isolate, propagated and sequenced independently. Alignments of the published RV50 genome (JQ685956) and the consensus sequence obtained in this study showed 100% identity. However, the 454 data identified a SNP at genome position 11,349 (L-gene), resulting in a non-synonymous coding change at residue 1982 resulting in conservative substitution Asn/Asp^{1982} (Figure 2c).

Table 3 Details of single nucleotide polymorphisms (SNPs) detected within the 454 data

Virus	Average depth	Maximum depth	Depth at SNP	% variant	Position (genome)	Gene (Position)	Nuc cons	Nuc variant	AA cons	AA variant	AA residue
Brain tissue samples											
RV437	44	103	60	8	2768	M (275)	T	C	Leu	Ser	92
			63	3	2781	M (288)	T	C	Val	Val	96
			55	51	3049	M (556)	T	C	Leu	Phe	186
			28	18	4403	G (1090)	T	C	Leu	Leu	364
			28	7	7855	L (2451)	C	T	Asn	Asn	817
RV2417	35.5	81	NA								
RV61	53.2	165	34	18	3503	G (209)	C	T	Ile	Thr	70
RV2508	17	53	17	23	9523	L (4128)	A	G	Ser	Ser	1376
RV2627	17.7	46	NA								
RV2516	16.5	43	NA								
RV50	8.6	16	11	27	11349	L (5944)	C	G	Asn	Asp	1982
RV2772	14.5	31	13	38	1293	N (1223)	A	G	Lys	Arg	408
			22	5	8179	L (2771)	T	A	Phe	Tyr	924
Cultured virus samples											
RV2324	65	272	47	11	3490	G (175)	G	A	Glu	Lys	59
			47	21	3491	G (176)	A	G	Glu	Gly	59
			42	31	3711	G (396)	C	A	His	Gln	132
			44	25	4103	G (788)	A	G	Gln	Arg	263
			36	11	8376	L (2967)	C	T	Asp	Asp	989
RV1787 combined	4.3	11	4	25	2679	M (170)	G	A	Gly	Glu	57
RV20 combined	156	277	107	2	346	N (276)	C	A/T	Tyr	*/Tyr	92
			87	2	1621	P (105)	T	C	Ser	Ser	35
			81	2	1669	P (153)	G	A	Glu	Glu	51
			77	3	1710	P (194)	A	T	Gln	Leu	65
			93	39	3143	M-G (NA)	C	T	NA	NA	NA
			103	2	3671	G (356)	G	A	Trp	*	119
			96	2	4403	G (1087)	T	C	Cys	Arg	363
			94	37	4584	G (1268)	T	C	Met	Thr	423
			95	2	6217	L (766)	T	C	Ser	Pro	256
			134	1	7910	L (2459)	T	C	Ile	Thr	820
			107	2	8244	L (2793)	C	A	Ser	Arg	931
			106	2	8473	L (3022)	A	G	Thr	Ala	1008

NA = not applicable as no SNPs were observed. * stop codon.

The frequency of this variant in the read data was 27% (Table 3).

RV2508 was sequenced both from the host brain tissue directly and from cell culture supernatant after 6 passages in BHK cells. Only one SNP was observed at position 9523 (L-gene) a synonymous coding change A/G^{9523}. Of the 17 original brain tissue reads covering this region, 23% were G^{9523}, whereas the tissue culture supernatant sample reads (n = 2) only had G^{9523}. To investigate this variation further, specific primers were designed to amplify this region from the original brain sample and the passaged TCSN. The original brain material PCR amplicon had a read depth of 3783 reads, 18.4% G^{9523} and 81.6% A^{9523}. The TCSN PCR amplicon had a read depth of 3280 reads, 97.5% G^{9523} and 2.5% A^{9523}, confirming the consensus sequence data, but indicating the A variant is still present after 6 passages, just at a much lower level. This deep sequence analysis is a useful tool to investigate certain SNPs of interest indicated from the consensus data.

RV437, a raccoon-dog RABV from Estonia, had good coverage across the genome (average read depth 44, maximum read depth 103) and also had the most SNPs

(a)

```
Genome Ref    Var    Total  Var
Pos    Nuc    Nuc    Depth  Freq
346    C      TA     107    2%

RV20 reference              TGCCCGGAAGATTGGACTAGTTACGGAATCAACATTGCTAAGAAAGG
                                                    *
Reads with Difference:

HHZJQVG08I8TPF              TGCCCGGAAGATTGGACTAGTTAAGGAATCAACATTGCTAAGAAAGG
HHZJQVG08JLZI2              TGCCCG-AAGATTGGACTAGTTATGGAATCAACATTGCTAAGAAAGG
                                                    *
```

(b)

```
Genome Ref    Var    Total  Var
Pos    Nuc    Nuc    Depth  Freq
3503   C      T      34     18%

R61 reference               TACATGGAACTTAAGGTGGGATACACCTCGGCCATAAAA
                                                  *
Reads with Difference:

HOTR63I03F6R8M              TACATGGAACTTAAGGTGGGATACATCTCGGCCATAAAA
HOTR63I03FNF8Y              TACATGGAACTTAAGGTGGGATACATCTCGGCCATAAAA
HOTR63I03G1IHY              TACATGGAACTTAAGGTGGGATACATCTCGGCCATAAAA
HNJBY7V01AJ3QA              TACATGGAACTTAAGGTGGGATACATCTCGGCCATAAA-
HNJBY7V01AXILR              TACATGGAACTTAAGGTGGGATACATCTCGGCCATAAA-
HOTR63I03GEUEP              TACATGGAACTTAAGGTGGGATACATCTCGGCCATAAAA
                                                  *
Published RV61 G-gene sequences
GU936881                    TACATGGAACTTAAGGTGGGATACACCTCGGCCATAAAA
JQ685981                    TACATGGAACTTAAGGTGGGATACATCTCGGCCATAAAA
                                                  *
```

(c)

```
Genome Ref    Var    Total  Var
 Pos   Nuc    Nuc    Depth  Freq

11349  A      G      11     27%

JQ685956_RV50 ref     TCGTGATAGTCTTCTCCAATAGGGTGTTCAATGTTTCGAAACCGCTGAC
                                                  *
Reads with Difference:

HG11MEB03HA8U5         TCGTGATAGTCTTCTCCAATAGGGTGTTCGATGTTTC
HG11MEB03F15NM         TCGTGATAGTCTTCTCCAATAGGGTGTTCGATGTTTCGAAACCGCTGAC
HG11MEB02EBHFL         TCGTGATAGTCTTCTCCAATAGGGTGTTCGATGTTTCGAAACCGCTGAC
                                                  *
```

Figure 2 Alignment of 454 reads for (a) RV20 consensus sequence and the reads with variations, (b) RV61 consensus sequence and published RV61 glycoprotein gene sequences and (c) RV50 published genome sequence; * indicating single nucleotide polymorphisms which are detailed at the top of each figure.

observed within a tissue sample (Table 3). Across the genome, 5 SNPs were detected, 3 within the M-gene (2 of which resulted in amino acid changes – Leu/Ser^{92} and Leu/Phe^{186}) and one synonymous change at genome position 2781. The remaining two were both silent, one in the glycoprotein gene (genome position 4403) and the other in the L-gene (genome position 7855) (Table 3).

Although a full genome contig was not obtained for RV1787, 6 contigs covering approximately 90% of the genome were generated. Full genome sequence had previously been generated using overlapping PCR products and Sanger sequencing [1]. Comparison between the Sanger and 454 generated sequences revealed 100% match apart from a single nucleotide polymorphism (SNP) at genome position 2679 (residue 57 in the Matrix Protein) which corresponded to substitution Gly/Glu (Table 3). Published sequences of all lyssavirus species (including the recently identified IKOV; [22] have Gly^{57}. The read coverage at this position was 6 reads, 5 of which contained the Glu^{57} SNP and 1 contained the original Gly^{57}, present in the Sanger sequencing and all other lyssaviruses.

Obtaining genomic termini

The genomic termini sequences from 454 data were obtained with varying success. The 5′ UTR was usually represented by reads more often than the 3′ UTR, the reasons

for which are unclear (Table 2 and Figure 1). Often *de novo* assembly failed to incorporate reads which contained the genomic termini. These reads were only incorporated after splicing the missing sequence, from a published similar genomic sequence, to the consensus sequence deduced from the *de novo* assembly and subsequently mapping reads against the spliced reference using GS mapper. Subsequent to this dataset, in an attempt to increase the population of genomic end viral reads, the 454 sample preparation methodology was modified by the addition of 1pmol of N165-146 and LRACEF2 at the hexamer cDNA synthesis stage, to enhance the cDNA population containing the genomic termini. This modification was trialed on a genome sample highly related to RV2772, where the genomic ends were obtained (data not shown). This simple modification will be used for future genome sequences.

For samples where the genomic termini were still absent, RT-PCR was performed on the depleted RNA, using primers designed against the highly conserved genomic termini to obtain PCR products which were sequenced directly using Sanger sequencing. Although this approach resulted in a genome sequence still lacking the sequencing relating to the primer sites, for known lyssavirus species isolates this is acceptable. However, for novel lyssavirus species prototypes the ends must be deduced either by RACE [1] or circularization of the RNA, followed by RT-PCR [2]. To this end IKOV genomic ends were obtained using circularization of the genomic RNA [22].

Confirming correct sequence and differentiating between indels/homopolymeric repeats

Whenever possible, consensus sequences where aligned together with genome sequences from the same lyssavirus species. Any potential insertions or deletions (indels) were investigated by analysis of the flowgrams in the GS program suite. A limitation of pyrosequencing processing of raw data is runs of homopolymers such as the GAAAAAAA poly A termination signal found at the end of each lyssavirus gene sequence (Figure 3a and b). Figure 3 illustrates the discrepancies between individual reads from the same sample in a homopolymeric run, where the signal strength for each base is plotted for each read. Detecting the correct signal in a homopolymer of four or more bases can be problematic, resulting in incorrect base calling. In general, the more reads covering the genome the more accurate the resulting consensus sequence (represented by contig00001 in Figure 3a). Where discrepancies cannot be resolved, or a new virus species is being sequenced, confirmation of the correct sequence using PCR products spanning the region in question sequenced using Sanger technologies may be required. In this dataset, only the proposed new lyssavirus species IKOV (RV2508) required additional PCR confirmation. Each intergenic region length and sequence was confirmed by RT-PCR, with 100% concurrence between the 454 consensus sequence and the Sanger sequence derived from PCR products (data not shown).

Discussion

Our goal was to establish a methodology which enabled sequencing from brain tissue, and cell cultured samples without the use of specific viral enhancement, ultracentrifugation or filtration to identify novel viruses, or obtain genome sequence for known virus species. It was important to develop a methodology which would not require prior knowledge of the virus being extracted and to use a methodology which is widely used to extract viral RNA. Instead of enriching for virus specific RNA, we depleted host genomic DNA and rRNA to increase the percentage of viral specific RNA in the sample. There are a large number of RNA extraction methodologies and kits available which can be broadly divided into two categories; 1) Phenol/choloform based extraction and 2) column or bead based extraction. We decided to compare TRIzol® as a phenol-based system and RNeasy® plus mini kit (Qiagen) as a column-based system. In our hands, the use of spin columns to extract the RNA from tissue, removed the majority of RNA in process, resulting in less than the 200 ng minimum quantity of total RNA recommended for fragmentation and more importantly, too few viral reads were obtained. The RNeasy® plus mini kit (Qiagen) was successfully used as part of the depletion protocol after initial extraction with TRIzol® with good results. In this study, less than 200 ng of RNA was used for a number of samples, and depending on the sample type less than 200 ng of RNA can provide full genome sequence. TRIzol® extracted tissue samples RV50 (153 ng RNA) and RV2772 (30 ng RNA), not only obtained sufficient reads to cover the genome, but also obtained enough depth of coverage to identify SNPs (Tables 2 and 3). Furthermore, for RV20 depleted tissue culture supernatant only 10 ng of RNA was fragmented, yet 1,054 viral specific reads were mapped equating to 41% of the total reads obtained. These observations are important to illustrate that the methods for library preparation can be further refined for virus sequencing.

The requirement to deplete tissue culture supernatant derived RNA was investigated using two 100 ml supernatant samples. Half the RNA was subjected to gDNA and rRNA depletion, whereas the other was not. The additional preparation time required to deplete the sample was not rewarded with a significant improvement of either proportion of viral-specific reads or read depth, therefore 454 sequencing directly from a TRIzol® extracted RNA is the simplest, yet most effective way to obtain viral genome sequence from PEG precipitated tissue culture supernatant samples. However, it is not always possible, or practical to culture viruses *in vitro*. Furthermore, any passaging of viruses in a cell culture system will apply selective pressures on the virus and,

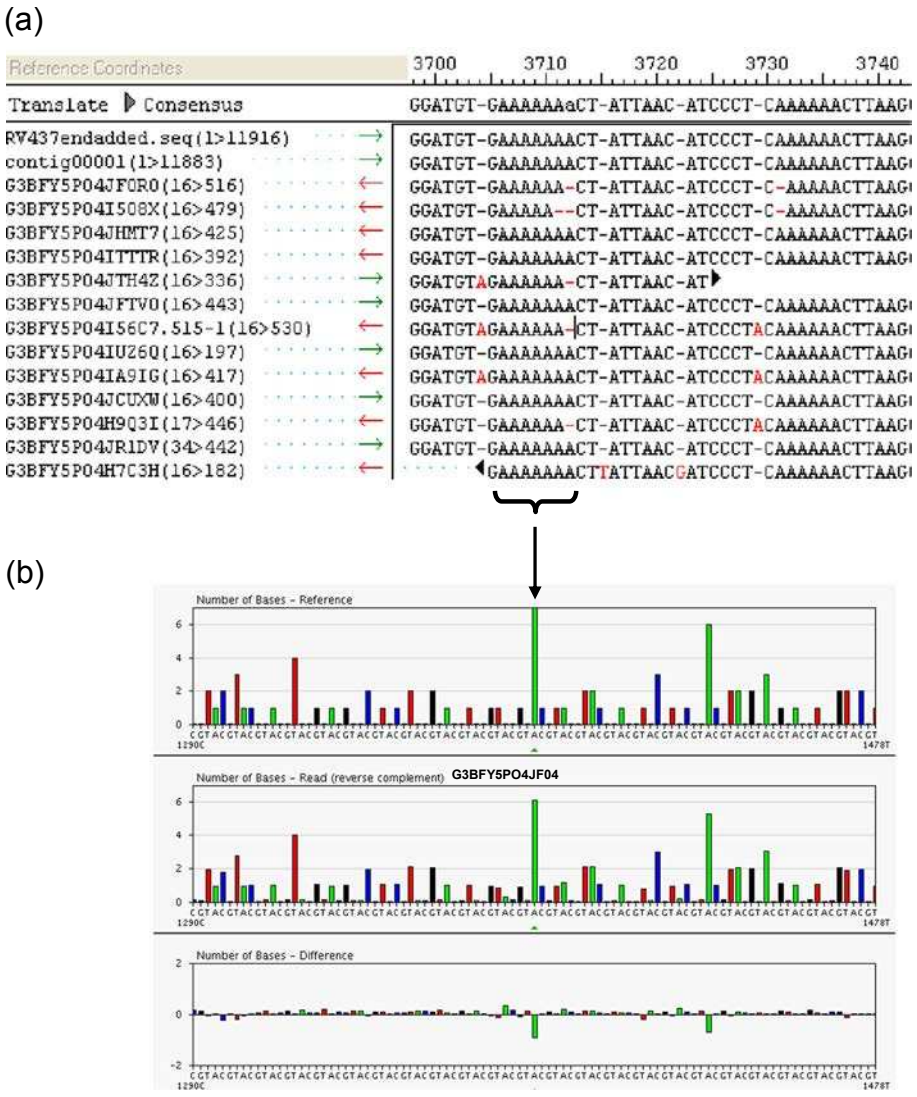


Figure 3 Typical variation of the reads over a homopolymeric repeat at the termination signal of the M-gene displayed as (a) an alignment of individual 454 reads with the consensus sequence (contig00001) and a reference sequence (RV437) and (b) a flowgram with reference sequence (RV437, top panel), read G3FY5P04JF0R0 (middle panel) and the difference between the reference sequence and read (bottom panel).

depending on the number of passages necessary to obtain virus at a high enough titer to harvest, will alter the population of virus being sequenced. Indeed, the number of SNPs observed in the cell cultured RNA preparations were greater than the tissue extracted samples (Table 3). Furthermore, the number of non-conservative substitutions observed (both nucleotide and amino acid) was significantly more in the cultured samples, including the presence of stop codons (Table 3). The RV2508 synonymous SNPs identified between the brain material and the TCSN passaged 6 times in BHK was investigated by designing specific primers to amplify across the SNP from both the original and cultured virus samples. 'Deep sequencing' of this amplicon determined that the dominant viral population changed from G^{9523} to A^{9523} after 6 passages, and although no A^{9523} SNP reads were observed in the TCSN consensus sequencing they were still present in the population, just at a low level (2.5%). Although for this example there is no functional relevance for the variation, this approach is useful to determine high proportion SNP changes within passaged virus samples using consensus sequencing which can then be targeted directly by deep sequencing the relevant genomic regions, rather than deep sequencing the entire genome.

The genomic termini are often under-represented in the viral reads obtained. It is likely this is due to a variety

of reasons including the process of trimming reads before assembly, the assembly parameters in *de novo* and mapping programmes as well as biological influences such as mRNA over-representing the viral RNA population relative to genomic RNA and the effect of defective interfering (DI) particles which are often truncated. The use of primers situated near the genomic termini during the cDNA synthesis stages have been shown to improve the number of termini sequences, but have not completely resolved the problem.

Sequences generated on the 454 platform, due to the way the incorporation of bases are detected by the intensity of light emission, inherently have issues with long homopolymeric repeats. On the whole, a consensus sequence with 10 or more reads, when mapped using an available genome from the same species, can be confidently confirmed by checking the flowgram data. However, novel viruses, or viruses with a limited number of reads may require additional confirmation by Sanger sequencing PCR products, which span the regions in question.

Conclusion

The application of this methodology to lyssaviruses from brain material and cell cultured samples has been shown to be highly successful. Moreover, we have successfully sequenced the majority of a flavivirus and hantavirus, demonstrating the applicability of the method to other families of viruses (data not shown). There is no reason why any cultured virus cannot be PEG precipitated, extracted and sequenced using the same methodology. We have shown for both a high titer virus (RV20) and a low titer virus (RV1787) 100 ml of supernatant is more than enough to obtain genome sequence, and to begin to investigate the presence of viral heterogeneity. Further investigations using 5-10 ml of tissue culture supernatant for a number of other cultured lyssaviruses have proved equally as successful (data not shown). Furthermore, extraction from the cell monolayer (often discarded when clarifying harvested supernatant) is an excellent source of viral RNA. The only disadvantage is the potential presence of non-functional virus variants presumably present in the cells as partially synthesized viral derivatives. In this study however, it was virus in the supernatant which appeared to contain defective virion particles, albeit at a low level.

The platform used in this study (Roche 454) has been shown to be reliable and consistent, however, the RNA can be prepared for any NGS platform. Indeed we have obtained full genome sequences from a small number of RNA samples prepared as described here from brain material on the Illumina platform successfully (data not shown). The application of NGS sequencing of viruses has begun a new era of virus discovery and characterization of novel viruses and will revolutionize phylogeographic studies in all fields of virology in the coming years.

Methods

Virus samples

Original clinical specimens, or once passaged mouse brain samples (Table 1), confirmed positive previously by FAT (Fluorescent Antibody Test) and by RT-PCR were used. Cultured viruses were passaged from original brain samples in Baby Hamster Kidney cells. All in vivo work was undertaken in BSL3/SAPO4 containment in AHVLA, following independent ethical review and complied with the Animal Scientific Procedures Act 1986.

PEG precipitation

Concentration of virus in tissue culture supernatant (TCSN) over 1 ml (specifically RV20 and RV1787) was achieved using 1 volume of PEG-it (System Biosciences) to 4 volumes of TCSN and following manufacturer's instructions. Pellets were combined and resuspended in tissue culture media to a final volume of 1 ml.

RNA extraction

TCSN samples were clarified to remove cell debris by centrifugation at 1,200 rpm for 5 mins and were extracted either directly (250 μl) or after PEG precipitation (see above) using TRIzol® LS following manufacturer's instructions and resuspended in 10 μl molecular grade water. All other samples were extracted using TRIzol® following manufacturer's instructions and eluting in 10 μl molecular grade water. For RNA extraction method comparisons, duplicate samples were extracted using RNeasy® plus mini kit (Qiagen), including the DNase treatment (see 2.4).

gDNA depletion

Genomic host DNA was depleted from the extracted RNA samples using the on-column DNase treatment in RNeasy® plus mini kit (Qiagen) following manufacturer's instructions, eluting in 30 μl molecular grade water.

rRNA depletion

Ribosomal RNA was depleted from the gDNA-depleted RNA samples using Terminator™ 5′-Phosphate-Dependent Exonuclease (Epicentre Biotechnologies). The reaction was performed according to the manufacturer's instructions, briefly, 30 μl of RNA was mixed with 3 μl of Buffer A, 0.5 μl RNAsin® Ribonuclease inhibitor (20–40 U/μl, Promega) and 1 μl Terminator (1 U/μl). The mixture was incubated at 30°C for 60 minutes. Thereafter the samples were subjected to subsequent round of purification using RNeasy® plus mini kit (Qiagen), without DNase digestion, following manufacturer's instructions, eluting in 30 μl molecular grade water.

RNA quantification and purity check

After extraction and then gDNA depletion, the RNA was quantified using NanoDrop® spectrophotometer (results in Table 2). After rRNA depletion the final RNA sample was quantified using RiboGreen® on a spectrophotometer before commencing the 454 RNA preparation protocols. For TCSN samples which did not undergo depletion, the RNA was diluted to 1000 ng/ul according to the nanodrop concentration, then quantified using RiboGreen® on a spectrophotometer.

RNA fragmentation, cDNA library amplification and sequencing

200 ng of RNA was fragmented using divalent cations ($ZnCl_2$) at 70°C for 30 s. Where less than 200 ng was available the maximum volume (20 μl) was used. The fragmented RNA was then purified using RNAclean XP (Beckman Coulter) magnetic beads and used as template for double-stranded cDNA synthesis using random hexamers (Roche) and a cDNA Synthesis System Kit (Roche) according to the manufacturer's instructions. To improve the population of viral reads at the genomic termini, two panlyssavirus specific primers were included in the random RT reaction: N165-146 (GCAGGGTAYTTRTACTCATA) previously described [25] and LRACEF2 (TGAGTCTRTCATCTCACTGG) at 1 pmol/μl. Ends were repaired and specific sequencing adapters with multiplex identifiers ligated using the Rapid Library Kit (Roche). The resultant fragments were purified and size-selected using Ampure XP (Beckman Coulter) magnetic beads and the libraries quantified with High Sensitivity DNA chips on a BioAnalyzer (Agilent).

Libraries were pooled as appropriate in equimolar concentrations. Pooled fragments were clonally amplified using emPCR kits (Roche) and sequenced on the Roche 454 GS FLX + instrument according to the manufacturer's instructions. Each sample was loaded onto 16th picotitre 454 plate.

Sequence data analysis

Initial assembly of the 454 reads was achieved with Newbler (V2.6) using the GS *de novo* assembly software (Roche). All contigs were exported from the software and were aligned with a reference genome using Seqman (DNAStar) whenever possible. The resulting consensus sequence was subsequently used in GS Reference Mapper (Roche) to obtain further sequence reads from the original raw data. If a suitable reference sequence was available, this was used in the GS Reference Mapper program with the 454 reads to obtain a consensus sequence without the initial use of the GS *de novo* assembly software. Where no suitable genome reference sequence was available (RV2508) a BLAST search with the contigs provided evidence for which contigs were viral sequence, using the megablast algorithm on the nucleotide database (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome). Due to the increased difficulty of mapping reads at the ends of contigs, it was occasionally necessary to splice the 3′ and/or 5′ UTR sequences from a reference genome sequence onto the termini of the contig sequence of interest. Using the spliced consensus sequence in the GS Reference Mapper program resulted in extra reads being mapped to the contig resulting in a complete, or almost complete genome sequence.

Obtaining genomic termini

The lyssavirus genomic termini are extremely conserved, in particular the first and last 9 nucleotides have complete complementarity across all lyssaviruses [1,2,5,22], furthermore the conservation remains stringent until residue 25 [2]. Previously described panlyssavirus primers situated at the leader and trailer extremities, leader: LYS001F (5′ACGCTTAACGAMAAA3′) and trailer: LYSEND (5′ACGCTTAACAAAWAAA3′) [2] were used with panlyssavirus primers JW6UNI (5′ CARTTVGCRCACATYTTRT3′) and LgeneFor (5′CTCACTGGATMAGRTTRATITACAA3′) respectively to obtain PCR products which were sequenced directly using the PCR primers as described elsewhere [26].

SNP 'deep sequencing'

Primers flanking the SNP identified for IKOV: RV2508-L-SNP-F (5′ CTGAAGCTTCGAGACTCTAC 3′) and RV2508-L-SNP-F (5′ CAGATGGATGACCCTATCAG 3′) were used in a standard RT-PCR reaction, using high fidelity enzymes in methods described previously [1]. The amplicons were purified using Ampure XP (Beckman Coulter) magnetic beads and specific sequencing adapters with multiplex identifiers were ligated using the Rapid Library Kit (Roche) and purified using Ampure XP (Beckman Coulter) magnetic beads. The purified ligated amplicons were size selected and quantified with High Sensitivity DNA chips on a BioAnalyzer (Agilent). Amplicons were pooled as appropriate in equimolar concentrations and clonally amplified using emPCR kits (Roche) and sequenced on the Roche 454 GS FLX + instrument according to the manufacturer's instructions. The RV2508 genome sequence was used in the GS Reference Mapper program with the 454 reads to determine the percentage of reads present for the SNP.

Abbreviations

NGS: Next generation sequencing; BHK: Baby hamster kidney; SNP: Single nucleotide polymorphism.

Competing interests
The authors declare they have no competing interests.

Authors' contributions
Funding was obtained by DM, LM, XDEL and ARF. DM carried out sequence analyses. DM, EW, SL carried out and optimised the experiments. DM, LM, RE, DH, XDEL and ARF designed the study. DD and RE provided reagents and tools. DM drafted the manuscript. All authors contributed to and approved the final manuscript.

Acknowledgements
The authors would like to thank Saira Crawthraw and Colin Black for technical assistance, Don King, Dirk Höper, Thomas Müller, Sylvia Grieson, Bhudipa Choudhury, Chad Fuller and Brandon Londt for fruitful discussions. We are grateful to our collaborators in the US, Estonia, Morocco, Tanzania, Iraq and Denmark, for providing the lyssavirus positive samples. This work was funded by the Department for Environment, Food and Rural Affairs (Defra), UK (grant ROAME SE0427) and by the EU FP7–funded Research Infrastructure Grant European Virus Archive (no.19 228292).

Author details
[1]Wildlife Zoonoses & Vector-Borne Diseases Research Group, Animal Health & Veterinary Laboratories Agency (AHVLA), New Haw, Addlestone, Surrey KT15 3NB, UK. [2]Central Sequencing Unit, Animal Health & Veterinary Laboratories Agency (AHVLA), New Haw, Addlestone, Surrey KT15 3NB, UK. [3]National Consortium for Zoonosis Research, Leahurst, Neston, Wirral, UK. [4]Rabies Laboratory, Kimron Veterinary Institute, Bet Dagan 50250, Israel. [5]Aix Marseille Univ, IRD French Institute of Research for Development, EHESP French School of Public Health, UMR_D 190, "Emergence de Pathologies Virales", 13005 Marseille, France.

Received: 26 March 2013 Accepted: 25 June 2013
Published: 4 July 2013

References
1. Marston DA, McElhinney LM, Johnson N, Muller T, Conzelmann KK, Tordo N, Fooks AR: **Comparative analysis of the full genome sequence of European bat lyssavirus type 1 and type 2 with other lyssaviruses and evidence for a conserved transcription termination and polyadenylation motif in the G-L 3′ non-translated region.** *J Gen Virol* 2007, **88**(Pt 4):1302–1314.
2. Kuzmin IV, Wu X, Tordo N, Rupprecht CE: **Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions.** *Virus Res* 2008, **136**(1–2):81–90.
3. Conzelmann KK, Cox JH, Schneider LG, Thiel HJ: **Molecular cloning and complete nucleotide sequence of the attenuated rabies virus SAD B19.** *Virology* 1990, **175**(2):485–499.
4. Tordo N, Poch O, Ermine A, Keith G, Rougeon F: **Walking along the rabies genome: is the large G-L intergenic region a remnant gene?** *Proc Natl Acad Sci USA* 1986, **83**(11):3914–3918.
5. Delmas O, Holmes EC, Talbi C, Larrous F, Dacheux L, Bouchier C, Bourhy H: **Genomic diversity and evolution of the lyssaviruses.** *PLoS One* 2008, **3**(4):e2057.
6. Kuzmin IV, Shi M, Orciari LA, Yager PA, Velasco-Villa A, Kuzmina NA, Streicker DG, Bergman DL, Rupprecht CE: **Molecular Inferences Suggest Multiple Host Shifts of Rabies Viruses from Bats to Mesocarnivores in Arizona during 2001–2009.** *PLoS Pathog* 2012, **8**(6):e1002786.
7. Bracho MA, Moya A, Barrio E: **Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity.** *J Gen Virol* 1998, **79**(Pt 12):2921–2928.
8. Schuster SC: **Next-generation sequencing transforms today's biology.** *Nat Methods* 2008, **5**(1):16–18.
9. Shendure J, Ji H: **Next-generation DNA sequencing.** *Nat Biotechnol* 2008, **26**(10):1135–1145.
10. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N: **Application of next-generation sequencing technologies in virology.** *J Gen Virol* 2012, **93**(Pt 9):1853–1868.
11. Tse H, Tsang AK, Tsoi HW, Leung AS, Ho CC, Lau SK, Woo PC, Yuen KY: **Identification of a novel bat papillomavirus by metagenomics.** *PLoS One* 2012, **7**(8):e43986.
12. Wu Z, Ren X, Yang L, Hu Y, Yang J, He G, Zhang J, Dong J, Sun L, Du J, *et al*: **Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces.** *J Virol* 2012, **86**(20):10999–11012.
13. Liu S, Vijayendran D, Bonning BC: **Next generation sequencing technologies for insect virus discovery.** *Viruses* 2011, **3**(10):1849–1869.
14. Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y, Shi Z: **Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China.** *J Virol* 2012, **86**(8):4620–4630.
15. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, Berlin AM, Malboeuf CM, Ryan EM, Gnerre S, *et al*: **Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection.** *PLoS Pathog* 2012, **8**(3):e1002529.
16. Hoper D, Hoffmann B, Beer M: **Simple, sensitive, and swift sequencing of complete H5N1 avian influenza virus genomes.** *J Clin Microbiol* 2009, **47**(3):674–679.
17. Cottam EM, Wadsworth J, Shaw AE, Rowlands RJ, Goatley L, Maan S, Maan NS, Mertens PP, Ebert K, Li Y, *et al*: **Transmission pathways of foot-and-mouth disease virus in the United Kingdom in 2007.** *PLoS Pathog* 2008, **4**(4):e1000050.
18. Ninomiya M, Ueno Y, Funayama R, Nagashima T, Nishida Y, Kondo Y, Inoue J, Kakazu E, Kimura O, Nakayama K, *et al*: **Use of illumina deep sequencing technology to differentiate hepatitis C virus variants.** *J Clin Microbiol* 2012, **50**(3):857–866.
19. Astrovskaya I, Tork B, Mangul S, Westbrooks K, Mandoiu I, Balfe P, Zelikovsky A: **Inferring viral quasispecies spectra from 454 pyrosequencing reads.** *BMC Bioinforma* 2011, **12**(Suppl 6):S1.
20. Tordo N, Poch O, Ermine A, Keith G, Rougeon F: **Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative-strand RNA viruses.** *Virology* 1988, **165**(2):565–576.
21. Freuling CM, Beer M, Conraths FJ, Finke S, Hoffmann B, Keller B, Kliemt J, Mettenleiter TC, Muhlbach E, Teifke JP, *et al*: **Novel lyssavirus in Natterer's bat, Germany.** *Emerg Infect Dis* 2011, **17**(8):1519–1522.
22. Marston DA, Ellis RJ, Horton DL, Kuzmin IV, Wise EL, McElhinney LM, Banyard AC, Ngeleja C, Keyyu J, Cleaveland S, *et al*: **Complete genome sequence of ikoma lyssavirus.** *J Virol* 2012, **86**(18):10242–10243.
23. Horton DL, McElhinney LM, Marston DA, Wood JL, Russell CA, Lewis N, Kuzmin IV, Fouchier RA, Osterhaus AD, Fooks AR, *et al*: **Quantifying antigenic relationships among the lyssaviruses.** *J Virol* 2010, **84**(22):11841–11848.
24. Marston DA, McElhinney LM, Banyard AC, Horton DL, Nunez A, Koser ML, Schnell MJ, Fooks AR: **Interspecies protein substitution to investigate the role of the lyssavirus glycoprotein.** *J Gen Virol* 2012, **94**(2):284–292.
25. Wakeley PR, Johnson N, McElhinney LM, Marston D, Sawyer J, Fooks AR: **Development of a real-time, TaqMan reverse transcription-PCR assay for detection and differentiation of lyssavirus genotypes 1, 5, and 6.** *J Clin Microbiol* 2005, **43**(6):2786–2792.
26. Marston DA, McElhinney LM, Ali YH, Intisar KS, Ho SM, Freuling C, Muller T, Fooks AR: **Phylogenetic analysis of rabies viruses from Sudan provides evidence of a viral clade with a unique molecular signature.** *Virus Res* 2009, **145**(2):244–250.

doi:10.1186/1471-2164-14-444
Cite this article as: Marston *et al.*: Next generation sequencing of viral RNA genomes. *BMC Genomics* 2013 **14**:444.

2.2.4: Article 5 Book Chapter published in 'Current Laboratory Techniques in Rabies Diagnosis, Research and Prevention'

Chapter 17: Next Generation Sequencing of Lyssaviruses

Summary: This book chapter is part of the 1st Edition of '*Current Laboratory Techniques in Rabies Diagnosis, Research and Prevention'* Handbook. The readership is researchers, students, health professionals, industry personnel and laboratory personnel in the vaccine, pharmaceutical and diagnostic industries. This chapter is the first rabies NGS book chapter written, therefore key to other rabies researchers intending on using NGS for diagnostic or research purposes.

This book chapter was written using article 4 and my written protocols as a basis.

CHAPTER SEVENTEEN

Next Generation Sequencing of Lyssaviruses

Denise A. Marston[1], Lorraine M. McElhinney[1,2], Emma Wise[1], Richard J. Ellis[3], Conrad M. Freuling[4], Thomas F. Müller[4], and Anthony R. Fooks[1,2]

[1]Wildlife Zoonosis and Vector-Borne Disease Research Group, Animal Health and Veterinary Laboratories Agency, Addlestone, Surrey, United Kingdom
[2]Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom
[3]Central Sequencing Unit, Animal Health and Veterinary Laboratories Agency, Addlestone, Surrey, United Kingdom
[4]Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Institute of Molecular Biology, Greifswald-Insel Riems, Germany

Chapter Contents

17.1 BACKGROUND

Lyssaviruses have small, simple genomes of approximately 12 Kb. Traditionally, this small size has facilitated the acquisition of a complete genome sequence via conventional Sanger sequencing of overlapping

C. Rupprecht & T. Nagarajan (Eds): Current Laboratory Techniques in Rabies Diagnosis, Research, and Prevention, Volume 1.
Doi: http://dx.doi.org/10.1016/B978-0-12-800014-4.00017-2

polymerase chain reaction (PCR) amplicons ("walking the genome"). However, this methodology is extremely time-consuming and expensive. Variation within and between the lyssaviruses species often demands considerable re-optimization of PCR primer sets, particularly for highly divergent or novel lyssaviruses. Moreover, PCR errors can be introduced during Sanger sequencing, often when amplicon cloning is undertaken or low-fidelity polymerases are employed.

Next Generation Sequencing (NGS) is a high-throughput sequencing methodology which generates millions of sequences simultaneously from one sample.[1] Multiple NGS platforms are available; two commonly used platforms are Roche 454 pyrosequencing (454 Life Science) and Illumina (Solexa).[2] These platforms were initially optimized and applied to large mammalian and bacterial genomes, where Sanger methods are prohibitively expensive. Obtaining the hundreds of nanograms of target nucleic acid required for NGS is relatively simple, as nucleic acid is in abundance. Due to the low percentage of viral RNA in relation to host RNA in any given sample, obtaining the quantity of viral RNA required by historical protocols has been problematic.

The advantages of NGS technologies in virology are numerous, and its use is becoming more commonplace,[3,4] particularly to detect and characterize pathogens without prior knowledge of their existence, and in the absence of specific primers.[5] Unfortunately, the virus is often at such low abundance in clinical samples that only one or two reads can be obtained, or at best, a number of unlinked contiguous sequences (contigs), for which traditional Sanger protocols are employed to complete the genome.[6]

In this chapter, we describe a broadly applicable approach to obtaining full genome sequences from clinical or cultured samples, which we have successfully applied to lyssavirus-infected samples.

17.2 METHODOLOGY

The following describes the protocols followed when using the Roche 454 FLX(+) Titanium. However, other platforms are also available.

17.2.1 Viral RNA Extraction

The proper preparation of the infected samples is the key to ensuring the success of obtaining viral genomes via NGS. Sample preparation protocols differ for the variety of tissue samples and viral culture samples employed due to their different properties, but the overall aim is to enrich for viral

RNA, either by depleting host genomic and ribosomal RNA, or by concentrating the culture samples prior to RNA extraction.

17.2.1.1 Tissue Samples

Extract RNA from the sample using TRIzol® (tissue) or TRIzol®LS (liquids) following manufacturer's instructions, and re-suspend in 10 µL molecular-grade water. Removal of cell debris in the sample using a stepwise filtration process involving 0.45 µm polyether sulfone membrane filters (diameter 13 mm) (Millipore) may be performed prior to RNA extraction if preferred, but is not essential if TRIzol or similar phase-separation methodologies are employed.

Make a 1 in 10 dilution of RNA with molecular grade water, and quantify by UV spectrophotometry using the Nanodrop ND-8000 or equivalent.

Perform depletion of genomic RNA and DNA using the RNeasy® plus mini kit (Qiagen), following manufacturer's instructions, eluting in 30 µL RNase-free water. DNA depletion is achieved by following the optional on-column DNase treatment step within the kit protocol.

Perform depletion of ribosomal RNA using Terminator™ 5′-Phosphate-Dependent Exonuclease (Epicentre Biotechnologies) according to the manufacturer's instructions, mixing the total volume (30 µL) of RNA with 3 µL of Buffer A, 0.5 µL RNAsin® Ribonuclease inhibitor (20–40 U/µL, Promega) and 1 µL Terminator (1 U/µL), then incubate at 30°C for 60 mins.

Cool samples on ice for 1 min, then perform one subsequent round of purification using RNeasy plus mini kit (Qiagen), without DNase digestion, following the manufacturer's instructions, and elute RNA in 30 µL RNase free water.

Quantify depleted RNA by UV spectrophotometry using the Nanodrop ND-8000 or equivalent.

RNA samples should be stored at −80°C until use.

Any depleted RNA left over after NGS should be retained and stored at −80°C for future use.

17.2.1.2 Virus Culture Samples

If the virus is at a low concentration, or if downstream procedures require it, an optional step of Polyeythylene glycol (PEG) precipitation may be performed to concentrate the virus prior to RNA extraction. Following manufacturer's instructions, use 1 volume of PEG-it (System Biosciences) to 4 volumes of sample. Re-suspend pellets in tissue culture medium and combine to achieve a final volume of 1 mL.

Extract RNA from the virus culture sample using TRIzol LS (following manufacturer's instructions). Make a 1 in 10 dilution of sample in molecular grade water, and quantify RNA as above.

17.2.2 Preparation of cDNA for Sequencing on the Roche 454 FLX(+) Titanium

The cDNA is generated from the enriched RNA using random or specific primers, and then sequencing primers are attached to each fragment. Individual fragments are then amplified clonally on beads prior to pyrosequencing.

17.2.2.1 Library Preparation

1. RNA sample quantitation is done by fluorometry using the Quant-iT™ RiboGreen® RNA Assay Kit (Life Technologies #R11490) according to the manufacturers instructions (1 µL sample diluted in 50 µL TE, add 50 µL 1:200 (v/v) RiboGreen reagent). Measure fluorescence using a fluorometer with a 480 nm excitation filter and 520 nm emission filter. Compare readings to standard curve.
2. Add up to 200 ng RNA to random primers as follows:
   a. 8 µL RNA sample ≤200 ng.
   b. 2 µL 400 µM Roche Primer "random" (Roche # 11034731001; alternatively, at this stage a cocktail of more specific primers could also be used to increase the relative abundance of the sequences of interest).
   c. Mix and heat to 70°C for 10 mins.
   d. Cool to 4°C.
3. Proceed to first strand synthesis using the Roche cDNA synthesis system (Roche # 11117831001). Add the following:
   a. 4 µL 5× RT-buffer AMV.
   b. 2 µL 0.1 M DTT.
   c. 2 µL 10 mM deoxynucleoside triphosphates (dNTPs).
   d. 0.5 µL Protector RNase Inhibitor (25 U/µL).
   e. 1 µL AMV reverse transcriptase (25 U/µL).
   f. Mix and incubate at 25°C for 10 mins and then at 42°C for 60 mins.
   g. Cool to 4°C.
4. Proceed immediately to second strand synthesis using the Roche cDNA synthesis system (Roche # 11117831001). Add the following:
   a. 15 µL 5× 2nd strand synthesis buffer.
   b. 36 µL Molecular Biology Grade Water.

   c. 1 µL 10 mM dNTPs.
   d. 3 µL 2nd strand enzyme.
   e. Mix and incubate at 16°C for 2 hrs.
   f. Add 10 µL T4 DNA polymerase.
   g. Mix and incubate at 16°C for a further 5 mins.
   h. Add 8.5 µL 0.2 M EDTA (pH 8).
5. Purify double-stranded cDNA:
   a. Add cDNA synthesis reaction to 150 µL Ampure XP beads (Beckman Coulter # A63880).
   b. Mix and incubate at room temperature for 5 mins.
   c. Place on a magnetic stand until solution clears, leaving beads against the side or bottom of the tube.
   d. Remove and discard supernatant without disturbing beads.
   e. Keeping the tube on the magnet, wash the beads 3 times with 500 µL 70% ethanol, removing all ethanol each time.
   f. Allow the beads to air dry for 5 mins.
   g. Remove the tube from the magnet and add 16 µL of 10 mM Tris-HCl, pH 7.5 and mix.
   h. Incubate at room temperature for 5 mins.
   i. Place on a magnetic stand until the solution clears, leaving beads against the side or bottom of the tube.
   j. Transfer supernatant containing cDNA to a new tube.
6. End-repair cDNA using a Roche Rapid Library Preparation Kit (Roche # 05608228001). Add the following to the cDNA:
   a. 2.5 µL RL 10× PNK Buffer.
   b. 2.5 µL RL ATP.
   c. 1 µL RL dNTP.
   d. 1 µL RL T4 Polymerase.
   e. 1 µL RL PNK.
   f. 1 µL RL Taq Polymerase.
   g. Mix and incubate at 25°C for 20 min, 72°C for 20 min followed by cooling to 4°C.
7. Ligate 454 sequencing primers using a Roche Rapid Library Preparation Kit (Roche # 05608228001), or using optional Multiplex identifiers (Roche # 05619211001). Add the following to the completed end-repair reaction:
   a. 1 µL of RL Adaptor or RL MID Adaptor.
   b. 1 µL of RL Ligase.
   c. Mix and incubate at 25°C for 10 mins.

8. Purify library fragments for sequencing as follows:
   a. Aliquot 125 µL Ampure XP beads into a 1.5 mL tube.
   b. Place on a magnetic stand until the solution clears, leaving the beads against the side or bottom of the tube.
   c. Remove and discard the supernatant without disturbing the beads.
   d. Add 26 µL TE and 140 µL sizing solution and mix.
   e. Add adapted library (~27 µL) from Step 7c.
   f. Mix and incubate at room temperature for 5 mins.
   g. Place on a magnetic stand until the solution clears, leaving the beads against the side or bottom of the tube.
   h. Remove and discard the supernatant without disturbing the beads.
   i. Add 53 µL TE and 140 µL sizing solution.
   j. Mix and incubate at room temperature for 5 mins.
   k. Place on a magnetic stand until the solution clears, leaving the beads against the side or bottom of the tube.
   l. Remove and discard the supernatant without disturbing the beads.
   m. Repeat Steps 8i to 8l.
   n. Keeping the tube on the magnet, wash the beads 3 times with 500 µL 70% ethanol, removing all ethanol each time.
   o. Allow the beads to air dry for 5 mins.
   p. Remove the tube from the magnet, add 22 µL of TE, and mix.
   q. Incubate at room temperature for 5 mins.
   r. Place on a magnetic stand until the solution clears, leaving the beads against the side or bottom of the tube.
   s. Transfer the supernatant containing the purified library to a new tube.
9. Assess library quality and concentration as follows:
   a. Run 1 µL aliquot on an Agilent Bioanalyzer High Sensitivity DNA chip.
   b. Based on the average fragment size and concentration determine the number of DNA molecules per µL, using the following formula: (Concentration [pg/µL] / Average size [bp]) × 9.17 × 10^8.
   c. Dilute libraries to 1 × 10^7 molecules per µL.
10. Pool up to 12 libraries (each with a different MID adapter added at Stage 7a) for efficient multiplexed sequencing.

17.2.2.2 Emulsion Polymerase Chain Reaction

DNA library fragments are amplified clonally on beads using Lib-L emPCR kits (Roche #05619149001). The DNA is mixed with the beads, which are then suspended in an oil emulsion, prior to amplification.

Typically, mixing three molecules of DNA per bead provides a suitable level of enrichment (maximizing the number of beads carrying amplified DNA, while reducing the possibility of having mixed molecules of DNA attached to the same bead), and leads to optimal sequence output. Manufacturers' instructions are followed for emPCR and bead recovery.

17.2.2.3 Sequencing on the Roche 454 FLX+

Recovered beads are deposited in PicoTiter plates and sequenced on the Roche 454 FLX+ instrument according to the manufacturers' instructions. Reads of up to 900 bp can be achieved using this approach. The instrument runs for approximately 24 hours, including the initial data processing. Reads are assembled using the GS Assembler (Newbler) software, using standard settings. This performs *de novo* assembly, and often this approach leads to a single viral consensus contig. Using the methods given here, viral sequencing reads account for between 0.5 and 10% of the output. Where a very closely related virus is available, it can be used as a reference to which sequence reads can be mapped with GS Mapper.

When the consensus contig sequence is used in GS Reference Mapper (Roche), further sequence reads may be obtained from the original raw data. If a suitable reference sequence is available, this is used in the GS Reference Mapper program with the 454 reads to obtain a consensus sequence without the initial use of the GS *de novo* assembly software. If no genome reference sequence is available, a BLAST search, with the individual contig sequences from the *de novo* assembly will indicate which contigs are viral sequences (use the megablast algorithm on the nucleotide database[7]). Due to the increased difficulty of mapping reads at the ends of contigs, it may be necessary occasionally to splice the 3′ and/or 5′ UTR sequences from a reference genome sequence onto the termini of the contig sequence of interest. Using the spliced consensus sequence in the GS Reference Mapper program may result in extra reads being mapped to the contig, resulting in a complete, or almost complete, genome sequence.

17.2.3 Analyzing and Annotating 454 Genome Sequences

The following protocols are undertaken using the software provided by Roche with the 454 sequencer and sequence editing programs provided in DNAStar package, Lasergene 10. Other editing programs are available.

17.2.3.1 Obtaining a Consensus Sequence from a De Novo Project

Open the *de novo* assembled project, click on "project" then "GS reads." This gives you the sff file ID number, the multiplex ID number, and an

indication of how successful the sequencing was accomplished (percentage of reads assembled should be 70% or higher).

Click on "alignment results" and look at the contig sequences listed on the left. If the top contig is approximately 11–12 kb in length (if analyzing a lyssavirus genome), this will be a single contig containing the virus genome. If the contigs are smaller, the viral reads may still be present, but in a number of smaller contigs.

Rename "454AllContigs.fna" file with the virus ID at the beginning.

Open SeqMan (DNAStar, Lasergene 10) and drag the .fna file into the "unassembled sequences" box to open the contig files. In "options," uncheck the "don't add single sequences" option, then click "assemble."

If the viral sequences are all in a single contig (always named contig00001), then simply rename contig0001 from the *de novo* folder and save the consensus sequence. If the viral sequences are in a number of contigs, add an appropriate reference genome sequence to bring them together into a single contig.

At this point you can enter the sequence into BLAST, if desired, to identify similar sequences.

The majority of sequences do not have the genomic ends covered by reads, and occasionally there are gaps in between contigs. To complete the genome, either re-sequence the sample on the NGS platform, to cover the regions without reads, or use a PCR-based strategy to amplify across the gaps, and use the Sanger sequencing technique to obtain the sequence.

17.2.3.2 Obtaining a Consensus Sequence from a Mapper Project

If a closely related sequence is available, this can be used instead of *de novo* analysis. Open the Mapper program, create a new mapper project, and call it "ID…mapped".

In "project," click on "reference" then on the "+" icon. Navigate to a relevant consensus sequence saved as a fasta file. Click on "GS reads," "+," and locate the sff folder relevant for your virus. Between one and four sff files will be available; use shift to highlight these and click "select." Check the "use multiplex filtering" box, click on the arrow next to "scheme, no multiplexing" and choose "RLMIDs." Uncheck all boxes except the RL used for the virus of interest. Click "OK."

Click the "Start" button, located far right. Analysis should take less than 1 min. When complete, the "alignment results" tab will be bold.

Click on "alignment results," then click on "contig0001." The top sequence is the imported reference sequence, and below will be the aligned NGS reads.

Scroll through the alignment to get an idea of the depth of coverage and if there are any gaps in the sequence (this will be obvious as there will be more than just the one contig00001 on the left). The project overview indicates the sff file ID, the multiplex ID, the percentage of viral specific reads, and the total number of viral specific reads.

Open the .fna file in EditSeq.

The above steps ensure optimal mapping has been carried out, and should be repeated any time the consensus sequence is modified (e.g. when genomic ends have been obtained by conventional PCR and sequencing and are incorporated into the genome), as reads may exist in the raw reads, but are not mapped. Be aware that the resulting contig will be the consensus sequence of the reads, therefore any corrections made may be lost on a subsequent mapped contig file and will need to be redone.

Save this reference sequence as a .fas file.

17.2.3.3 Checking Gene and Intergenic Lengths

Print out:

1. The consensus sequence in EditSeq to annotate.
2. If the lyssavirus is a known species, perform an alignment of the consensus sequence with other genomes from the same species. Annotate the start and stop regions of both using annotation from the Genbank file of the known lyssavirus genome.

Using the print outs from above, annotate the Editseq file with start and stop codons. Highlight each region in Editseq and note the lengths. Gene lengths MUST be divisible by 3. If the gene length differs from other viruses of same species, then:

1. Look at the genome alignment (or individual gene alignment) and identify indels (indicated by dashes).
2. In EditSeq, copy the coding sequence and paste it into a new DNA file. Select all and translate, and find where stop codons are introduced to determine where the problem is.

If intergenic regions differ, first check that the termination signal is GAAAAAAA. Often this is only 5 or 6 As because NGS platforms cannot read polymers easily. Look at raw data reads and identify number of 7 As, and look at flowgrams. Otherwise, look at the genome alignment to determine indels. It is common to have different length intergenic regions (Table 17.1).

Table 17.1 Lengths of Coding and Non-Coding Regions of Lyssavirus Genomes

	RABV	LBV	MOKV	DUVV	EBLV-1	EBLV-2	ABLV	ARAV	KHUV	IRKV	WCBV	SHIBV	BBLV	IKOV
3′ UTR⋆	70	70	70	70	70	70	70	70	70	70	70	70	70	70
N protein	**1353**	**1353**	**1353**	**1356**	**1356**	**1356**	**1353**	**1356**	**1356**	**1356**	**1353**	**1353**	**1356**	**1353**
N–P	90-1	101	100-102	90	90	101	94	85	95	93	64	98	91	66
P protein	**894**	**918**	**912**	**897**	**897**	**894**	**894**	**894**	**894**	**897**	**894**	**918**	**894**	**870**
P–M	88	75	80	83	83	88	89	85	72	82	133	76	86	74
M protein	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**
M–G	211-5	204	203-204	191	211	210 (205)	207-209	210	208	214	206	205	210	209
G protein	**1575**	**1569**	**1569**	**1602**	**1575**	**1575**	**1578-81**	**1581**	**1581**	**1575**	**1578**	**1569**	**1575**	**1575**
G–L	522	578-588	546-563	562-3	560	512	508-509	514	504	569	862	613	496	569
L protein	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6381**
5′ UTR	131	145	112–114	130–131	131	131	131	130	130	131	125	150	129	126
Genome	**11 923-8**	**12006-16**	**11 940-57**	**11 975-6**	**11 966**	**11 930**	**11 918**	**11918**	**11903**	**11980**	**12278**	**12045**	**11900**	**11902**

Full genomes of BBLV (Bokeloh Bat Lyssavirus, JF311903), RABV (Rabies Virus, M31046; EU293111; EU293115; EU293113; EU293116), DUVV (Duvenhage Virus, EU293120; EU293119), EBLV-1 (European Bat Lyssavirus type-1, EU293109, EU293112, EF157976), IKOV (Ikoma Lyssavirus, JX193798), ABLV (Australian Bat lyssavirus, NC_003243; AF081020), KHUV (Khujand Virus, EF614261), IRKV (Irkut Virus, FJ905105, EF614260), EBLV-2 (European Bat Lyssavirus type-2, EF157977, EU293114), ARAV (Aravan Virus, EF614259), LBV (Lagos Bat Virus, EU293108; EU293110), MOKV (Mokola Virus, EU293118, EU293117), SHIBV (Shimoni Bat Lyssavirus, GU170201) and WCBV (West Caucasian Bat Virus, EF614258) were derived from NCBI Genbank.
⋆6387 in SHBRV.

Check any positions queried by lower case letters in the GS program. Edit the EditSeq file according to your findings and save.

17.2.3.4 Obtaining Genomic Termini

The lyssavirus genomic termini are extremely conserved; in particular the first and last 9 nucleotides have complete complementarity across all lyssaviruses,[8] furthermore the conservation remains stringent until residue 25.[9] Pan-lyssavirus primers situated at the leader and trailer extremities, leader: LYS001F (5′ACGCTTAACGAMAAA3′) and trailer: LYSEND (5′ACGCTTAACAAAWAAA3′)[9] can be used with panlyssavirus primers JW6UNI (5′CARTTVGCRCACATYTTRT3′) and LgeneFor (5′CTCACTGGATMAGRTTRATITACAA3′), respectively, to obtain PCR products which can be sequenced directly using Sanger sequencing.[10] Use the remaining depleted RNA as the template.

17.2.4 Depth of Coverage Analysis

The depth of viral read coverage from which a single genome contig is obtained can vary depending on the starting material and sample preparation employed. Tissue culture supernatant prepared samples achieved the highest read coverage and depth of coverage, as the sample is amplified virus. However, culturing virus is not always possible. Furthermore, culturing virus in mammalian cells will result in adaptation of the virus resulting in sequence modification.

Sequences generated on the 454 platform, due to the way the incorporation of bases are detected by the intensity of light emission, inherently have issues with long homopolymeric repeats. On the whole, a consensus sequence with 10 or more reads, when mapped using an available genome from the same species, can be confirmed confidently by checking the flowgram data. However, novel viruses, or viruses with a limited number of reads, may require additional confirmation by Sanger sequencing PCR products, which span the regions in question.

17.3 DISCUSSION

The methodology described above enables the sequencing from brain tissue, and cell cultured samples without the use of specific viral enhancement, ultracentrifugation or filtration to identify novel viruses, or obtain genome sequences for known virus species. This methodology does not require prior knowledge of the virus being extracted, and uses a

methodology that is widely used to extract viral RNA (so archival RNA samples can be investigated retrospectively). Instead of enriching for virus-specific RNA, host genomic DNA and rRNA is depleted to increase the percentage of viral specific RNA in the sample.

There are a large number of RNA extraction methodologies and kits available that can be divided broadly into two categories: (1) Phenol/chloroform phase separation-based extraction, and (2) Column or bead based extraction. We have compared TRIzol as a phenol-based system and RNeasy plus mini kit (Qiagen) as a column-based system. In our hands, the use of spin columns to extract the RNA from tissue removed the majority of RNA in process, resulting in less than the 200 ng minimum quantity of total RNA recommended for fragmentation and, more importantly, too few viral reads were obtained to provide useful contigs. The RNeasy plus mini kit (Qiagen) was used successfully as part of the depletion protocol after initial extraction with TRIzol with good results.[4] Complete genome sequence can be obtained by NGS even when less than the recommended 200 ng of RNA is available for library preparation in TRIzol-extracted tissue samples. These observations are important to illustrate that the methods for library preparation can be refined further for virus sequencing.

The requirement to deplete tissue culture supernatant-derived RNA was investigated using two 100 mL supernatant samples.[4] Half the RNA was subjected to gDNA and rRNA depletion, whereas the other was not. The additional preparation time required to deplete the sample was not rewarded with a significant improvement of either proportion of viral-specific reads or read depth. Therefore, 454 sequencing directly from a TRIzol-extracted RNA is the simplest and most effective way to obtain a viral genome sequence from PEG-precipitated tissue culture supernatant samples.

In addition to genome sequencing, NGS technologies can be applied to "Deep Sequence" clinical material for a plethora of uses, including determining viral heterogeneity, or the effect of the immune system or pharmaceutical drugs on viruses.[11] These deep sequencing methodologies employ the use of viral PCR amplicons to obtain a depth of coverage sufficient to detect variants which occur down to 0.1% frequency, and may be applied to lyssavirus-infected material to gain a greater understanding of the role that viral heterogeneity ("quasispecies") plays in virus adaptation and evolution.

The platform described in this chapter (Roche 454) has been shown to be reliable and consistent. However, the RNA prepared using the methodologies described here can be used for any NGS platform and

have been used successfully on the Illumina platform. The application of NGS sequencing of viruses has begun a new era of virus discovery and characterization of novel viruses, and will revolutionize phylogeographic studies in all fields of virology in the coming years.

All viral sequences should be published, particularly if included in a peer-reviewed journal (e.g. NCBI Nucleotide[12]). For future submissions, we strongly recommend using the most recent version of the standardized sequence submission software, Sequin, obtained from the NCBI website,[13] which will harmonize sequence submissions and encourage submitting laboratories to include all essential data, including collection date, virus species, host species, and country of isolation.

ACKNOWLEDGMENTS

This work was funded by the Department for Environment, Food and Rural Affairs (Defra), UK (grant ROAME SE0427), and by the EU FP7-funded Research Infrastructure Grant European Virus Archive (No. 19 228292).

REFERENCES

1. Schuster SC. Next-generation sequencing transforms today's biology. *Nat Methods* 2008;**5**(1):16–18.
2. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008;**26**(10):1135–45.
3. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. *J Gen Virol* 2012;**93**(Pt 9):1853–68.
4. Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, et al. Next Generation Sequencing of viral RNA genomes. *BMC Genomics* 2013;**14**(1):444.
5. Tse H, Tsang AK, Tsoi HW, Leung AS, Ho CC, Lau SK, et al. Identification of a novel bat papillomavirus by metagenomics. *PLoS One* 2012;**7**(8):e43986.
6. Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y, Shi Z. Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* 2012;**86**(8):4620–30.
7. Basic Local Assignment Search Tool (BLAST): <http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome> [accessed 23.02.14].
8. Marston DA, Ellis RJ, Horton DL. Kuzmin IV, Wise EL, McElhinney LM, et al. Complete genome sequence of Ikoma lyssavirus. *J Virol* 2012;**86**(18):10242–3.
9. Kuzmin IV Wu X, Tordo N, Rupprecht CE. Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions. *Virus Res* 2008;**136**(1–2):81–90.
10. Marston DA, McElhinney LM, Ali YH, Intisar KS, Ho SM, Freuling C, et al. Phylogenetic analysis of rabies viruses from Sudan provides evidence of a viral clade with a unique molecular signature. *Virus Res* 2009;**145**(2):244–50.
11. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, et al. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* 2012;**8**(3):e1002529.
12. NCBI, Nucleotide: <http://www.ncbi.nlm.nih.gov/nucleotide/> [accessed 23.02.14].
13. NCBI, Sequin: <www.ncbi.nlm.nih.gov/projects/Sequin/> [accessed 23.02.14].

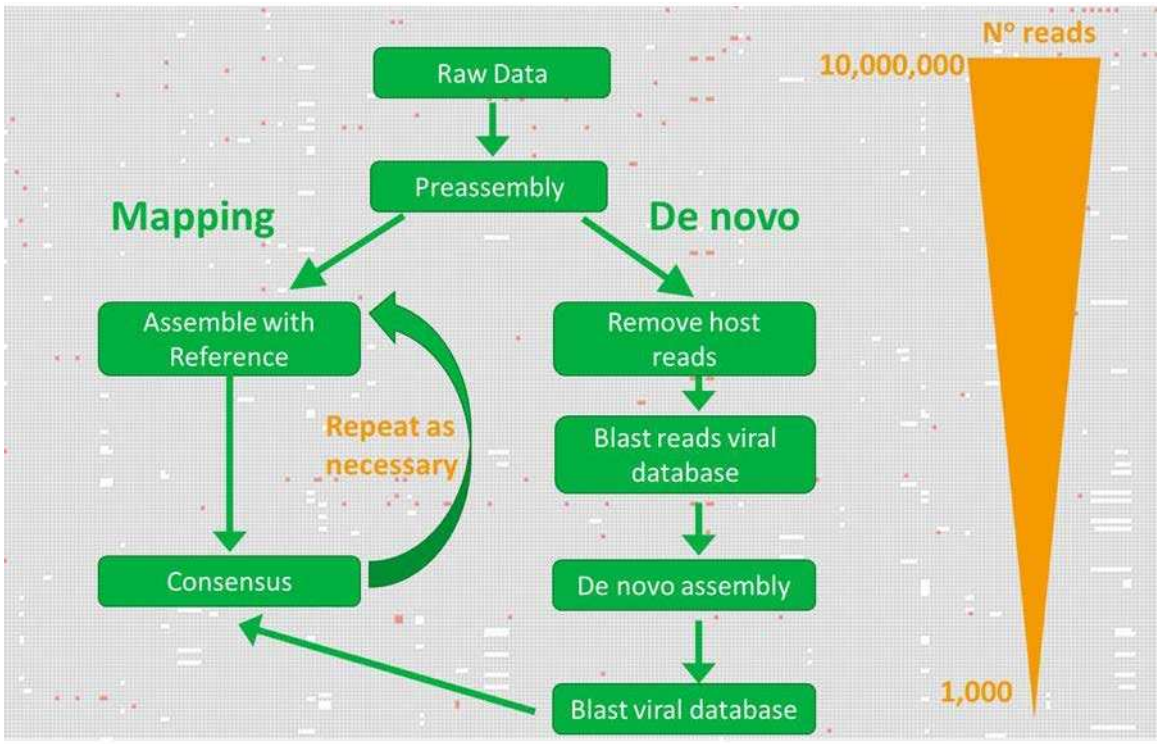
2.3: Development of bioinformatics to obtain full length viral genomes

2.3.1: Introduction

Once the RNA preparation protocol was optimised, I applied these preparation protocols to obtain full length genome sequences for a number of important viruses for which no full genome sequence was available. The analysis of Roche 454 data was achieved using a suite of programs developed by Roche, specifically for data from the 454 sequencing machine. Illumina NGS machines however were not supported with a bioinformatics platform, therefore required development. Regardless of the NGS platform, the pipeline used to assemble the reads, depended on whether a suitable reference sequence was available to map the reads against. Figure 11 is a schematic of the bioinformatics pipeline developed for obtaining viral genome sequences regardless of the NGS platform used. This pipeline has been developed by the Sequencing Facility at APHA, using a command line linux platform. I contributed to the development of this pipeline and the 'iteration' process. This iteration or repeat mapping of the reads to intermediate consensus sequences was a direct result of my findings that mapping reads to a reference sequence, which had indels compared to the virus being sequenced, resulted in an incorrect consensus. However, if the iteration 1 consensus sequence is used as the reference sequence to remap the NGS reads, then the next consensus is more representative of the virus of interest. It was shown that four iterations were usually enough to revolve any indel issues and result in a reliable consensus sequence.

The articles in Chapters 3 and 4 have been published to provide complete genome sequences of important viruses to make them accessible to the scientific community. Other articles throughout this thesis have also used the methodologies described in this chapter. Unless stated otherwise, the APHA bioinformatics pipeline described above has been used to obtain the initial consensus sequence. The genome sequences have been obtained using either the mapping pipeline if a suitable reference sequence is available (Chapter 3 – virus genome characterisation) or using the de novo pipeline if no suitable reference is available (Chapter 4 – Virus discovery). The latter ultimately is processed through the mapping pipeline once (a) a suitable reference is identified on Genbank using the Sherlock pipeline, or (b) a consensus sequence is obtained using the de novo assembly software, creating a reference sequence to remap the original reads against. This process can take many weeks to complete and usually highly variable regions require traditional PCR amplicon sequencing to span gaps. A final mapping process is always performed using the complete virus consensus sequence to obtain the most comprehensive coverage of reads across the genome.

Figure 11: Schematic of the bioinformatics pipeline developed at APHA for viral samples.



2.3.2: Obtaining complete virus genome sequences using mapping pipeline

Of the two pipelines, the mapping pipeline is a simpler, less time consuming pipeline. A suitable reference sequence is used to map the viral reads from the sample of interest and a consensus sequence is obtained. This intermediate consensus sequence is used to remap the reads and this process in repeated four times. Due to this iterative process, the reference sequence can differ from the virus of interest (including the length) but there are limits to this divergence. This process successfully resolved a 5 bp deletion in the M-G region of an EBLV-2 genome when mapped against a reference sequence with the five bases present. A single mapping step contained the 5bp sequence despite the absence of these bases in the reads spanning this region. However, experience with an influenza virus containing a 78 bp deletion in comparison to the reference sequence resulted in incorrect consensus sequence after four iterations, most likely due to the length of the insertion. The maximum size of an indel that can be correctly called will depend on the read length and mapping parameters, however a rule of thumb that an indel up to 1/3 of the read length can be tolerated.

However, this size is reduced if more than one indel is present in the same read. Analysis of the resulting consensus sequence is always essential, including analysis of the alignment file to identify any indels which usually are minority, but in the above examples was majority, and any important SNPs. Manually altering the reference sequence to reflect what is observed in the sample, or performing PCR and Sanger sequencing across the variable region then splicing this into the consensus sequence obtained has been necessary to resolve difficult hypervariable regions. Usually SNPs which are approximately 50:50 are identified as 'N' by the mapping software, which subsequently require the operator to investigate the position and decide the correct base. A number of examples are given in this thesis where indels and SNPs were manually resolved to obtain a correct final consensus sequence.

2.3.3: Obtaining complete virus genome sequences using de novo pipeline

De novo pipelines are essential for viruses where no suitable reference sequence exists. De novo assembly is accomplished by assembling individual reads together to create longer contigs (contiguous sequences) without using a reference sequence. This process is time consuming and more prone to errors, particularly with short read platforms such as Illumina as repeat regions cannot be resolved. In addition, for viral samples where the viral reads are present at extremely low proportions in comparison to the host reads, de novo assembly can be even more problematic. If the host is known and a suitable host genome library is available, mapping of the sample reads to the host genome and removal of these reads effectively reduces the reads by 80% or more, resulting in a more manageable amount of reads to denovo assemble.

The choice of program to de novo assemble viral reads can make a significant difference to the number of reads assembled into contigs. BWA is the most robust and widely used program for NGS analysis, but was designed for human and bacterial assembly. Due to the higher amount of heterogeneity present in viral samples, BWA and other standard assembly programs, reject reads with too many mis-matches resulting in a reduced number of viral contigs, or in the worst case scenario, not viral contigs at all. There are a number of de novo assembly programs developed particularly for virus assembly including Velvet, IVA, Tanoti and Vicuna. In the pipeline described in Figure 11 BWA is used as our first de novo assembly program, but if unsuccessful, one of the alternative viral specific programs is used. In the articles described below Velvet and IVA, were successfully used in addition to BWA.

After de novo assembly, the contigs are blasted against the viral sequence database (ViPR) and those with hits are taken forward to the mapping pipeline. Chapter 4 describes the de novo assembled genomes obtained during this thesis.

2.4: Conclusions

This chapter focuses on the development of NGS methodologies suitable for viruses. There are continual improvements to this methodology as better virus enrichment processes are tested and improved programs released to accommodate the additional complexities of sequencing viruses. Fundamentally our methodologies is hugely beneficial because it can be applied to any RNA or indeed DNA virus (if DNAse I treatment is removed).

Within APHA, full genome sequences of over 11 virus genera have been successfully sequenced, including important pathogens such as avian influenza, coronaviruses, rift valley fever virus and indeed over 70 lyssaviruses. The following chapters describe the application of these methodologies to a wide range of viruses with excellent results. Furthermore, the analysis of the genome sequences, including phylogeography, virus transmission and deep sequence virus heterogeneity analysis all are made possible due to the refinement of these methodologies.

Chapter 3: Virus Characterisation

Chapter 3: Virus Characterisation

3.1: Introduction

Chapter 2 described the bioinformatics developed to obtain full genome sequences of viruses. The majority of times this is completed for viruses which are not novel, rather they are of interest due to a new geographic location, host or other unique feature. Usually they form part of a larger study (see Chapters 5- 7) but occasionally are published in their own right to inform researchers or be used by others in larger studies. Below are a number of examples of individual genomes published using the mapping bioinformatics pipeline.

3.2: Article 6 published in Genome Announcements

Complete Genomic Sequence of Rabies Virus from an Ethiopian Wolf

DOI: 10.1128/genomeA.00157-15

Summary: This genome was a significant sequence to make available for two reasons. 1) The number of complete genomes available from Africa is extremely limited, although this is now slowly being addressed. When published, this genome represented the first RABV genome available from Ethiopia and the first Africa 1a genome sequence from East Africa. Indeed, only one Africa 1a genome had previously been published from Morocco. 2) Ethiopian wolves are highly endangered and obtaining the complete genome sequence of the RABV causing devastating outbreaks in these canids enable tracking of virus transmission and inform policy makers (see 6.5 for further details).

This genome sequence was obtained while the iteration process was being optimised. Indeed, analysis of the consensus sequence obtained from mapping the reads to a related RABV genome (Moroccan cow KF155001) with 96.2% identity. Just one iteration of mapping resulted in 2,023 reads mapped, but there were 13 incorrect base calls, including one 10bp deletion. Development of repeat mapping (4 iterations) resulted in 4,750 mapped reads (a 235% increase in number of reads mapped). Perhaps more critically, the errors observed after a single round of mapping were all resolved, including the 10 bp deletion. Interestingly, the reference sequence had the 10bp sequence, so the absence of the 10bp sequence after the first round of mapping was due to the low coverage of reads at that position, not due the reference sequence being used regardless of the read data. However, the iterative mapping improved the coverage across that region resulting in the 10bp region being correctly recognised.

genomeA Journals.ASM.org

Complete Genomic Sequence of Rabies Virus from an Ethiopian Wolf

Denise A. Marston,[a] Emma L. Wise,[a] Richard J. Ellis,[b] Lorraine M. McElhinney,[a,c] Ashley C. Banyard,[a] Nicholas Johnson,[a] Asefa Deressa,[d] Fekede Regassa,[e] Xavier de Lamballerie,[f] Anthony R. Fooks,[a,c,g] Claudio Sillero-Zubiri[h,i]

Wildlife Zoonoses & Vector-Borne Diseases Research Group,[a] Specialist Scientific Support Department,[b] Animal and Plant Health Agency (APHA), Surrey, United Kingdom; National Consortium for Zoonosis Research, University of Liverpool, Liverpool, United Kingdom[c]; Ethiopian Public Health Institute, Addis Ababa, Ethiopia[d]; Ethiopian Wildlife conservation Authority, Addis Ababa, Ethiopia[e]; UMR_D 190, Emergence de Pathologies Virales, Aix Marseille Université, Marseille, France[f]; Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, United Kingdom[g]; Wildlife Conservation Research Unit, The Recanati-Kaplan Centre, Zoology, University of Oxford, United Kingdom[h]; Ethiopian Wolf Conservation Programme, Dinsho, Ethiopia[i]

Ethiopian wolves are the rarest canid in the world, with only 500 found in the Ethiopian highlands. Rabies poses the most immediate threat to their survival, causing epizootic cycles of mass mortality. The complete genome sequence of a rabies virus (RABV) derived from an Ethiopian wolf during the most recent epizootic is reported here.

Received 23 February 2015 **Accepted** 6 March 2015 **Published** 26 March 2015

Citation Marston DA, Wise EL, Ellis RJ, McElhinney LM, Banyard AC, Johnson N, Deressa A, Regassa F, de Lamballerie X, Fooks AR, Sillero-Zubiri C. 2015. Complete genomic sequence of rabies virus from an Ethiopian wolf. Genome Announc 3(2):e00157-15. doi:10.1128/genomeA.00157-15.

Address correspondence to Denise A. Marston, denise.marston@apha.gsi.gov.uk.

Rabies virus (RABV) is the type species of the *Lyssavirus* genus and the causative agent of rabies. Throughout Africa, domestic dogs are the principal reservoir for RABV (1), transmitting the virus to a wide range of terrestrial mammals, including wildlife, such as the Ethiopian wolf (*Canis simensis*) (2, 3). The dominant RABV lineage circulating within domestic dogs in Ethiopia is Africa 1a, which is part of the cosmopolitan lineage (4). Long-term population monitoring of Ethiopian wolves, in combination with laboratory testing of all wolf carcasses recovered, indicates that rabies is not endemic in the wolf population; rather, rabies epizootics occur in locations with a high density of wolves (5). Previous outbreaks of rabies in Ethiopian wolves have been effectively controlled with parenteral vaccination (6). A comparison of the complete virus genomes can enable fine mapping of disease transmission events, such as those that trigger Ethiopian wolf RABV epizootics. However, publically available complete genomes for African RABVs are limited. Of the 6 Africa 1 clade genomes available, only one is an Africa-1a bovine isolate from Morocco (accession no. KF155001). Here, we describe full-genome sequencing of an Africa-1a RABV obtained from an infected Ethiopian wolf from the Sanetti Plateau, Bale Mountains, southeastern Ethiopia (RV2985), in August 2014.

RNA from brain material stored in glycerol was prepared for next-generation sequencing on the MiSeq platform. Briefly, TRIzol-extracted viral RNA was depleted of host genomic DNA and rRNA, as described previously (7). Double-stranded (ds)-cDNA was synthesized from 50 ng of RNA using a random cDNA synthesis system (Roche), according to the manufacturer's instructions. The ds-cDNA was purified using AMPure XP magnetic beads (Beckman Coulter), and 1 ng was used for the Nextera XT DNA sample preparation kit (Illumina). A sequencing library was prepared according to the manufacturer's instructions and sequenced on an Illumina MiSeq with 2 × 150-bp paired-end reads, according to standard Illumina protocols. The total reads (5,333,189) were mapped to a reference sequence (accession no. KF15501) using the Burrows-Wheeler Aligner (BWA) (version 0.7.5a-r405) (8) and were visualized in Tablet (9). A modified SAMtools/vcfutils (10) script was used to generate an intermediate consensus sequence in which any indels relative to the original reference sequence were appropriately called. The intermediate consensus was used as the reference for subsequent iteration of mapping and consensus calling. The total number of assembled viral reads was 4,750 (0.09% of the total reads). Despite the low proportion of viral sequence detected within the total data set, coverage of the entire genome was obtained (average read depth, 43.7×).

The genetic organization of the Ethiopian wolf RABV genome was consistent with that of other RABV genomes, with a complete genome size of 11,926 nucleotides. The lengths of the individual coding regions were conserved; however, indels were observed in the G-L intergenic region. The Ethiopian wolf RABV genome has the closest homology to the Moroccan RABV genome sequence (accession no. KF155001, 95%), indicating that it belongs to the Africa-1a lineage. Future analysis with domestic dog RABV genomes from Ethiopia will further our understanding of the cross-species transmissions that lead to epizootic cycles observed in the Ethiopian wolf population.

Nucleotide sequence accession number. The complete genomic sequence of RV2985 has been deposited in GenBank under the accession no. KP723638.

ACKNOWLEDGMENTS

This work was supported by the Department for Environment, Food and Rural Affairs, UK (SE0427), and by the EU FP7–funded Research Infrastructure Grant European Virus Archive (no. 19 228292).

The Ethiopian Wolf Conservation Programme [http://www.ethiopianwolf.org] thanks the Ethiopian Willdife Conservation Authority and the Bale Mountains National Park for permission to work in Bale and the Born Free Foundation and the Wildlife Conservation Network for funding.

REFERENCES

1. **Talbi C, Holmes EC, de Benedictis P, Faye O, Nakouné E, Gamatié D, Diarra A, Elmamy BO, Sow A, Adjogoua EV, Sangare O, Dundon WG, Capua I, Sall AA, Bourhy H.** 2009. Evolutionary history and dynamics of dog rabies virus in western and central Africa. J Gen Virol **90:**783–791. http://dx.doi.org/10.1099/vir.0.007765-0.
2. **Johnson N, Mansfield KL, Marston DA, Wilson C, Goddard T, Selden D, Hemson G, Edea L, van Kesteren F, Shiferaw F, Stewart AE, Sillero-Zubiri C, Fooks AR.** 2010. A new outbreak of rabies in rare Ethiopian wolves (*Canis simensis*). Arch Virol **155:**1175–1177. http://dx.doi.org/10.1007/s00705-010-0689-x.
3. **Sillero-Zubiri C, King AA, Macdonald DW.** 1996. Rabies and mortality in Ethiopian wolves (*Canis simensis*). J Wildl Dis **32:**80–86. http://dx.doi.org/10.7589/0090-3558-32.1.80.
4. **Kissi B, Tordo N, Bourhy H.** 1995. Genetic polymorphism in the rabies virus nucleoprotein gene. Virology **209:**526–537. http://dx.doi.org/10.1006/viro.1995.1285.
5. **Marino J, Sillero-Zubiri C, Gottelli D, Johnson PJ, Macdonald DW.** 2013. The fall and rise of Ethiopian wolves: lessons for conservation of long-lived, social predators. Anim Conserv **16:**621–632. http://dx.doi.org/10.1111/acv.12036.
6. **Haydon DT, Randall DA, Matthews L, Knobel DL, Tallents LA, Gravenor MB, Williams SD, Pollinger JP, Cleaveland S, Woolhouse ME, Sillero-Zubiri C, Marino J, Macdonald DW, Laurenson MK.** 2006. Low-coverage vaccination strategies for the conservation of endangered species. Nature **443:**692–695. http://dx.doi.org/10.1038/nature05177.
7. **Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR.** 2013. Next generation sequencing of viral RNA genomes. BMC Genomics **14:**444. http://dx.doi.org/10.1186/1471-2164-14-444.
8. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26:**589–595. http://dx.doi.org/10.1093/bioinformatics/btp698.
9. **Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D.** 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform **14:**193–202. http://dx.doi.org/10.1093/bib/bbs012.
10. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup.** 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25:**2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.

3.3: Article 7 published in Genome Announcements

Complete Genomic Sequence of Six South African Rabies Viruses

DOI:10.1128/genomeA.01085-15

Summary: This genome was written in collaboration with colleagues from South Africa.

The RNA from six RABVs from South Africa and Zimbabwe were sent by Dr. Claude Sabeta as part of an EU- funded Horizon 2020 project to obtain full genome sequences from important virus isolates. Similar to the Ethiopian wolf sequence (Article 6), these sequences are the first full length genomes from South Africa and Zimbabwe and provide important references for these regions. The RNA was depleted and sequenced on the 454 platform. Due to the low number of viral reads, individual consensus sequences for each virus could not be obtained. However, combining the reads from the three related Zimbabwe and three related RSA viruses together resulted in a complete genome consensus sequence from each country. More RNA was requested and sequenced on the Illumina platform. All six genomes were completed and published as first representative canine RABVs circulating in Southern Africa.

genomeA Journals.ASM.org

Complete Genome Sequences of Six South African Rabies Viruses

Claude Sabeta,[a,b] Baby Phahladira,[a] Denise A. Marston,[c] Emma L. Wise,[c] Richard J. Ellis,[c] Anthony R. Fooks[c,d]

Agricultural Research Council-Onderstepoort Veterinary Institute, OIE Rabies Reference Laboratory, Pretoria, Republic of South Africa[a]; University of Pretoria, Veterinary Tropical Diseases Department, Pretoria, Republic of South Africa[b]; Animal and Plant Health Agency, Weybridge, Surrey, United Kingdom[c]; Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, United Kingdom[d]

South African rabies viruses (RABVs) from dogs and jackals (canid viruses) are highly related and most likely originated from a single progenitor. RABV is the cause of most global human rabies cases. The complete genome sequences of 3 RABVs from South Africa and Zimbabwe are reported here.

Received 8 August 2015 **Accepted** 21 August 2015 **Published** 1 October 2015

Citation Sabeta C, Phahladira B, Marston DA, Wise EL, Ellis RJ, Fooks AR. 2015. Complete genome sequences of six South African rabies viruses. Genome Announc 3(5):e01085-15. doi:10.1128/genomeA.01085-15.

Address correspondence to Claude Sabeta, sabetac@arc.agric.za.

Rabies virus (RABV) is the type species of the *Lyssavirus* genus (*Rhabdoviridae* family), which causes a fatal zoonotic disease characterized by encephalomyelitis late in infection. In South Africa, RABV infects many host species. Rabies virus transmission cycles are independently sustained in specific carnivore host species, mainly canine species (*Canidae*), and also mongoose species (principally *Herpestidae*). The domestic dog is the main reservoir and vector species for the transmission of rabies virus in this subregion (similar to other regions in Africa and Asia) (1). Furthermore, in South Africa, in the absence of domestic dogs, RABV (canid biotype) is maintained in wildlife carnivores, such as the black-backed jackal *Canis mesomelas* (2) and bat-eared fox *Otocyon megalotis* (3). The side-striped jackal species, *Canis adustus*, however, cannot maintain RABV independent of disease cycles in dogs (4). Genetic analysis of the rabies viruses from Ƶ mbabwe and South Africa demonstrated a high degree of sequence similarity (≥96.5% sequence identity) between virus isolates originating from dogs and wildlife (jackals), regardless of the enzootic area from which the field specimens were obtained (5, 6). These data emphasize a common progenitor and an ease of exchange of viruses across species barriers. Full-length genome sequences from representative RABVs globally, and in particular from Africa, are essential for understanding these host/viral interactions.

Total viral RNA was extracted from original brain tissues (for South African viruses) and mouse-passaged materials (for Ƶ mbabwean viruses) and prepared for next-generation sequencing (NGS) using the MiSeq platform. Briefly, TRIzol-extracted viral RNA was depleted of host genomic DNA and rRNA, as described previously (7). Double-stranded (ds)-cDNA was synthesized from 50 ng of RNA using a random cDNA synthesis system (Roche), according to the manufacturer's instructions. The ds-cDNA was purified using AMPure XP magnetic beads (Beckman Coulter), and 1 ng was used for the Nextera XT DNA sample preparation kit (Illumina). A sequencing library was prepared according to the manufacturer's instructions and sequenced on an Illumina MiSeq with 2 × 150-bp paired-end reads, according to standard Illumina protocols. The total reads were mapped to a reference sequence (accession no. JX473840) using Burrows-Wheeler Aligner (BWA) (v0.7.5a-r405) and were visualized in Tablet, as described previously (7). A modified SAMtools/vcfutils script was used to generate an intermediate consensus sequence in which any indels relative to the original reference sequence were appropriately called. The intermediate consensus was used as the reference for the subsequent iteration of mapping and consensus calling, as described previously (7). The genetic organization of South African canid viral genomes was consistent with that of other previously characterized lyssaviruses (7, 8). In this analysis, the complete genome sequences of 11,923 nucleotides were obtained, with conserved lengths of individual coding regions of the nucleoprotein (N), matrix protein (M), phosphoprotein (P), and the RNA-dependent polymerase (L). Phylogenetic analysis of the sequences revealed geographical grouping. Continued investigations and future analysis of complete RABV genomes will further the understanding of viral evolution and host interaction.

TABLE 1 Epidemiological information of the canine viruses sequenced in this study

Virus no.	Lab reference no.	Location of origin	Yr of isolation	No. of reads	% viral reads in brain sample	GenBank accession no.
1	20034	Mutare, Ƶ mbabwe	1991	19,349	0.43	KT336433
2	21057	Muzarabani, Ƶ mbabwe	1992	186,102	1.3	KT336434
3	21467	Goromonzi, Ƶ mbabwe	1993	96,501	2.13	KT336435
4	33512	Gauteng, South Africa	2012	121,722	3.67	KT336436
5	34312	Vhembe, South Africa	2012	11,882	0.31	KT336437
6	55612	North West, South Africa	2012	117,358	4.86	KT336432

Nucleotide sequence accession numbers. The complete genome sequences of the six South African RABVs have been deposited in GenBank under the accession numbers listed in Table 1.

ACKNOWLEDGMENTS

This work was supported by the UK Department for Environment, Food and Rural Affairs (SE0427) and by Horizon 2020 - Research & Innovation H2020 - grant agreement (no. 653316) EVAg.

We acknowledge the Central Veterinary Laboratory (Harare) for providing virus isolates to include in this analysis.

REFERENCES

1. **Bingham J, Foggin CM, Wandeler AI, Hill FW**. 1999. The epidemiology of rabies in Ƶ mbabwe. 1. Rabies in dogs (*Canis familiaris*). Onderstepoort J Vet Res **66:**1–10.
2. **Bingham J, Foggin CM, Wandeler AI, Hill FW**. 1999. The epidemiology of rabies in Ƶ mbabwe. 2. Rabies in jackals (*Canis adustus* and *Canis mesomelas*). Onderstepoort J Vet Res **66:**11–23.
3. **Sabeta CT, Mansfield KL, McElhinney LM, Fooks AR, Nel LH**. 2007. Molecular epidemiology of rabies in bat-eared foxes (*Otocyon megalotis*) in South Africa. Virus Res **129:**1–10. http://dx.doi.org/10.1016/j.virusres.2007.04.024.
4. **Rhodes CJ, Atkinson RP, Anderson RM, Macdonald DW**. 1998. Rabies in Ƶ mbabwe: reservoir dogs and the implications for disease control. Phil Trans R Soc Lond B Biol Sci **353:**999–1010. http://dx.doi.org/10.1098/rstb.1998.0263.
5. **Sabeta CT, Bingham J, Nel LH**. 2003. Molecular epidemiology of canid rabies in Zimbabwe and South Africa. Virus Res **91:**203–211. http://dx.doi.org/10.1016/S0168-1702(02)00272-1.
6. **Nel LH, Thomson GR, Von Teichman BF**. 1993. Molecular epidemiology of rabies virus in South Africa. Onderstepoort J Vet Res **60:**301–306.
7. **Marston DA, Wise EL, Ellis RJ, McElhinney LM, Banyard AC, Johnson N, Deressa A, Regassa F, de Lamballerie X, Fooks AR, Sillero-Zubiri C**. 2015. Complete genomic sequence of rabies virus from an Ethiopian wolf. Genome Announc 3(2):e00157-15. http://dx.doi.org/10.1128/genomeA.00157-15.
8. **Kissi B, Tordo N, Bourhy H**. 1995. Genetic polymorphism in the rabies virus nucleoprotein gene. Virology **209:**526–537. http://dx.doi.org/10.1006/viro.1995.1285.

3.4: Article 8 published in Genome Announcements

Complete Genomic Sequence of European Bat Lyssavirus 1, Isolated from *Eptesicus isabellinus* in Spain

DOI:10.1128/genomeA.01518-14

Summary: This paper announces the first complete genome sequence from the EBLV-1 lineage that circulates in *Eptesicus isabellinus*. Partial genome analysis has previously shown that the virus circulating in *Eptesicus isabellinus* is genetically distinct from those circulating in *Eptesicus serotinus*. Complete genomes from the other two EBLV-1 lineages have already been deduced.

The RNA extracted in Spain and sent to APHA in 2008. Our original plan was to use primer walking to obtain complete genome sequence, but due to other priorities the work was not undertaken. Instead the RNA was prepared for Illumina sequencing and a complete genome sequence was obtained directly from the bat brain RNA sample.

genomeA
Journals.ASM.org

Complete Genomic Sequence of European Bat Lyssavirus 1, Isolated from *Eptesicus isabellinus* in Spain

Denise A. Marston,[a] Sonia Vázquez-Morón,[b,g] Richard J. Ellis,[c] Emma L. Wise,[a] Lorraine M. McElhinney,[a,d] Xavier de Lamballerie,[e] Anthony R. Fooks,[a,d,f] Juan E. Echevarría[b,g]

Wildlife Zoonoses & Vector-Borne Diseases Research Group,[a] Specialist Scientific Support Department,[c] Animal and Plant Health Agency (APHA), Surrey, United Kingdom; Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain[b]; National Consortium for Zoonosis Research, University of Liverpool, Liverpool, United Kingdom[d]; UMR_D 190,"Emergence de Pathologies Virales," Aix Marseille University, Marseille, France[e]; Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, United Kingdom[f]; Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain[g]

All members of the lyssavirus genus cause the disease rabies. European bat lyssavirus 1 (EBLV-1) viruses are divided genetically into three groups according to geographic location and host reservoir. We report here the first genome sequence for an EBLV-1 isolated from *Eptesiscus isabellinus* in the Iberian Peninsula, Spain.

Received 15 December 2014 **Accepted** 5 January 2015 **Published** 12 February 2015

Citation Marston DA, Vá uez Morón S, Ellis RJ, Wise EL, McElhinney LM, de Lamballerie X, Fooks AR, Echevarría JE. 2015. Complete genomic sequence of European bat lyssavirus 1, isolated from *Eptesicus isabellinus* in Spain. Genome Announc 3(1):e01518-14. doi:10.1128/genomeA.01518-14.

Address correspondence to Denise A. Marston, denise.marston@apha.gsi.gov.uk.

There are 14 classified lyssavirus species. In Europe, bat rabies cases are dominated by two lyssavirus species: European bat lyssavirus 1 (EBLV-1) and European bat lyssavirus 2 (EBLV-2) (1). EBLV-1 molecular epidemiological investigations separated the species into 2 sublineages (1a and 1b) (2). Analysis of EBLV-1 viruses from *Eptesiscus isabellinus* in the Iberian Peninsula has demonstrated that these viruses are genetically similar, yet distinct from other EBLV-1 circulating in *Eptesiscus serotinus* (3). Full-genome sequences are available for five EBLV-1a and a single EBLV-1b virus. Here, we describe full-genome sequencing of an EBLV-1 virus (292R07, RV2416), isolated from an infected *E. isabellinus* bat from the Iberian peninsula (Granada), Spain, in 2007.

RNA from mouse brain homogenate (passage 1) was prepared for next-generation sequencing on the MiSeq platform. Briefly, TRIzol extracted viral RNA was depleted of host genomic DNA and rRNA as described previously (4). Double-stranded (ds) cDNA was synthesized from 50 ng RNA by use of a random cDNA synthesis system (Roche), according to the manufacturer's instructions. The ds cDNA was purified using Ampure XP magnetic beads (Beckman Coulter), and 1 ng was used for the Nextera XT DNA sample preparation kit (Illumina). A sequencing library was prepared according to the manufacturers' instructions and sequenced on an Illumina MiSeq with 2 × 150-bp paired-end reads following standard Illumina protocols. The total reads (2,032,092) were mapped to a reference sequence (EU293112) using BWA (v0.7.5a-r405) (5) and was visualized in Tablet (6). A modified samtools/vcfutils (7) script was used to generate an intermediate consensus sequence in which any indels relative to the original reference sequence were appropriately called. The intermediate consensus was used as the reference for subsequent iteration of mapping and consensus calling. The total number of assembled viral reads was 5,600 (0.27% of the total reads). Despite the low proportion of viral sequence detected within the total data set, adequate coverage of the entire genome was obtained (average read depth of 58.16), with the exception of the highly conserved complementary untranslated regions (UTRs). The genomic termini sequence was obtained by using primers described previously (4).

The genetic organization of the *E. isabellinus* EBLV-1 genome was similar to those of other EBLV-1 genomes, with a complete genome size of 11,964 nucleotides. The coding region lengths were conserved; however, indels were observed in a number of intergenic regions. A 6-bp indel in the N-P intergenic region (1478 to 1483), which was previously identified within the EBLV-1 species (8), was observed. Interestingly, only the EBLV-1b full genome from *E. serotinus* from France had the 6-bp insertion; all other EBLV-1 genomes, including the *E. isabellinus* EBLV-1, did not. Two further indels were identified, (i) A^{2420} in the P-M region and (ii) C^{5249} in the G-L region. Both nucleotides were absent in the *E. isabellinus* EBLV-1 in comparison to the six EBLV-1 genomes available. It is unlikely that the observed indels have any biological influence; however, they support the independent evolution of the Iberian peninsula EBLV-1 viruses to other EBLV-1 viruses, due to geographical and host separation.

Nucleotide sequence accession number. The complete genomic sequence of RV2416 (292R07) has been deposited in GenBank under the accession number KP241939.

ACKNOWLEDGMENTS

This work was supported by grants from the Department for Environment, Food and Rural Affairs (SE0423 and SE0427) and by the European Union FP7-funded Research Infrastructure Grant European Virus Archive (19 228292).

REFERENCES

1. **McElhinney LM, Marston DA, Leech S, Freuling CM, van der Poel WH, Echevarria J, Vázquez-Moron S, Horton DL, Müller T, Fooks AR.** 2013. Molecular epidemiology of bat lyssaviruses in Europe. Zon oses Public Health **60:**35–45. http://dx.doi.org/10.1111/zph.12003.
2. **Davis PL, Holmes EC, Larrous F, Van der Poel WH, Tjørnehøj K, Alonso WJ, Bourhy H.** 2005. Phylogeography, population dynamics, and molecular evolution of European bat lyssaviruses. J Virol **79:**10487–10497. http://dx.doi.org/10.1128/JVI.79.16.10487-10497.2005.

3. **Vázquez-Moron S, Juste J, Ibáñez C, Berciano JM, Echevarria JE.** 2011. Phylogeny of European bat lyssavirus 1 in *Eptesicus isabellinus* bats, Spain. Emerg Infect Dis **17:**520–523. http://dx.doi.org/10.3201/eid1703.100894.
4. **Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR.** 2013. Next generation sequencing of viral RNA genomes. BMC Genomics **14:**444. http://dx.doi.org/10.1186/1471-2164-14-444.
5. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. BioInformatics **26:**589–595. http://dx.doi.org/10.1093/bioinformatics/btp698.
6. **Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D.** 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform **14:**193–202. http://dx.doi.org/10.1093/bib/bbs012.
7. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup.** 2009. The Sequence Alignment/Map format and SAMtools. BioInformatics **25:**2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.
8. **Johnson N, Freuling C, Marston DA, Tordo N, Fooks AR, Müller T.** 2007. Identification of European bat lyssavirus isolates with short genomic insertions. Virus Res **128:**140–143. http://dx.doi.org/10.1016/ j.virusres.2007.04.012.

3.5: Article 9 published in Genome Announcements

First Complete Genomic Sequence of a Rabies Virus from the Republic of Tajikistan obtained directly from an FTA card.

DOI:10.1128/genomeA.00515-17

Summary: Transporting infectious viruses from many counties is problematic. As an alternative we use Flinders Technologies Association (FTA) cards to safely transport viral nucleic acid from country of origin to APHA for testing. FTA cards are filter papers impregnated with chemicals that inactivate the virus and preserve the nucleic acid. Here it is shown that RABV positive brain homogenates can successfully be stored on FTA cards, shipped from Tajikistan to APHA and the nucleic acid be eluted from the FTA card. gDNA was removed, and the sample sequenced on the MiSeq platform. A complete genome sequence was obtained.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

VIRUSES

First Complete Genomic Sequence of a Rabies Virus from the Republic of Tajikistan Obtained Directly from a Flinders Technology Associates Card

H. Goharriz,[a] D. A. Marston,[a] F. Sharifzoda, R. J. Ellis,[b] D. L. Horton,[e] T. Khakimov,[d] A. Whatmore,[c] K. Khamroev,[d] A. N. Makhmadshoev,[d] M. Bazarov,[d] A. R. Fooks,[a,f] A. C. Banyard[a]

Wildlife Zoonoses & Vector-Borne Diseases Research Group, Animal and Plant Health Agency, Surrey, United Kingdom[a]; Specialist Scientific Support Department, Animal and Plant Health Agency, Surrey, United Kingdom[b]; Bacterial Characterization Team, Animal and Plant Health Agency, Surrey, United Kingdom[c]; National Center for Veterinary Diagnostics, Dushanbe, Republic of Tajikistan[d]; School of Veterinary Medicine, University of Surrey, Guildford, United Kingdom[e]; Institute for Infection and Immunity, St. George's Hospital Medical School, University of London, London, United Kingdom[f]

ABSTRACT A brain homogenate derived from a rabid dog in the district of Tojikobod, Republic of Tajikistan, was applied to a Flinders Technology Associates (FTA) card. A full-genome sequence of rabies virus (RABV) was generated from the FTA card directly without extraction, demonstrating the utility of these cards for readily obtaining genetic data.

Received 24 April 2017 **Accepted** 9 May 2017 **Published** 6 July 2017

Citation Goharriz H, Marston DA, Sharifzoda F, Ellis RJ, Horton DL, Khakimov T, Whatmore A, Khamroev K, Makhmadshoev AN, Bazarov M, Fooks AR, Banyard AC. 2017. First complete genomic sequence of a rabies virus from the Republic of Tajikistan obtained directly from a Flinders Technology Associates card. Genome Announc 5:e00515-17. https://doi.org/10.1128/genomeA.00515-17.

Address correspondence to A. C. Banyard, Ashley.Banyard@apha.gsi.gov.uk.

Rabies is one of the most neglected zoonotic diseases, causing an estimated 60,000 human deaths annually (1). Accurate diagnostics, as with many viral diseases, relies on sample collection and reliable transportation, which can require maintenance of a cold chain to preserve the integrity of the sample for diagnostic evaluation. In rabies, areas endemic for the disease have been identified as a bottleneck to diagnosis. The use of Flinders Technology Associates (FTA) filter paper to preserve nucleic acids in samples has proven invaluable under field conditions for numerous pathogens, including lyssaviruses (2, 3). However, extraction of viral nucleic acid was performed before downstream applications, such as molecular diagnostics or phylogenetic analysis.

One hundred microliters of a clarified brain homogenate that tested positive at the National Centre for Veterinary Diagnostics (NCVD), Dushanbe, Tajikistan, by fluorescent antibody test (FAT) was placed onto an FTA card and allowed to air dry before transportation to the Animal and Plant Health Agency (APHA) for molecular characterization.

RNA was eluted from a 2-mm punch obtained from the FTA card in 50 μl of molecular-grade water on ice, with periodic agitation. The eluted nucleic acid was depleted of host genomic DNA (4), and double-stranded (ds) cDNA was synthesized using a random cDNA synthesis system (Roche). The ds cDNA was purified using AMPure XP magnetic beads (Beckman Coulter, Inc.) and 1 ng used for the Nextera XT DNA sample preparation kit (Illumina). A sequencing library was prepared and sequenced on an Illumina MiSeq with 2 $\times$ 150 bp paired-end reads according to standard Illumina protocols. The total reads (6,540,291) were mapped to a reference sequence (accession no. JQ944705) using BWA (version 0.7.5a-r405) (5) and was visualized in Tablet (6). A modified SAMtools/vcfutils (7) script was used to generate an intermediate consensus sequence in which any indels relative to the original reference sequence were appropriately called. The intermediate consensus was used as the reference for subsequent iteration of mapping and consensus calling. The total number of assembled

viral reads was 87,971 (1.35% of the total reads). Despite the low proportion of viral sequence detected within the total data set, coverage of the entire genome was obtained (average read depth, 926.86×). The complete genome sequence was 11,922 nucleotides and most closely related to accesion no. LN879480.1, an RABV genome sequence from Republic of Azerbaijan. The present sequence represents the first full-genome sequence available for an RABV from the Republic of Tajikistan.

Described here is a simple methodology to successfully obtain a complete viral genome sequence from an FTA card without RNA extraction. The application of this methodology to other viruses could significantly aid the use of FTA cards to obtain viral nucleic acids without extraction for multiple uses, including molecular diagnostics and full-genome sequencing.

Accession number(s). The sequence data have been deposited in GenBank under accession number KY765901.

ACKNOWLEDGMENTS

This project is funded by Ministry of Defence and DEFRA through EXGOG1308 and SE0431 grants and technical support from European Virus Archive Global (EVAg).

REFERENCES

1. Fooks AR, Banyard AC, Horton DL, Johnson N, McElhinney LM, Jackson AC. 2014. Current status of rabies and prospects for elimination. Lancet 384:1389–1399. https://doi.org/10.1016/S0140-6736(13)62707-5.
2. Picard-Meyer E, Barrat J, Cliquet F. 2007. Use of filter paper (FTA) technology for sampling, recovery and molecular characterisation of rabies viruses. J Virol Methods 140:174–182. https://doi.org/10.1016/j.jviromet.2006.11.011.
3. Montmayeur AM, Ng TF, Schmidt A, Zhao K, Magaña L, Iber J, Castro CJ, Chen Q, Henderson E, Ramos E, Shaw J, Tatusov RL, Dybdahl-Sissoko N, Endegue-Zanga MC, Adeniji JA, Oberste MS, Burns CC. 2017. High-throughput next-generation sequencing of polioviruses. J Clin Microbiol 55:606–615. https://doi.org/10.1128/JCM.02121-16.
4. Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR. 2013. Next generation sequencing of viral RNA genomes. BMC Genomics 14:444. https://doi.org/10.1186/1471-2164-14-444.
5. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.
6. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D. 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform 14:193–202. https://doi.org/10.1093/bib/bbs012.
7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/MAP format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

3.6: Article 10: published in BMC Veterinary Research

First detection of European bat lyssavirus type 2 (EBLV-2) in Norway

DOI:10.1186/s12917-017-1135-z

Summary: In Europe, bat rabies is primarily attributed to European bat lyssavirus type 1 (EBLV-1) and European bat lyssavirus type 2 (EBLV-2), both have a restricted host reservoir. Approximately thirty cases of infection with EBLV-2 in Daubenton's bats (Myotis daubentonii) and pond bats (M. dasycneme) have been reported, but none in Norway. In October 2015, EBLV-2 was detected in a Daubenton's bat. Viral RNA was detected in brain tissues by RT-PCR. Surprisingly, fluorescent antibody test (FAT) performed in the EU Reference laboratory in France, on brain impressions was negative. RNA was sent to APHA and a full genome sequence was obtained.

Moldal et al. BMC Veterinary Research (2017) 13:216
DOI 10.1186/s12917-017-1135-z

BMC Veterinary Research

CASE REPORT Open Access

First detection of *European bat lyssavirus* type 2 (EBLV-2) in Norway

Torfinn Moldal[1*], Turid Vikøren[1], Florence Cliquet[2], Denise A. Marston[3], Jeroen van der Kooij[4], Knut Madslien[1] and Irene Ørpetveit[1]

Abstract

Background: In Europe, bat rabies is primarily attributed to *European bat lyssavirus* type 1 (EBLV-1) and *European bat lyssavirus* type 2 (EBLV-2) which are both strongly host-specific. Approximately thirty cases of infection with EBLV-2 in Daubenton's bats (*Myotis daubentonii*) and pond bats (*M. dasycneme*) have been reported. Two human cases of rabies caused by EBLV-2 have also been confirmed during the last thirty years, while natural spill-over to other non-flying mammals has never been reported. Rabies has never been diagnosed in mainland Norway previously.

Case presentation: In late September 2015, a subadult male Daubenton's bat was found in a poor condition 800 m above sea level in the southern part of Norway. The bat was brought to the national Bat Care Centre where it eventually displayed signs of neurological disease and died after two days. EBLV-2 was detected in brain tissues by polymerase chain reaction (PCR) followed by sequencing of a part of the nucleoprotein gene, and lyssavirus was isolated in neuroblastoma cells.

Conclusions: The detection of EBLV-2 in a bat in Norway broadens the knowledge on the occurrence of this zoonotic agent. Since Norway is considered free of rabies, adequate information to the general public regarding the possibility of human cases of bat-associated rabies should be given. No extensive surveillance of lyssavirus infections in bats has been conducted in the country, and a passive surveillance network to assess rabies prevalence and bat epidemiology is highly desired.

Keywords: Rabies, Daubenton's bat (Myotis daubentonii), European bat lyssavirus type 2 (EBLV-2), fluorescent antibody test (FAT), polymerase chain reaction (PCR), rabies tissue culture infection test (RTCIT)

Background

Rabies is a fatal zoonotic neurological disease caused by RNA viruses belonging to the genus *Lyssavirus* in the family *Rhabdoviridae*. *Rabies virus* (RABV) is the prototype of the genus and causes approximately 59,000 deaths in humans yearly [1]. Red fox (*Vulpes vulpes*) is the main reservoir for RABV in Europe, where the disease has been eliminated in many member states of the European Union due to oral vaccination programs [2]. Rabies in non-flying mammals has never been diagnosed in mainland Norway [3]. However, in the Svalbard archipelago, which is under Norwegian jurisdiction, several occasional detections and two rabies outbreaks during the last 35 years have been reported and are related to arctic strains of RABV in Arctic fox (*Vulpes lagopus*), Svalbard reindeer (*Rangifer tarandus platyrhyncus*) and one ringed seal (*Pusa hispida*) [4–6].

The genus Lyssavirus consists of 14 recognized species [7], and all but two have been isolated from bats [8]. The different species display distinct features regarding geographical distribution and host specificity. In Europe, bat rabies is primarily attributed to *European bat lyssavirus* type 1 (EBLV-1) and *European bat lyssavirus* type 2 (EBLV-2) [9]. Most reported bat rabies cases are caused by EBLV-1 which is almost exclusively found in serotine bats (*Eptesicus serotinus*) and Isabelline serotine bats (*E. isabellinus*), while EBLV-2 is detected mainly in Daubenton's bats (*Myotis daubentonii*) and to a lesser extent in pond bats (*M. dasycneme*) [9]. Evolutionary studies suggest that EBLV-2 diverged from other lyssaviruses more than 8000 years ago and that the current

* Correspondence: torfinn.moldal@vetinst.no
[1]Norwegian Veterinary Institute, Postbox 750, Sentrum, 0106 Oslo, Norway
Full list of author information is available at the end of the article

diversity of EBLV-2 has built up during the last 2000 years [10].

Hitherto, EBLV-2 has been detected in bats in the United Kingdom, the Netherlands, Germany, Switzerland, Finland and Denmark [11–16], while antibodies against EBLV-2 have been found in bats in Sweden [17]. Two cases of human rabies associated with EBLV-2 have been reported in Finland [18] and in Scotland [19]. To date, no cases of natural spill-over to other non-flying mammals have been reported, but EBLV-2 has successfully been transferred to sheep and foxes in experimental studies, even though the susceptibility seems to be low in these species [20, 21].

Among the 12 species of bats regularly reported in Norway, the northern bat *(E. nilssonii*), the soprano pipistrelle (*Pipistrellus pygmaeus*) and the Daubenton's bat are the most common and widespread species [22–24]. The serotine bat has been recorded once and is not considered as a resident species [25]. The range of the Daubenton's bat reaches to 63° N in Norway [24]. The results of ringing experiments in other countries suggest that this species does not migrate over long distances [26], which is likely also the case in Norway. The Daubenton's bat is specialized in hunting insects and spiders, which it mainly catches with its feet from the water surface [27], but it is also found searching for food in forests [24]. Daubenton's bats seldom occur above the timberline and roost mainly in tree cavities, but also use crevices in bridges and cliffs. Roosts in buildings are very rare [24, 28] and so are contacts with the general public [29].

There is no official surveillance program for lyssavirus in bats in Norway. According to the annals of the Norwegian Veterinary Institute (NVI), a total of 27 bats (swabs from 18 bats sampled alive and brain tissues from nine carcasses) were examined for lyssavirus in the period 1998–2015. Here, we report the first detection of EBLV-2 in Norway.

Case presentation

Clinical signs and treatment

On September 29th 2015, a bat was discovered close to a cabin in Valdres in the county of Oppland. The locality is situated just below the timberline at around 800 m above sea level at 60° 59′ N in the southern part of Norway. The landscape is undulating with open spaces and mixed forest dominated by Norway spruce (*Picea abies*) and mountain birch (*Betula pubescens ssp. czerepanovii*), lakes and small meadows. The cabin owner discovered the animal clinging to a stone when she removed a tarpaulin. Three days later, the bat was still there, and the cabin owner called the National Bat Helpline, which advised her to bring the bat to the national Bat Care Centre of the Norwegian Zoological Society (NZI). The centre is approved and partly funded by the Norwegian Environmental Agency (NEA) and the Norwegian Food Safety Authority (NFSA) [29].

The bat arrived at the Bat Care Centre the next day, on October 3rd 2016, and was identified as a male Daubenton's bat based on external characteristics, i.e. short tragus, large feet and the attachment of the wing membrane at the middle of the foot, and as a subadult individual based on the brownish face and the presence of a chin-spot (Fig. 1) [30]. The bat weighed 6.6 g on arrival and had a forearm length of 38.3 mm, hence the body condition index (mass g/forearm mm) was 0.17, which indicated low energy reserves [31]. It was weak, and the tongue was dark red and swollen. It had dark plaque on its teeth, and there was a heavy load of mites of the genus *Spinturnix* on the wing membrane surface. As a safety precaution, the diseased bat was routinely placed in a clean plastic container (35x45x40 cm) separated from other bats in the facility.

The bat drank and ate mealworm (*Tenebrio molitor*) on arrival. It walked around inside the container and on one occasion lunged against the bat carer. After a few hours the bat grew stronger, groomed itself and produced some excrement. The following day, on October 4th 2016, it gradually refused to eat, displayed difficulty when swallowing water and got weaker. The bat was force-fed several times, received fluid subcutaneously and eventually had increased locomotor activity. In the late evening it displayed mild ataxia. The following morning, the bat's condition deteriorated. The bat walked around, but its locomotion was less agile and it was not able to groom itself. The bat became even weaker towards the evening in spite of treatment with

Fig. 1 The bat was identified as a subadult male Daubenton's bat based on external characteristics. Photographer: Jeroen van der Kooij

antibiotics and fluid, and when force-fed with water or fluids, it showed signs of vomiting and spread both wings and legs. It also tried to bite, but without force. The bat died in the evening of October 5th 2016 and was subsequently frozen.

Necropsy, sampling and histology

On October 6th 2015, the bat was submitted to the NVI in Oslo for laboratory examinations on suspicion of rabies. The necropsy revealed that the bat was in a poor condition with no visible body fat. The tongue was dark red and dry. The internal organs were congested with no specific gross findings.

Brain impressions on glass slides were fixed in acetone baths for fluorescent antibody test (FAT), and tissues from different parts of the brain were put on empty tubes and tubes with lysis buffer for rabies tissue culture infection test (RTCIT) and detection of viral RNA by reverse transcription and polymerase chain reaction (RT-PCR), respectively. The remains of the brain and the head including salivary glands, lung, heart, liver, kidney, spleen, gastrointestinal tract and pancreas were fixed in 10% buffered formalin for two weeks, routinely processed and embedded in paraffin before cutting ultrathin sections that were mounted on glass slides and stained with haematoxylin-eosin for histological examination.

Only small pieces of brain tissue were available for histology, and no lesions were detected. In the lungs, multifocal areas of foreign material aspiration with infiltration of inflammatory cells were seen, and a similar small focal inflammation was found in the salivary gland tissue. A few parasitic structures without any associated inflammation were detected in the liver.

FAT and RTCIT

The laboratory tests used for rabies diagnosis are recommended by WHO and OIE. FAT, in which virus antigens in brain tissues are detected, is the gold standard for diagnosing rabies [32]. Fixed brain impressions were stained with FITC Anti-Rabies Monoclonal Globulin (Fujirebio Diagnostics Inc.) according to the manufacturer's instructions. A brain impression from a rabid cow served as a positive control, while a brain impression from a healthy fox served as a negative control. All slides were investigated in a fluorescence microscope (Leitz) under ultraviolet light. Virus antigen was detected in the positive control, but not in the negative control or any of the brain impressions from the bat.

The detection of infectious particles in brain homogenates was performed by RTCIT on neuroblastoma cells with an incubation period of 48 h at 37 °C as previously described [33]. The staining was performed using a rabies anti-nucleocapsid FITC-conjugated antibody (BioRad). The brain tissues from the bat were found positive for the presence of infectious virus by RTCIT.

Nucleic acid extraction, RT-PCR and partial sequencing of the nucleoprotein gene

Total nucleic acids were extracted using the NucliSens® easyMAG™ (bioMerieux Inc.) according to the manufacturer's instructions for the off-board protocol. Nucleic acids were eluted in 55 μl buffer. Primers targeting the nucleoprotein gene were applied in a two-step RT-PCR. RT was performed with Invitrogen SuperScript® III Reverse Transcriptase (Thermo Fisher Scientific) with the primer LYSSA-NMA (5′-ATGTAACACCYCTACAATG-‘3) that is modified from primer JW12 in [34].

PCR was performed with Qiagen HotStarTaq® DNA Polymerase (Qiagen) according to the manufacturer's instructions. The forward primer, LYSSA-NMB (5′-ATG TAACACCYCTACAATGGA-‘3) was used in two separate PCR reactions with either LYSSA-NGA (5’-TGACTC CAGTTRGCRCACAT-‘3) or LYSSA-NGC (5’-GGGTAC TTGTACTCATAYTGRTC-‘3) as reverse primers, yielding amplicons of 612 bp and 108 bp respectively. The primers are modified from the primers JW12, JW6 and SB1 in [34] respectively. Amplicons were separated on 1% agarose gel at 90 V for 90 min and visualized by DNA Gel Loading Dye (Thermo Fisher Scientific).

The PCR products yielding bands of the expected size in the agarose gel electrophoresis were further analysed by DNA sequencing. Following enzymatic PCR clean up with illustra™ ExoStar™ (GE Healthcare Life Sciences), DNA sequencing was performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The products were analysed in an ABI PRISM® 3100 Genetic Analyzer (Thermo Fisher Scientific) according to the manufacturer's instructions and with Sequencher® version 5.3 sequence analysis software (Gene Codes Corporation).

A sequence with 567 nucleotides (GenBank accession number KX644889) was obtained with the primers LYSSA-NMB and LYSSA-NGA. BLAST search against the GenBank database showed 94–96% identity with EBLV-2 (Table 1). The sequence was aligned with sequences for EBLV-2 published in GenBank to evaluate the genetic diversity and the relationship to other bat-associated lyssaviruses, and phylogenetic trees were constructed with MEGA 6 based on an alignment of 391 nucleotides applying the Maximum Likelihood and Neighbor-Joining algorithms with 1000 bootstrap replicates and different substitution models (Fig. 2) [35]. The trees showed similar results, and although the node is not strongly supported, the results indicate that the sequence from the Norwegian virus did not cluster closely with any other published sequences of EBLV-2.

Table 1 The site and year of collection, host species and GenBank accession number for sequences for a part of the nucleoprotein gene of *European bat lyssavirus* 2 used to generate a phylogenetic tree

Place	Country	Year	Host species	GenBank accession number	Reference(s)
Helsinki	Finland	1985	*Homo sapiens*	AY062091	[18, 61]
Wommels	Netherlands	1986	*Myotis dasycneme*	U22847	[12, 62]
Tjerkwerd	Netherlands	1987	*Myotis dasycneme*	U89480	[12, 62]
Andijk	Netherlands	1989	*Myotis dasycneme*	U89481	[12, 62]
Plaffeien	Switzerland	1992	*Myotis daubentonii*	AY212117	[14, 46]
Roden	Netherlands	1993	*Myotis dasycneme*	U89482	[12, 62]
Versoix	Switzerland	1993	*Myotis daubentonii*	U89479	[62]
Sussex	United Kingdom	1996	*Myotis daubentonii*	U89478	[11, 62]
Lancashire	United Kingdom	2002	*Myotis daubentonii*	AY212120	[46]
Angus	United Kingdom	2002	*Homo sapiens*	AY247650	[19]
Geneva	Switzerland	2002	*Myotis daubentonii*	AY863408	[63]
Surrey	United Kingdom	2004	*Myotis daubentonii*	JQ796807	[9, 47]
Lancashire	United Kingdom	2004	*Myotis daubentonii*	JQ796808	[9, 64]
Oxfordshire	United Kingdom	2006	*Myotis daubentonii*	JQ796809	[9, 48]
Magdeburg	Germany	2006	*Myotis daubentonii*	JQ796805	[9, 50]
Schwansee	Germany	2006	*Myotis daubentonii*	KF826115	[65]
Bad Buchau	Germany	2007	*Myotis daubentonii*	GU227648	[13]
Shropshire	United Kingdom	2007	*Myotis daubentonii*	JQ796810	[9, 49]
Surrey	United Kingdom	2008	*Myotis daubentonii*	JQ796811	[9]
Shropshire	United Kingdom	2008	*Myotis daubentonii*	JQ796812	[9, 66]
West Lothian	United Kingdom	2009	*Myotis daubentonii*	JQ796806	[9, 67]
Turku	Finland	2009	*Myotis daubentonii*	GU002399	[15]
Gießen	Germany	2013	*Myotis daubentonii*	KF826149	[65]
Valdres	Norway	2015	*Myotis daubentonii*	KX644889	This study

Whole genome sequencing

Total RNA was depleted of host genomic DNA (gDNA) and ribosomal RNA (rRNA) following methods described previously [36, 37]. Briefly, gDNA was depleted using the on-column DNase digestion protocol in RNeasy plus mini kit (Qiagen) following manufacturer's instructions, eluting in 30 μl molecular grade water. Subsequently, rRNA was depleted, using Terminator 5′-phosphate-dependent exonuclease (Epicentre Biotechnologies). Briefly, 30 μl of gDNA depleted RNA was mixed with 3 μl of Buffer A, 0.5 μl of RNAsin Ribonuclease inhibitor (20–40 U/μl) and incubated at 30 °C for 60 min. The depleted RNA was purified to remove the enzyme using the RNeasy plus mini kit as above, without the DNase digestion, eluting in 30 μl of molecular grade water. Double stranded cDNA (ds-cDNA) was synthesised using random hexamers, and a cDNA synthesis kit (Roche) following manufacturer's instructions. The resulting ds-cDNA was purified using AMPure XP magnetic beads (Beckman Coulter), quantified using Quantifluor (Promega) and approximately 1 ng of the ds-cDNA library was used in a 'tagmentation' reaction mix using a Nextera XT DNA sample preparation kit (Illumina) following manufacturer's instructions – without the bead normalization step. The DNA library was quantified using Quantifluor (Promega) and sequenced as 2 × 150 bp paired-end reads on an Illumina MiSeq platform.

Short reads were mapped to the most genetically related full length EBLV-2 genome available (Germany 2012 – GenBank accession number KY688149). Reads were mapped using the Burrow-Wheeler Aligner (BWA version 0.7.5a–r405) [38] and visualized in Tablet [39]. A modified SAMtools/vcfutils [40] script was used to generate an intermediate consensus sequence in which any indels and SNPs relative to the original reference sequence were appropriately called. The intermediate consensus sequence was used as the reference for four subsequent iterations of mapping and consensus calling, described previously [41].

A total of 38,288 EBLV-2 specific reads were mapped from 5,839,126 total reads (0.66%). The average depth of coverage was 373.6 with a maximum depth of coverage of 1900. Sequencing resulted in complete genomic

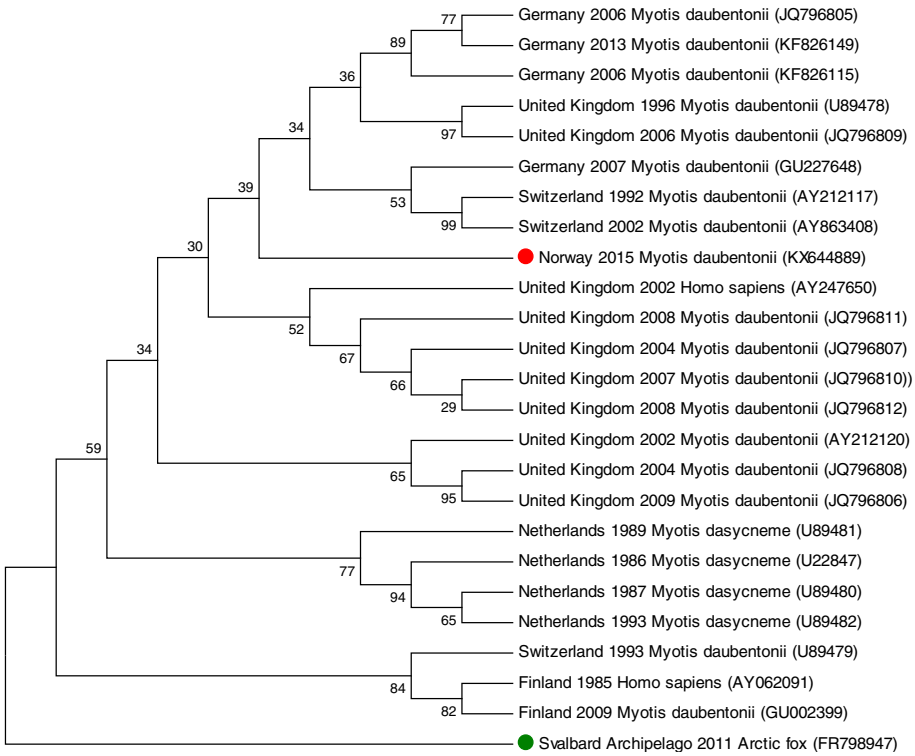


Fig. 2 Phylogenetic relationship between the Norwegian EBLV-2 sequence (●) and other EBLV-2 sequences based on 391 nucleotides from the nucleoprotein gene. The tree was constructed with the Maximum Likelihood algorithm with 1000 bootstrap replicates and the Tamura-Nei substitution model. The number at each branch of the phylogenetic tree represents the likelihood in percentage that the sequences cluster together. RABV from an Arctic fox is used as outgroup (●). The country and year of collection, host species and GenBank accession number for sequences are given

coverage, apart from the first 9 nucleotides. However, the lyssavirus genomic ends are highly conserved between species and palindromic [42, 43]. Therefore the first 9 nucleotides of the 3′ UTR sequence was deduced using the 5′ end sequence and matched 100% with other EBLV-2 genomes available. The total genomic length is 11,928 nucleotides (Table 2), conforming to other EBLV-2 genome sequences (11,924–30), where all regions are identical in length except for the non-coding regions of M-G and G-L. The complete genome sequence was submitted to GenBank (accession number KY688154).

Discussion and conclusions

Here, we report the first detection of EBLV-2 in Norway. The detection of EBLV-2 in a Daubenton's bat is in accordance with previous studies that have revealed that EBLV-2 is mainly found in this species [9]. As the bat was a subadult and ringing experiments suggest that this species does not migrate over long distances [26], it can be assumed that the bat was native to the area where it was found and that EBLV-2 is present in Norway. The latter is supported by the phylogenetic analyses, which indicate that the Norwegian isolate differs from other published isolates, further suggesting that the Norwegian isolate may have evolved separately from a common European ancestor [10]. Very few bats have been examined for rabies in Norway, and we have no knowledge of the prevalence and epidemiology of EBLV. In neighbouring countries,

Table 2 Length of coding (bold) and non-coding regions of EBLV-2 in nucleotides (nts)

Region	Length (nts)
3′ UTR	70
N protein	1356
N-P	101
P protein	894
P-M	88
M protein	609
M-G	210
G protein	1575
G-L	510
L protein	6384
5′ UTR	131
Total	**11,928**

EBLV-2 was detected in diseased Daubenton's bats in Finland in 2009 and 2016 [15, 44], while active surveillance has revealed viral RNA from mouth swabs from Daubenton's bat in Denmark and seropositive Daubenton's bats in both Sweden and Finland [16, 17, 45].

The FAT performed on brain impressions at the NVI was negative, and unfortunately, due to insufficient material, it could not be repeated at the Nancy OIE/WHO/EU Laboratory for Rabies and Wildlife (EURL). However, the EURL confirmed the presence of lyssavirus in the RTCIT. The negative result for FAT does not concur with findings in other Daubenton's bats naturally infected with EBLV-2 [11, 13–15, 44, 46–51]. The annual proficiency tests for rabies diagnosis techniques organized by the EURL have demonstrated the difficulty for laboratories to reliably detect EBLV strains when using the FAT [52], with results depending on the rabies virus antibody conjugate and even the batch used [53]. However, the same batch used in the FAT for the present case has successfully been applied at the NVI for detection of this virus species in two proficiency tests (unpublished results). A polyclonal antibody was used to confirm the presence of lyssavirus in the RTCIT, and a difference between the antibodies regarding the ability to detect the current virus cannot be excluded.

The aspiration pneumonia might have been caused by a dysfunctional swallow reflex or as a result of the force-feeding. The bat had a low state of nutrition, displayed several signs of weakness and, in the end, an inability to fly. This could be caused by the EBLV-2 infection or due to general weakness as a consequence of either shortage of feed at the high altitude or inexperience in hunting since being a subadult. Also, a poor body condition could have made the bat more susceptible for EBLV-2 infection. Weight loss, the inability to fly and death within 14 days followed by detection of viral antigen in brain tissue is reported in Daubenton's bats after intracerebral inoculation with EBLV-2 [54]. In that study the bats were fed ad libitum, so the weight loss could not be related to a decrease in hunting success. During a natural EBLV-2 infection, the course of the disease is probably longer than that found in the referred experiment, and thus the ability to hunt and eat could deteriorate over some time span, resulting in gradual loss of body condition. Poor body condition in bats can also be influenced by heavy parasite loads [55]. Other possible signs of the disease, like the inability to fly, were registered only up to 48 h before death in the experiment [54]. Most of the bats which were inoculated either intramuscularly or subdermally survived the study period of 123 days [54]. These findings highlight that there is a high intra-species barrier in transmission of EBLV-2 and that the incubation period under natural conditions is probably several weeks to months.

Two human cases of rabies caused by EBLV-2 have been reported, in Finland in 1985 and in Scotland in 2002 [18, 19]. Both persons were in close contact with bats over time and had been bitten several times without any history of immunization prior to exposure or post-exposure prophylaxis. Spill-over to other non-flying mammals under natural conditions has not been reported, and the detection of EBLV-2 does not pose an immediate risk to humans as Daubenton's bats are rarely in contact with the public [29]. However, the possibility of human cases of bat-associated rabies acquired in Norway, which is considered free of rabies, cannot be neglected; hence adequate information should be given to the general public as well as to people and professionals who come into contact with bats.

All bat species in Norway are strictly protected under both national laws and international commitments to preserve bats [56, 57]. Culling bats for lyssavirus testing is therefore not a feasible strategy. Passive surveillance by testing bats found sick or dead is the most appropriate way of assessing the incidence of rabies as compared with active surveillance by testing swabs and/or sera from live bats from natural populations [58, 59]. Passive surveillance of bat rabies is therefore strongly recommended by international organizations [58, 60].

Monitoring of EBLV in bats has so far not been prioritized by the Norwegian authorities, probably due to few human cases of bat-associated rabies being reported in Europe. According to the annals of the NVI, only seven carcasses of bats including the EBLV-2 positive case were examined during the last three years. In 2014 and 2015, the NVI in collaboration with the NZS collected and tested nasopharyngeal and cloacal swabs from totally 16 free-ranging Norwegian bats for lyssaviruses with a negative result (unpublished data). The experience gained in this pilot study and from surveillance in other countries is important for future monitoring of bat rabies in Norway.

The authors believe that testing bats found sick or dead should get priority through a national rabies surveillance network in line with the recent recommendations of EUROBATS and EFSA [58, 60]. Additionally, an active surveillance program, over a limited number of years and in defined areas to obtain baseline data on the prevalence of bat rabies in Norway, is also recommended. Finally, we emphasize the importance of good cooperation between bat biologists and laboratory workers to ensure access to representative samples of good quality.

Abbreviations

EBLV-1: *European bat lyssavirus* type 1; EBLV-2: *European bat lyssavirus* type 2; EURL: Nancy OIE/WHO/EU Laboratory for Rabies and Wildlife; FAT: Fluorescent antibody test; gDNA: Genomic DNA; NEA: Norwegian Environmental Agency; NFSA: Norwegian Food Safety Authority;

NVI: Norwegian Veterinary Institute; NZI: Norwegian Zoological Society; PCR: Polymerase chain reaction; RABV: *Rabies virus*; rRNA: Ribosomal RNA; RT: Reverse transcription; RTCIT: Rabies tissue culture infection test

Acknowledgements
The authors want to thank the cabin owner who found the bat and brought it to the Bat Care Centre and Anke Kirkeby for practical care and additional information on the behaviour of the bat. We are grateful to Marianne Heum, Kristin Stangeland Soetaert, Lene Hermansen, Faisal Suhel, Rosa Ferreira Fristad, Britt Gjerset and Hilde Sindre at the NVI, Alexandre Servat and Evelyne Picard-Meyer at the EURL and Richard J. Ellis at Animal and Plant Health Agency.

Funding
The Bat Help Line and the Bat Care Centre receive financial funding from the NEA. The necropsy, histological examination, FAT, RT-PCR and the partial sequencing of the nucleoprotein gene were funded by the NVI, while the RTCIT and the whole genome sequencing were funded by the EURL and the European Union H2020-grant 'European Virus Archive Global (EVAg)' (H2020-grant agreement number 653316) respectively.

Availability of data and materials
The sequences for a part of the nucleoprotein gene and the complete genome are submitted to GenBank (https://www.ncbi.nlm.nih.gov/genbank) and assigned accession numbers KX644889 and KY688154 respectively, while the formalin-fixed, paraffin embedded tissues and propagated virus are stored at the NVI and the EURL respectively.

Authors' contributions
TM was responsible for the RT-PCR and the partial sequencing of the nucleoprotein gene and coordinated the drafting of the manuscript, TV performed the necropsy, sampling and histological examination, FC was responsible for the RTCIT, DM was responsible for the whole genome sequencing, JvdK had the primary contact with the cabin owner and was responsible for the handling of the bat at the Bat Care Centre, KM coordinated the submission of the bat for necropsy, and IØ supervised the RT-PCR and the partial sequencing of the nucleoprotein gene. All authors took part in writing the manuscript and have approved the final version.

Ethics approval
The bat was handled according to standard procedures at the NZS's Bat Care Centre that is approved by the NEA and the NFSA.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Norwegian Veterinary Institute, Postbox 750, Sentrum, 0106 Oslo, Norway. [2]Nancy OIE/WHO/EU Laboratory for Rabies and Wildlife, French Agency for Food, Environmental and Occupational Health & Safety, CS 40009, 54220 Malzéville, France. [3]Animal and Plant Health Agency, New Haw, Addlestone, Surrey KT15 3NB, UK. [4]Norwegian Zoological Society's Bat Care Centre, Rudsteinveien 67, 1480 Slattum, Norway.

Received: 11 December 2016 Accepted: 28 June 2017
Published online: 11 July 2017

References

1. Hampson K, Coudeville L, Lembo T, Sambo M, Kieffer A, Attlan M, et al. Estimating the global burden of endemic canine rabies. PLoS Negl Trop Dis. 2015; doi:10.1371/journal.pntd.0003709.
2. Müller T, Freuling CM, Wysocki P, Roumiantzeff M, Freney J, Mettenleiter TC, et al. Terrestrial rabies control in the European Union: historical achievements and challenges ahead. Vet J. 2015;203:10–7.
3. Heier BT, Lange H, Hauge K, Hofshagen M. Norway 2014 – Trends and source of zoonoses and zoonotic agents in humans foodstuffs, animals and feedingstuffs. Oslo: Norwegian Veterinary Institute; 2015.
4. Ødegaard ØA, Krogsrud J. Rabies in Svalbard: infection diagnosed in arctic fox, reindeer and seal. Vet Rec. 1981;109:141–2.
5. Ørpetveit I, Ytrehus B, Vikøren T, Handeland K, Mjøs A, Nissen S, et al. Rabies in an Arctic fox on the Svalbard archipelago, Norway, January 2011. Euro Surveill. 2011;16:2–3.
6. Mørk T, Bohlin J, Fuglei E, Åsbakk K, Tryland M. Rabies in the arctic fox population, Svalbard. Norway J Wildl Dis. 2011;47:945–57.
7. International Committee on Taxonomy of Viruses (ICTV). Virus Taxonomy: 2015 Release. https://talk.ictvonline.org/taxonomy/. Accessed 03 Aug 2016.
8. Banyard AC, Evans JS, Luo TR, Fooks AR. Lyssaviruses and bats: emergence and zoonotic threat. Viruses. 2014;6:2974–90.
9. McElhinney LM, Marston DA, Leech S, Freuling CM, van der Poel WHM, Echevarria J, et al. Molecular epidemiology of bat lyssaviruses in Europe. Zoonoses Public Health. 2013;60:35–45.
10. Jakava-Viljanen M, Nokireki T, Sironen T, Vapalahti O, Sihvonen L, Liisa S, et al. Evolutionary trends of European bat lyssavirus type 2 including genetic characterization of Finnish strains of human and bat origin 24 years apart. Arch Virol. 2015;160:1489–98.
11. Whitby JE, Heaton PR, Black EM, Wooldridge M, McElhinney LM, Johnstone P. First isolation of a rabies-related virus from a Daubenton's bat in the United Kingdom. Vet Rec. 2000;147:385–8.
12. Van der Poel WHM, Van der Heide R, Verstraten ERAM, Takumi K, Lina PHC, Kramps JA. European bat lyssaviruses, The Netherlands. Emerg Infect Dis. 2005;11:1854–9.
13. Freuling C, Grossmann E, Conraths FJ, Schameitat A, Kliemt J, Auer E, et al. First isolation of EBLV-2 in Germany. Vet Microbiol. 2008;131:26–34.
14. Megali A, Yannic G, Zahno ML, Brügger D, Bertoni G, Christe P, et al. Surveillance for European bat lyssavirus in Swiss bats. Arch Virol. 2010;155:1655–62.
15. Jakava-Viljanen M, Lilley T, Kyheroinen EM, Huovilainen A. First encounter of European bat lyssavirus type 2 (EBLV-2) in a bat in Finland. Epidemiol Infect. 2010;138:1581–5.
16. Rasmussen TB, Chriél M, Baagøe HJ, Fjederholt E, Kooi EA, Belsham GJ, et al. Detection of European bat lyssavirus type 2 in Danish Daubenton's bats using a molecular diagnostic strategy. In Proceedings of 8th annual EPIZONE meeting. Copenhagen; 2014.
17. Carlsson U, Lahti E, Elvander M. Surveillance of infectious diseases in animals and humans in Sweden 2012. Uppsala: National Veterinary Institute; 2013.
18. Lumio J, Hillbom M, Roine R, Ketonen L, Haltia M, Valle M, et al. Human rabies of bat origin in Europe. Lancet. 1986;1:378.
19. Fooks AR, McElhinney LM, Pounder DJ, Finnegan CJ, Mansfield K, Johnson N, et al. Case report: isolation of a European bat lyssavirus type 2a from a fatal human case of rabies encephalitis. J Med Virol. 2003;71:281–9.
20. Brookes SM, Klopfleisch R, Müller T, Healy DM, Teifke JP, Lange E, et al. Susceptibility of sheep to European bat lyssavirus type-1 and -2 infection: a clinical pathogenesis study. Vet Microbiol. 2007;125:210–23.
21. Cliquet F, Picard-Meyer E, Barrat J, Brookes SM, Healy DM, Wasniewski M, et al. Experimental infection of foxes with European Bat Lyssaviruses type-1 and 2. BMC Vet Res. 2009;5:19.
22. Syvertsen P, Isaksen K. Rare and potentially new bat species in Norway. Fauna. 2007;60:109–19.
23. Sunding M. The most common bat species in Norway. Fauna. 2007;60:104–8.
24. Isaksen K, Klann M, van der Kooij J, Michaelsen T, Olsen K, Starholm T, et al. Flaggermus i Norge. 2009.
25. Michaelsen T. Movements of bats in western Norway. Fauna. 2011;64:31–43.
26. Hutterer R, Ivanova T, Meyer-Cords C, Rodrigues L. Bat Migrations in Europe: A Review of Banding Data and Literature. 2005.
27. Kalko E, Schnitzler U. The echolocation and hunting behaviour of Daubenton's bat, *Myotis daubentonii*. Behav Ecol Sociobiol. 1989;24:225–38.
28. Michaelsen T. Bat activity along a Norwegian fiord. Fauna. 2011;64:80–3.
29. Olsen K. Kunnskapsstatus for flaggermus i Norge. 1996.
30. van der Kooij J. The Norwegian Zoological Society's bat care centre – five years of practise. Fauna. 2007;60:183–9.
31. Dietz C, Kiefer, A. Die Fledermäuse Europas. Kennen, bestimmen, schützen. Stuttgart: Kosmos Verlag; 2014.
32. Kokurewicz T. Sex and age related habitat selection and mass dynamics of daubenton's bats *Myotis daubentonii* (Kuhl, 1817) hibernating in natural conditions. Acta Chiropterologica. 2004;6:121–44.

33. The World Organisation for Animal Health (OIE). OIE Terrestrial Manual. Chapter 2.1.13 Rabies. 2013.
34. Servat A, Picard-Meyer E, Robardet E, Muzniece Z, Must K, Cliquet F. Evaluation of a Rapid Immunochromatographic Diagnostic Test for the detection of rabies from brain material of European mammals. Biologicals. 2012;40:61–6.
35. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6. Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30:2725–9.
36. Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, et al. Next generation sequencing of viral RNA genomes. BMC Genomics. 2013;14:444.
37. Brunker K, Marston DA, Horton DL, Cleaveland S, Fooks AR, Kazwala R, et al. Elucidating the phylodynamics of endemic rabies virus in eastern Africa using whole-genome sequencing. Virus Evol. 2015; doi:10.1093/ve/vev011.
38. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95.
39. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, et al. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform. 2013;14:193–202.
40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
41. Marston DA, Wise EL, Ellis RJ, McElhinney LM, Banyard AC, Johnson N, et al. Complete genomic sequence of rabies virus from an ethiopian wolf. Genome Announc. 2015; doi:10.1128/genomeA.00157-15.
42. Marston DA, McElhinney LM, Johnson N, Müller T, Conzelmann KK, Tordo N, et al. Comparative analysis of the full genome sequence of European bat lyssavirus type 1 and type 2 with other lyssaviruses and evidence for a conserved transcription termination and polyadenylation motif in the G-L 3′ non-translated region. J Gen Virol. 2007;88:1302–14.
43. Kuzmin IV, Wu X, Tordo N, Rupprecht CE. Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions. Virus Res. 2008;136:81–90.
44. Tiina N. Rabies (EBLV-2), bat - Finland: Daubenton's bat. International Society for Infectious Diseases. Archive Number: 20161018.4568558. 2016.
45. Nokireki T, Huovilainen A, Lilley T, Kyheroinen EM, Ek-Kommonen C, Sihvonen L, et al. Bat rabies surveillance in Finland. BMC Vet Res. 2013;9:174.
46. Johnson N, Selden D, Parsons G, Healy D, Brookes SM, McElhinney LM, et al. Isolation of a European bat lyssavirus type 2 from a Daubenton's bat in the United Kingdom. Vet Rec. 2003;152:383–7.
47. Fooks AR, McElhinney LM, Marston DA, Selden D, Jolliffe TA, Wakeley PR, et al. Identification of a European bat lyssavirus type 2 in a Daubenton's bat found in Staines, Surrey, UK. Vet Rec. 2004;155:434–5.
48. Fooks AR, Marston D, Parsons G, Earl D, Dicker A, Brookes SM. Isolation of EBLV-2 in a Daubenton's bat (*Myotis daubentonii*) found in Oxfordshire. Vet Rec. 2006;159:534–5.
49. Harris SL, Mansfield K, Marston DA, Johnson N, Pajamo K, O'brien N, et al. Isolation of European bat lyssavirus type 2 from a Daubenton's bat (*Myotis daubentonii*) in Shropshire. Vet Rec. 2007;161:384–6.
50. Freuling CM, Kliemt J, Schares S, Heidecke D, Driechciarz R, Schatz J, et al. Detection of European bat lyssavirus 2 (EBLV-2) in a Daubenton's bat (*Myotis daubentonii*) from Magdeburg, Germany. Berl Münch Tierärztl Wschr. 2012;125:255–8.
51. Johnson N, Goddard TM, Goharriz H, Wise E, Jennings D, Selden D, et al. Two EBLV-2 infected Daubenton's bats detected in the north of England. Vet Rec. 2016;179:311–2.
52. Robardet E, Picard-Meyer E, Andrieu S, Servat A, Cliquet F. International interlaboratory trials on rabies diagnosis: an overview of results and variation in reference diagnosis techniques (fluorescent antibody test, rabies tissue culture infection test, mouse inoculation test) and molecular biology techniques. J Virol Methods. 2011;177:15–25.
53. Robardet E, Andrieu S, Rasmussen TB, Dobrostana M, Horton DL, Hostnik P, et al. Comparative assay of fluorescent antibody test results among twelve European National Reference Laboratories using various anti-rabies conjugates. J Virol Methods. 2013;191:88–94.
54. Johnson N, Vos A, Neubert L, Freuling C, Mansfield KL, Kaipf I, Denzinger A, Hicks D, Nunez A, Franka R, et al. Experimental study of European bat lyssavirus type-2 infection in Daubenton's bats (*Myotis daubentonii*). J Gen Virol. 2008;89:2662–72.
55. Lucan RK. Relationships between the parasitic mite *Spinturnix andegavinus* (*Acari*: *Spinturnicidae*) and its bat host, *Myotis daubentonii* (*Chiroptera*: *Vespertilionidae*): seasonal, sex- and age-related variation in infestation and possible impact of the parasite on the host condition and roosting behaviour. Folia Parasitol (Praha). 2006;53:147–52.
56. Lov om jakt og fangst av vilt (viltloven). https://lovdata.no/dokument/NL/lov/1981-05-29-38. Accessed 03 Aug 2016.
57. Agreement on the Conservation of Populations of European Bats. http://www.eurobats.org. Accessed 03 Aug 2016.
58. Cliquet F, Freuling C, Smreczak M, Van der Poel WHM, Horton D, Fooks AR, et al. Development of harmonised schemes for monitoring and reporting of rabies in animals in the European Union. 2010.
59. Schatz J, Fooks AR, McElhinney L, Horton D, Echevarria J, Vazquez-Moron S, et al. Bat rabies surveillance in Europe. Zoonoses Public Health. 2013;60:22–34.
60. Battersby J. Guidelines for Surveillance and Monitoring of European Bats. Bonn: UNEP/EUROBATS; 2010.
61. Johnson N, McElhinney LM, Smith J, Lowings P, Fooks AR. Phylogenetic comparison of the genus Lyssavirus using distal coding sequences of the glycoprotein and nucleoprotein genes. Arch Virol. 2002;147:2111–23.
62. Amengual B, Whitby JE, King A, Cobo JS, Bourhy H. Evolution of European bat lyssaviruses. J Gen Virol. 1997;78:2319–28.
63. Davis PL, Holmes EC, Larrous F, Van der Poel WHM, Tjørnehøj K, Alonso WJ, et al. Phylogeography, population dynamics, and molecular evolution of European bat lyssaviruses. J Virol. 2005;79:10487–97.
64. Fooks AR, Selden D, Brookes SM, Johnson N, Marston DA, Jolliffe TA, et al. Identification of a European bat lyssavirus type 2 in a Daubenton's bat found in Lancashire. Vet Rec. 2004;155:606–7.
65. Schatz J, Freuling CM, Auer E, Goharriz H, Harbusch C, Johnson N, et al. Enhanced passive bat rabies surveillance in indigenous bat species from Germany - a retrospective study. PLoS Negl Trop Dis. 2014; doi:10.1371/journal.pntd.0002835.
66. Banyard AC, Johnson N, Voller K, Hicks D, Nunez A, Hartley M, et al. Repeated detection of European bat lyssavirus type 2 in dead bats found at a single roost site in the UK. Arch Virol. 2009;154:1847–50.
67. Horton DL, Voller K, Haxton B, Johnson N, Leech S, Goddard T, et al. European bat lyssavirus type 2 in a Daubenton's bat in Scotland. Vet Rec. 2009;165:383–4.

3.7: Article 11 published in Genome Announcements

First Complete Genomic Sequence of Canine Distemper Virus from an Ethiopian Wolf.

DOI:10.1128/genomeA.00621-17

Summary: Ethiopian wolves are a critically endangered species with less than 500 worldwide. Through a close collaboration with the Ethiopian Wolf Conservation Trust, organs are received from dead Ethiopian wolves to diagnose rabies (see 2.3.2.1) and CDV. This is the first CDV genome from Ethiopia and phylogenetic analysis indicates this virus more closely related to Asian strains than those circulating in nearby Tanzania.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

VIRUSES

Complete Genomic Sequence of Canine Distemper Virus from an Ethiopian Wolf

Denise A. Marston,[a,b] Jemma Watson,[a] Emma L. Wise,[a]* Richard J. Ellis,[a] Eric Bedin,[c,d] Girma Ayalew,[e] Muktar Abute,[d] Xavier de Lamballerie,[b] Anthony R. Fooks,[a,f] Claudio Sillero-Zubiri,[c,d] Ashley C. Banyard[a]

Animal and Plant Health Agency (APHA), Surrey, United Kingdom[a]; UMR "Émergence des Pathologies Virales" (EPV) Aix-Marseille University, IRD 190, INSERM 1207, EHESP, IHU Méditerranée Infection, Marseille, France[b]; Wildlife Conservation Research Unit, The Recanati-Kaplan Centre, Zoology, University of Oxford, Oxford, United Kingdom[c]; Ethiopian Wolf Conservation Programme, Dinsho, Ethiopia[d]; Ethiopian Wildlife Conservation Authority, Addis Ababa, Ethiopia[e]; Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom[f]

ABSTRACT Canine distemper virus (CDV) has been implicated in population declines of wildlife, including many threatened species. Here we present the full genome of CDV from an Ethiopian wolf, *Canis simensis*, the world's rarest and most endangered canid.

Received 19 May 2017 **Accepted** 20 May 2017 **Published** 20 July 2017

Citation Marston DA, Watson J, Wise EL, Ellis RJ, Bedin E, Ayalew G, Abute M, de Lamballerie X, Fooks AR, Sillero-Zubiri C, Banyard AC. 2017. Complete genomic sequence of canine distemper virus from an Ethiopian wolf. Genome Announc 5:e00621-17. https://doi.org/10.1128/genomeA.00621-17.

Address correspondence to Denise A. Marston, denise.marston@apha.gsi.gov.uk.

* Present address: Emma L. Wise, Public Health England, Porton Down, Salisbury, United Kingdom.

Canine distemper virus (CDV), a member of the *Morbillivirus* genus, has a negative-strand nonsegmented RNA genome. CDV has recently been implicated in population declines of numerous wildlife species, including lions (*Panthera. leo*) (1), black-backed jackals (*Canis mesomelas*) (2), fennec foxes (*Vulpes zerda*) (3), spotted hyenas (*Crocuta crocuta*) (4), and rhesus monkeys (*Macaca mulatta*) (5), and aquatic species, including Lake Baikal seals (*Phoca siberica*) and Caspian seals (*Phoca caspia*) (6). Significantly, CDV has had an impact on several threatened species, including the world's most endangered felid, the Iberian lynx (*Lynx pardinus*) (7), and the Amur tiger (*Panthera tigris. altaica*) (8), the Santa Catalina Island fox (*Urocyon littoralis catalinae*) (9), and the Ethiopian wolf (*Canis simensis*) (10). Historically, different strains of CDV were grouped geographically (11); however, an increased interest in CDV and the potential for translocation of infected animals has revealed that CDV phylogenies are complex, much like those of the related measles virus, presenting a multilineage global distribution (12). Here we describe full-genome sequencing of a CDV derived from an infected subadult male Ethiopian wolf from the Sodota pack, found dead during an outbreak in the Web Valley, Bale Mountains, southeastern Ethiopia, in September 2015 (CDV06) (10).

RNA from spleen stored in glycerol was prepared for next-generation sequencing on the MiSeq platform. Briefly, TRIzol-extracted viral RNA was depleted of host genomic DNA and rRNA as described previously (13). Double-stranded (ds) cDNA was synthesized from 50 ng RNA, using a random cDNA synthesis system (Roche), according to the manufacturers' instructions. The ds cDNA was purified using Ampure XP magnetic beads (Beckman Coulter, Inc.), and 1 ng was used for the Nextera XT DNA sample preparation kit (Illumina). A sequencing library was prepared and sequenced on an Illumina MiSeq with 2- × 150-bp paired-end reads. The total reads (5,354,417) were mapped to a reference sequence (GenBank accession number JN896331) using BWA (v 0.7.5a-r405) (14), and visualized in Tablet (15). A modified SAMtools/vcfutils (16) script was used to generate an intermediate consensus sequence in which any indels relative to the original reference sequence were appropriately called. The intermediate consensus was used as the reference for subsequent iteration of mapping and consensus calling. The total number of assembled viral reads was 387,043 (7.23% of the total

reads). Despite the low proportion of viral sequence detected within the total data set, coverage of the entire genome was obtained (average read depth of 1,798).

The genetic organization of the Ethiopian wolf CDV genome was consistent with those of other CDV genomes, with a complete genome size of 15,690 nucleotides. The genome has closest homology (99%) to a sequence from China (GenBank accession number JN896331) clustering with Asia-1 sequences rather than African CDV sequences, including those from neighboring Tanzania.

Accession number(s). The complete genomic sequence of CDV06 has been deposited in GenBank under accession number MF041963.

ACKNOWLEDGMENTS

This work was supported by the European Union H2020-funded Research Infrastructure Grant "European Virus Archive Global (EVAg)" (H2020 — grant agreement 653316). The Ethiopian Wolf Conservation Programme (http://www.ethiopianwolf.org) thanks the Ethiopian Wildlife Conservation Authority and the Bale Mountains National Park for permission to work in Bale and the Born Free Foundation, the Wildlife Conservation Network, and the U.S. Fish and Wildlife Service for funding.

REFERENCES

1. Roelke-Parker ME, Munson L, Packer C, Kock R, Cleaveland S, Carpenter M, O'Brien SJ, Pospischil A, Hofmann-Lehmann R, Lutz H, Mwamengele GL, Mgasa MN, Machange GA, Summers BA, Appel MJ. 1996. A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). Nature 379:441–445. https://doi.org/10.1038/379441a0.
2. Bellan SE, Cizauskas CA, Miyen J, Ebersohn K, Küsters M, Prager KC, Van Vuuren M, Sabeta C, Getz WM. 2012. Black-backed jackal exposure to rabies virus, canine distemper virus, and *Bacillus anthracis* in Etosha National Park, Namibia. J Wildl Dis 48:371–381. https://doi.org/10.7589/0090-3558-48.2.371.
3. Woo GH, Jho YS, Bak EJ. 2010. Canine distemper virus infection in fennec fox (*Vulpes zerda*). J Vet Med Sci 72:1075–1079. https://doi.org/10.1292/jvms.09-0510.
4. Haas L, Hofer H, East M, Wohlsein P, Liess B, Barrett T. 1996. Canine distemper virus infection in Serengeti spotted hyenas. Vet Microbiol 49:147–152. https://doi.org/10.1016/0378-1135(95)00180-8.
5. Sun Z, Li A, Ye H, Shi Y, Hu Z, Zeng L. 2010. Natural infection with canine distemper virus in hand-feeding rhesus monkeys in China. Vet Microbiol 141:374–378. https://doi.org/10.1016/j.vetmic.2009.09.024.
6. Kuiken T, Kennedy S, Barrett T, Van de Bildt MW, Borgsteede FH, Brew SD, Codd GA, Duck C, Deaville R, Eybatov T, Forsyth MA, Foster G, Jepson PD, Kydyrmanov A, Mitrofanov I, Ward CJ, Wilson S, Osterhaus AD. 2006. The 2000 canine distemper epidemic in Caspian seals (*Phoca caspica*): pathology and analysis of contributory factors. Vet Pathol 43:321–338. https://doi.org/10.1354/vp.43-3-321.
7. Millán J, Candela MG, Palomares F, Cubero MJ, Rodríguez A, Barral M, de la Fuente J, Almería S, León-Vizcaíno L. 2009. Disease threats to the endangered Iberian lynx (*Lynx pardinus*). Vet J 182:114–124. https://doi.org/10.1016/j.tvjl.2008.04.005.
8. Seimon TA, Miquelle DG, Chang TY, Newton AL, Korotkova I, Ivanchuk G, Lyubchenko E, Tupikov A, Slabe E, McAloose D. 2013. Canine distemper virus: an emerging disease in wild endangered Amur tigers (*Panthera tigris altaica*). mBio 4. https://doi.org/10.1128/mBio.00410-13.
9. Timm SF, Munson L, Summers BA, Terio KA, Dubovi EJ, Rupprecht CE, Kapil S, Garcelon DK. 2009. A suspected canine distemper epidemic as the cause of a catastrophic decline in Santa Catalina Island foxes (*Urocyon littoralis catalinae*). J Wildl Dis 45:333–343. https://doi.org/10.7589/0090-3558-45.2.333.
10. Gordon CH, Banyard AC, Hussein A, Laurenson MK, Malcolm JR, Marino J, Regassa F, Stewart AM, Fooks AR, Sillero-Zubiri C. 2015. Canine distemper in endangered Ethiopian wolves. Emerg Infect Dis 21:824–832. https://doi.org/10.3201/eid2105.141920.
11. Barrett T. 1999. Morbillivirus infections, with special emphasis on morbilliviruses of carnivores. Vet Microbiol 69:3–13. https://doi.org/10.1016/S0378-1135(99)00080-2.
12. Muñoz-Alía MÁ, Muller CP, Russell SJ. 2017. Antigenic drift defines a new D4 subgenotype of measles virus. J Virol 91:e00209-17. https://doi.org/10.1128/JVI.00209-17.
13. Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR. 2013. Next generation sequencing of viral RNA genomes. BMC Genomics 14:444. https://doi.org/10.1186/1471-2164-14-444.
14. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.
15. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D. 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform 14:193–202. https://doi.org/10.1093/bib/bbs012.
16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

3.8: Article 12 published in Genome Announcements

Louping Ill Virus Genome Sequence Derived from the Spinal Cord of an Infected Lamb.

DOI:10.1128/genomeA.00454-13

Summary: LIV is a tick-borne flavivirus which is endemic in certain parts of the UK. This complete genome is the first obtained directly from a clinical sample and only the second to be published. The sequence was obtained from this sample in 2012 using the 454 platform.

genomeA
Journals.ASM.org

Louping Ill Virus Genome Sequence Derived from the Spinal Cord of an Infected Lamb

Denise A. Marston,[a] Karen L. Mansfield,[a] Rebecca Mearns,[b] Richard J. Ellis,[c] Anthony R. Fooks,[a,d] Nicholas Johnson[a]

Wildlife Zoonosis and Vector-Borne Diseases Research Group, Animal Health and Veterinary Laboratories Agency, Addlestone, Surrey, United Kingdom[a]; Animal Health and Veterinary Laboratories AgencyP enrith, Merrythought, Calthwaite, Penrith, United Kingdom[b]; Specialist Scientific Support Department, Animal Health and Veterinary Laboratories Agency, Addlestone, Surrey, United Kingdom[c]; University of Liverpool, Department of Clinical Infection, Microbiology and Immunology, Liverpool, United Kingdom[d]

Louping ill virus (LIV) is a zoonotic virus causing fatal encephalitis in young sheep and grouse. We have recovered the complete genome sequence from a spinal cord sample prepared from a lamb that was naturally infected with LIV. This is only the second LIV genome sequence reported and the first prepared from a clinical sample.

Received 24 May 2013 **Accepted** 17 June 2013 **Published** 18 July 2013

Citation Marston DA, Mansfield KL, Mearns R, Ellis RJ, Fooks AR, Johnson N. 2013. Louping ill virus genome sequence derived from the spinal cord of an infected lamb. Genome Announc. 1(4):e00454-13. doi:10.1128/genomeA.00454-13.

Address correspondence to Nicholas Johnson, nick.johnson@ahvla.gsi.gov.uk.

The tick-borne *Flavivirus* genome consists of a positive-sense RNA, approximately 10,800 bp in length (1). This encodes a single polyprotein that is posttranslationally cleaved to form three structural proteins and seven nonstructural proteins (2). Louping ill virus (LIV) is genetically similar to tick-borne encephalitis virus (TBEV), and it has been proposed that they form a single species with four viral types (3). LIV is distinct from TBEV in that it is prevalent in the United Kingdom, in the absence of TBEV, and is virulent in sheep, causing fatal encephalitis. TBEV does not cause disease in sheep but is highly virulent in humans. A single LIV genome sequence is currently available in GenBank (that of LIV strain 369/T2; accession no. NC_001809). This was obtained from a virus isolated from a Scottish *Ixodes ricinus* tick in 1963 and was sequenced after a long passage history in the late 1990s (4). In contrast to LIV, TBEV has almost 50 complete genome sequences available. The generation of additional genomic data for LIV will assist further investigations into the properties of this virus.

Total RNA was extracted using the TRIzol method from a spinal cord sample prepared from a young sheep that died suddenly near Penrith, England, in 2009. LIV infection in the sheep was confirmed independently by immunohistochemistry and detection of flavivirus by reverse transcription-PCR. RNA for whole-genome sequencing using pyrosequencing was prepared as previously described (5). Briefly, RNA was depleted of host genomic DNA using RNase-free DNase (Qiagen), and host rRNA was depleted using Terminator 5′-phosphate-dependent exonuclease (Epicentre Biotechnologies). Depleted RNA was fragmented, and a random-primed cDNA library was prepared and sequenced with the Roche 454 GS FLX system. Sequence data were assembled in the GS mapping assembly software (Roche) using the currently available LIV sequence (accession no. NC_001809). This approach recovered a partial genome sequence with only 77 viral reads (0.1% of total reads), reflecting the relatively small proportion of virus genome in relation to host nucleic acids. The remaining sequence was obtained using directed PCR and Sanger sequencing of genomic gaps. The contigs from both approaches were assembled using SeqMan (DNAStar).

The LIV Penrith genome is 10,875 nucleotides long with 95.6% identity to the existing LIV genome. The polyprotein-coding sequence is 10,245 nucleotides and contains 3,415 codons. The 5′ and 3′ untranslated regions (UTR) are 131 and 499 nucleotides, respectively. A comparison with the existing LIV genome sequence (that of strain 369/T2) suggests conservation within the polyprotein region, which is the same length and is 97.5% identical at the amino acid level. Both the 5′ UTR and the 3′ UTR of LIV 369/T2 are 2 nucleotides longer than those of LIV Penrith. This is the first complete LIV genome sequence from England and the first LIV sequence from primary diagnostic material. Additional sequences from LIV isolates and samples from across its geographical range should enhance our understanding of this neglected virus.

Nucleotide sequence accession number. The complete genome sequence of LIV Penrith has been deposited in GenBank under the accession no. KF056331.

ACKNOWLEDGMENTS

This work was supported by grants from the Department for Environment, Food and Rural Affairs (SE4112) and by the EU FP7-funded Research Infrastructure Grant European Virus Archive (no. 19 228292).

REFERENCES

1. **Mansfield KL, Johnson N, Phipps LP, Stephenson JR, Fooks AR, Solomon T.** 2009. Tick-borne encephalitis virus— review of an emerging zoonosis. J. Gen. Virol. **90:**1781–1794.
2. **Lindenbach BD, Rice CM.** 2003. Molecular biology of flaviviruses. Adv. Virus Res. **59:**23–61.

3. **Grard G, Moureau G, Charrel RN, Lemasson JJ, Gonzalez JP, Gallian P, Gritsun TS, Holmes EC, Gould EA, de Lamballerie X.** 2007. Genetic characterization of tick-borne flaviviruses: new insights into evolution, pathogenetic determinants and taxonomy. Virology **361:**80–92.
4. **Gritsun TS, Venugopal K, Zanotto PM, Mikhailov MV, Sall AA, Holmes EC, Polkinghorne I, Frolova TV, Pogodina VV, Lashkevich VA, Gould EA.** 1997. Complete sequence of two tick-borne flaviviruses isolated from Siberia and the UK: analysis and significance of the 5′ and 3′-UTRs. Virus Res. **49:**27–39.
5. **Marston DA, McElhinney LM, Ellis RJ, Horton DJ, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR.** Next generation sequencing of RNA viral genomes. BMC Genomics, in press.

3.9: Article 13 published in Genome Announcements

Complete Genome Sequence of Rift Valley Fever Virus Strain Lunyo.

DOI:10.1128/genomeA.00170-16

Summary: RVFV lacks full genome sequences, particularly for the Lunyo strain. As part of a PhD, another student in the laboratory, Sarah Lumley, required the complete genome sequence for a RVFV strain Lunyo.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Complete Genome Sequence of Rift Valley Fever Virus Strain Lunyo

Sarah Lumley,[a,b,c] Daniel L. Horton,[b] Denise A. Marston,[c] Nicholas Johnson,[c] Richard J. Ellis,[d] Anthony R. Fooks,[c,e] Roger Hewson[a,f]

Virology and Pathogenesis Group, Microbiology Services Division, Public Health England, Wiltshire, United Kingdom[a]; Department of Pathology and Infectious Disease, School of Veterinary Medicine, University of Surrey, Guildford, United Kingdom[b]; Wildlife Zoonoses and Vector-Borne Diseases Research Group, Animal and Plant Health Agency, Addlestone, Surrey, United Kingdom[c]; Specialist Scientific Support Department, Animal and Plant Health Agency, Addlestone, Surrey, United Kingdom[d]; Department of Clinical Infection, Microbiology and Immunology, Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom[e]; National Institute for Health Research, Health Protection Research Unit in Emerging and Zoonotic Infections, University of Liverpool, Liverpool, United Kingdom[f]

Using next-generation sequencing technologies, the first complete genome sequence of Rift Valley fever virus strain Lunyo is reported here. Originally reported as an attenuated antigenic variant strain from Uganda, genomic sequence analysis shows that Lunyo clusters together with other Ugandan isolates.

Received 22 February 2016 **Accepted** 24 February 2016 **Published** 14 April 2016

Citation Lumley S, Horton DL, Marston DA, Johnson N, Ellis RJ, Fooks AR, Hewson R. 2016. Complete genome sequence of Rift Valley fever virus strain Lunyo. Genome Announc 4(2):e00170-16. doi:10.1128/genomeA.00170-16.

Address correspondence to Sarah Lumley, sarah.lumley@phe.gov.uk.

Rift Valley fever virus (RVFV) is a *Phlebovirus*, in the family *Bunyaviridae*, and was first isolated in 1930. It primarily affects ruminants, causing abortions and mortality, and in humans it causes a febrile illness, which is severe in 1% of cases (1). The virus cycles between its mosquito vector and mammalian hosts. The strain Lunyo was isolated from a pool of *Aedes* mosquitoes in the Lunyo Forest, Uganda, in 1955 (2). It was described as a variant strain due to varying reports of hemagglutinin and antibody neutralization (2–5). Pathogenicity in mice was initially low but increased with serial passage (2, 3), except in some strains of mice, hamsters, gerbils, and rats, where the virus remains attenuated (5, 6). Genome characterization is essential to improve understanding of the role sequence variation has on virulence.

Lunyo mouse brain suspension ($P \geq 11$) was passaged once in Vero E6 cells. Viral RNA was extracted from supernatant using the QIAamp viral RNA minikit (Qiagen). Double-stranded (ds) cDNA was synthesized from 50 ng of RNA using the cDNA Synthesis System (Roche) in accordance with the manufacturer's instructions. ds-cDNA was purified using Agencourt AMPure XP system (Beckman Coulter). A sequencing library was prepared from 1 ng of DNA using the Nextera XT-DNA library preparation kit (Illumina) according to the manufacturer's instructions. Sequencing was performed following standard Illumina protocols on the Illumina MiSeq with 150-bp paired-end reads. Total reads (8,556,470) were mapped to reference sequences (EU312121, DQ380217, DQ375419) using BWA version 0.7.5a-r405 (7) and visualized in Tablet (8). Consensus sequence was generated in modified samtools/vcfutils (9), and intermediate sequence was used as a reference for subsequent iteration mapping (10). A total of 67,991 reads were mapped (0.79%), resulting in near-complete small (S) (18,889 reads), complete medium (M) (22,833), and large (L) (26,269) segments with an average read depth of 1,208, 623, and 443, respectively. The S segment intergenic region nucleotides were resolved by amplicon-based next-generation sequencing using primers binding to terminal regions (11).

Phylogenetic analyses of the S, M, and L segments were performed in MEGA6. All complete RVFV sequences in GenBank (n = 160, 99, and 94 for the S, M and L segments, respectively) were aligned by ClustalW, and a maximum likelihood phylogenetic tree was constructed for each genome segment. Analysis of novel sequence data showed that all three Lunyo segments cluster with strains originating from Uganda, within the previously defined lineage E (12). Genetic similarity to a closely related strain isolated from mosquitoes in Entebbe, Uganda, in 1944 (DQ 80156, DQ 80191, DQ 75429) demonstrated 97.5%, 98.6%, and 98.3% nucleotide identities in the S, M, and L segments, respectively. Results show 4 nonsynonymous substitutions in the nucleoprotein, 3 in the nonstructural protein, 8 in the glycoproteins, and 8 in the L protein. A single-nucleotide insertion in the intergenic region of the S segment conserves the segment size range of 1,690 to 1,962 (12). The contribution of a further complete genome to RVFV investigations will improve understanding of viral evolution and host interactions.

Nucleotide sequence accession numbers. Complete genomic sequences of RVFV strain Lunyo S, M, and L segments have been deposited in GenBank under accession numbers KU167025 to KU167027.

ACKNOWLEDGMENTS

We thank Public Health England's Genomics of Rare and Emerging Human Pathogens department for technical support.

FUNDING INFORMATION

This work was supported by Public Health England (PHE), the University of Surrey, and the Animal and Plant Health Agency (APHA) internal Ph.D. grant (project SC1402). Next-generation sequencing was partially funded from the European Commission Seventh Framework Programme under ANTIGONE (project 278976).

REFERENCES

1. **Mansfield KL, Banyard AC, McElhinney L, Johnson N, Horton DL, Hernández-Triana LM, Fooks AR.** 2015. Rift Valley fever virus: a review of diagnosis and vaccination, and implications for emergence in Europe. Vaccine **33:**5520–5531. http://dx.doi.org/10.1016/j.vaccine.2015.08.020.
2. **Weinbren MP, Williams MC, Haddow AJ.** 1957. A variant of Rift Valley fever virus. S Afr Med J **31:**951–957.
3. **Tomori O, Kasali O.** 1979. Pathogenicity of different strains of Rift Valley fever virus in Swiss albino mice. Br J Exp Pathol **60:**417–422.
4. **Swanepoel R, Struthers JK, Erasmus MJ, Shepherd SP, McGillivray GM, Shepherd AJ, Hummitzsch DE, Erasmus BJ, Barnard BJ.** 1986. Comparative pathogenicity and antigenic cross-reactivity of Rift Valley fever and other African phleboviruses in sheep. J Hyg (Lond) **97:**331–346.
5. **Peters CJ, Anderson GW, Jr.** 1981. Pathogenesis of Rift Valley fever. Contrib Epidemiol Biostat **3:**21–41.
6. **Battles JK, Dalrymple JM.** 1988. Genetic variation among geographic isolates of Rift Valley fever virus. Am J Trop Med Hyg **39:**617–631.
7. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics **26:**589–595. http://dx.doi.org/10.1093/bioinformatics/btp698.
8. **Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D.** 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform **14:**193–202. http://dx.doi.org/10.1093/bib/bbs012.
9. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1000 Genome Project Data Processing.** 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25:**2078–2079.
10. **Marston DA, Vázquez-Morón S, Ellis RJ, Wise EL, McElhinney LM, de Lamballerie X, Fooks AR, Echevarría JE.** 2015. Complete genomic sequence of European Bat lyssavirus 1, isolated from *Eptesicus isabellinus* in Spain. Genome Announc **3**(1):e01518-14. http://dx.doi.org/10.1128/genomeA.01518-14.
11. **Weingartl HM, Zhang S, Marszal P, McGreevy A, Burton L, Wilson WC.** 2014. Rift Valley fever virus incorporates the 78 kDa glycoprotein into virions matured in mosquito C6/36 cells. PLoS One **9:**e87385. http://dx.doi.org/10.1371/journal.pone.0087385.
12. **Bird BH, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST.** 2007. Complete genome analysis of 33 ecologically and biologically diverse Rift Valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. J Virol **81:**2805–2816. http://dx.doi.org/10.1128/JVI.02095-06.

3.10: Article 14 published in Virology Journal

Detection and genetic characterization of Seoul Virus from commensal brown rats in France.

DOI:10.1186/1743-422X-11-32

Summary: This paper described the methodology used by our collaborators to obtain complete genome sequence from hantaviruses. It also describes our efforts to do the same, but using Roche 454 and amplicons to complete missing sequence. The methodology used by our collaborators involves a more thorough depletion of host nucleic acid – which for hantaviruses is more critical as the virus is present at negligible levels.

At APHA, RNA was sequenced using the 454 platform. Due to the low number of reads available there were gaps. Primers were designed and Sanger sequencing was used to cover the gaps. The complete genome from one of the positive rats (LY0852) was obtained.

Random amplification was attempted using a protocol developed at AMU using Phi29 (Nelson, 2014). The resultant DNA was sequenced on the 454 but only yielded two virus specific reads, so this method was not taken forward.

In the meantime, our French collaborators developed their own Illumina protocols and obtained complete genome sequence using a depletion method using nuclease after freeze-thawing and filtering the tissue homogenate, before RNA extraction. rRNA was then depleted before library preparation. Ultimately this method was much better suited to hantavirus samples.

VIROLOGY JOURNAL

RESEARCH Open Access

Detection and genetic characterization of Seoul Virus from commensal brown rats in France

Tatiana Dupinay[1†], Kieran C Pounder[2†], Florence Ayral[1], Maria-Halima Laaberki[1], Denise A Marston[3], Sandra Lacôte[4], Catherine Rey[5], Fabienne Barbet[5], Katja Voller[3], Nicolas Nazaret[5], Marc Artois[1], Philippe Marianneau[4], Joel Lachuer[5], Anthony R Fooks[3,6], Michel Pépin[1*], Catherine Legras-Lachuer[5] and Lorraine M McElhinney[3,6*]

Abstract

Background: Hantaviruses are single-stranded RNA viruses, which are transmitted to humans primarily via inhalation of aerosolised virus in contaminated rodent urine and faeces. Whilst infected reservoir hosts are asymptomatic, human infections can lead to two clinical manifestations, haemorrhagic fever with renal syndrome (HFRS) and hantavirus cardiopulmonary syndrome (HCPS), with varying degrees of clinical severity. The incidence of rodent and human cases of Seoul virus (SEOV) in Europe has been considered to be low, and speculated to be driven by the sporadic introduction of infected brown rats (*Rattus norvegicus*) via ports.

Methods: Between October 2010 and March 2012, 128 brown rats were caught at sites across the Lyon region in France.

Results: SEOV RNA was detected in the lungs of 14% (95% CI 8.01 – 20.11) of brown rats tested using a nested pan-hantavirus RT-PCR (polymerase gene). Phylogenetic analysis supports the inclusion of the Lyon SEOV within Lineage 7 with SEOV strains originating from SE Asia and the previously reported French & Belgian SEOV strains. Sequence data obtained from the recent human SEOV case (Replonges) was most similar to that obtained from one brown rat trapped in a public park in Lyon city centre. We obtained significantly improved recovery of virus genome sequence directly from SEOV infected lung material using a simple viral enrichment approach and NGS technology.

Conclusions: The detection of SEOV in two wild caught brown rats in the UK and the multiple detection of SEOV infected brown rats in the Lyon region of France, suggests that SEOV is circulating in European brown rats. Under-reporting and difficulties in identifying the hantaviruses associated with HFRS may mask the public health impact of SEOV in Europe.

Keywords: Hantavirus, SEOV, France, Brown rat, *Rattus norvegicus*, Next generation sequencing, Viral enrichment

Introduction

Hantaviruses (family *Bunyaviridae*, genus Hantavirus) are single-stranded RNA viruses. Unlike other members of the *Bunyaviridae*, hantaviruses are not transmitted by arthropods but primarily by rodents of the families Cricetidae and Muridae, although insectivore and bat hosts have also been reported [1,2]. Each hantavirus appears to be adapted and largely restricted to an individual reservoir host species, implying that they have co-evolved, although phylogenetic analyses suggests that this apparent co-evolution may be more attributed to recent preferential host switching and local adaptation [3].

Transmission to humans is primarily via inhalation of aerosolised virus in contaminated rodent urine and faeces. Whilst infected reservoir hosts are asymptomatic, human infections can lead to two clinical manifestations,

* Correspondence: michel.pepin@vetagro-sup.fr; Lorraine.McElhinney@ahvla.gsi.gov.uk
†Equal contributors
[1]Université de Lyon, VetAgro Sup, USC 1233/Equipe « Pathogènes émergents et rongeurs sauvages (PERS), F-69280 Marcy-L'Etoile, France
[3]Wildlife Zoonoses and Vector-borne Diseases Research Group, Animal Health and Veterinary Laboratories Agency (AHVLA), New Haw, Addlestone, Surrey KT15 3NB, UK
Full list of author information is available at the end of the article

haemorrhagic fever with renal syndrome (HFRS) and hantavirus cardiopulmonary syndrome (HCPS), with varying degrees of morbidity and mortality [4]. Surveillance in Europe has detected six rodent-borne hantaviruses; Dobrava-Belgrade virus (DOBV), Saaremaa virus (SAAV), Seoul virus (SEOV), Puumala virus (PUUV), Tatenale virus (TATV) and Tula virus (TULV) plus two insectivore-borne hantaviruses; Seewis virus (SWSV) and Nova virus (NVAV) [4-7]. The relative geographic distribution of each hantavirus is defined by their reservoir host [7]. The most common and widespread hantavirus across northern, central and eastern Europe is PUUV, which is associated with the mildest form of HFRS [4].

Unlike other hantaviruses, SEOV has a global distribution due to the worldwide dispersal of its carrier host (*Rattus sp*). Confirmed human SEOV infections have been reported in Asia (Japan [8], South Korea [9], China [10,11]) and the Americas (USA [12], Brazil [13]). Norwegian/brown rats (*Rattus norvegicus*) are a cosmopolitan species and represent an emerging and widely distributed host of hantavirus in China, where, a total of 1,557,622 cases of HFRS were reported in humans between 1950–2007 with 46,427 deaths (3%) [11,14]. Historically, the presence of Seoul virus in Europe was considered anecdotal and speculated to be driven by the sporadic introduction of infected brown rats via ports [4]. Previously, a single HFRS case near the port city of Lyon, France, had only been confirmed serologically by SEOV FRNT [15] and SEOV antibodies had been reported in brown rats in France (10-78.9%) and Belgium (27.1%) [15,16]. However, more recently the virus has been isolated from wild brown rats in the UK [17] and pet rats in the UK and Sweden [18-20]. In addition, SEOV associated HFRS has been reported in four cases in the UK and France, all of which were clinically severe and involved renal impairment [17,21,22].

This study aimed to determine the presence of SEOV in wild rats (*R. norvegicus*) trapped in and around Lyon, France and analyse any resulting molecular epidemiological data. The study also determined the optimal approach to obtaining SEOV genomic sequence data directly from infected lung tissue by comparing different sample preparation techniques and next generation sequencing (NGS) platforms.

Results and discussion

Detection of SEOV virus

A total of 128 brown rats were caught from 23 sites in and around Lyon. Seoul hantavirus RNA was detected in 14% (18/128) of the rat lung samples tested in triplicate (95% CI 8.01 – 20.11). Positive rats were detected in 6 out of the 23 sites of capture (see Additional file 1).

There was a male bias of 2:1 in the infected individuals: 11 adult males, one juvenile male, one pregnant female and five adult females. The proportion of all males infected was larger than females, 16.4% and 11.3% respectively, but this was not significant (Pearson's Chi-squared test, $\chi 2 = 0.6568$, df = 1, P = 0.4177).

A male biased ratio amongst SEOV infected rats is not uncommon, and has been reported on several occasions [23-25]. Whilst neither male nor female rats are believed to be more susceptible to Seoul virus infection, males do shed the virus for a longer duration in their urine, faeces and saliva [25] and so the viral RNA may be detectable for longer in the host tissues. In addition, the primary route of transmission between adult males is thought to be through wounds [26], so it has been suggested that the likelihood of males acquiring the Seoul virus is greater due to them having more aggressive encounters [25].

All 18 RT-PCR positive rats were selected for genetic analysis and partial sequences of the L segment (317 bp) were recovered. Eight variable sites were located within this partial sequence. Phylogenetic analysis resolved the Lyon SEOV into three clusters (Lyon I, II and III; Figure 1, Additional file 1) reflecting their disparate trapping locations. The Lyon I, II and III variants were

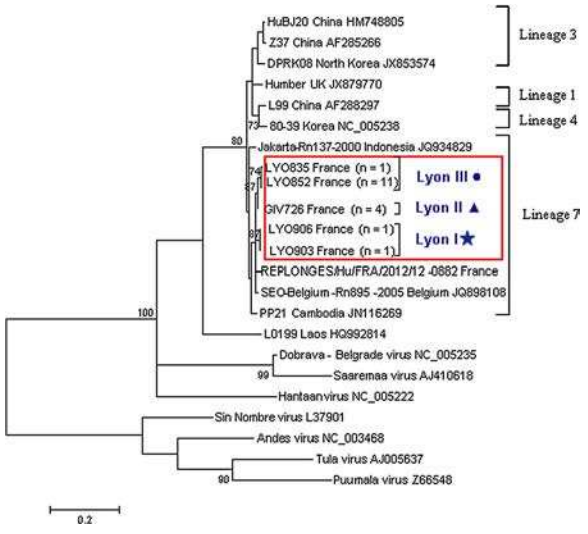


Figure 1 Maximum likelihood tree using the model T92 + Gamma [28] for SEOV partial L segment sequences n = 23 in MEGA5 [35]. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. The scale bar indicates nucleotide substitutions per site. Only bootstrap support of >70% are shown. Positions with less than 95% site coverage were eliminated. There were a total of 317 positions in the final dataset. The phylogenetic positions of groups Lyon I, II and III are shown in relation to representative Seoul strains (identical sequences removed for clarity). GIV726 partial L sequence was identical to GIV733, GIV737 and GIV757 (Lyon II). LYO852 partial L sequence was identical to LYO799, 837, 838, 839, 843, 845, 848, 853, 884 and 871 (Lyon III). Genbank accession numbers are shown next to taxa names.

detected at 1 (n = 2 rats), 2 (n = 4 rats) and 3 (n = 12) locations respectively (See Additional file 2A). No co-circulation of variants was observed. Lyon I and II were geographically restricted, whereas Lyon III was the most frequently detected and widespread of the variants (Additional file 1). All Lyon SEOV partial L sequences, including that derived from the recent human case (Replonges) showed highest identity to the Belgium SEO/Belgium/Rn895/2005 strain (JQ898108). Lyon I partial L sequences (LYO903 and LYO906) were more divergent from the other Lyon sequences (0.5-1.4%), but they were the closest Lyon SEOV to the nearby severe HFRS case in Replonges (~97.8%) [21]. All 18 RT-PCR L sequences clustered with previously described Lineage 7 sequences within Phylogroup A [14,27], with moderate bootstrap support. Despite their disparate isolation, most SEOV variants published to date are genetically homogenous [11,14] making it difficult to determine the source of introduction. However, at a local level the higher degree of sequence homology can result in geographical clustering as observed in China [14], the UK [18] and for the Lyon SEOV in this study. To further study the molecular epidemiology of a Lyon SEOV strain in the context of global SEOV, we obtained full genome sequence of a representative sample. The strong and non-degraded SEOV positive lung tissue sample LYO852 was chosen as it represented the most frequent and widespread variant detected (Lyon III).

Due to the low abundance of viral sequences relative to total host nucleic acids, we optimised the procedure to obtain complete SEOV genome using two next generation sequencing platforms and differing sample preparation approaches.

Roche 454 output and assembly statistics

Viral specific reads from LYO852 were obtained directly from lung tissue on the Roche 454 NGS platform without the use of viral enrichment or ultra centrifugation. *De Novo* assembly of the 454 reads yielded 59 contigs (consisting of 73,105 reads, totalling 24,730,464 bp) representing 82% total reads, with a mean length of 702 bp (ranging between 105-2920 bp). There were 15 contigs ≥500 bp. Based on BLAST identity searches, all contigs were host or mycoplasma sequences. Mapping of the reads using GS Reference Mapper (Roche) with published SEOV genome sequences identified 44 (0.03%) SEOV specific reads yielding 9 contigs in total for LYO852. Two partial nucleocapsid (S) gene contigs were retrieved of 715 and 786 bp. Three partial glycoprotein (M) gene contigs were retrieved of 612, 987 and 1,735 bp. Four partial polymerase (L) gene contigs were retrieved, of 459, 603, 740 and 1,564 bp. Following alignment, the total 454 coverage for each of the three segments of LYO852 was 84.8% (S), 91.3% (M) and 51.5% (L).

Viral enrichment methodology

To improve upon the genome coverage obtained using the Roche-454 platform, we compared several purification procedures and employed the Illumina NGS platform (Figure 2). We combined the homogenization step with or without freeze-thaw cycles, with or without sample filtration (to remove cells and mitochondria), and with or without either of 2 nuclease digestion protocols (to degrade DNA and RNA host contaminants) (data not shown). We observed that the nuclease digestion for

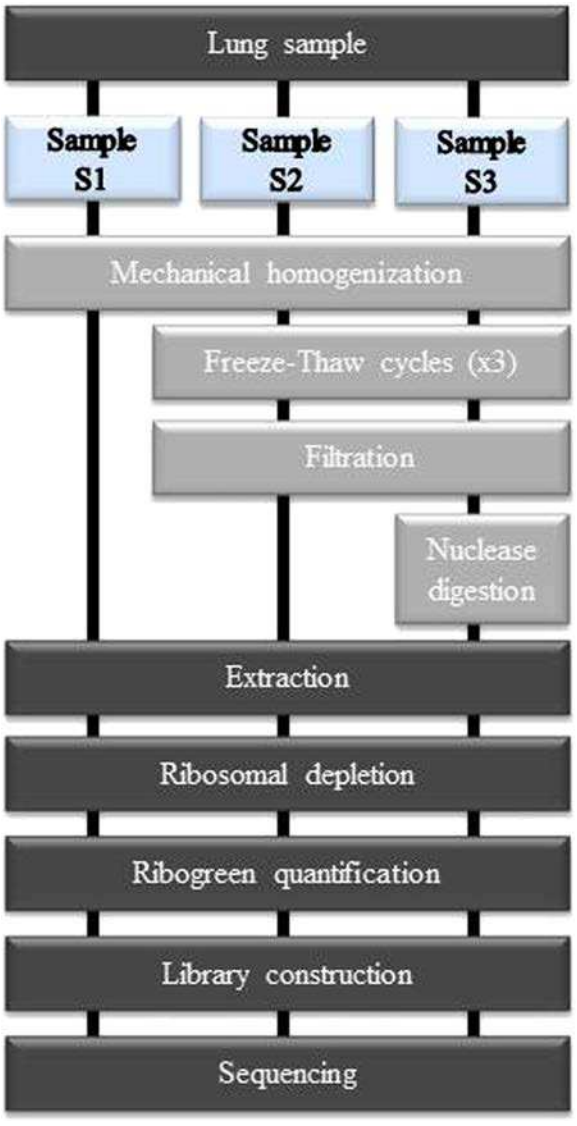


Figure 2 Workflow for the preparation of lungs tissue samples for next generation sequencing. All samples were extracted using RNeasy mini kit (Qiagen) and treated by ScriptSeq complete golg kit (Epicentre) and submitted to Illumina sequencing. Comparison of viral enrichment methods: no enrichment procedure was performed for the S1 condition, a filtration step was included for S2 condition and a filtration step with nuclease digestion were combined for the condition S3.

90 min was not sufficient to remove all rRNA and therefore we performed a ribosomal depletion. We compared the various approaches using qPCR assays for GAPDH, β-actin cDNA and viral RNA (data not shown) in comparison to the non-enriched sample (S1). The two optimal enrichment protocols involved homogenisation of the tissue with a micropestle in cold HBSS followed by dry ice freeze thaw cycles and a centrifugation/filtration step, without (S2) or with a subsequent 2 step-digestion (S3) (Figure 2). The 3 resultant samples (S1, S2 and S3) were then used to carry out next-generation sequencing using the Illumina platform.

Illumina outputs and assembly statistics

Without viral enrichment (S1), 513 Illumina contigs were generated for a total length of 189,884 bp. There were 31 contigs ≥1000 bp. Of all the contigs, 13 corresponded to SEOV genome (0.34%).

Illumina sequencing for each of the three samples generated between 62 and 91 million sequence reads but 17% to 29% of the reads were discarded after quality filtering (Table 1). As shown in Table 1, the sample S1 presented a larger number of reads that aligned with the reference rat genome sequence (88%) than for the virus enriched samples S2 and S3 (approximately 71 and 47% respectively). Notably, the viral reads were 6 times more abundant in the S3 sample (2.20%) in comparison to the S1 sample (0.34%). Furthermore, the S3 sample appeared more enriched in SEOV reads than the two other samples (Table 1) and that obtained using the Roche 454 platform. Mapping sequence reads revealed complete or near complete (>99%) coverage of the SEOV reference genomes for the virally enriched samples (S3). Complete SEOV genome sequences were recovered from the LYO852 sample and the SEOV consensus sequences of the three samples were identical. The SEOV sequences have been deposited in GenBank under accession numbers KF387723 to KF387725. Hence, we report the significantly improved chances of successfully obtaining complete viral genome sequences by NGS following simple viral enrichment steps. The S3 enrichment approach will be assessed for future NGS analysis on the Roche-454 platform.

Genetic and phylogenetic analysis

We report the complete genomic sequence of a SEOV strain isolated from *R. norvegicus* in France.

The S-segment has a total of 1755nt with a deduced coding sequence of 1290nt. The putative encoded nucleoprotein (N) (AGZ59811) is 429 amino acids for a predicted 48KDa protein. The S-segment complete coding sequence shared the highest identity (98%) with the complete coding sequences of Vietnamese strains [29,30], the two Singaporean strains Rn41 and Rn46 [31], the SEOV Belgian Rn895 strain and the French Replonges strain [21] (respective GenBank accession numbers: AB618112 to AB618126, GQ274944 and GQ274945, JQ898106 and KC902522). All these strains originated from wild *R. norvegicus* but the Replonges strain was obtained from a patient. The putative N protein was identical to the deduced N protein of strains that originated from South Korea, China, Vietnam, the United Kingdom and the French Replonges strain (NP_942556, ADE34611, BAL46798, AGB05597 and AGL45258 respectively). It is identical to the 91 amino acid long partial sequence of the France 90 strain (CAI47594), implying that the 2 substitutions at the nucleotide level were silent. It also presents 99.76% nucleotide identity and 100% amino acid similarity with the complete coding region of Belgium/Rn895 strain (AFN11574).

The M-segment sequence (3638nt long) has a deduced coding sequence of 3402nt encoding for a putative polyprotein of 125KDa (AGZ59810). This putative protein precursor presented the conserved cleavage site (WAASA) that is required to give rise to the Gn and Gc membrane glycoproteins [32]. The full M-segment sequence shared, at the nucleotide and protein levels, the highest identity with the strains originating from Vietnam and Singapore. The putative polyprotein is identical to the predicted partial protein sequence of the France90 strain (CAI47595) confirming that the 3 substitutions are silent. It also presents 6 substitutions (5 of which are synonymous substitutions) when compared with the sequence of the Belgium/Rn895 strain (JQ898107, AFN11575).

The L-segment sequence obtained is 6511nt long. The deduced coding sequence (6456nt) encodes the putative

Table 1 Overview of the sequence reads and mapped SEOV sequences obtained using the Illumina NGS platform for sample preparations S1-S3

Illumina Sample Prep	Read data							Reads from SEOV Segments			Reads from SEOV	% reads from SEOV
	Total reads	PF	PF aligned	% PF aligned	% Ribosomal	Viral reads	% viral reads	Segment S	Segment M	Segment L		
S1	91,661,298	75,522,658	66,580,460	88.16%	0.99%	255,911	0.34%	12,675	30,211	13,617	56,503	22%
S2	83,920,361	59,434,134	42,659,876	71.78%	30.95%	434,756	0.73%	10,610	37,111	43,345	91,066	20.90%
S3	62,479,875	46,105,731	21,974,935	47.66%	21.55%	1,016,600	2.20%	112,152	363,008	381,521	856,681	84.29%

PF: Pass Filter (Quality Check). PF Aligned – sequences aligned with rat genome (see methods).

RNA-dependent RNA polymerase (AGZ59809) whose size is predicted at 246KDa. When compared exclusively to full length coding sequences, the LYO852 strain shared the highest identity with the L-segment of the China Z37 strain (96%). However, when considering partial sequences, it presented the highest identity (99%) with the Belgium/Rn895 strain partial L-segment sequence (JQ898108).

From the 3 reconstructed phylogenetic trees (Figure 3), the LYO852 strain is resolved in the SEOV clade within the South East Asian virus lineage, also referred to as lineage 7 [27,33]. According to the nucleotide and protein analysis and the phylogenetic reconstruction of the full length S-segment, the LYO852 strain shared the most evolutionary relatedness with the strains previously detected either in Belgium (Belgium/Rn895) or in France (Replonges strain). Altogether, these findings are consistent with the earlier genetic description of the Belgium/Rn895 strain [33]. It supports further the hypothesis of a SEOV introduction in Europe due to the migration of its carrier, the brown rat, during trade between China and Europe [14]. However, the Lyon SEOV (LYO852) strain is clearly distinct from the strains isolated in the United Kingdom (IR461, Humber and Banbury strains) which appear to represent a distinct lineage. As the Replonges strain had been detected in a patient, our results raise concerns regarding the circulation of the SEOV virus in the Lyon area. Unfortunately, the wider distribution of SEOV in France is currently unknown. In particular, SEOV prevalence should be investigated in other large cities such as Marseille or Paris where commensal rodent populations are significant.

We report the presence of multiple foci of SEOV infected wild brown rats trapped in disparate locations in and around the large French city of Lyon. The recent detection of SEOV in wild brown rats in the UK [17] and pet rats in the UK and Sweden [18-20] may suggest SEOV emergence. However, there has been limited surveillance for hantaviral RNA in brown rats in Europe. It is likely that future surveillance will identify similar foci of infection in *R.Norvegicus* in other European countries. The contribution of SEOV to European HFRS cases should be further investigated to estimate the public health impact posed by commensal brown rats.

Materials and methods

Sample collection

Between October 2010 and March 2012, 128 brown rats *(Rattus norvegicus*) were trapped in and around the city of Lyon, France. Rats were trapped using small (28 cm × 9 cm × 9 cm) or large (50 cm × 15 cm × 15 cm) single catch rat traps. Captured rats were transported to the laboratory where live rats were immediately anaesthetised using Isofluorane and sacrificed by cervical dislocation. Each rat was aseptically dissected. Lung tissues were collected from different lobes.

Rats provided for this study were trapped for the purpose of pest control (agreement no. 69-1810). They were euthanized and used (agreement no. 69-020931) according to ethical rules supervised by the ethical committee of VetAgroSup and European regulation (Directive EU 86/609).

Screening for Hantavirus RNA

Immediately after collection, lung tissue was stored at -80°C pending further analysis. Approximately 50 to 100 mg of lung tissue was homogenised in 1 ml TRIzol® Reagent (Invitrogen, Life Technologies, Paisley, UK) with QIAGEN Stainless steel beads (5 mm) using a QIAGEN TissueLyser (Qiagen, UK) for 2 mins at 30 Hz. RNA was extracted from the homogenate according to the manufacturer's instructions (Invitrogen, Life Technologies, Paisley, UK). The RNA samples were reverse transcribed using random hexamers and screened for hantavirus as previously described [5] employing a pan-hantavirus nested RT-PCR directed against partial polymerase (L) gene sequences [34].

Phylogenetic analysis of partial L gene sequences

Multiple nucleotide sequence alignments of the 18 partial polymerase gene sequences and available published SEOV sequences were generated in MEGA5 [35]. Sequence identities were compared using Geneious 5.6.5. Optimum substitution models were estimated and maximum likelihood phylogenetic trees generated in MEGA5 [35] with bootstrap replications of 10,000 [36].

Roche-454 and Illumina platform sequencing

Initially, the Roche-454 was employed to obtain genome sequence for a representative Lyon SEOV following previous optimisation and success in obtaining complete genome coverage for lyssaviruses using this platform [37]. Viral specific reads from LYO852 were obtained directly from lung tissue on the Roche 454 platform without the use of viral enrichment or ultra centrifugation. Briefly, for performing 454 Roche sequencing, the TRIzol® extracted viral RNA obtained during hantavirus screening was depleted of host genomic DNA using RNase-free DNAse (Qiagen, UK) and host ribosomal RNA was depleted using Terminator™ 5'-Phosphate-Dependent Exonuclease (Epicentre Biotechnologies) as described previously [37]. The RNA was fragmented, a random-primed cDNA library was made and run using the Roche 454 GS FLX System. The sequencing data were initially assembled in the GS *de novo* assembly software (Roche). Subsequently, previously published SEOV sequences were used to map specific reads from the original raw data using GS Reference Mapper (Roche).

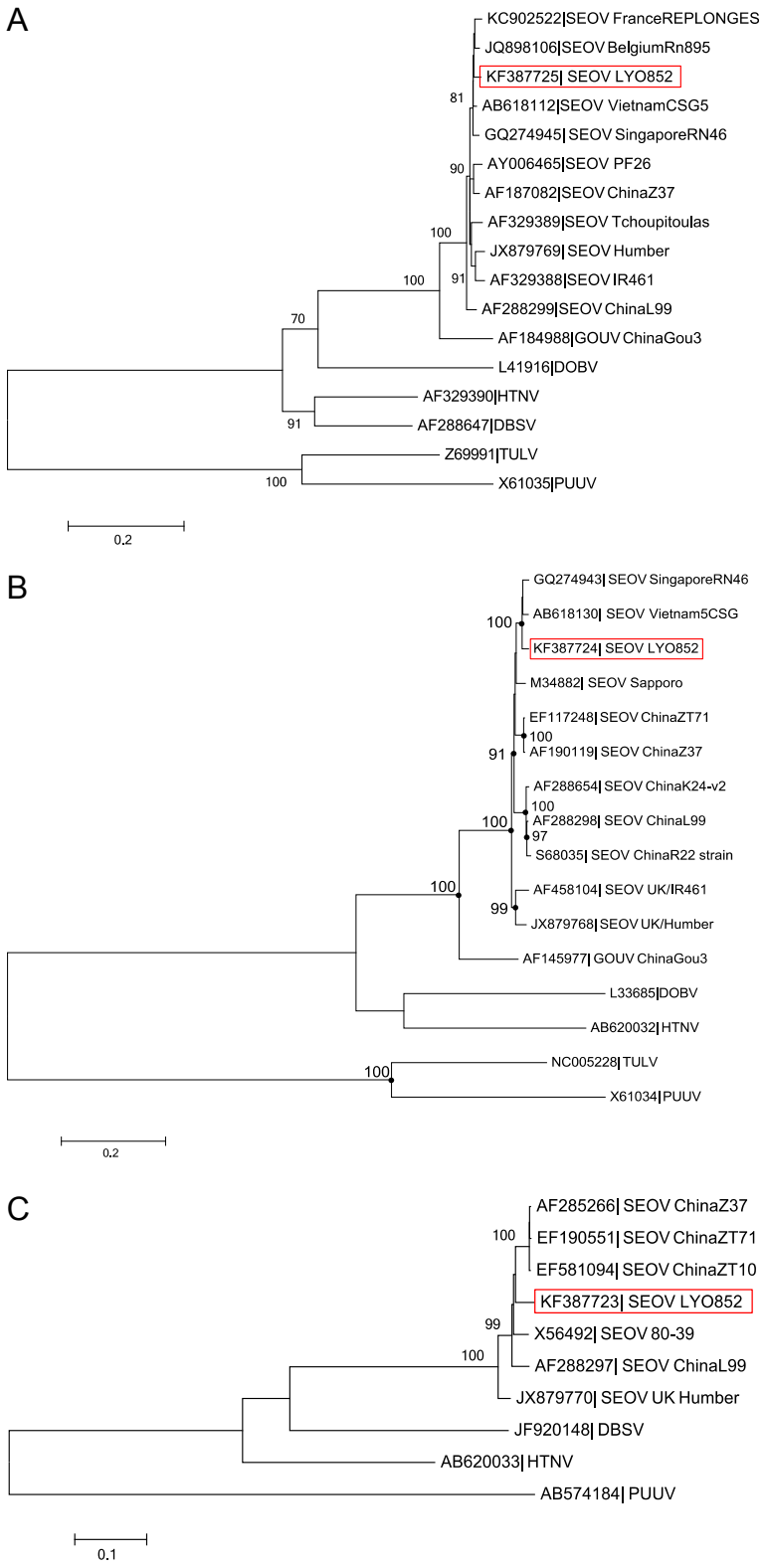


Figure 3 (See legend on next page.)

(See figure on previous page.)
Figure 3 Phylogenetic tree of hantaviruses based on the S, M and L segment sequences (complete coding region). All analyses were performed using the MEGA software [35]. **A**. S segment: analysis was performed applying the generalized time reversible model (GTR) using a Gamma distribution with five rate categories and invariant site (+G + I) Only bootstrap percentages ≥70% (from 1000 resamplings) are indicated. **B**. M segment: analysis was performed applying the generalized time reversible model (GTR) using a Gamma distribution with five rate categories (+G). For clarity purpose, the nodes corresponding to bootstrap percentages ≥70% (from 1000 resamplings) are indicated by dots. **C**. L segment: analysis was performed applying the generalized time reversible model (GTR) using a Gamma distribution with five rate categories (+G). Only bootstrap percentages ≥70% (from 500 resamplings) are indicated. The scale bars indicate nucleotide substitution per site. The red boxes highlight the LYO852 strain described in the present study. Accession numbers are indicated for each strain in the corresponding taxa name.

Subsequently Illumina sequencing was assessed. Total RNA extractions were performed using RNeasy mini kit (Qiagen) according to the manufacturer's instructions. The remaining rRNA was depleted using ScriptSeq complete gold kit Human/Mouse/Rat (Epicentre) following manufacturer instructions and controlled with the 2100 Bioanalyzer using "Eukaryote total RNA Pico Assay" (Agilent Technologies). 3.5-5 ng of depleted RNA was prepared for NGS using the Illumina protocol where 10 cycles of PCR were performed and standard TruSeq adapters and TruSeq barcoded primers were used. A final size selection was performed by native agarose gel electrophoresis to yield a library of inserts 250-350 bases in length. The library was extracted from the agarose gel using purification columns. The prepared library was then loaded onto an Illumina HiSeq 2500 v3 single read flow cell, standard cluster generation was performed on a Cbot and sequenced for 50 bases.

Sequence reads

Reads were processed using CASAVA 1.8.2 and demultiplexed based on index sequences. The FastQC was used for Quality Check. Sequences were aligned, first, using TopHat 2.0.6 to the Rat genome and the unaligned reads were aligned using Bowtie 2.0.2 software against known viruses.

Viral enrichment and Illumina Hiseq 2500 v3 sequencing

The different approaches for virus enrichment were also evaluated in this study for the infected lung tissue sample (LYO852) (Figure 2). Briefly, a piece of lung tissue was immersed in 1 ml of HBSS 1X and homogenized with micropestle and then placed on dry ice for approximately two minutes until frozen, and thawed quickly before returning to ice. Homogenization followed by freezing and thawing was repeated a further two times to disrupt the cells. Samples were then spun at 1500×g for 5 minutes at 4°C to pellet the nuclei and large cellular aggregates. The resulting supernatant was transferred to a new tube and 2 different treatments were applied: for the condition S2 a step-wise filtration process involving 0.45 μm polyethersulfone membrane filters (diameter 13 mm) (Millipore) was performed before RNA extraction. Condition S3 combined a step-wise filtration and a 2-step digestion with 25U of RNase I at 37°C for 90 min in 1× RNase I buffer and DNA is removed on-column (Qiagen). Following treatment, we extracted viral encapsidated RNA and residual host nucleic acids using the RNeasy mini extraction kit (Qiagen). Viral RNA was eluted to a final volume of 30 μl. Total RNA concentration was quantified with Quant-iT ribogreen RNA kit (Invitrogen).

The remaining rRNA was depleted using ScriptSeq complete gold kit Human/Mouse/Rat (Epicentre) following the manufacturer's instructions and quantified using the 2100 Bioanalyzer using "Eukaryote total RNA Pico Assay" (Agilent Technologies).

Reverse transcription and qPCR quantification

cDNA was generated with random hexamers using the iScript™cDNA Synthesis kit (Bio-Rad). RNA (30 ng) from each sample was incubated in the presence of 5× iScript reaction mix (containing iScript Reverse transcriptase) and nuclease-free water added to bring the final reaction volume to 20 μl. This volume was incubated at 25°C for 5 min, at 42°C for 30 min, at 85°C for 5 min. To quantify the enrichment of viral RNA, we performed various real-time PCRs targeting cDNA of Seoul Hantavirus, GAPDH and β-actin (see Additional file 2 for primer details). We calculated the fold enrichment in viral RNAs by comparing the proportion of encapsidated viral RNA CT (threshold cycle) values between the control and each treatment.

Genetic and phylogenetic analysis of genomic segments

The deduced amino acid sequences of the 3 genomic segments of the LYO852 strain were obtained using the Serial Cloner 2.6.1 software. The complete coding sequence of the S, M and L segments and the predicted protein sequences were compared to the NCBI database using the BLAST program (http://blast.ncbi.nlm.nih.gov/).

Multiple sequence alignments of coding sequences were carried out using ClustalW algorithm in the MEGA 5.2.2 software (default parameters) [35]. Phylogenetic reconstructions were performed using the Maximum Likelihood statistical method. Bootstrapping (1000 or 500 resamplings) was applied according to the best-fit substitution model recommended.

Additional files

Additional file 1: The locations of the trapping sites (circles) within a) France and b) Rhône-Alps department. SEOV positive variants 'Lyon I, II and III' are represented by a star, triangle and blocked out circles, respectively.

Additional file 2: A-Distribution of the SEOV variants detected in 6 of the 23 sites sampled. B-Oligonucleotides used in this study. vRNA: viral RNA.

Competing interests

All of the authors declare that they have no competing interests with respect to the publication of this manuscript.

Authors' contributions

TD, KCP, DAM and KV were involved in the PCR screening and NGS work. SL, CR, FB and NN were involved in viral enrichment and Illumina sequencing. MHL was involved in phylogenetic analysis. PM, MA, MP, ARF, JL, CLL and LMM were involved in project conception, data analysis and logistical support. FA coordinated the trapping and post-mortems of the rodents as part of the WildTech consortium (http://www.wildtechproject.com/wildtech/). All authors contributed to the writing of this manuscript. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge financial support from the European Commission under the Food, Agriculture and Fisheries, and Biotechnology theme of the 7th Framework Program for Research and Technological Development, FP7 'WildTech' project grant agreement no. 222633 and FP7 'European Virus Archive' project – grant agreement no 19 228292. We also acknowledge funding from the UK Defra surveillance projects SV3037 & FZ2100. Technical support and assistance was also provided by Emma Wise, Richard Ellis and Charlotte Featherstone (AHVLA). We acknowledge special financial support from VetagroSup for the post-doctoral grant (TD) and we also acknowledge Pr Etienne Benoit (USC1233) for his constant support.

Author details

[1]Université de Lyon, VetAgro Sup, USC 1233/Equipe « Pathogènes émergents et rongeurs sauvages (PERS), F-69280 Marcy-L'Etoile, France. [2]Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK. [3]Wildlife Zoonoses and Vector-borne Diseases Research Group, Animal Health and Veterinary Laboratories Agency (AHVLA), New Haw, Addlestone, Surrey KT15 3NB, UK. [4]Anses-Laboratoire de Lyon, Unité Virologie, F-69347 Lyon, France. [5]Viroscan 3D / Profilexpert, Faculté de Médecine et de Pharmacie, 3453 CNRS-US7 Inserm, 8 avenue Rockefeller, aile C2, 69373, Lyon Cedex 08, France. [6]National Consortium for Zoonosis Research, University of Liverpool, Leahurst, Neston, South Wirral CH64 7TE, UK.

Received: 30 October 2013 Accepted: 27 January 2014
Published: 20 February 2014

References

1. Klempa B, Fichet-Calvet E, Lecompte E, Auste B, Aniskin V, Meisel H, Barrière P, Koivogui L, ter Meulen J, Krüger DH: **Novel hantavirus sequences in shrew, Guinea.** *Emerg Infect Dis* 2007, **13:**520–522.
2. Weiss S, Witkowski PT, Auste B, Nowak K, Weber N, Fahr J, Mombouli JV, Wolfe ND, Drexler JF, Drosten C, Klempa B, Leendertz FH, Krüger DH: **Hantavirus in bat, Sierra Leone.** *Emerg Infect Dis* 2012, **18:**159–161.
3. Ramsden C, Holmes EC, Charleston MA: **Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence.** *Mol Biol Evol* 2009, **26:**143–153.
4. Vaheri A, Henttonen H, Voutilainen L, Mustonen J, Sironen T, Vapalahti O: **Hantavirus infections in Europe and their impact on public health.** *Rev Med Virol* 2012, **23:**35–49.
5. Pounder KC, Begon M, Sironen T, Henttonen H, Watts PC, Voutilainen L, Vapalahti O, Klempa B, Fooks AR, McElhinney LM: **Novel hantavirus in field vole, United Kingdom.** *Emerg Infect Dis* 2013, **19:**673–675.
6. Heyman P, Ceianu CS, Christova I, Tordo N, Beersma M, João Alves M, Lundkvist A, Hukic M, Papa A, Tenorio A, Zelená H, Essbauer S, Visontai I, Golovljova I, Connell J, Nicoletti L, Van Esbroeck M, Dudman SG, Aberle SW, Avšič-Županc T, Korukluoglu G, Nowakowska A, Klempa B, Ulrich RG, Bino S, Engler O, Opp M, Vaheri A: **A five-year perspective on the situation of haemorrhagic fever with renal syndrome and status of the hantavirus reservoirs in Europe, 2005-2010.** *Euro Surveill* 2011, **16:**19961.
7. Olsson GE, Leirs H, Henttonen H: **Hantaviruses and their hosts in Europe: reservoirs here and there, but not everywhere?** *Vector Borne Zoonotic Dis* 2010, **10:**549–561.
8. Kariwa H, Yoshimatsu K, Araki K, Chayama K, Kumada H, Ogino M, Ebihara H, Murphy ME, Mizutani T, Takashima I, Arikawa J: **Detection of hantaviral antibodies among patients with hepatitis of unknown etiology in Japan.** *Microbiol Immunol* 2000, **44:**357–362.
9. Kim YS, Ahn C, Han JS, Kim S, Lee JS, Lee PW: **Hemorrhagic fever with renal syndrome caused by the Seoul virus.** *Nephron* 1995, **71:**419–427.
10. Song G: **Epidemiological progresses of hemorrhagic fever with renal syndrome in China.** *Chin Med J* 1999, **112:**472–477.
11. Zhang YZ, Zhang FX, Wang JB, Zhao ZW, Li MH, Chen HX, Zou Y, Plyusnin A: **Hantaviruses in rodents and humans, Inner Mongolia Autonomous Region, China.** *Emerg Infect Dis* 2009, **15:**885–891.
12. Glass GE, Watson AJ, LeDuc JW, Childs JE: **Domestic cases of hemorrhagic fever with renal syndrome in the United States.** *Nephron* 1994, **68:**48–51.
13. Iversson LB, da Rosa AP, Rosa MD, Lomar AV, Sasaki MG, LeDuc JW: **Human infection by hantavirus in southern and southwestern Brazil.** *Rev Assoc Med Bras* 1994, **40:**85–92.
14. Lin XD, Guo WP, Wang W, Zou Y, Hao ZY, Zhou DJ, Dong X, Qu YG, Li MH, Tian HF, Wen JF, Plyusnin A, Xu JG, Zhang YZ: **Migration of Norway rats resulted in the worldwide distribution of Seoul hantavirus today.** *J Virol* 2012, **86:**972–981.
15. Heyman P, Plyusnina A, Berny P, Cochez C, Artois M, Zizi M, Pirnay JP, Plyusnin A: **Seoul hantavirus in Europe: first demonstration of the virus genome in wild *Rattus norvegicus* captured in France.** *Eur J Clin Microbiol Infect Dis* 2004, **23:**711–717.
16. Heyman P, Baert K, Plyusnina A, Cochez C, Lundkvist A, Esbroeck MV, Goossens E, Vandenvelde C, Plyusnin A, Stuyck J: **Serological and genetic evidence for the presence of Seoul hantavirus in *Rattus norvegicus* in Flanders, Belgium.** *Scand J Infect Dis* 2009, **41:**51–56.
17. Jameson LJ, Logue CH, Atkinson B, Baker N, Galbraith SE, Carroll MW, Brooks T, Hewson R: **The continued emergence of hantaviruses: isolation of a Seoul virus implicated in human disease, United Kingdom, October 2012.** *Euro Surveill* 2013, **18:**4–7.
18. Jameson LJ, Taori SK, Atkinson B, Levick P, Featherstone CA, van der Burgt G, McCarthy N, Hart J, Osborne JC, Walsh AL, Brooks T, Hewson R: **Pet rats as a source of hantavirus in England and Wales, 2013.** *Euro Surveill* 2013, **18:**20415.
19. Lundkvist A, Verner-Carlsson J, Plyusnina A, Forslund L, Feinstein R, Plyusnin A: **Pet rat harbouring Seoul hantavirus in Sweden, June 2013.** *Euro Surveill* 2013, **18:**20521.
20. Featherstone CA, Wyllie S, Frost AJ, McElhinney L, Fooks AR: **Hantavirus and pet rodents.** *Vet Rec* 2013, **172:**370–370.
21. Macé G, Feyeux C, Mollard N, Chantegret C, Audia S, Rebibou JM, Spagnolo G, Bour JB, Denoyel GA, Sagot P, Reynes JM: **Severe Seoul hantavirus infection in a pregnant woman, France, October 2012.** *Euro Surveill* 2013, **18:**20464.
22. Taori SK, Jameson LJ, Campbell A, Drew PJ, McCarthy ND, Hart J, Osborne JC, Sudhanva M, Brooks TJG: **UK hantavirus, renal failure, and pet rats.** *Lancet* 2013, **381:**1070–1070.
23. Cueto GR, Cavia R, Bellomo C, Padula PJ, Suárez OV: **Prevalence of hantavirus infection in wild *Rattus norvegicus* and *R. rattus* populations of Buenos Aires City, Argentina.** *Trop Med Int Health* 2008, **13:**46–51.
24. Glass GE, Childs JE, Korch GW, LeDuc JW: **Association of intraspecific wounding with hantaviral infection in wild rats (*Rattus norvegicus*).** *Epidemiol Infect* 1988, **101:**459–472.
25. Klein SL, Bird BH, Glass GE: **Sex differences in Seoul virus infection are not related to adult sex steroid concentrations in Norway rats.** *J Virol* 2000, **74:**8213–8217.
26. Hinson ER, Shone SM, Zink MC, Glass GE, Klein SL: **Wounding: the primary mode of Seoul virus transmission among male Norway rats.** *Am J Trop Med Hyg* 2004, **70:**310–317.
27. Plyusnina A, Ibrahim IN, Winoto I, Porter KR, Gotama IBI, Lundkvist A, Vaheri A, Plyusnin A: **Identification of Seoul hantavirus in *Rattus norvegicus* in Indonesia.** *Scand J Infect Dis* 2004, **36:**356–359.

28. Tamura K: **Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases.** *Mol Biol Evol* 1992, **9:**678–687.
29. Luan VD, Yoshimatsu K, Endo R, Taruishi M, Huong VT, Dat DT, Tien PC, Shimzu K, Koma T, Yasuda SP, Nhi L, Vu TQH, Arikawa J: **Studies on hantavirus infection in small mammals captured in southern and central highland area of Vietnam.** *J Vet Med Sci* 2012, **74:**1155–1162.
30. Truong TT, Yoshimatsu K, Araki K, Lee BH, Nakamura I, Endo R, Shimizu K, Yasuda SP, Koma T, Taruishi M, Okumura M, Truong UN, Arikawa J: **Molecular epidemiological and serological studies of hantavirus infection in Northern Vietnam.** *J Vet Med Sci* 2009, **71:**1357–1363.
31. Johansson P, Yap G, Low HT, Siew CC, Kek R, Ng LC, Bucht G: **Molecular characterization of two hantavirus strains from different rattus species in Singapore.** *Virol J* 2010, **7:**15.
32. Löber C, Anheier B, Lindow S, Klenk HD, Feldmann H: **The Hantaan virus glycoprotein precursor is cleaved at the conserved pentapeptide WAASA.** *Virology* 2001, **289:**224–229.
33. Plyusnina A, Heyman P, Baert K, Stuyck J, Cochez C, Plyusnin A: **Genetic characterization of seoul hantavirus originated from norway rats (*Rattus norvegicus*) captured in Belgium.** *J Med Virol* 2012, **84:**1298–1303.
34. Klempa B, Fichet-Calvet E, Lecompte E, Auste B, Aniskin V, Meisel H, Denys C, Koivogui L, ter Meulen J, Krüger DH: **Hantavirus in African wood mouse, Guinea.** *Emerg Infect Dis* 2006, **12:**838–840.
35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28:**2731–2739.
36. Felsenstein J: **Confidence-limits on phylogenies - an approach using the bootstrap.** *Evolution* 1985, **39:**783–791.
37. Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR: **Next generation sequencing of viral RNA genomes.** *BMC Genomics* 2013, **14:**444.

doi:10.1186/1743-422X-11-32
Cite this article as: Dupinay *et al.*: Detection and genetic characterization of Seoul Virus from commensal brown rats in France. *Virology Journal* 2014 **11**:32.

3.11: Conclusions

With the advent of NGS, and its subsequent uptake in the scientific community, there has been a definitive shift away from the acceptance of partial gene data for virus characterisation and analysis, towards requirement of complete genome sequence. For viruses such as lyssaviruses, present in clinical samples at high levels that can be relatively easily cultured, this is a welcome advancement, which benefits science and research. For other viruses such as hantaviruses and coronaviruses which are present at low viral loads, are difficult to culture, and have additional technical issues with assembly of reads, the change to using whole genome sequences routinely will be more tentative. What is clear is that NGS is now embedded in the scientific community and is being ustilised to answer many interesting questions including evolutionary history, transmission, host interactions as well as fundamental mechanisms such as virus pathogenicity.

Chapter 4: Virus Discovery and Characterisation

Chapter 4: Virus Discovery and Characterisation

4.1: Introduction

Until recently the traditional discovery of viruses has been at a relatively constant level. This involved many eminent virologists acquiring clinical and field samples from across the globe. These newly acquired viruses were propagated and classified using standard virological and serological methodologies. The diligence and expertise of these scientists has resulted in a valuable resource of viruses, which still exist today. However, with the advent of NGS technologies, metagenomics enables the screening of hundreds of samples resulting in an exponential increase in the diversity of viruses. This new area of 'Viromics', where the total viral genome contents found in certain biological sampling, including humans (Moustafa et al., 2017), rodents (Chen et al., 2017) and invertebrates (Shi et al., 2016). As powerful a tool NGS is, a number of issues arise due to this huge increase in data. The fundamental problem is the lack of 'Koch's postulates', previously adhered to, as the majority of the viruses discovered using NGS techniques do not cause disease, nor can they be cultured. Furthermore many of the novel viruses discovered cannot be classified under the ICTV guidelines due to novel gene arrangements, including lack of essential genes.

Chapter 2 described the bioinformatics developed to obtain full genome sequences of viruses at APHA. Currently, we are focussed on clinical samples where disease is suspected, or a virus is confirmed as present by another diagnostic test (such as RT-PCR). The de novo pipeline has been developed to obtain complete genome sequences for novel viruses.

4.2: Lyssaviruses

Identification of novel viruses is usually achieved by utilising accredited diagnostic techniques. For lyssaviruses, the gold standard OIE prescribed test is the FAT or dRIT. RT-PCR, although not considered a diagnostic test, is used in most OIE reference laboratories as a confirmatory test. The advantage of RT-PCRs is the ability to sequence the amplicon to obtain genetic data to a) confirm the presence of a lyssavirus, and 2) compare the sequence with previously published sequences to type the virus, or in the case of RABV to confirm the origin of the virus. Sequences from diagnostic RT-PCRs were the first indication for IKOV and GBLV that a novel virus was identified. For example, IKOV was identified in an African civet; which have previously been found to be infected with mongoose and canine variants of RABV, therefore it was assumed that the African civet was infected with RABV. However, the sequence

from the RT-PCR was completely different to other RABV sequences and was subsequently shown to be a novel lyssavirus (4.2.1).

During this PhD I have been fortunate enough to be directly involved in the discovery of two lyssaviruses: Ikoma lyssavirus (IKOV) and Gannoruwa bat lyssavirus (GBLV). In addition, I have been involved in the characterisation of both IKOV and GBLV and also a third virus Lleida bat virus (LLEBV), which was discovered by collaborators in Spain. As part of the characterisation, our NGS methodologies were used to obtained full genome sequence for all three viruses (see articles 16, 19 and 21). In order for ICTV to recognise these viruses as new lyssavirus species, proposals are submitted for consideration at the ICTV annual meeting. The proposals for all three lyssaviruses were prepared as part of this PhD. IKOV has been accepted (2013) and LLEBV and GBLV were submitted in 2017 for consideration after been ratified by the Rhabdoviridae study group.

The following articles describe the process involved in elucidating the characteristics of the three novel lyssaviruses.

4.2.1: Article 15 published in Emerging and Infectious Diseases

Ikoma Lyssavirus, highly divergent novel lyssavirus in an African civet

DOI:10.3201/eid1804.111553

Summary: Case report of the discovery of IKOV including clinical signs and diagnostic results. Phylogenetic analysis of partial N gene data obtained from the diagnostic hnRT-PCR revealed that IKOV was the most divergent lyssavirus identified.

Ikoma Lyssavirus, Highly Divergent Novel Lyssavirus in an African Civet[1]

Denise A. Marston, Daniel L. Horton, Chanasa Ngeleja, Katie Hampson, Lorraine M. McElhinney, Ashley C. Banyard, Daniel Haydon, Sarah Cleaveland, Charles E. Rupprecht, Machunde Bigambo, Anthony R. Fooks, and Tiziana Lembo

Evidence in support of a novel lyssavirus was obtained from brain samples of an African civet in Tanzania. Results of phylogenetic analysis of nucleoprotein gene sequences from representative *Lyssavirus* species and this novel lyssavirus provided strong empirical evidence that this is a new lyssavirus species, designated Ikoma lyssavirus.

Eleven *Lyssavirus* species have been classified: *Rabies virus* (RABV), *Lagos bat virus* (LBV), *Mokola virus* (MOKV), *Duvenhage virus* (DUVV), *European bat lyssavirus* types -1 and -2, *Australian bat lyssavirus*, *Aravan virus*, *Khujand virus*, *Irkut virus*, and *West Caucasian bat virus* (WCBV) (*1*). All these viruses except MOKV have been detected in bats. Two newly identified lyssaviruses, Shimoni bat virus (SHIBV) (*2*) and Bokeloh bat lyssavirus (*3*), both detected in bats, have not yet been classified. The presence of numerous lyssaviruses in bat species has led to increasing research efforts toward lyssavirus discovery in bat populations globally. However, lyssavirus surveillance in terrestrial mammals remains limited across most of Africa.

Of the 13 lyssaviruses, 5 circulate in Africa (RABV, LBV, MOKV, DUVV, and SHIBV). LBV, MOKV, DUVV, and SHIBV are detected exclusively in Africa, whereas RABV is detected worldwide. The predominant RABV variants circulating in Africa are the mongoose and canine biotypes. In South Africa, canine RABV is considered to have been introduced in the eastern Cape Province after importation of an infected dog from England in 1892 and subsequently spread, infecting domestic and wild carnivores (*4*). Separate introductions of canine RABV (particularly in northern Africa) have been suggested (*5*). In addition, molecular clock analysis indicates that mongoose RABV was present in southern Africa ≈200 years before the introduction of canine RABV (*6*).

In Tanzania, canine RABV is endemic and widespread throughout the country. In the Serengeti ecosystem, detailed studies have shown a single variant of canine RABV circulating in multiple host species (*7*). However, annual mass rabies vaccination campaigns have been conducted for dogs in villages surrounding Serengeti National Park since 2003, and rabies has not been detected in the park since 2000 (*8*). Enhanced laboratory-based surveillance in support of this canine rabies elimination program has been running concurrently in the region.

The Study

On May 11, 2009, an African civet (*Civettictis civetta*) displaying clinical signs consistent with rabies was killed by rangers in Ikoma Ward within Serengeti National Park (Figure 1). Rangers were contacted because the civet had bitten a child on the right leg in an unprovoked attack. The wound was washed with soap and water, and the child received postexposure rabies vaccination but no rabies immunoglobulin. Brain samples from the civet were tested multiple times (as part of a training course) at the Central Veterinary Laboratory in Tanzania. Results of the fluorescent antibody test and a direct rapid immunohistochemistry test were positive for lyssavirus-specific antigen. When testing was complete, the samples were sent to the Animal Health and Veterinary Laboratories Agency (AHVLA, Weybridge, UK) for additional confirmation of results and molecular analysis.

RNA was extracted by using TRIzol Reagent (Invitrogen, Paisley, UK), and a pan-lyssavirus reverse transcription PCR yielded a specific 606-bp amplicon (*9*). The amplicon (GenBank accession no. JN800509) was sequenced by using standard primers and protocols (*10*). Bayesian reconstructions were used for phylogenetic analysis of the nucleoprotein gene region (405 bp) and included representatives of all species from the lyssavirus genus; results showed that the sequence was unique and most closely related to WCBV (Figure 2). A canine RABV biotype from Tanzania and a mongoose RABV biotype from southern Africa were included in the dataset. Nucleotide comparisons indicated similar divergence from all

Author affiliations: Animal Health and Veterinary Laboratories Agency, Addlestone, UK (D.A. Marston, D.L. Horton, L.M. McElhinney, A.C. Banyard, A.R. Fooks); Central Veterinary Laboratory, Dar es Salaam, Tanzania (C. Ngeleja); University of Glasgow, Glasgow, Scotland, UK (K. Hampson, D. Haydon, S. Cleaveland, T. Lembo); National Consortium for Zoonosis Research, Neston, UK (L.M. McElhinney, A. R. Fooks); Centers for Disease Control and Prevention, Atlanta, Georgia, USA (C.E. Rupprecht); and Lincoln Park Zoo Tanzania Program, Arusha, Tanzania (M. Bigambo)

DOI: http://dx.doi.org/10.3201/eid1804.111553

[1]These data were presented in part at the XXII Rabies in the Americas meeting in San Juan, Puerto Rico, October 16–21, 2011.

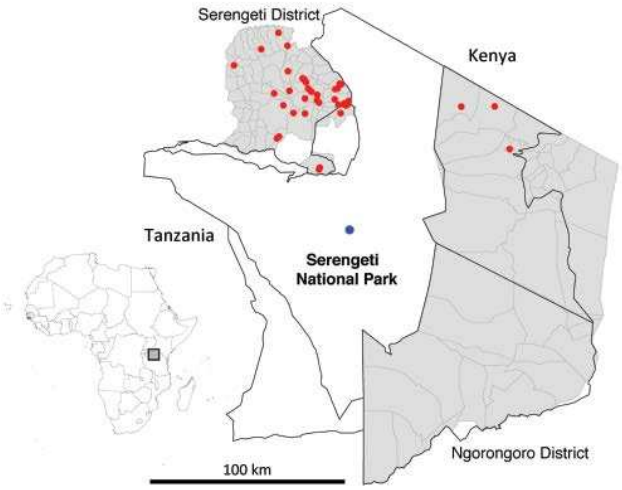


Figure 1. Serengeti National Park and surrounding districts (Serengeti and Ngorongoro). Blue dot indicates location of Ikoma lyssavirus–infected African civet within Ikoma Ward in northwest Tanzania. Red dots indicate cases of rabies confirmed during 2003–2011. Top left, map of Africa indicating study area in Tanzania (gray box).

lyssavirus species (minimum identity 62.2% *Australian bat lyssavirus*, maximum identity 68.6% WCBV), including 12 canine RABV sequences from domestic and wild animals in the Serengeti ecosystem (64.1%–65.1% identity). The posterior probabilities indicated that the IKOV and WCBV grouping was strongly supported, despite low sequence identity. Further phylogenetic analysis of representatives from other rhabdoviruses demonstrated that IKOV is a member of the *Lyssavirus* genus (41.6%–50.9% identity to representative rhabdovirus sequences that are available for this region of the genome) (data not shown).

Conclusions

We describe evidence, based on genomic sequences obtained from the brain sample of an African civet with clinical signs consistent with rabies, for the existence of a novel lyssavirus designated IKOV (Ikoma lyssavirus). The sample was frozen and thawed several times before being sent to AHVLA, which had a detrimental effect on the sample quality and resulted in viral RNA degradation and loss in viral viability. The results of confirmatory FATs performed at AHVLA were inconclusive, and attempts to isolate virus by using the rabies tissue culture inoculation test and the mouse inoculation test were unsuccessful. Despite the lack of isolated virus, the pan-lyssavirus sensitivity and specificity of the fluorescent antibody test (the test prescribed by the World Organisation for Animal Health as the standard for rabies testing) and direct rapid immunohistochemistry test support the assertion that a novel lyssavirus exists in the region. In addition, despite the poor quality of the sample, molecular techniques identified a lyssavirus-specific amplicon that was confirmed to be unique by phylogenetic analysis and to be highly divergent from known circulating RABV strains. A real-time PCR also detected this unique lyssavirus sequence, confirming that both molecular tests are pan-lyssavirus specific and are sufficient for the detection of highly divergent novel lyssaviruses (*11*).

The child who was bitten by the African civet received appropriate wound care and postexposure rabies vaccination. At the time of this report, the child remained well. We cannot, however, draw any conclusions as to whether the African civet was shedding virus when it bit the child or whether postexposure vaccinations are effective against IKOV.

This case of rabies in an African civet in the center of Serengeti National Park was highly unexpected. Since

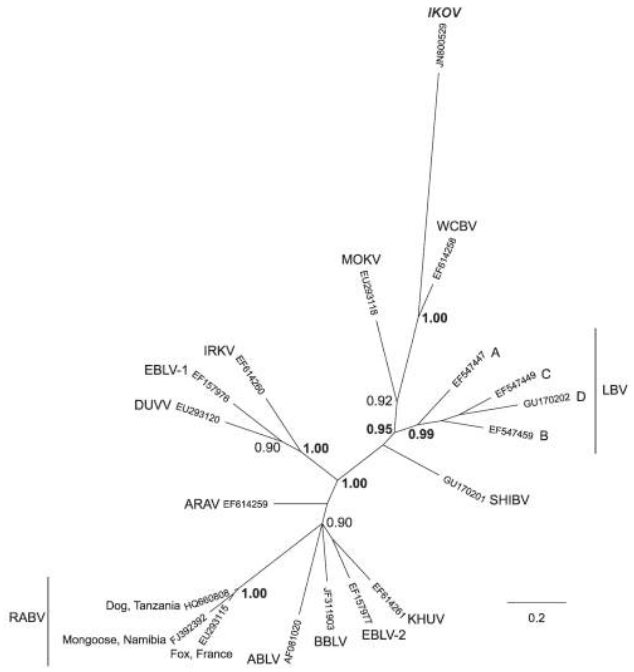


Figure 2. Phylogenetic relationships between all currently identified lyssaviruses compared with Ikoma lyssavirus (IKOV; shown in ***boldface italics***), as determined on the basis of partial nucleoprotein gene sequences (405 bp). Relationships are presented as an unrooted phylogram based on Bayesian Markov chain Monte Carlo analysis. Posterior probability values represent the degree of support for each node on the tree: only values ≥0.90 are shown; values ≥0.95 are shown in **boldface**. Scale bar indicates branch length, expressed as the expected number of substitutions per site. ARAV, Aravan virus; ABLV, Australian bat lyssavirus; BBLV, Bokeloh bat lyssavirus; DUVV, Duvenhage virus; EBLV-1 and EBLV-2, European bat lyssavirus type 1 and 2; IRKV, Irkut virus; KHUV, Khujand virus; LBV, Lagos bat virus (lineages A, B, C and D); MOKV, Mokola virus; RABV, rabies virus; SHIBV, Shimoni bat virus; WCBV, West Caucasian bat virus. Complete methods are described online (wwwnc.cdc.gov/EID/article/18/4/11-1553-F2.htm).

2000, the park had been free of rabies and no cases had been detected within a 30-km radius. This case of rabies in wildlife implied a major breach in the vaccination program. However, subsequent molecular characterization demonstrated that this case of rabies had not been caused by a RABV from a canine source. Thus, a breach had not occurred; instead, a novel lyssavirus with an unknown reservoir had caused the infection. Previously published data on lyssavirus infection in African civets (n = 6) was restricted to the RABV mongoose lineage (*12*). Although African civets can be infected with RABV and IKOV, infrequent detection of lyssaviruses in this species suggests that they are more likely to be incidental hosts. The nocturnal, opportunistic foraging behaviors of African civets imply that contact with bats is possible, particularly at roosts where interactions with a grounded rabid bat are more likely to occur. In the absence of virus isolates, the origin of IKOV is difficult to determine. Surveillance for rabies in bats and other mammals in Tanzania and typing of all lyssavirus-positive samples is necessary to determine the distribution and prevalence of IKOV.

The detection of WCBV cross-reacting neutralizing antibodies in gregarious *Miniopterus* spp. bats in neighboring Kenya could be informative, given the strong posterior probability values on the grouping of IKOV and WCBV in the Bayesian analysis (Figure 2) (*13*). Additional genomic and evolutionary analysis is underway to support IKOV as a new *Lyssavirus* species, potentially grouping with WCBV in phylogroup III (*14*), and to determine the antigenic diversity of the glycoprotein (*15*). Given that IKOV is highly distinct from RABV (more genetically distinct than WCBV from RABV) and that a human has been bitten by an infected animal, the effectiveness of current rabies vaccines needs to be further investigated.

Acknowledgments

We are grateful to the Tanzania Wildlife Research Institute, Tanzania Commission for Science and Technology, Tanzania National Parks, the Ngorongoro Conservation Area Authority, and the National Institute for Medical Research for permissions. We thank Tanzania National Parks, Tanzania Wildlife Research Institute, and Frankfurt Zoological Society for field support and the Ministry of Livestock and Fisheries Development for laboratory support in Tanzania. We also thank Sarah Devery, Stacey Leech, Emma Wise, Colin Black, Lillian Orciari, and Michael Niezgoda for their technical support.

This work was partially supported by The Department for Environment, Food and Rural Affairs ROAME SE0423/SV3500, by the EU FP7–funded Research Infrastructure Grant European Virus Archive (no. 228292), and by the Research and Policy for Infectious Disease Dynamics program of the Science and Technology Directorate (US Department of Homeland Security), and the Fogarty International Center. In Tanzania, the work was funded by Lincoln Park Zoo (Chicago, IL, USA), the Wellcome Trust, UBS Optimus Foundation, and the Medical Research Council.

Mrs Marston is a research scientist at AHVLA. Her research interests include viral genetics, molecular epidemiology, and virus–host interactions.

References

1. Dietzgen R, Calisher CH, Kurath G, Kuzmin IV, Rodriguez LL, Stone DM, et al. *Rhabdoviridae*. In: King A, Adams MJ, Carstens EB, Lefkowitz, EJ, editors. Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses. San Diego (CA): Elsevier; 2011. p. 654–81.
2. Kuzmin IV, Mayer AE, Niezgoda M, Markotter W, Agwanda B, Breiman RF, et al. *Shimoni bat virus*, a new representative of the *Lyssavirus* genus. Virus Res. 2010;149:197–210. http://dx.doi.org/10.1016/j.virusres.2010.01.018
3. Freuling CM, Beer M, Conraths FJ, Finke S, Hoffmann B, Keller B, et al. Novel lyssavirus in Natterer's bat, Germany. Emerg Infect Dis. 2011;17:1519–22.
4. Swanepoel R, Barnard BJ, Meredith CD, Bishop GC, Bruckner GK, Foggin CM, et al. Rabies in southern Africa. Onderstepoort J Vet Res. 1993;60:325–46.
5. Talbi C, Lemey P, Suchard MA, Abdelatif E, Elharrak M, Nourlil J, et al. Phylodynamics and human-mediated dispersal of a zoonotic virus. PLoS Pathog. 2010;6:e1001166. http://dx.doi.org/10.1371/journal.ppat.1001166
6. Van Zyl N, Markotter W, Nel LH. Evolutionary history of African mongoose rabies. Virus Res. 2010;150:93–102. http://dx.doi.org/10.1016/j.virusres.2010.02.018
7. Lembo T, Haydon DT, Velasco-Villa A, Rupprecht CE, Packer C, Brandao PE, et al. Molecular epidemiology identifies only a single rabies virus variant circulating in complex carnivore communities of the Serengeti. Proc Biol Sci. 2007;274:2123–30. http://dx.doi.org/10.1098/rspb.2007.0664
8. Lembo T, Hampson K, Haydon DT, Craft M, Dobson A, Dushoff J, et al. Exploring reservoir dynamics: a case study of rabies in the Serengeti ecosystem. J Appl Ecol. 2008;45:1246–57. http://dx.doi.org/10.1111/j.1365-2664.2008.01468.x
9. Heaton PR, Johnstone P, McElhinney LM, Cowley R, O'Sullivan E, Whitby JE. Heminested PCR assay for detection of six genotypes of rabies and rabies-related viruses. J Clin Microbiol. 1997;35:2762–6.
10. Hayman DT, Johnson N, Horton DL, Hedge J, Wakeley PR, Banyard AC, et al. Evolutionary history of rabies in Ghana. PLoS Negl Trop Dis. 2011;5:e1001. http://dx.doi.org/10.1371/journal.pntd.0001001
11. Hayman DT, Banyard AC, Wakeley PR, Harkess G, Marston D, Wood JL, et al. A universal real-time assay for the detection of *Lyssaviruses*. J Virol Methods. 2011;177:87–93. http://dx.doi.org/10.1016/j.jviromet.2011.07.002
12. Sabeta CT, Shumba W, Mohale DK, Miyen JM, Wandeler AI, Nel LH. Mongoose rabies and the African civet in Zimbabwe [letter]. Vet Rec. 2008;163:580. http://dx.doi.org/10.1136/vr.163.19.580
13. Kuzmin IV, Niezgoda M, Franka R, Agwanda B, Markotter W, Beagley JC, et al. Possible emergence of West Caucasian bat virus in Africa. Emerg Infect Dis. 2008;14:1887–9. http://dx.doi.org/10.3201/eid1412.080750
14. Fooks A. The challenge of new and emerging lyssaviruses. Expert Rev Vaccines. 2004;3:333–6. http://dx.doi.org/10.1586/14760584.3.4.333

15. Horton DL, McElhinney LM, Marston DA, Wood JL, Russell CA, Lewis N, et al. Quantifying antigenic relationships among the lyssaviruses. J Virol. 2010;84:11841–8. http://dx.doi.org/10.1128/JVI.01153-10

Address for correspondence: Anthony R. Fooks, Wildlife Zoonoses and Vector-borne Diseases Research Group, Animal Health and Veterinary Laboratories Agency, New Haw, Addlestone, Surrey, KT15 3NB, UK; email: tony.fooks@ahvla.gsi.gov.uk

4.2.2: Article 16 published in Journal of Virology

Complete Genomic Sequence of Ikoma Lyssavirus

DOI:10.1128/JVI.01628-12

Summary: Ikoma lyssavirus was the first novel lyssavirus discovered in our laboratory. It was the first virus we attempted de novo assembly, and it was sequenced using the Roche 454 platform. In fact due to the longer read lengths obtained using 454 NGS, the de novo assembly was fairly successful, the issue was that each polyadenylation site (homopolymeric A) between each gene was poorly resolved. Revolving homopolymeric regions was a known issue for the 454 platform and therefore problematic for lyssavirus which have a polyadenylation site (AAAAAAAA) at the end of each gene. Therefore, ach intergenic region required Sanger sequencing to confirm the sequence. 3' and 5' RACE was attempted, but unsuccessful. Instead, RNA ligation and nested RT-PCR was undertaken by a collaborator (Dr. Ivan Kuzmin) to confirm the genomic termini.

JVI Journals.ASM.org

GENOME ANNOUNCEMENT

Complete Genome Sequence of Ikoma Lyssavirus

Denise A. Marston,[a] Richard J. Ellis,[b] Daniel L. Horton,[a] Ivan V. Kuzmin,[c] Emma L. Wise,[a] Lorraine M. McElhinney,[a,d] Ashley C. Banyard,[a] Chanasa Ngeleja,[e] Julius Keyyu,[f] Sarah Cleaveland,[g] Tiziana Lembo,[g] Charles E. Rupprecht,[c] and Anthony R. Fooks[a,d]

Wildlife Zoonoses & Vector-Borne Diseases Research Group, Animal Health & Veterinary Laboratories Agency, Addlestone, Surrey, United Kingdom[a]; Specialist Scientific Support Department, Animal Health & Veterinary Laboratories Agency, Addlestone, Surrey, United Kingdom[b]; Rabies Program, Poxvirus and Rabies Branch, Division of High-Consequence Pathogens and Pathology, Centers for Disease Control and Prevention, Atlanta, Georgia, USA[c]; University of Liverpool, National Consortium for Zoonosis Research, Leahurst, Neston, South Wirral, United Kingdom[d]; Central Veterinary Laboratory, Dar es Salaam, Tanzania[e]; Tanzania Wildlife Research Institute, Arusha, Tanzania [f]; and MRC-University of Glasgow, Centre for Virus Research, Glasgow, United Kingdom[g]

Lyssaviruses (family *Rhabdoviridae*) constitute one of the most important groups of viral zoonoses globally. All lyssaviruses cause the disease rabies, an acute progressive encephalitis for which, once symptoms occur, there is no effective cure. Currently available vaccines are highly protective against the predominantly circulating lyssavirus species. Using next-generation sequencing technologies, we have obtained the whole-genome sequence for a novel lyssavirus, Ikoma lyssavirus (IKOV), isolated from an African civet in Tanzania displaying clinical signs of rabies. Genetically, this virus is the most divergent within the genus *Lyssavirus*. Characterization of the genome will help to improve our understanding of lyssavirus diversity and enable investigation into vaccine-induced immunity and protection.

The lyssavirus genome consists of a single-stranded, negative-sense RNA, approximately 12 kb long. The RNA serves as a template for the production of five proteins; nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G), and RNA-dependent RNA polymerase protein (L). The International Committee on Taxonomy of Viruses (ICTV) has classified 12 lyssavirus species, with a further two awaiting classification; Bokeloh bat lyssavirus (BBLV) and Ikoma lyssavirus (IKOV) (1, 2, 4).

IKOV was isolated from the brain of a rabid African civet (*Civettictis civetta*), killed following an unprovoked attack on a child in Tanzania in 2009 (4). Initial phylogenetic analysis of a partial N sequence confirmed that this virus was most closely related to members of the genus *Lyssavirus*, albeit highly divergent. Whole-genome sequencing has now been undertaken to further characterize the isolate. RNA was prepared for next-generation sequencing from brain tissue stored in RNAlater (Ambion). Briefly, TRIzol-extracted viral RNA was depleted of host genomic DNA using RNase-free DNase (Qiagen, United Kingdom), and host rRNA was depleted using Terminator 5′-phosphate-dependent exonuclease (Epicentre Biotechnologies). The RNA was fragmented, and two random-primed cDNA libraries were made and run concurrently using the Roche 454 GS FLX system. The sequencing data were assembled in the GS *de novo* assembly software (Roche). Four viral contigs were identified and aligned with the genomic terminus sequences using Seqman (DNAStar), and the resulting consensus sequence was used in GS Reference Mapper (Roche) to obtain further sequence reads from the original raw data. The total number of assembled viral reads was 2,429, equating to 0.64% of the total reads. Despite the low proportion of viral sequence detected within the total data set, adequate coverage of the entire genome (with the exception of the 3′ untranslated region [UTR]) was obtained. The genomic termini were obtained using methods described previously (3). Briefly, genomic RNA was subjected to ligation by T4 RNA ligase (Promega, Madison, WI) using the manufacturer's instructions. The ligated RNA was subjected to nested PCR using primers located in the 5′ end of the L gene and 3′ end of the N gene. Amplicons were cloned using standard techniques and sequenced. Further Sanger sequencing of noncoding regions and the G gene confirmed the 454 sequence data.

The IKOV genome is 11,902 nucleotides long with 61% to 62.1% sequence identity compared to the other lyssavirus whole-genome sequences. Open reading frame lengths were similar to previously deduced lengths from other lyssavirus species: 3′ UTR, 70 nt; N gene, 1,353 nt; N P region, 66 nt; P gene, 870 nt; P-M region, 74 nt; M gene, 609 nt; M-G region, 209 nt; G gene, 1575 nt; G-L region, 569 nt; L gene, 6381 nt; and 5′ UTR, 126 nt. Analysis of glycoprotein antigenic sites revealed little conservation, suggesting that the currently available vaccine would confer limited or no cross protection. Interestingly, the L-protein has a deletion at residue 2, resulting in the protein being 2,127 residues rather than 2,128 as all other lyssavirus species. The initial and terminal 9 nt are complementary and identical to those in all other lyssaviruses. These data will contribute to our understanding of lyssavirus diversity and evolution and further our knowledge of vaccine-induced immunity and protection.

Nucleotide sequence accession number. The complete genomic sequence of IKOV has been deposited in GenBank under accession number JX193798.

ACKNOWLEDGMENTS

This work was supported by grants from The Department for Environ-

Received 26 June 2012 **Accepted** 27 June 2012

Address correspondence to Anthony R. Fooks, Tony.Fooks@ahvla.gsi.gov.uk.

doi:10.1128/JVI.01628-12

ment, Food and Rural Affairs (SE0424 and SE0427) and by the EU FP7–funded Research Infrastructure Grant European Virus Archive (no. 19 228292).

REFERENCES

1. **Banyard AC, Hayman D, Johnson N, McElhinney L, Fooks AR.** 2011. Bats and lyssaviruses. Adv. Virus Res. **79**:239–289.
2. **Freuling CM, et al.** 2011. Novel lyssavirus in Natterer's bat, Germany. Emerg. Infect. Dis. **17**:1519–1522.
3. **Kuzmin IV, Wu X, Tordo N, Rupprecht CE.** 2008. Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions. Virus Res. **136**:81–90.
4. **Marston DA, et al.** 2012. Ikoma lyssavirus, highly divergent novel lyssavirus in an African civet. Emerg. Infect. Dis. **18**:664–667.

4.2.3: Article 17 published in Journal of General Virology

Antigenic and genetic characterization of a divergent African virus, Ikoma lyssavirus

DOI:10.1099/vir.0.061952-0

Summary: The advent of sensitive molecular techniques has led to a rapid increase in the discovery of novel viruses. Detecting viral sequence alone however, only allows for prediction of phenotypic characteristics and not their measurement. In the present study we describe the in-vitro and in-vivo characterisation of IKOV, demonstrating that it is (1) pathogenic by peripheral inoculation in an animal model, (2) antigenically distinct from current rabies vaccine strains, and (3) is poorly neutralised by sera from humans and animals immunised against rabies. In a laboratory mouse model, no protection was elicited by a licensed rabies vaccine. We also investigated the role of bats as reservoirs of IKOV. We found no evidence for IKOV infection among 483 individuals of at least 13 bat species sampled across sites in the Serengeti and Southern Kenya.

Journal of General Virology (2014), **95**, 1025–1032 **DOI** 10.1099/vir.0.061952-0

Antigenic and genetic characterization of a divergent African virus, Ikoma lyssavirus

Daniel L. Horton,[1] Ashley C. Banyard,[1] Denise A. Marston,[1] Emma Wise,[1] David Selden,[1] Alejandro Nunez,[1] Daniel Hicks,[1] Tiziana Lembo,[2] Sarah Cleaveland,[2] Alison J. Peel,[3] Ivan V. Kuzmin,[4] Charles E. Rupprecht[4,5] and Anthony R. Fooks[1,6]

Correspondence
Anthony R. Fooks
tony.fooks@ahvla.gsi.gov.uk

[1]Animal Health and Veterinary Laboratories Agency (AHVLA), Weybridge, UK

[2]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK

[3]Disease Dynamics Unit, Department of Veterinary Medicine, University of Cambridge, UK

[4]Global Alliance for Rabies Control, Manhattan, KS, USA

[5]Ross University School of Veterinary Medicine, St Kitts

[6]Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, UK

In 2009, a novel lyssavirus (subsequently named Ikoma lyssavirus, IKOV) was detected in the brain of an African civet (*Civettictis civetta*) with clinical rabies in the Serengeti National Park of Tanzania. The degree of nucleotide divergence between the genome of IKOV and those of other lyssaviruses predicted antigenic distinction from, and lack of protection provided by, available rabies vaccines. In addition, the index case was considered likely to be an incidental spillover event, and therefore the true reservoir of IKOV remained to be identified. The advent of sensitive molecular techniques has led to a rapid increase in the discovery of novel viruses. Detecting viral sequence alone, however, only allows for prediction of phenotypic characteristics and not their measurement. In the present study we describe the *in vitro* and *in vivo* characterization of IKOV, demonstrating that it is (1) pathogenic by peripheral inoculation in an animal model, (2) antigenically distinct from current rabies vaccine strains and (3) poorly neutralized by sera from humans and animals immunized against rabies. In a laboratory mouse model, no protection was elicited by a licensed rabies vaccine. We also investigated the role of bats as reservoirs of IKOV. We found no evidence for infection among 483 individuals of at least 13 bat species sampled across sites in the Serengeti and Southern Kenya.

Received 5 December 2013
Accepted 28 January 2014

INTRODUCTION

The discovery of novel viruses has flourished with the advent of highly sensitive molecular detection techniques (Bexfield & Kellam, 2011; Lipkin & Firth, 2013). These include metagenomic studies of the viral flora of healthy animals aimed at predicting transmission risks to other species, and the detection of pathogens in clinical samples and excreta for diagnosis (Bodewes *et al.*, 2013; Ge *et al.*, 2012; Li *et al.*, 2011; Phan *et al.*, 2011; van den Brand *et al.*, 2012). Such studies provide valuable information on the presence and variability of viral pathogens in different animal populations. However, the detection of viral nucleic acid alone provides limited information on the zoonotic and pathogenic potential of such viruses.

Two supplementary tables and supplementary methods are available with the online version of this paper.

Recent expansion of the genus *Lyssavirus*, within the family *Rhabdoviridae*, provides an example of rapid increase in the number of novel viruses identified, with the number of lyssavirus species doubling in the past 10 years. There are 12 lyssavirus species classified by the International Committee on Taxonomy of Viruses, with two awaiting classification (Dietzgen *et al.*, 2011) and genetic evidence for a further putative lyssavirus detected in Spain (Aréchiga Ceballos *et al.*, 2013). Lyssaviruses are divided into at least two phylogroups, based on genetic and antigenic distance (Badrane *et al.*, 2001; Evans *et al.*, 2012; Hanlon *et al.*, 2005; Horton *et al.*, 2010; Table S1, available in the online Supplementary Material). All lyssaviruses cause rabies, which remains untreatable once clinical signs develop (Johnson *et al.*, 2010) with a near 100 % fatality rate. Therefore, prevention of disease relies upon pre- and post-exposure prophylaxis through administration of vaccine and immune globulin. All licensed vaccines and immune globulin

treatments are based on inactivated preparations or antibodies directed against 'classical' rabies virus (RABV). These preparations elicit protective immunological responses to all RABV variants tested, but have variable efficacy against several other lyssaviruses (Badrane *et al.*, 2001; Brookes *et al.*, 2005a; Hanlon *et al.*, 2005; Lodmell *et al.*, 1995). Although difficult to quantify, this variation is related to antigenic distances of such lyssaviruses from vaccine strains (Badrane *et al.*, 2001; Evans *et al.*, 2012; Hanlon *et al.*, 2005; Horton *et al.*, 2010). A strong immunological response to standard rabies vaccine has been shown to protect against lyssaviruses that are classified within phylogroup I, but not against viruses classified within other phylogroups (Fooks, 2004).

In 2009, an African civet (*Civettictis civetta*), with clinical signs of rabies, was killed in the Serengeti National Park (SNP), Tanzania, after biting a child. The child received appropriate wound care and post-exposure rabies prophylaxis and remained healthy. Direct antigen-detection tests on brain samples from the civet confirmed infection with a lyssavirus, and genetic analysis demonstrated that the causative agent (Ikoma lyssavirus, IKOV) was highly divergent from all lyssaviruses characterized previously (Marston *et al.*, 2012a, b). This finding was particularly significant as dogs are not permitted in the SNP, and the area of the park where the civet was encountered had been free from canine rabies for over 8 years (Lembo *et al.*, 2008). Prediction of antigenic distance and by extrapolation, degree of protection, provided by rabies vaccine strains using sequence data alone is not precise, but the degree of genetic divergence suggested that currently available vaccines would not be likely to be able to confer protection against IKOV (Evans *et al.*, 2012; Horton *et al.*, 2010).

The African civet, a solitary scavenger, does not fill the ecological niche nor have the population dynamics typical of a carnivore rabies host (Cleaveland & Dye, 1995; Estes, 1992). The few published reports of rabies in civets have been either dog- or mongoose-associated RABV variants, suggestive of spillover rather than maintenance of infection (Sabeta *et al.*, 2008). Despite dogs being responsible for the majority of human rabies cases worldwide, a variety of RABV lineages and numerous diverse lyssavirus species have been detected in wildlife, and particularly in a range of bat species (Banyard *et al.*, 2011; Hayman *et al.*, 2012; Kuzmin *et al.*, 2010). Antibodies to a phylogenetically related lyssavirus, West Caucasian bat virus (WCBV), have been detected in *Miniopterus* spp. bats in Kenya close to the border with Tanzania (Kuzmin *et al.*, 2008a), but there are no published studies showing evidence for lyssavirus antibodies in bats in the SNP. Therefore, with evidence for bat reservoirs for closely related viruses, it is possible that the natural host and reservoir for IKOV is a bat species.

The detection of genetic material from a new virus causing rabies in a potential spillover host, with no known reservoir, requires further investigation. Here we describe the isolation and characterization of viable IKOV *in vivo* and *in vitro*, assess the degree of serological cross-neutralization of IKOV with RABV and WCBV, and assess the efficacy of rabies vaccination against IKOV in an animal model. We also present results of surveillance for IKOV in bats living in close proximity to the location of the index case.

RESULTS

In vitro characterization

IKOV was not easily isolated from clinical material. Repeated blind passage of civet brain homogenate in neuroblastoma cell culture failed to amplify viable virus. IKOV was initially isolated through intracranial (IC) inoculation of 4-week-old CD-1 mice and stocks of virus were then generated by six serial passages in murine fibroblast (baby hamster kidney, BHK-21) cells. The fluorescent antibody test (FAT) was performed on brain smears post-mortem. Viral antigens were detected using FITC-conjugated anti-rabies nucleocapsid protein antibodies. This preparation of antibodies has demonstrated high sensitivity and specificity in detecting lyssaviruses (Robardet *et al.*, 2012). Virus titres at passage five ($10^{4.42}$, $TCID_{50}$ ml^{-1}) and passage six ($10^{4.8}$ $TCID_{50}$ ml^{-1}) were comparable to those seen with other lyssaviruses (Brookes *et al.*, 2005a; Horton *et al.*, 2010; Koraka *et al.*, 2012).

A modified fluorescent antibody virus neutralization (mFAVN) test was used to assess the degree of cross-neutralization between IKOV and other lyssaviruses (Brookes *et al.*, 2005b; Cliquet *et al.*, 1998). Sera from four humans and ten dogs vaccinated with commercial rabies vaccines demonstrated negligible levels of cross-neutralization against IKOV (Table 1). Three of these individuals had exceptionally high reciprocal antibody titres against RABV (Challenge Virus Standard, CVS) of over 1 : 30 000, and yet had reciprocal titres of less than 1 : 8 against IKOV, indistinguishable from negative controls and therefore effectively negative. In addition, neat human rabies immune globulin (HRIG), at a concentration of 270 IU ml^{-1}, showed no detectable neutralization of IKOV. Sera from mice inoculated with IKOV that showed reciprocal antibody titres of 1 : 27 to 1 : 420 against IKOV failed to neutralize CVS in standard FAVN tests. Furthermore, rabbit serum with a neutralizing antibody titre of over 1 : 1000 against WCBV failed to neutralize IKOV (Table 1). This lack of detectable cross-neutralization precluded accurate positioning of IKOV on an antigenic map developed previously (Horton *et al.*, 2010).

In vivo pathogenesis

All mice inoculated IC with 0.03 ml tissue culture passaged IKOV at high dose ($10^{4.8}$ $TCID_{50}$ ml^{-1}) or a 10-fold dilution ($10^{3.8}$ $TCID_{50}$ ml^{-1}) succumbed to challenge with an incubation period of between 4.5 and 6 days, with no apparent dose effect. There was, however, a detectable dose effect (albeit not statistically significant at the 95 % level) for the intramuscularly (IM) challenged mice, with 5/5 of the mice challenged IM with 0.03 ml IKOV at $10^{4.8}$ $TCID_{50}$

Table 1. Cross-neutralization of sera from rabies vaccinated humans and animals against IKOV

Neutralizing antibody titres against RABV (Challenge virus standard, CVS) and IKOV in reciprocal 50 % end point titres. International units (IU) are given for rabies by comparison to a standard control (not applicable to IKOV). Mice 3, 4 and 5 were inoculated IM with IKOV at $10^{3.8}$ $TCID_{50}$ ml^{-1} and survived. Rabbit 821 is rabbit anti-WCBV serum with a reciprocal titre of 1448 against WCBV. ND, not done.

	Virus		
	CVS (100 $TCID_{50}$)		IKOV (100 $TCID_{50}$)
Sample ID	IU ml^{-1}	Reciprocal titre	Reciprocal titre
HUMAN - 1	23	959	<8
HUMAN - 2	31	1 263	<8
HUMAN - 3	1 093	34 092	<8
HUMAN - 4	1 093	34 092	<8
DOG 1	53	1 263	<8
DOG 2	41	960	<8
DOG 3	1 094	25 904	<8
DOG 4	122	2 878	<8
DOG 5	365	8 635	<8
DOG 6	1	16	<8
DOG 7	1 094	34 109	<8
DOG 8	364	11 364	<8
DOG 9	14	421	<8
DOG 10	41	1 263	<8
HRIG	270	ND	<16
Mouse 3	<0.2	<16	27
Mouse 4	<0.2	<16	420
Mouse 5	ND	ND	81
Rabbit 821	<0.2	<16	<8

ml^{-1} succumbing between 6.5 and 9 days, but only 2/5 of the group receiving IKOV at $10^{3.8}$ $TCID_{50}$ ml^{-1} succumbing, with incubation periods of 7 and 11 days (Fischer's exact test, $P=0.08$; Fig. 1). The remaining three mice inoculated IM with the lower dose had no detectable lyssavirus antigens in the brain post-mortem, but had seroconverted when subjected to euthanasia at 28 days (Table 1). Prodromal clinical signs were similar to those recorded for other lyssaviruses in mice, including reduced appetite, ruffled fur and hunched posture (Healy *et al.*, 2013). Mice then progressed to either hind limb paralysis, or hyperexcitability and convulsions, with a predominance of the latter, and were subjected to euthanasia.

Pathology

Mice inoculated with IKOV and subject to euthanasia at a clinical score of 2 to 3 had developed non-suppurative encephalitis (Fig. 2a). The inflammatory changes were very mild, with occasional perivascular cuffing but without notable gliosis or degenerative changes in neurons. IKOV antigens were observed using immunohistochemistry (IHC) in the perikaryon and neuropil in all brain regions examined, with the majority of staining in the medulla and only rare and dispersed antigens in the cortex and thalamus (Fig. 2b).

Vaccine challenge experiments

Nineteen mice were vaccinated with one dose of a commercial rabies vaccine, and 18 of these had seroconverted against RABV by day 21 (Titre range 0.70–256 IU ml^{-1}), confirming an adequate response to vaccination. Mice vaccinated under identical conditions were previously demonstrated to be protected against IC challenge with RABV (Brookes *et al.*, 2005a). All 19 mice challenged IC with IKOV in a modified National Institutes of Health (NIH) test developed rabies, as did all the unvaccinated controls. The presence of IKOV antigens and RNA was confirmed in the brains of challenged mice.

Genetic characterization

Comparison of the full genome of IKOV with the full genome of representatives of other lyssaviruses allows interpretation of genetic relationships among the lyssaviruses (Fig. 3). Analysis of the full genomes supports the previously reported relationship of WCBV and IKOV. Lagos bat virus (LBV), Shimoni Bat virus (SHIBV) and Mokola virus (MOKV), all previously characterized as phylogroup II, comprise a strongly supported separate monophyletic group. The European bat lyssaviruses have separate ancestors. European bat lyssavirus type-1 (EBLV-1) shares a common ancestor with Duvenhage virus (DUVV) and Irkut virus (IRKV), separate from the ancestor common to European bat lyssavirus type-2 (EBLV-2) and the remaining lyssavirus species [RABV, Aravan virus (ARAV), Khujand virus (KHUV), Bokeloh bat lyssavirus (BBLV) and Australian bat lyssavirus (ABLV)]. IKOV and WCBV form a monophyletic group outside of the phylogroup I and phylogroup II lyssaviruses with significant bootstrap support, although they are separated by long genetic distance (63.4 % nucleotide identity for the concatenated coding gene sequences).

Sampling of bats in Africa for antibodies to IKOV

The common and ubiquitous free-tailed bats (*Chaerephon pumilus*) in the immediate vicinity of the index IKOV case in the SNP were roosting in the roof space of the few human dwellings, used by park employees and research scientists. These were therefore the priority for sampling in Tanzania, and were the only species encountered during sampling in SNP in 2012. Sera were available from 25 of these free-tailed bats (*C. pumilus*) sampled within SNP, and in two settlements in close proximity to the north-western boundary of the SNP. Two other species (*Hipposideros* sp., $n=16$, and *Nycteris* sp., $n=1$) were also sampled in these settlements. In Kenya, 441 individuals from at least 10

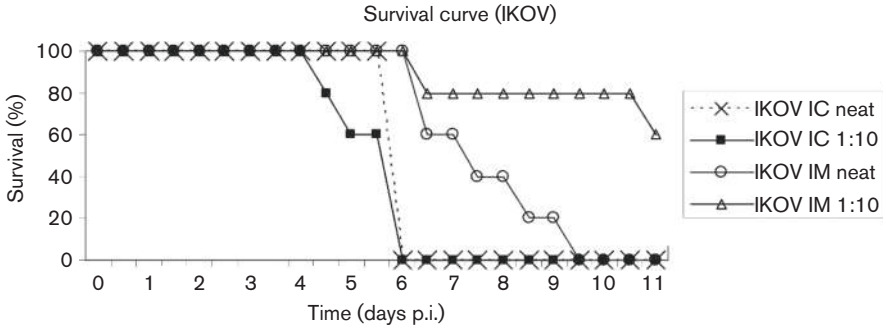


Fig. 1. Survival chart showing percentage survival in days post-inoculation (p.i) for groups of five mice inoculated either intracranially (IC) or intramuscularly (IM) with neat IKOV ($10^{4.8}$ $TCID_{50}$ ml^{-1}) or a 1/10 dilution ($10^{3.8}$ $TCID_{50}$ ml^{-1}). Mice not clinically affected after 11 days were still healthy when euthanized at 28 days and negative for lyssavirus antigens post-mortem.

different bat species were caught in multiple locations during 2011, in the framework of the Global Disease Detection Program of the US Centers for Disease Control and Prevention (CDC) (Table S2). These included 229 *Miniopterus* spp., of which 48 neutralized WCBV (titre range 1.3–3.1 $\log_{10}$ ED_{50}). All 483 of these sera from Tanzania and Kenya failed to show any neutralizing activity against IKOV (see Table S2 and Methods provided in the online Supplementary Material).

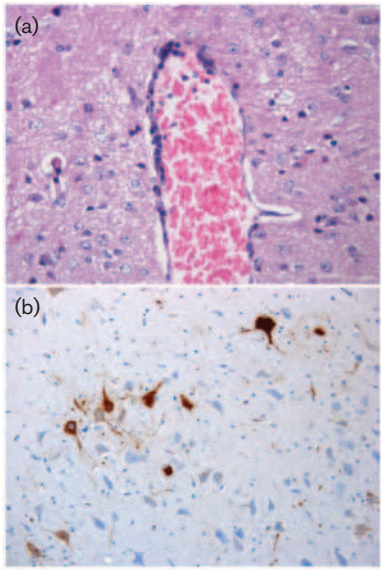


Fig. 2. Histopathological examination of brain from IKOV-infected mice. (a) Perivascular cuffing in the cortex composed of few inflammatory cells. Inflammatory changes were minimal to mild and perivascular cuffs rare and not prominent. Image shows haematoxylin and eosin staining at 400× magnification. (b). IHC demonstration of viral antigens in the medulla (brown labelling). Image shows IHC at 400× magnification.

DISCUSSION

The discovery of a novel lyssavirus (IKOV) causing rabies in an African civet, in a wildlife-rich area with potential for wildlife–human interaction, required further investigation to assess its public and animal health significance. Here, we have demonstrated that peripheral pathogenicity of IKOV is comparable to that of RABV in a rodent model. The virus is antigenically distinct to all other lyssaviruses, and vaccination with rabies vaccine produced no cross-neutralizing antibodies in humans or animals, and did not elicit protection in an animal challenge model.

The G protein is the immunodominant lyssavirus protein, and comparison of the G coding sequence of IKOV with other lyssavirus species demonstrates 46–50 % and 52–55 % similarity at the nucleotide and amino acid levels, respectively (Evans *et al.*, 2012). Previous quantitative assessments of the effect of these genetic differences on antigenicity suggested that 72 % amino acid identity along the G ectodomain is a threshold for efficient cross-neutralization (Badrane *et al.*, 2001), and that a 4.8 % difference in amino acid sequence identity over the ectodomain region would cause, on average, a twofold difference in antibody titre against a virus (Horton *et al.*, 2010). Therefore, with only 47 % amino acid identity with CVS, we would expect to see no cross-neutralization between RABV and IKOV. Here we were able to confirm this hypothesis, demonstrating a complete lack of cross-neutralization in sera of humans and animals vaccinated from rabies, even if they demonstrated very high neutralizing titres against CVS. Similar studies of phylogroup I lyssaviruses showed reduced but significant neutralization of EBLVs relative to CVS (Brookes *et al.*, 2006; Malerczyk *et al.*, 2009), and even studies of the most divergent lyssavirus prior to this discovery, WCBV, showed a limited neutralization by sera of rabbits that had high neutralizing titres against CVS (Hanlon *et al.*, 2005; Horton *et al.*, 2010). In this study, we have demonstrated a lack of cross-neutralization between IKOV and WCBV, despite the viruses having a monophyletic relationship in phylogenetic reconstructions. Our

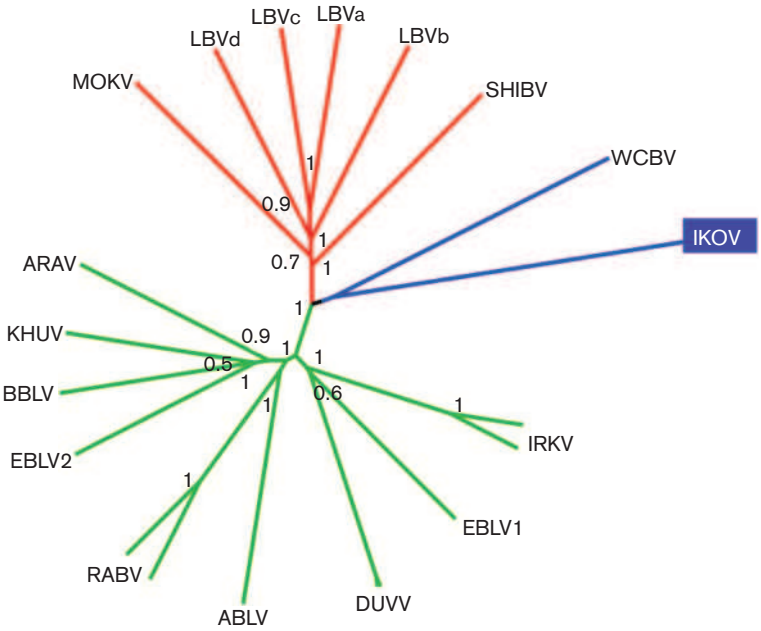


Fig. 3. Neighbour-joining phylogenetic tree of concatenated coding areas of five genes (N+P+M+G+L) of IKOV compared to representative lyssaviruses from all other species (Table S1). Branches are coloured by phylogroup: green, phylogroup one; red, phylogroup two; and blue, uncharacterized. Boostrap values (proportion of 1000 replicates) are given at each node.

observations suggest that demarcation of lyssaviruses into phylogroups may be more complex than a comparison of glycoprotein identity values and amount of cross-neutralization as proposed initially (Badrane *et al.*, 2001).

Crucially, HRIG at 270 IU ml^{-1} failed to show significant neutralization of IKOV (Table 1). In all category three exposures, HRIG is recommended as defined by the WHO (Anon, 2009), but these data demonstrated it is unlikely that HRIG would be effective against IKOV. This lack of protection afforded by rabies biologics reinforces the value of thorough wound cleaning and antiseptic treatment in cases of exposure to carnivores and bats (Anon, 2009).

The current gold standard test for assessment of rabies vaccine potency is the NIH test, which uses an IC challenge in mice vaccinated with serial dilutions of vaccine (Wilbur & Aubert, 1996). The test has been successfully modified for the assessment of protection against other viruses by using a single vaccine dilution, and mice vaccinated using this technique have previously been protected against IC challenge with RABV (Brookes *et al.*, 2005a; Wilbur & Aubert, 1996). However, all 19 mice vaccinated with commercial rabies vaccine and challenged with IKOV succumbed to disease, demonstrating that no protection was conferred. These results confirm the predictions obtained from the genetic and *in vitro* antigenic studies that rabies vaccines do not elicit protection against IKOV in mice and are therefore unlikely to provide protection in other mammals, including humans.

Analysis of full concatenated coding gene sequences of representative lyssaviruses showed similar topology to previous analyses based on partial N-gene sequences and glycoprotein sequences, suggesting that IKOV and WCBV form a monophyletic group, albeit with deeply rooted divergence (Evans *et al.*, 2012; Marston *et al.*, 2012a). One difference seen here is that in the previous partial N-gene analyses, MOKV, WCBV and IKOV formed a monophyletic group separate from LBV, which is not the case in this full genome comparison.

Knowledge of the reservoir of zoonotic diseases is important, to inform prevention of spillover infections. IKOV was detected in an African civet, an unlikely reservoir for a lyssavirus. Civets are mostly solitary scavengers, with presumed low contact rates (Estes, 1992). Consistent with this ecology is the detection of varied canid and mongoose variants of RABV in the few reported cases of rabies in civets, suggesting that civets are spillover hosts (Sabeta *et al.*, 2008). No further cases caused by IKOV have been detected, despite high levels of tourist and research-related activity in the area that have been sufficient to detect RABV in another wild carnivore since the index IKOV case. This suggests that the civet case did not result in a sustained chain of transmission. Nonetheless, a civet reservoir cannot be ruled out definitively, so vigilance for rabies in civets and other wild carnivore species still needs to be maintained, including virus characterization of any positive cases.

In light of the association of lyssaviruses with bats (Banyard *et al.*, 2011), attention has inevitably focused on bat populations in and around SNP to find the reservoir of IKOV. The IKOV-positive civet was encountered on the edge of the Serengeti plains, with few suitable roosting places for bats. Previous studies demonstrated up to 70 % seroprevalence of bats to various lyssaviruses (Harris *et al.*, 2009; Hayman *et al.*, 2012; Kuzmin *et al.*, 2010, 2008a, b; Turmelle *et al.*, 2009) but all bat sera tested in this study were negative for antibodies to IKOV. The limited sample

size, species coverage and narrow time frame preclude robustly ruling out IKOV infection in bats. However, based on these data, and seroprevalence levels demonstrated in reservoir bat populations to other lyssaviruses, it is unlikely that IKOV was circulating in the local free-tailed bat population at the time of sampling. However, the presence of other bat species, the variable success of bat capture techniques for different species, and the possibility of long-distance bat movements, suggests that the reservoir of IKOV is still likely to be in a bat species. The most recently detected putative lyssavirus species, Lleida bat lyssavirus, was detected in a *Miniopterus schreibersii* bat in Spain, and is phylogenetically related to WCBV and IKOV (Aréchiga Ceballos *et al.*, 2013). *Miniopterus* spp. bats are also a putative reservoir for WCBV in Kenya (Kuzmin *et al.*, 2008a). There was a notable absence of *Miniopterus* spp. captured in the SNP, despite the species being detected in large numbers in neighbouring Kenya. All *Miniopterus* spp. bat sera from Kenya were negative for IKOV antibodies, but there are at least 12 species of *Miniopterus* bats described in Africa (IUCN, 2013), and some species roost mainly in caves and can travel long distances to feed. Sampling of cave-dwelling bats in and around the SNP, together with serological screening of other wildlife samples from the SNP, would be a logical next step to investigating potential reservoirs.

In the experimental animal model used in this study, the incubation period, clinical signs and pathology of IKOV infection were consistent with those caused by other lyssaviruses (Healy *et al.*, 2013), and these data therefore confirm that IKOV causes rabies. The peripheral pathogenicity of IKOV concurs with recent studies showing that substitution at a key domain of the glycoprotein (K/R333D) which is shared by IKOV and phylogroup II lyssaviruses, does not result in lower peripheral pathogenicity (Badrane *et al.*, 2001; Kgaladi *et al.*, 2013; Kuzmin *et al.*, 2010, 2008b; Markotter *et al.*, 2008), in contrast to initial data based on fewer isolates (Badrane *et al.*, 2001). These results are corroborated by the fact that IKOV caused clinical rabies in the civet from which it was originally isolated, and therefore has potential to cause encephalitis in other hosts. This raises the question as to why spillover events like this do not happen more frequently. One obvious possibility is that they do, but remain undetected. IKOV was identified during dedicated studies on rabies dynamics in Tanzania, but virus characterization is not routine in most regions of Africa, with positive cases assumed to be RABV. Alternatively, species constraints might limit the number of effective spillover infections with IKOV as was suggested for other bat lyssaviruses (Banyard *et al.*, 2011). Continued enhanced surveillance, including laboratory-based confirmation of diagnosis and virus characterization, is necessary to assess and mitigate the public and animal health threats posed by IKOV and other emerging viruses (Banyard *et al.*, 2013).

METHODS

In vitro. Civet brain material was stored frozen at the Tanzania National Parks veterinary field laboratory in the Serengeti from 2009 until 2011 and then shipped to AHVLA for virus isolation and characterization. Brain homogenates from the first mouse passage were used to inoculate BHK-21 cell cultures, and the virus was passaged to a high titre for *in vivo* analysis during six passages: 0.5 ml clarified 10 % mouse brain homogenate in PBS was added to 2 ml of a BHK-21 cell suspension at 2×10^5 cells ml^{-1}, and incubated at 37 °C for 20 min with intermittent agitation. The infected cells were then added to a cell culture flask with fresh Glasgow minimal essential medium supplemented with 10 % FBS and 10 % tryptose phosphate broth. Every 3–5 days the cells were disrupted using antibiotic–trypsin–versene and transferred to a new flask with fresh media, and uninfected cells at a 1 : 1 ratio of infected to uninfected cells. Virus titre was assessed at passages five and six using techniques described previously (Aubert, 1996) and calculated with the Spearman–Kärber method. Viral antigens were visualized in acetone-fixed cells by the direct fluorescent antibody test (FAT) using standard techniques (Dean *et al.*, 1996) with an FITC-conjugated antibody (FITC anti-rabies monoclonal globulin; Fujirebio Diagnostics).

A modified fluorescent antibody neutralization test (mFAVN) was developed and optimized for IKOV using tissue culture passaged virus supernatant (TCSN) (Horton *et al.*, 2010). A standard quantity of virus (100 $TCID_{50}$ per 50 µl) was added to serial twofold dilutions of serum in duplicate, with the quantity of virus checked by back-titration on each test. Sera and virus were incubated with BHK-21 cells for 48 h before fixing in acetone and staining with FITC-conjugated antibody (Fujirebio Diagnostics). The 50 % end point serum dilution was calculated with the Spearman–Kärber method (Aubert, 1996). Mouse serum from the second mouse passage of IKOV was used as a positive control, and serum from uninfected mice was used as a negative control. A panel of sera from animals and human vaccinees, with proven high serum neutralizing antibody levels to RABV, were tested for their ability to neutralize IKOV (Table 1). A rabbit anti-WCBV serum (rabbit 821), which neutralized WCBV at a reciprocal titre of 1 : 1448, was also tested in the same assay (Horton *et al.*, 2010).

In vivo. All experimental work in animal models was undertaken under Home Office Licence after independent ethical review. Virus was first isolated from clinical material in 4-week-old OF1 mice, after unsuccessful isolation attempts in cell culture despite repeated passages using the rabies tissue culture inoculation test (RTCIT; Marston *et al.*, 2012a). Clarified civet brain homogenate (30 µl) was inoculated into mice (n=5) IC. Mice were monitored twice daily using a clinical scoring system from 1 to 5 (Healy *et al.*, 2013) and subjected to euthanasia as soon as they progressed beyond clinical score 1. Brains of mice were tested by FAT. To investigate the virulence of IKOV, four groups of five OF1 mice were challenged with either neat TCSN (high dose, neat $10^{4.8}$ $TCID_{50}$ ml^{-1}) or a 10-fold dilution (low dose $10^{3.8}$ $TCID_{50}$ ml^{-1}) by either IC or IM inoculation, followed by twice daily observation and euthanasia when clinical score progressed beyond 1. IHC demonstration of RABV antigens in post-mortem samples from brains of mice was performed as described previously (Hicks *et al.*, 2009).

To assess protection provided by vaccination, a group of 19, 4-week-old OF1 mice were vaccinated with 0.5 ml of a commercially available vaccine (VERORAB, Sanofi Pasteur MSD) by intraperitoneal injection, as standard. A group of five control mice were left unvaccinated. After 21 days a blood sample was taken from the tail vein of each mouse and the level of RABV neutralizing antibodies was assessed using a CVS pseudotype assay (Wright *et al.*, 2009). The mice were then challenged IC with IKOV TCSN at $10^{4.8}$ $TCID_{50}$ ml^{-1},

observed twice daily and subjected to euthanasia at the first sign of disease.

Molecular analyses. Nucleic acids were extracted using Trizol (Invitrogen) according to the manufacturer's instructions. A pan-lyssavirus real-time PCR assay using iScript (Bio-Rad) was then used to test for the presence of lyssavirus RNA in mouse brain (Hayman *et al.*, 2011).

Phylogenetic analyses. The full genome of IKOV was previously determined directly from clinical civet brain material using a combination of next generation sequencing and Sanger sequencing methods (GenBank accession number JX193798; Marston *et al.*, 2012b). In this report, concatenated gene sequences for all genes (N+P+M+G+L) from representatives from all lyssavirus species were compared to those of IKOV using neighbour-joining analysis, thereby avoiding the potential issue of different evolutionary rates between genes. Complete concatenated sequences were aligned using the CLUSTAL W algorithm in MEGA version 4, a neighbour-joining tree was constructed using p-distances with 1000 boostrap replicates, and visualized using FigTree(v1.2).

Field sampling of bats in Africa. To investigate the possibility of a local bat reservoir for IKOV, after ethical review and under permit from local authorities, bats were sampled both in the immediate vicinity of the index IKOV case, and in areas of human habitation (Bunda town and Fort Ikoma village) close to the north-western border of the SNP in Tanzania (n=42) in July 2012. Sera were also analysed from bats sampled in multiple locations in Southern Kenya (n=441) as part of a separate study (Table S2). All 483 serum samples from at least 11 bat species were tested for the presence of neutralizing antibodies to IKOV using mFAVN and modified rabies fluorescent focus inhibition test (see Methods section in the online Supplementary Material).

ACKNOWLEDGEMENTS

The authors thank the Tanzania Commission for Science and Technology, Tanzania National Parks, Tanzania Wildlife Research Institute, Machunde Bigambo and Frankfurt Zoological Society for field support, and the Ministry of Livestock and Fisheries Development for laboratory support in Tanzania. We also thank the Central Sequencing Unit, Stuart Ackroyd, Trudy Goddard and Hooman Goharriz for their excellent technical support at AHVLA. This work was partially supported by Defra ROAMEs SE0426/SE0427/SV3500 and by the EU FP7-funded Research Infrastructure Grant 'European Virus Archive (EVA)' (grant number 228292), and EUFP7-funded project ANTIGONE (grant number 278976) and in Tanzania by Lincoln Park Zoo, the Wellcome Trust, the UBS Optimus Foundation and the Medical Research Council. The University of Cambridge, the UK Department for Environment, Food and Rural Affairs (Defra-grant VT0105), The Charles Slater Trust, Zebra Foundation for Veterinary Zoological Education and Isaac Newton Trust provided grants to A. J. P.

REFERENCES

Anon (2009). *Current WHO Guide for Rabies Pre and Post-exposure Prophylaxis in Humans*. Geneva: World Health Organisation.

Aréchiga Ceballos, N., Vázquez Morón, S., Berciano, J. M., Nicolás, O., Aznar López, C., Juste, J., Rodríguez Nevado, C., Aguilar Setién, A. & Echevarría, J. E. (2013). Novel lyssavirus in bat, Spain. *Emerg Infect Dis* **19**, 793–795.

Aubert, M. (1996). Methods for the calculation of titres. In *Laboratory Techniques in Rabies*, pp. 445–459. Edited by F. X. Meslin, M. M. Kaplan & H. Koprowski. Geneva: World Health Organisation.

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

Banyard, A. C., Hayman, D., Johnson, N., McElhinney, L. & Fooks, A. R. (2011). Bats and lyssaviruses. *Adv Virus Res* **79**, 239–289.

Banyard, A. C., Horton, D. L., Freuling, C., Müller, T. & Fooks, A. R. (2013). Control and prevention of canine rabies: the need for building laboratory-based surveillance capacity. *Antiviral Res* **98**, 357–364.

Bexfield, N. & Kellam, P. (2011). Metagenomics and the molecular identification of novel viruses. *Vet J* **190**, 191–198.

Bodewes, R., van der Giessen, J., Haagmans, B. L., Osterhaus, A. D. & Smits, S. L. (2013). Identification of multiple novel viruses, including a parvovirus and a hepevirus, in feces of red foxes. *J Virol* **87**, 7758–7764.

Brookes, S. M., Parsons, G., Johnson, N., McElhinney, L. M. & Fooks, A. R. (2005a). Rabies human diploid cell vaccine elicits cross-neutralising and cross-protecting immune responses against European and Australian bat lyssaviruses. *Vaccine* **23**, 4101–4109.

Brookes, S. M., Aegerter, J. N., Smith, G. C., Healy, D. M., Jolliffe, T. A., Swift, S. M., Mackie, I. J., Pritchard, J. S., Racey, P. A. & other authors (2005b). European bat lyssavirus in Scottish bats. *Emerg Infect Dis* **11**, 572–578.

Brookes, S. M., Healy, D. M. & Fooks, A. R. (2006). Ability of rabies vaccine strains to elicit cross-neutralising antibodies. *Dev Biol (Basel)* **125**, 185–193.

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

Cliquet, F., Aubert, M. & Sagné, L. (1998). Development of a fluorescent antibody virus neutralisation test (FAVN test) for the quantitation of rabies-neutralising antibody. *J Immunol Methods* **212**, 79–87.

Dean, D. J., Abelseth, M. K. & Atanasiu, P. (1996). The fluorescent antibody test. In *Laboratory Techniques in Rabies,* pp. 88–93. Edited by F. X. Meslin, M. M. Kaplan & H. Koprowski. Geneva: World Health Organisation.

Dietzgen, R. G., Calisher, C. H., Kurath, G., Kuzmin, I. V., Rodriguez, L. L., Stone, D. M., Tesh, R. B., Tordo, N., Walker, P. J. & other authors (2011). Rhabdoviridae. In *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Edited by M. J. A. Andrew, M. Q. King, E. B. Carstens & E. J. Lefkowitz. Oxford: Elsevier.

Estes, R. D. (1992). Chapter 19: Family Viverridae. In *The Behavior Guide to African Mammals*, 1st edn. California: University of California Press.

Evans, J. S., Horton, D. L., Easton, A. J., Fooks, A. R. & Banyard, A. C. (2012). Rabies virus vaccines: is there a need for a pan-lyssavirus vaccine? *Vaccine* **30**, 7447–7454.

Fooks, A. (2004). The challenge of new and emerging lyssaviruses. *Expert Rev Vaccines* **3**, 333–336.

Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y. & Shi, Z. (2012). Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* **86**, 4620–4630.

Hanlon, C. A., Kuzmin, I. V., Blanton, J. D., Weldon, W. C., Manangan, J. S. & Rupprecht, C. E. (2005). Efficacy of rabies biologics against new lyssaviruses from Eurasia. *Virus Res* **111**, 44–54.

Harris, S. L., Aegerter, J. N., Brookes, S. M., McElhinney, L. M., Jones, G., Smith, G. C. & Fooks, A. R. (2009). Targeted surveillance for

European bat lyssaviruses in English bats (2003–06). *J Wildl Dis* **45**, 1030–1041, 1041.

Hayman, D. T., Banyard, A. C., Wakeley, P. R., Harkess, G., Marston, D., Wood, J. L., Cunningham, A. A. & Fooks, A. R. (2011). A universal real-time assay for the detection of Lyssaviruses. *J Virol Methods* **177**, 87–93.

Hayman, D. T., Fooks, A. R., Rowcliffe, J. M., McCrea, R., Restif, O., Baker, K. S., Horton, D. L., Suu-Ire, R., Cunningham, A. A. & Wood, J. L. N. (2012). Endemic Lagos bat virus infection in *Eidolon helvum*. *Epidemiol Infect* **140**, 2163–2171.

Healy, D. M., Brookes, S. M., Banyard, A. C., Núñez, A., Cosby, S. L. & Fooks, A. R. (2013). Pathobiology of rabies virus and the European bat lyssaviruses in experimentally infected mice. *Virus Res* **172**, 46–53.

Hicks, D. J., Nuñez, A., Healy, D. M., Brookes, S. M., Johnson, N. & Fooks, A. R. (2009). Comparative pathological study of the murine brain after experimental infection with classical rabies virus and European bat lyssaviruses. *J Comp Pathol* **140**, 113–126.

Horton, D. L., McElhinney, L. M., Marston, D. A., Wood, J. L., Russell, C. A., Lewis, N., Kuzmin, I. V., Fouchier, R. A., Osterhaus, A. D. & other authors (2010). Quantifying antigenic relationships among the lyssaviruses. *J Virol* **84**, 11841–11848.

IUCN (2013). IUCN Red List of Threatened Species. Version 2013.1.

Johnson, N., Vos, A., Freuling, C., Tordo, N., Fooks, A. R. & Müller, T. (2010). Human rabies due to lyssavirus infection of bat origin. *Vet Microbiol* **142**, 151–159.

Kgaladi, J., Nel, L. H. & Markotter, W. (2013). Comparison of pathogenic domains of rabies and African rabies-related lyssaviruses and pathogenicity observed in mice. *Onderstepoort J Vet Res* **80**, 511.

Koraka, P., Martina, B. E., Roose, J. M., van Thiel, P. P., van Amerongen, G., Kuiken, T. & Osterhaus, A. D. (2012). In vitro and in vivo isolation and characterization of Duvenhage virus. *PLoS Pathog* **8**, e1002682.

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

Kuzmin, I. V., Niezgoda, M., Franka, R., Agwanda, B., Markotter, W., Beagley, J. C., Urazova, O. Y., Breiman, R. F. & Rupprecht, C. E. (2008a). Possible emergence of West Caucasian bat virus in Africa. *Emerg Infect Dis* **14**, 1887–1889.

Kuzmin, I. V., Niezgoda, M., Franka, R., Agwanda, B., Markotter, W., Beagley, J. C., Urazova, O. Y., Breiman, R. F. & Rupprecht, C. E. (2008b). Lagos bat virus in Kenya. *J Clin Microbiol* **46**, 1451–1461.

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

Lembo, T., Hampson, K., Haydon, D. T., Craft, M., Dobson, A., Dushoff, J., Ernest, E., Hoare, R., Kaare, M. & other authors (2008). Exploring reservoir dynamics: a case study of rabies in the Serengeti ecosystem. *J Appl Ecol* **45**, 1246–1257.

Li, L., Shan, T., Wang, C., Côté, C., Kolman, J., Onions, D., Gulland, F. M. & Delwart, E. (2011). The fecal viral flora of California sea lions. *J Virol* **85**, 9909–9917.

Lipkin, W. I. & Firth, C. (2013). Viral surveillance and discovery. *Curr Opin Virol* **3**, 199–204.

Lodmell, D. L., Smith, J. S., Esposito, J. J. & Ewalt, L. C. (1995). Cross-protection of mice against a global spectrum of rabies virus variants. *J Virol* **69**, 4957–4962.

Malerczyk, C., Selhorst, T., Tordo, N., Moore, S. & Müller, T. (2009). Antibodies induced by vaccination with purified chick embryo cell culture vaccine (PCECV) cross-neutralize non-classical bat lyssavirus strains. *Vaccine* **27**, 5320–5325.

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

Marston, D. A., Horton, D. L., Ngeleja, C., Hampson, K., McElhinney, L. M., Banyard, A. C., Haydon, D., Cleaveland, S., Rupprecht, C. E. & other authors (2012a). Ikoma lyssavirus, highly divergent novel lyssavirus in an African civet. *Emerg Infect Dis* **18**, 664–667.

Marston, D. A., Ellis, R. J., Horton, D. L., Kuzmin, I. V., Wise, E. L., McElhinney, L. M., Banyard, A. C., Ngeleja, C., Keyyu, J. & other authors (2012b). Complete genome sequence of Ikoma lyssavirus. *J Virol* **86**, 10242–10243.

Phan, T. G., Kapusinszky, B., Wang, C., Rose, R. K., Lipton, H. L. & Delwart, E. L. (2011). The fecal viral flora of wild rodents. *PLoS Pathog* **7**, e1002218.

Robardet, E., Demerson, J. M., Andrieu, S. & Cliquet, F. (2012). First European interlaboratory comparison of tetracycline and age determination with red fox teeth following oral rabies vaccination programs. *J Wildl Dis* **48**, 858–868.

Sabeta, C. T., Shumba, W., Mohale, D. K., Miyen, J. M., Wandeler, A. I. & Nel, L. H. (2008). Mongoose rabies and the African civet in Zimbabwe. *Vet Rec* **163**, 580.

Turmelle, A. S., Allen, L. C., Jackson, F. R., Kunz, T. H., Rupprecht, C. E. & McCracken, G. F. (2009). Ecology of rabies virus exposure in colonies of Brazilian free-tailed bats (*Tadarida brasiliensis*) at natural and man-made roosts in Texas. *Vector Borne Zoonotic Dis* **10**, 165–175.

van den Brand, J. M., van Leeuwen, M., Schapendonk, C. M., Simon, J. H., Haagmans, B. L., Osterhaus, A. D. & Smits, S. L. (2012). Metagenomic analysis of the viral flora of pine marten and European badger feces. *J Virol* **86**, 2360–2365.

Wilbur, L. A. & Aubert, M. (1996). NIH test for potency. In *Laboratory Techniques in Rabies*, 4th edn, pp. 360–368. Edited by F. X. Meslin, M. M. Kaplan & H. Koprowski. Geneva: World Health Organisation.

Wright, E., McNabb, S., Goddard, T., Horton, D. L., Lembo, T., Nel, L. H., Weiss, R. A., Cleaveland, S. & Fooks, A. R. (2009). A robust lentiviral pseudotype neutralisation assay for in-field serosurveillance of rabies and lyssaviruses in Africa. *Vaccine* **27**, 7178–7186.

4.2.4: Article 18 published on ICTV website

IKOV ICTV classification document

https://data.ictvonline.org/proposals/2012.009aV.A.v4.Lyssavirus_sp.pdf

Summary: The complete genome sequence, in vitro and in vivo characteristics were combined in a new species ICTV document for consideration by the Rhabdovirus study group. The proposal was approved and IKOV was accepted as a virus in the new lyssavirus species *Ikoma lyssavirus*.

This form should be used for all taxonomic proposals. Please complete all those modules that are applicable (and then delete the unwanted sections). For guidance, see the notes written in blue and the separate document "Help with completing a taxonomic proposal"

Please try to keep related proposals within a single document; you can copy the modules to create more than one genus within a new family, for example.

MODULE 1: **TITLE, AUTHORS, etc**

Code assigned:	***2012.009aV***	(to be completed by ICTV officers)

Short title: create 1 new species in the genus *Lyssavirus*
(e.g. 6 new species in the genus *Zetavirus*)

Modules attached (modules 1 and 9 are required)					
	1 ☒	**2** ☒	**3** ☐	**4** ☐	**5** ☐
	6 ☐	**7** ☐	**8** ☐	**9** ☒	

Author(s) with e-mail address(es) of the proposer:

Tony Fooks (Tony.Fooks@ahvla.gsi.gov.uk) Denise Marston (Denise.Marston@ahvla.gsi.gov.uk),

List the ICTV study group(s) that have seen this proposal:

A list of study groups and contacts is provided at http://www.ictvonline.org/subcommittees.asp . If in doubt, contact the appropriate subcommittee chair (fungal, invertebrate, plant, prokaryote or vertebrate viruses)	Rhabdoviridae Study Group

ICTV-EC or Study Group comments and response of the proposer: Supported with some modifications, which have all been completed.

Date first submitted to ICTV: 21/06/2013
Date of this revision (if different to above):

MODULE 2: **NEW SPECIES**

creating and naming one or more new species.

If more than one, they should be a group of related species belonging to the same genus. All new species must be placed in a higher taxon. This is usually a genus although it is also permissible for species to be "unassigned" within a subfamily or family. Wherever possible, provide sequence accession number(s) for one isolate of each new species proposed.

Code	***2012.009aV***	(assigned by ICTV officers)

To create 2 new species within:

Genus:	***Lyssavirus***
Subfamily:	
Family:	***Rhabdoviridae***
Order:	***Mononegavirales***

Fill in all that apply.

- If the higher taxon has yet to be created (in a later module, below) write "**(new)**" after its proposed name.
- If no genus is specified, enter "**unassigned**" in the genus box.

And name the new species:	**GenBank sequence accession number(s) of reference isolate:**
Ikoma lyssavirus	JX193798

Reasons to justify the creation and assignment of the new species:

- Explain how the proposed species differ(s) from all existing species.
  - If species demarcation criteria (see module 3) have previously been defined for the genus, **explain how the new species meet these criteria**.
  - If criteria for demarcating species need to be defined (because there will now be more than one species in the genus), please state the proposed criteria.
- Further material in support of this proposal may be presented in the Appendix, Module 9

In general, demarcation criteria for lyssavirus species include (Dietzgen et al., 2011):

1. Genetic distances, with the threshold of 80–82% nucleotide identity for the complete N gene, that provides a better quantitative resolution compared to other genes, or 80–81% nucleotide identity for concatenated coding regions of N+P+M+G+L genes. Globally, all isolates belonging to the same species have higher identity values than the threshold, except the viruses currently included into the LBV species. For that reason some authors suggested that LBV be subdivided into several genotypes. However, as these LBV representatives are segregated into a monophyletic cluster in the majority of phylogenetic reconstructions, in the absence of other sufficient demarcation characters there is currently no possibility to subdivide LBV into several viral species.

2. Topology and consistency of phylogenetic trees, obtained with various evolutionary models.

3. Antigenic patterns in reactions with anti-nucleocapsid monoclonal antibodies (preceded by serologic cross-reactivity and definition of lyssavirus serotypes, using polyclonal antisera).

4. Whenever available, additional characters, such as ecological properties, host and geographic range, pathological features are recruited.

Phylogenetic and serological relationships correlate, which has enabled the delineation of two major phylogroups within the genus *Lyssavirus* (Badrane et al., 2001). Phylogroup I, includes rabies virus (RABV), Duvenhage (DUVV), European bat lyssaviruses type 1 and 2 (EBLV-1 and -2), Australian bat lyssavirus (ABLV), Aravan virus (ARAV), Khujand virus (KHUV), and Irkut virus (IRKV). Phylogroup II, includes Lagos bat virus (LBV), Mokola virus (MOKV), and Shimoni bat virus (SHIBV). West Caucasian bat virus (WCBV) does not cluster in either phylogroup and has been suggested to be a representative of Phylogroup III. There is significant cross-neutralization within phylogroups, but very limited, to no cross-neutralization detected between phylogroups.

Based on the criteria above, Ikoma lyssavirus belongs to the *Lyssavirus* genus but cannot be included in any of the existing species. A new species should be created for this virus. The evidence is documented (Marston et al., 2012a and b).

Ikoma Lyssavirus (IKOV)

- Isolated from an African Civet (*Civetticus civetta*) in Tanzania in 2009. The African civet was displaying clinical signs consistent with rabies and had attacked a child without provocation.
- Pathogenic to laboratory mice via intracranial and peripheral inoculation, causing acute progressive fatal encephalitis (rabies).
- During infection, IKOV forms typical to lyssavirus, intracytoplasmic inclusions, detected by staining with FITC-conjugated anti-nucleocapsid monoclonal antibodies.
- The complete genome (Genbank JX193798) of IKOV consists of 11,902 nucleotides and includes 5 genes: Nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and RNA-dependent RNA polymerase (L) (see Annex Table 2 for comparison of gene and intergenic region sizes for all lyssaviruses).
- IKOV demonstrates 63.1-63.5% identity to other lyssaviruses using concatenated 5 gene sequences, much lower than the threshold of 80-81% (Annex Table 1)
- Phylogenetically IKOV belongs to the *Lyssavirus* genus (Annex Figure 1). Within the genus it is placed ancestrally to all other lyssaviruses and is the most divergent lyssavirus described to date. It groups with WCBV outside of both phylogroup I and II using Neighbour-joining, Bayesian and Maximum likelihood models (Annex Figure 1).
- Cross-neutralizing assays using a panel of human and dog sera with proven high serum neutralizing antibodies to RABV (CVS) clearly show little/no neutralization of the vaccinated sera to IKOV (Table 3)
- IKOV does not readily cross-react with WCBV (most closely phylogenetically related lyssavirus). Miniopterus bat sera neutralizing WCBV did not neutralize IKOV.

MODULE 9: **APPENDIX**: supporting material

additional material in support of this proposal

References:

Bourhy H, Kissi B, Tordo N (1993) Molecular diversity of the Lyssavirus genus. Virology 194: 70-81.

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

Delmas O, Holmes EC, Talbi C, Larrous F, Dacheux L, et al. (2008) Genomic diversity and evolution of the lyssaviruses. PLoS One 3: e2057.

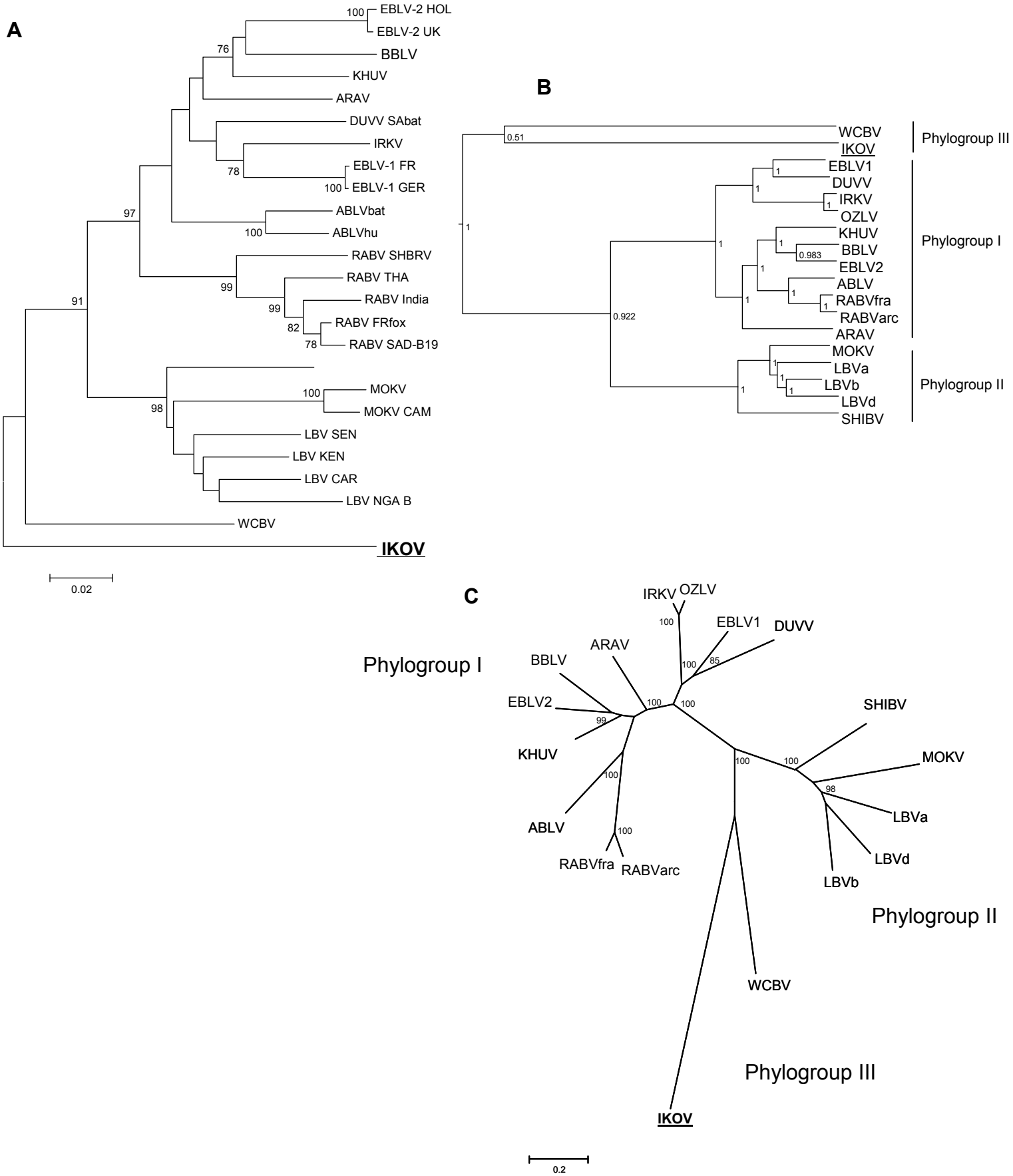
Dietzgen, R. G., Calisher, C. H., Kurath, G., Kuzmin, I. V., Rodriguez, L. L., Stone, D. M., Tesh, R. B., Tordo, N., Walker, P. J., Wetzel, T. and Whitfield, A. E. (2011). Rhabdoviridae. In Andrew M. Q. King, Michael J. Adams, Eric B. Carstens and Elliot J. Lefkowitz (Ed.), *Virus taxonomy: Ninth report of the International Committee on Taxonomy of Viruses* (pp. 654-681) Oxford, United Kingdom: Elsevier.

Marston DA, Horton DL, Ngeleja C, Hampson K, McElhinney LM, et al. (2012) Ikoma lyssavirus, highly divergent novel lyssavirus in an african civet. Emerg Infect Dis 18: 664-667.

Marston DA et al., (2012). Complete genomic sequence of Ikoma lyssavirus. J. Virol. 86 (18): 10242-10243.

Annex:

Include as much information as necessary to support the proposal, including diagrams comparing the old and new taxonomic orders. The use of Figures and Tables is strongly recommended but direct pasting of content from publications will require permission from the copyright holder together with appropriate acknowledgement as this proposal will be placed on a public web site. For phylogenetic analysis, try to provide a tree where branch length is related to genetic distance.


A
EBLV-2 HOL
EBLV-2 UK
BBLV
KHUV
ARAV
DUVV SAbat
IRKV
EBLV-1 FR
EBLV-1 GER
ABLVbat
ABLVhu
RABV SHBRV
RABV THA
RABV India
RABV FRfox
RABV SAD-B19
MOKV
MOKV CAM
LBV SEN
LBV KEN
LBV CAR
LBV NGA B
WCBV
IKOV
0.02
B
WCBV
IKOV
EBLV1
DUVV
IRKV
OZLV
KHUV
BBLV
EBLV2
ABLV
RABVfra
RABVarc
ARAV
MOKV
LBVa
LBVb
LBVd
SHIBV
Phylogroup III
Phylogroup I
Phylogroup II
C
Phylogroup I
IRKV
OZLV
EBLV1
DUVV
ARAV
BBLV
EBLV2
KHUV
ABLV
RABVfra
RABVarc
SHIBV
MOKV
LBVa
LBVd
LBVb
Phylogroup II
WCBV
Phylogroup III
IKOV
0.2

Annex Figure 1. Phylogenetic reconstructions of the *Lyssavirus* genus (A), neighbour-joining using MEGA 5 and (B), Bayesian, using Mr Bayes based on full N-gene sequences, and (C) full genome sequence with maximum likelihood evolutionary models using MEGA 5. Significant bootstrap values or posterior probabilities are shown for key nodes.

Species	RABV	LBV	MOKV	DUVV	EBLV-1	EBLV-2	ABLV	ARAV	KHUV	IRKV	WCBV	SHIBV	**BBLV**	**IKOV**
RABV	81.6-92.3													
LBV	67.0-67.7	76.2												
MOKV	66.9-67.3	73.3-74.1	86.6											
DUVV	71.2-71.8	67.4-67.6	67.0-67.1	98.9										
EBVL-1	71.7-72.4	68.1-68.5	67.3-67.7	76.1	95.6-98.1									
EBLV-2	72.7-73.8	67.3-68.1	67.8-68.0	73.1-73.3	74.2-74.4	98.2								
ABLV	73.2-73.8	67.0-67.4	66.4-66.8	71.2	72.2-72.3	73.9								
ARAV	72.9-73.2	68.2-68.3	67.7-68.1	73.4-73.5	75.4-75.5	76.9	73.6							
KHUV	72.9-73.4	67.5-68.0	67.1-67.3	73.5-73.6	74.7	78.7-78.9	74.5	77.5						
IRKV	71.5-72.3	67.9-68.5	67.7-68.3	74.3-74.4	76.3-76.5	73.9	71.6	73.6-74.2	73.6-74.3	91.9				
WCBV	64.8-65.5	65.8-66.0	65.3-65.5	65.8	65.5-65.7	65.5	65.2	65.7	65.4	65.2				
SHIBV	67.2-67.7	73.8-75.1	71.9-72.0	67.7-67.8	68.2-68.3	68.1	67.2	68.1	67.9	68.7	66.4			
BBLV	72.7-73.6	67.5-67.9	67.6-68.1	73.1	74.2-74.3	78.2	74.3	76.3	78.4	73.6	65.1	68.7		
IKOV	61.9-62.5	62.7-62.9	62.3-62.5	62.5-62.6	62.6-62.7	62.8	62.3	62.6	62.4	62.4	63.2	63.5	62.5	

Table 1. Nucleotide identity values for concatenated coding regions (N, P, M, G and L genes) of IKOV in comparison with lyssaviruses from all identified species. Full genomes of BBLV (JF311903), RABV (M31046; EU293111; EU293115; EU293113; EU293116), DUVV (EU293120; EU293119), EBLV-1 (EU293109, EU293112, EF157976), IKOV (JX193798), ABLV (NC_003243; AF081020), KHUV (EF614261), IRKV (FJ905105, EF614260), EBLV-2 (EF157977, EU293114), ARAV (EF614259), LBV (EU293108; EU293110), MOKV (EU293118, EU293117), SHIBV (GU170201) and WCBV (EF614258) were derived from NCBI Genbank. Concatenated sequences were aligned using ClustalW and a distance matrix was calculated as implemented in BioEdit.

Table 2: The lengths of coding and non-coding regions of lyssavirus genomes.

	RABV	LBV	MOKV	DUVV	EBLV-1	EBLV-2	ABLV	ARAV	KHUV	IRKV	WCBV	SHIBV	**BBLV**	**IKOV**
3' UTR*	70	70	70	70	70	70	70	70	70	70	70	70	70	70
N protein	**1353**	**1353**	**1353**	**1356**	**1356**	**1356**	**1353**	**1356**	**1356**	**1356**	**1353**	**1353**	**1356**	**1353**
N–P	90-1	101	100-102	90	90	101	94	85	95	93	64	98	91	66
P protein	**894**	**918**	**912**	**897**	**897**	**894**	**894**	**894**	**894**	**897**	**894**	**918**	**894**	**870**
P–M	88	75	80	83	83	88	89	85	72	82	133	76	86	74
M protein	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**
M–G	211-5	204	203-204	191	211	210 (205)	207-209	210	208	214	206	205	210	209
G protein	**1575**	**1569**	**1569**	**1602**	**1575**	**1575**	**1578-1581**	**1581**	**1581**	**1575**	**1578**	**1569**	**1575**	**1575**
G–L	522	578-588	546-563	562-563	560	512	508-509	514	504	569	862	613	496	569
L protein	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6381**
5' UTR	131	145	112-114	130-131	131	131	131	130	130	131	125	150	129	126
Genome	**11 923-8**	**12006-16**	**11 940-57**	**11 975-6**	**11 966**	**11 930**	**11 918**	**11918**	**11903**	**11980**	**12278**	**12045**	**11900**	**11902**

Table 3. Cross neutralization of sera from rabies vaccinated humans and animals against IKOV in comparison with RABV. International units (IU) are given for RABV by comparison to a standard control (not applicable to IKOV). Reciprocal titres of over 10,000 are considered exceptionally high, and less than 20 would be considered below the detectable threshold and therefore effectively negative.

	VIRUS		
		CVS	**IKOV**
Sample ID	IU/ml	Reciprocal titre	Reciprocal titre
HUMAN - 1	23	959	<8
HUMAN - 2	31	1263	<8
HUMAN - 3	1093	34092	<8
HUMAN - 4	1093	34092	<8
DOG 1	53	1263	<8
DOG 2	41	960	<8
DOG 3	1094	25904	8
DOG 4	122	2878	<8
DOG 5	365	8635	<8
DOG 6	1	16	<8
DOG 7	1094	34109	<8
DOG 8	364	11364	<8
DOG 9	14	421	<8
DOG 10	41	1263	<8
HRIG	270	nd	<16

4.2.5: Article 19 published in Emerging and Infectious Diseases

Lyssavirus in Indian Flying Foxes, Sri Lanka

DOI:10.3201/eid2208.151986

Summary: Gannoruwa bat lyssavirus (GBLV) was the second novel lyssavirus discovered in our laboratory. The complete genome sequence was obtained from four GBLV positive bats using the Illumina platform. The de novo assembly of host depleted reads using Velvet resulted in 4 genomes which varied in the number of viral read between 86,013 and 3,123,481.uring the assembly process, the viral contigs were used to 'blast' the viral database, and three different RABV genomes were identified as the top match for the 4 samples (AB569299 – RABV Sri Lanka, KF154999 - RABV Sri Lanka and JQ944706 – RABV Russia), none of which had high homology. Therefore, the resulting consensus sequences, obtained mapping the viral reads against these reference genomes using BWA, were almost unusable, despite the apparent high coverage. The four genome sequences were aligned and primers designed to span the regions with gaps/errors. The assumption was made that all four GBLV positive samples will contain highly related viral genome sequences and therefore a reference sequence was spliced together using the the majority sequence. Combined with the Sanger amplicon sequences, this modified reference sequence was used to map the reads, resulting in complete genomes for all four isolates.

Lyssavirus in Indian Flying Foxes, Sri Lanka

Panduka S. Gunawardena,[1] Denise A. Marston,[1] Richard J. Ellis, Emma L. Wise, Anjana C. Karawita, Andrew C. Breed, Lorraine M. McElhinney, Nicholas Johnson, Ashley C. Banyard, Anthony R. Fooks

A novel lyssavirus was isolated from brains of Indian foxes (*Pteropus medius*) in Sri Lanka. Phylogenetic analysis of complete virus genome sequences, and geographic location and host species, provides strong evidence that this virus is a putative new lyssavirus species, designated as Gannoruwa bat lyssavirus.

There are 14 recognized species in the genus *Lyssavirus*: rabies virus (RABV), Lagos bat virus, Mokola virus (MOKV), Duvenhage virus, European bat lyssavirus types 1 and 2, Australian bat lyssavirus (ABLV), Aravan virus (ARAV), Khujand virus, Irkut virus, Shimoni bat virus, Bokeloh bat lyssavirus, West Caucasian bat virus, and Ikoma lyssavirus (IKOV) (*1*). RABV has a global distribution and is the dominant lyssavirus circulating in nonvolant (incapable of mammals across Asia, including Sri Lanka. Bats are known reservoir hosts of all lyssaviruses except MOKV and IKOV. Discovery of new lyssaviruses in bats has stimulated research and surveillance efforts to identify additional members of this genus in bat populations (*2*).

Although lyssaviruses circulate in bats in Asia (*2*), RABV in bats in Asia remains Irkut virus was the bat lyssavirus in China (*3*). ARAV, Khujand virus, and West Caucasian bat virus have been isolated exclusively from insectivorous bats in Eurasia. Pathogen discovery in insectivorous and hematophagous bats is progressing. However, surveillance for lyssaviruses in fruit bats remains limited, particularly across Asia. Frugivorous bats in the Americas, which are distant genetically from bats of the family *Pteropodidae*, are independent reservoirs of RABV (*4*).

Although several regions contain fruit bats of the genus *Pteropus*, only pteropid bats in Australia have been as reservoirs for a lyssavirus species, ABLV, which has been isolated from all 4 *Pteropus* species in Australia. Moreover, ABLV has also been detected in at least 1 insectivorous bat (*Saccolaimus*) (*5*). Although antibodies have been detected in bats from several countries in Asia (*2*), the only lyssaviruses reportedly isolated from fruit bats in Asia have not been characterized (*6*,*7*).

In Sri Lanka, lyssavirus surveillance has focused on canine RABV as the primary public health concern. The Indian fox (*P. medius*, formerly known as *P. giganteus)*, is a large frugivorous and nectarivorous bat that lives in forest, urban, and rural areas and is one of the most persecuted (e.g., cutting down of roosting trees and hunting) bats in southern Asia (*8*). These bats can long distances (≤150 km) to forage and have a wide distribution (India, China, Bangladesh, Bhutan, Myanmar, the Maldives, Nepal, Pakistan, and Sri Lanka). We report of a

The Study

Ethical clearance was obtained from the ethics committee of the Faculty of Veterinary Medicine and Animal Science at the University of Peradeniya (Peradeniya, Sri Lanka) and the Animal and Plant Health Agency (Addlestone, UK). Specimens were collected under permit no. WL/3/2/62/14 from the Sri Lanka Department of Wildlife Conservation.

During January 1, 2014–October 31, 2015, a total of 62 grounded bats were collected in an area inhabited by a long-established roost of ≈20,000 Indian foxes in Gannoruwa, Peradeniya, Sri Lanka (7°16′N, 80°36′E), which is located 600 m above sea level. Most bats were found dead. One bat (AK-42), which had clinical signs of illness, died shortly after capture (Table).

The bat collected (AK-15) was a fresh carcass of a mature male that weighed 1.5 kg. A detailed necropsy showed that the animal had been healthy and had well-developed pectoral muscles. Except for a few multifocal hemorrhages in the lungs and mild, diffuse hyperemia and edema in the brain, gross pathologic were unremarkable. However, Negri bodies of various sizes were in the brain (Figure 1, panel A). Numerous aggregations of lyssavirus nucleocapsid antigen were observed in brain smears subjected to a direct antibody test (dFAT) (Figure 1, panel B). Histopathologic examination of brain and spinal cord showed mild nonsuppurative lesions, leptomeningitis, and encephalomyelitis. Three additional dFAT-positive samples were from the 62

Author University of Peradeniya, Peradeniya, Sri Lanka (P.S. Gunawardena, A.C. Karawita); Animal and Plant Health Agency, Addlestone, UK (D.A. Marston, R.J. Ellis, E.L. Wise, A.C. Breed, L.M. McElhinney, N. Johnson, A.C. Banyard, A.R. Fooks); University of Liverpool, Liverpool, UK (L. M. McElhinney, A.R. Fooks)

DOI: http://dx.doi.org/10.3201/eid2208.151986

[1]These authors contributed equally to this article.

Table. Characteristics of 4 Indian flying foxes infected with Gannoruwa bat lyssavirus, Sri Lanka*

Original ID no.	APHA ID no.	Collection date	Location	Weight, g/age/sex	Clinical signs/condition	GenBank accession no.
AK-15	RV3266	2014 Sept 17	Peradeniya	1,500/mature/M	Dead at collection	KU244266
AK-40	RV3267	2015 May 8	Peradeniya	350/immature/F	Dead at collection	KU244267
AK-42	RV3268	2015 May 25	Peradeniya	500/immature/M	Cachectic, paresis, unable to fly, nystagmus, intermittent seizures (≈10 s), spontaneous vocalization, aggressiveness, biting, died shortly after capture	KU244268
AK-74	RV3269	2015 Sep 11	Gannoruwa	212.5/immature/F	Dead at collection	KU244269

*APHA, Animal and Plant Health Agency; ID, identification.

bats tested (Table). Subsequent virus isolation and molecular analysis were conducted for these 4 brain samples.

Virus was isolated by using N2A cells (*9*). After 5 days of incubation, 3 of 4 samples were positive for virus. Two of the isolates, RV3267 and RV3269, were subsequently cultured in BHK cells. RNA was extracted by using TRIzol reagent (Invitrogen, Paisley, UK). A pan-lyssavirus reverse transcription PCR yielded a 606-bp amplicon for the virus nucleoprotein gene (*10*). Results for a differential real-time reverse transcription PCR with a TaqMan probe for RABV showed no for the 4 RNA samples. A 145-bp amplicon was visualized after electrophoresis on a 2% agarose gel. Thus, pan-lyssavirus primers used in the real-time assay detected this virus, but the probe did not bind to the amplicon, which suggested presence of a non-RABV lyssavirus.

Complete genome sequences (GenBank accession nos. KU244266–9) were obtained from brain RNA samples by using next-generation sequencing according to previous methods (*11*,*12*). Phylogenetic analysis of complete genome sequences, including representatives of all lyssavirus species, showed that sequences of the new non-RABV lyssavirus clustered with each other and had a common ancestor with ABLV and RABV in phylogroup 1 (Figure 2). This novel virus was designated as Gannoruwa bat lyssavirus (GBLV).

Representative canine and golden palm civet RABV sequences from Sri Lanka were included in the dataset, but those sequences clustered with other RABVs, distinct from the GBLV sequence (nucleotide identity 78%). Nucleotide identity across the complete genome ranged from 61% (IKOV) to 76.5% (ABLV), which showed that GBLV is a member of the genus *Lyssavirus* but is distinct from viruses circulating in nonvolant mammals in Sri Lanka.

Conclusions

We report isolation of a novel non-RABV lyssavirus (GBLV) that is most closely related to RABV and ABLV. GBLV is pathogenic; it caused fatal disease in 4 Indian foxes, and clinical signs for these foxes were similar to those observed in other bat lyssavirus infections (Table). Diagnostic tests ed Negri bodies by staining with hematoxylin and eosin and lyssavirus antigens by dFAT in brain and spinal cord tissue (Figure 1). Molecular techniques lyssavirus nucleic acid, and full-genome analysis indicated that GBLV was divergent from known RABVs circulating in Sri Lanka (Figure 2).

Although rabies is prevalent in Sri Lanka, and a number of wildlife species have been as being rabid, most of the RABVs involved have not been genetically typed. Furthermore, over a 12-year period, only 1 bat tested for RABV was shown to be uninfected (*13*).

We report a novel non-RABV lyssavirus in Sri Lanka, which indicates that Indian foxes are a reservoir for lyssaviruses on the Indian subcontinent and nearby regions. Indian foxes are widespread in urban and rural areas and occasionally come in contact with humans and domestic dogs, which provides opportunities for virus spillover. Indian foxes are also reservoirs for

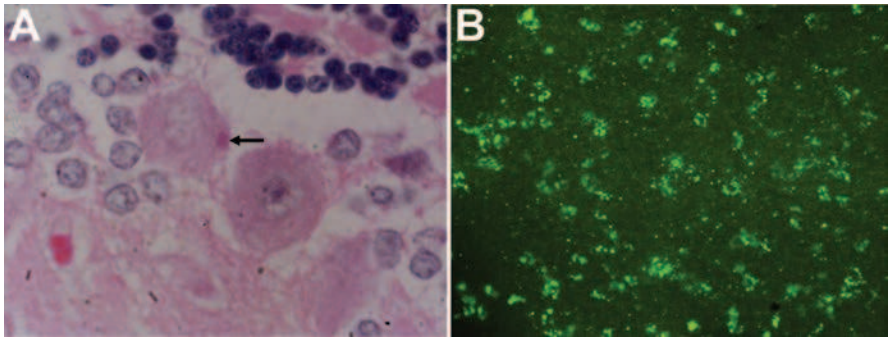


Figure 1. Negri bodies and lyssavirus antigens in brain tissue from an Indian fox, Sri Lanka. A) Degenerate Purkinje's cell with an eosinophilic, intracytoplasmic inclusion body and a Negri body (arrow). Hematoxylin and eosin stain, original ×1,000. B) Fluorescence indicative of lyssavirus nucleoprotein in a brain smear subjected to a direct antibody test with isothiocyanate–conjugated monoclonal antibody. Original ×100. A color version of this is available online (http://wwwnc.cdc.gov/EID/article/22/8/15-1986-F1.htm).

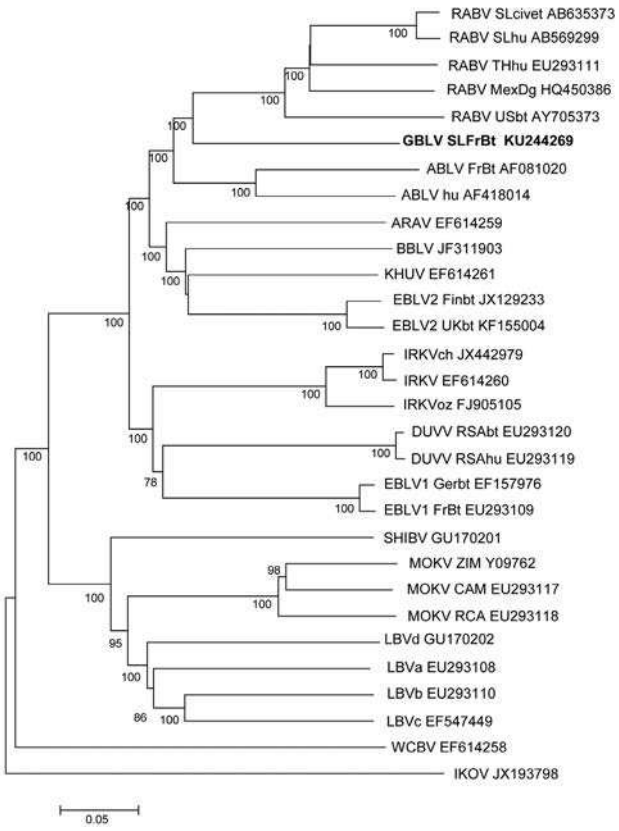


Figure 2. Phylogenetic relationships between representatives from all lyssaviruses and novel Gannoruwa bat lyssavirus (GBLV) on the basis of complete genome sequences. All 4 GBLV sequences form a monophyletic clade and are >99.9% identical across the genome; therefore, only 1 sequence (in bold) is shown. Relationships are shown as an unrooted phylogram, which was constructed by using the maximum-likelihood method and a general time reversible plus gamma distribution plus proportion of invariable sites model, and are by using the model test implemented in MEGA6 (http://www.megasoftware.net). Bootstrap values ≥70 (1,000 replicates) are indicated next to branches; sequences are listed with GenBank accession numbers. RABV, rabies virus; ABLV, Australian bat lyssavirus; ARAV, Aravan virus; BBLV, Bokeloh bat lyssavirus; KHUV, Khujand virus; EBLV, European bat lyssavirus; IRKV, Irkut virus; DUVV, Duvenhage virus; SHIBV, Shimoni bat virus; MOKV, Mokola virus; LBV, Lagos bat virus; WCBV, West Caucasian bat virus; IKOV, Ikoma lyssavirus. Scale bar indicates nucleotide substitutions per site.

Nipah virus in Bangladesh and India, where transmission to humans has resulted in outbreaks and human deaths. Other bat species, including insectivorous bats, might also be reservoir hosts for lyssaviruses in the study region. Thus, further surveillance is required to understand the role that bats play in the epidemiology of lyssaviruses in Asia.

Continued and extended surveillance of bats and other mammalian species is necessary to determine the distribution and prevalence of GBLV. Detailed phylogenetic analysis and monoclonal typing and antigenic mapping will help clarify the evolutionary relationship between GBLV and other lyssaviruses, in particular RABV and ABLV. In vitro and in vivo cross-neutralization and protection studies will elucidate properties of GBLV and provide information on protection from this virus by available prophylaxis.

Acknowledgments

We thank O. Perera, F.A. Leighton, D.R.T.G. Ratnayake, and T. Wijayathilaka for providing guidance and facilitating testing of samples submitted for virus characterization; Waruni Jayaweera, David Selden, and Hooman Goharriz for providing expert technical

This study was supported by Department for Environment, Food and Rural Affairs, the Scottish Government and Welsh Government through grants SE0426, SE0427, and SV3500; the European Union Horizon 2020 Research and Innovation Program under grant agreement no. 653316 (European Virus Archive Global); and the International Development Research Centre of Canada (Project Building Research Excellence in Wildlife and Human Health in Sri Lanka).

Dr. Gunawardena is a senior lecturer and head of the Division of Veterinary Pathology, Faculty of Veterinary Medicine and Animal Science, University of Peradeniya, Peradeniya, Sri Lanka. His main research interests include disease etiology, pathogenesis, and control of neurologic diseases in animals.

References

1. Dietzgen R, Calisher, CH, Kurath, G, Kuzmin, IV, Rodriguez, LL, Stone, DM, et al. Rhabdoviridae. In: King A, Adams, MJ, Carstens, EB, Lefkowitz, EJ, editors. Virus taxonomy:

   International Committee on Taxonomy of Viruses. San Diego: Elsevier; 2011. p. 654–81.
2. Banyard AC, Hayman D, Johnson N, McElhinney L, Fooks AR. Bats and lyssaviruses. Adv Virus Res. 2011;79:239–89. http://dx.doi.org/10.1016/B978-0-12-387040-7.00012-3
3. Wang L, Tang Q, Liang G. Rabies and rabies virus in wildlife in mainland China, 1990–2013. Int J Infect Dis. 2014;25:122–9. http://dx.doi.org/10.1016/j.ijid.2014.04.016
4. Escobar LE, Peterson AT, Favi M, Yung V, Medina-Vogel G. Bat-borne rabies in Latin America. Rev Inst Med Trop Sao Paulo. 2015; 57:63–72. http://dx.doi.org/10.1590/S0036-46652015000100009
5. Gould AR, Kattenbelt JA, Gumley SG, Lunt RA. Characterisation of an Australian bat lyssavirus variant isolated from an insectivorous bat. Virus Res. 2002;89:1–28. http://dx.doi.org/10.1016/S0168-1702(02)00056-4
6. Pal SR, Arora B, Chhuttani PN, Broor S, Choudhury S, Joshi RM, *Pteropus policephalus* in Chandigarh, northern India. Trop Geogr Med. 1980;32:265–7.
7. Smith PC, Lawhaswasdi K, Vick WE, Stanton JS. Isolation of rabies virus from fruit bats in Thailand. Nature. 1967;216:384. http://dx.doi.org/10.1038/216384a0
8. Molur S, Srinivasulu C, Bates P, Francis C. *Pteropus giganteus*. The IUCN red list of threatened species, 2008 [cited 2016 Jan 12]. http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T18725A8511108.en

9. Webster WA, Casey GA. Virus isolation in neuroblastoma cell culture. In: Meslin F-X, Kaplan MM, Koprowski H, editors. Laboratory techniques in rabies. 4th ed. Geneva: World Health Organization; 1996. p. 96–104.
10. Heaton PR, Johnstone P, McElhinney LM, Cowley R, O'Sullivan E, Whitby JE. Heminested PCR assay for detection of six genotypes of rabies and rabies-related viruses. J Clin Microbiol. 1997;35:2762–6.
11. Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, et al. Next generation sequencing of viral RNA genomes. BMC Genomics. 2013;14:444. http://dx.doi.org/10.1186/1471-2164-14-444
12. Atkinson B, Marston DA, Ellis RJ, Fooks AR, Hewson R. Complete genomic sequence of Issyk-Kul virus. Genome Announc. 2015;3:e00662–15. http://dx.doi.org/10.1128/genomeA.00662-15
13. Karunanayake D, Matsumoto T, Wimalaratne O, Nanayakkara S, Perera D, Nishizono A, et al. Twelve years of rabies surveillance in Sri Lanka, 1999–2010. PLoS Negl Trop Dis. 2014;8:e3205. http://dx.doi.org/10.1371/journal.pntd.0003205

Address for correspondence: Anthony R. Fooks, Wildlife Zoonoses and Vector-Borne Diseases Research Group, Animal and Plant Health Agency, Weybridge, New Haw, Addlestone, Surrey KT15 3NB, UK; email: tony.fooks@apha.gsi.gov.uk

4.2.6: Article 20 published on ICTV website

GBLV ICTV classification document

https://talk.ictvonline.org/files/proposals/animal_dsrna_and_ssrna_viruses/m/animal_rna_minus_ec_approved/6949

Summary: The complete genome sequence, phylogenetic analysis, and antibody neutralisation data were combined in a new species ICTV document which has been approved and is awaiting ratification in early 2018.

This form should be used for all taxonomic proposals. Please complete all those modules that are applicable.
For guidance, see the notes written in blue and the separate document "Help with completing a taxonomic proposal"

Please try to keep related proposals within a single document.

Part 1: **TITLE, AUTHORS, etc**

Code assigned:		(to be completed by ICTV officers)
Short title: One new species (*Gannaruwa bat lyssavirus*) in the genus *Lyssavirus*		
Modules attached (Modules 1, 4 and either 2 or 3 are required.	**1** ☒ **2** ☒ **3** ☐ **4** ☒	

Author(s):

Denise A. Marston, Thomas Mueller, Conrad Freuling, Ashley C. Banyard and Anthony R Fooks

Corresponding author with e-mail address:

Denise.Marston@apha.gsi.gov.uk

List the ICTV study group(s) that have seen this proposal:

A list of study groups and contacts is provided at http://www.ictvonline.org/subcommittees.asp . If in doubt, contact the appropriate subcommittee chair (there are six virus subcommittees: animal DNA and retroviruses, animal ssRNA-, animal ssRNA+, fungal and protist, plant, bacterial and archaeal)	***Rhabdoviridae*** **Study Group**

ICTV Study Group comments (if any) and response of the proposer:

Date first submitted to ICTV:
Date of this revision (if different to above):

ICTV-EC comments and response of the proposer:

Part 2: **PROPOSED TAXONOMY**
Present the proposed new taxonomy on accompanying spreadsheet

Name of accompanying spreadsheet: TP_Template_Excel_module_2017_GBLV

Please display the taxonomic changes you are proposing on the accompanying spreadsheet module 2017_TP_Template_Excel_module. Submit both this and the spreadsheet to the appropriate ICTV Subcommittee Chair.

Part 4: APPENDIX: supporting material
additional material in support of this proposal

References:

DIETZGEN, R., CALISHER, CH, KURATH, G, KUZMIN, IV, RODRIGUEZ, LL, STONE, DM, TESH, RB, TORDO, N, WALKER, PJ, WETZEL, T AND WHITfiELD, A E (2011) Rhabdoviridae. In Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Ed A. KING, ADAMS, MJ, CARSTENS, EB AND LEFKOWITZ, EJ, San Diego: Elsevier. pp 654-681

GUNAWARDENA, P. S., MARSTON, D. A., ELLIS, R. J., WISE, E. L., KARAWITA, A. C., BREED, A. C., MCELHINNEY, L. M., JOHNSON, N., BANYARD, A. C. & FOOKS, A. R. (2016) Lyssavirus in Indian Flying Foxes, Sri Lanka. Emerg Infect Dis 22, 1456-1459

HORTON, D. L., BANYARD, A. C., MARSTON, D. A., WISE, E., SELDEN, D., NUNEZ, A., HICKS, D., LEMBO, T., CLEAVELAND, S., PEEL, A. J., KUZMIN, I. V., RUPPRECHT, C. E. & FOOKS, A. R. (2014) Antigenic and genetic characterization of a divergent African virus, Ikoma lyssavirus. J Gen Virol 95, 1025-1032

SCHNEIDER, L. G., BARNARD, B. J. H., SCHNEIDER, H. P., SCHNEIDER, L. G., ØDEGAARD, Ø. A., MUELLER, J., SELIMOV, M., SCHNEIDER, L. G., COX, J. H., WANDELER, A. I., BLANCOU, J. & MEYER, S. (1985) Application of Monoclonal Antibodies for Epidemiological Investigations and Oral Vaccination Studies. In Rabies in the Tropics. Eds E. KUWERT, C. MÉRIEUX, H. KOPROWSKI, K. BÖGEL. Berlin, Heidelberg, Springer Berlin Heidelberg. pp 47-59

Annex:

Please explain the reasons for the taxonomic changes you are proposing and provide evidence to support them. The following information should be provided, where relevant:

- **Species demarcation criteria**: Explain how new species differ from others in the genus and demonstrate that these differences meet the criteria previously established for demarcating between species. If no criteria have previously been established, and if there will now be more than one species in the genus, please state the demarcation criteria you are proposing.
- **Higher taxa**:
  - There is no formal requirement to state demarcation criteria when proposing new genera or other higher taxa. However, a similar concept should apply in pursuit of a rational and consistent virus taxonomy.
  - Please indicate the **origin of names** assigned to new taxa at genus level and above.
  - For each new genus a **type species** must be designated to represent it. Please explain your choice.
- **Supporting evidence**: The use of Figures and Tables is strongly recommended (note that copying from publications will require permission from the copyright holder). For phylogenetic analysis, try to provide a tree where branch length is related to genetic distance.

In general, demarcation criteria for species in the genus *Lyssavirus* include (Dietzgen et al., 2011):

1. Genetic distances, with the threshold of 80–82% nucleotide identity for the complete N gene, that provides a better quantitative resolution compared to other genes, or 80–81% nucleotide identity for concatenated coding regions of N+P+M+G+L genes. Globally, all isolates belonging to the same species have higher identity values than the threshold, except the viruses currently included into the LBV species. For that reason some authors suggested that LBV be subdivided into several genotypes. However, as these LBV representatives are segregated into a monophyletic cluster in the majority of phylogenetic reconstructions, in the absence of other sufficient demarcation characters there is currently no possibility to subdivide LBV into several viral species.

2. Topology and consistency of phylogenetic trees, obtained with various evolutionary models.

3. Antigenic patterns in reactions with anti-nucleocapsid monoclonal antibodies (preceded by serologic cross-reactivity and definition of lyssavirus serotypes, using polyclonal antisera).

4. Whenever available, additional characters, such as ecological properties, host and geographic range, pathological features are recruited.

Gannaruwa bat lyssavirus

Isolated from four Indian flying-foxes (*Pteropus medius*) in Sri Lanka in 2014-15. The bats were collected from the ground, in an area inhabited by a long-established roost of approximately 20,000 Indian flying-foxes (Gunawardena *et al* 2016).

1. Genetic distances, with the threshold of 80–81% nucleotide identity for concatenated coding regions of N+P+M+G+L genes to demarcate species.

- The complete genome (Genbank KU244266-9) of GBLV consists of 11,919 nucleotides and includes 5 genes: nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and RNA-dependent RNA polymerase (L) (see Annex Table 1 for comparison of protein coding and non-coding region sizes for all lyssaviruses)
- GBLV demonstrates a range of identity to other lyssaviruses (concatenated genes range from 62.9-77.1% IKOV and ABLV, respectively). The two most closely related lyssaviruses are RABV and ABLV with 76.2-76.6% identity to RABV and 75.8-77.1% identity to ABLV, which is below the threshold of within species identity suggesting that GBLV is a virus belonging to a new lyssavirus species (see Annex Table 2).

2. Topology and consistency of phylogenetic trees, obtained with various evolutionary models.

- Phylogenetic analysis performed on concatenated gene sequences demonstrated limited relatedness between GBLV and all other lyssaviruses using neighbour-joining, and maximum likelihood reconstructions with Kimura-2 parameter and GTR+G+I models respectively. Regardless of the model or reconstruction used, the resulting tree topologies did not alter (Annex Figure 1).

3. Antigenic patterns in reactions with anti-nucleocapsid monoclonal antibodies (preceded by serologic cross-reactivity and definition of lyssavirus serotypes, using polyclonal antisera).

- Antigenic patterns in reactions with anti-nucleocapsid monoclonal antibodies clearly distinguish GBLV from other lyssaviruses, most importantly from RABV and ABLV which are the most closely related genetically (Schneider et al., 1985; Annex Table 3)
- Cross-neutralizing assays using a panel of vaccinated human and dog sera with a range of neutralizing antibodies including proven high serum neutralizing antibodies to a laboratory adapted RABV strain (CVS), indicate that vaccination using a RABV vaccine strain cross-neutralises GBLV which is true for all phylogroup I viruses. All sera tested cross-neutralised GBLV as well as or less than RABV (methodology described in Horton et al., 2014). These data indicate GBLV is a member of phylogroup I, distinct from RABV. (Annex Table 4).

4. Additional characters, such as ecological properties, host and geographic range, pathological features are recruited.

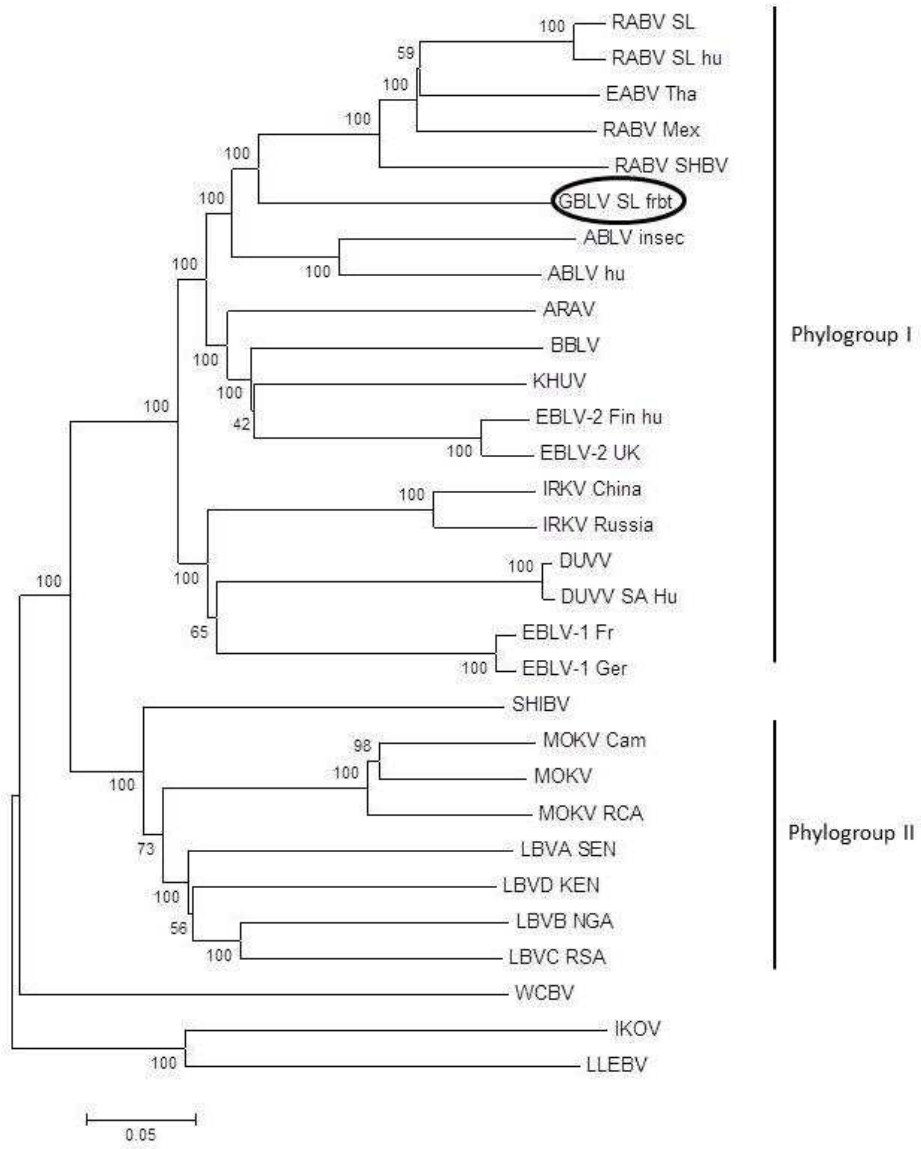
- Isolated from four Indian flying-foxes (*Pteropus medius*) in Sri Lanka in 2014-15. This is the first recorded lyssavirus isolation from *P. medius*.
- RABV has not been identified in a bat species in the 'old world' and ABLV has not been detected outside of Australia.
- 3 bats were dead on collection, the 4th, bat AK-42, displayed clinical signs consistent with rabies, including aggressiveness, biting, spontaneous vocalization and inability to fly.
- Pathogenic to laboratory mice via intracranial and peripheral inoculation causing acute progressive fatal encephalitis (rabies).
- During in vitro and in vivo infection GBLV forms intracytoplasmic inclusions, detected by staining with FITC-conjugated anti-nucleocapsid monoclonal antibodies (Fujirebio).

Taken together, these data suggest that GBLV represents a new species in the *Lyssavirus* genus '*Gannoruwa bat lyssavirus'*.

Annex Figures and Tables:

Annex Figure 1: Phylogenetic reconstructions of members of the *Lyssavirus* genus using concatenated gene sequences including the unclassified GBLV (KY006983) in (A) a neighbour-joining using Kimura-2 parameter implemented in MEGA 6, (B) maximum likelihood using GTR G+I evolutional model implemented in MEGA 6. Significant bootstrap values are shown. Full genomes of BBLV (Bokeloh bat lyssavirus, JF311903), RABV (rabies virus, HQ450386; EU293111; AB635373, AB589299), DUVV (Duvenhage virus, EU293120; EU293119), EBLV-1 (European bat lyssavirus 1, EU293109, EF157976), IKOV (Ikoma virus, JX193798), ABLV (Australian bat lyssavirus, NC_003243; AF081020), KHUV (Khujand virus, EF614261), IRKV (Irkut virus, FJ905105, JX193798), EBLV-2 (European bat lyssavirus 2, JX129233, KF155004), ARAV (Aravan virus, EF614259), LBV (Lagos bat virus, EU293108; EU293110, GU170202), MOKV (Mokola virus, EU293118, EU293117, NC_006439), SHIBV (Shimoni bat virus, GU170201), LLEBV and WCBV (West Caucasian bat virus, EF614258) were derived from NCBI Genbank.

(A)


RABV SL
RABV SL hu
EABV Tha
RABV Mex
RABV SHBV
GBLV SL frbt
ABLV insec
ABLV hu
ARAV
BBLV
KHUV
EBLV-2 Fin hu
EBLV-2 UK
IRKV China
IRKV Russia
DUVV
DUVV SA Hu
EBLV-1 Fr
EBLV-1 Ger
SHIBV
MOKV Cam
MOKV
MOKV RCA
LBVA SEN
LBVD KEN
LBVB NGA
LBVC RSA
WCBV
IKOV
LLEBV
0.05
Phylogroup I
Phylogroup II

(B)

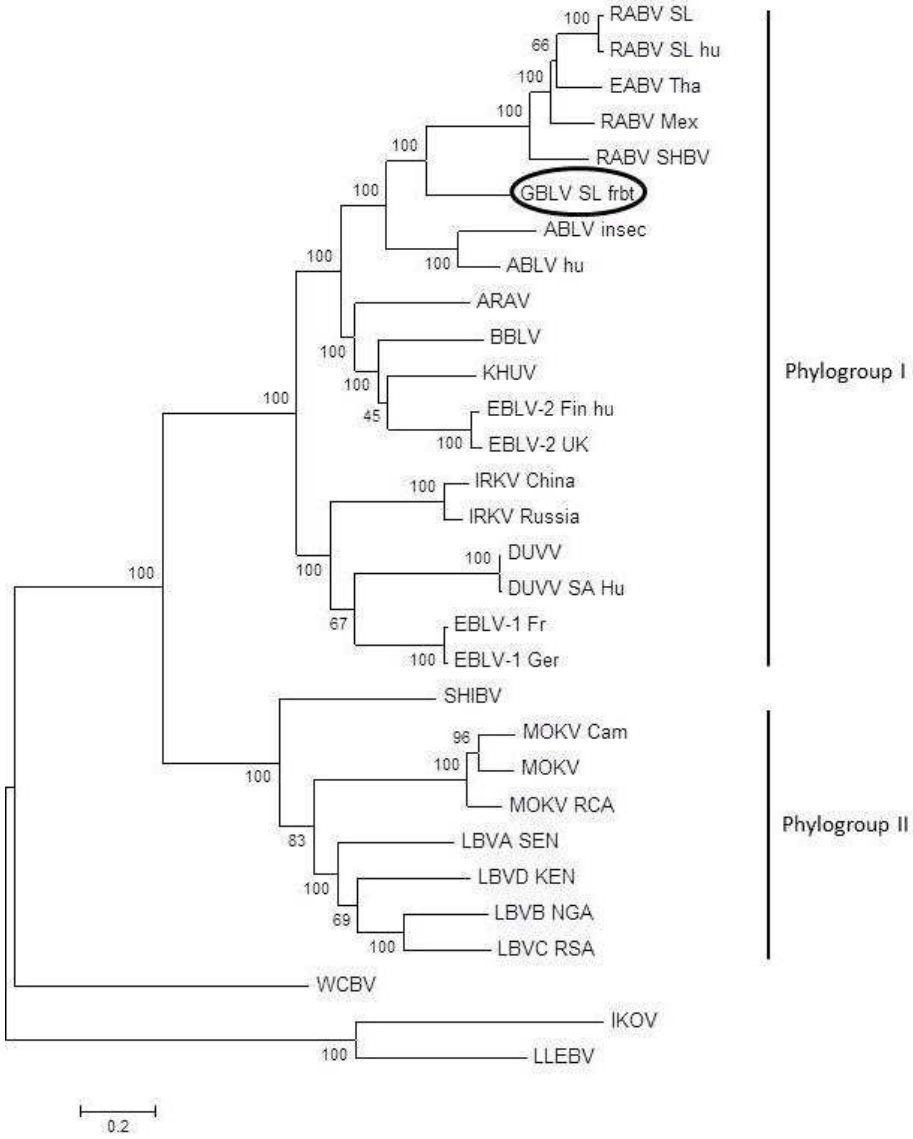

RABV SL
RABV SL hu
EABV Tha
RABV Mex
RABV SHBV
GBLV SL frbt
ABLV insec
ABLV hu
ARAV
BBLV
KHUV
EBLV-2 Fin hu
EBLV-2 UK
IRKV China
IRKV Russia
DUVV
DUVV SA Hu
EBLV-1 Fr
EBLV-1 Ger
SHIBV
MOKV Cam
MOKV
MOKV RCA
LBVA SEN
LBVD KEN
LBVB NGA
LBVC RSA
WCBV
IKOV
LLEBV
Phylogroup I
Phylogroup II
0.2

Table 1: Protein coding, non-coding and genome lengths for all lyssaviruses with GBLV highlighted in grey.

	Phylogroup I										Phylogroup II			Phylogroup III		
	RABV	ABLV	GBLV	EBLV-1	EBLV-2	BBLV	DUVV	ARAV	KHUV	IRKV	SHIBV	LBV	MOKV	WCBV	IKOV	LLEBV
3' UTR*	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70
N protein	**1353**	**1353**	**1353**	**1356**	**1356**	**1356**	**1356**	**1356**	**1356**	**1356**	**1353**	**1353**	**1353**	**1353**	**1353**	**1353**
N–P	90-1	94	92	90	101	93	90	85	95	92	98	101-103	100-102	64	66	68
P protein	**894**	**894**	**894**	**897**	**894**	**894**	**897**	**894**	**894**	**897**	**918**	**918**	**912**	**894**	**870**	**870**
P–M	87-89	89	88	83	88	86	83	85	72	83	76	73-76	80-81	133	74	74
M protein	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**
M–G	211-5	207-209	212	211	210 (205)	210	191	210	208	214	205	204	203-204	206	209	198
G protein	**1575**	**1578-81**	**1581**	**1575**	**1575**	**1575**	**1602**	**1581**	**1581**	**1575**	**1569**	**1569**	**1569**	**1578**	**1575**	**1578**
G–L	516-522	508-509	505	560	512	496	562-563	514	504	569	613	574-588	546-563	862	569	608
L protein	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6381**	**6381**
5' UTR	130-133	131	131	131	131	129	130-131	130	130	131	150	143-146	112-114	125	126	122
Genome	**11,923-8**	**11,918**	**11,919**	**11,966**	**11,930**	**11,902**	**11,975-6**	**11,918**	**11,903**	**11,980**	**12,045**	**12003-16**	**11,940-57**	**12,278**	**11,902**	**11,931**
Concat	**10,815**	**10,821**	**10,821**	**10,821**	**10,818**	**10,818**	**10,848**	**10,824**	**10,824**	**10,821**	**10,833**	**10,833**	**10,827**	**10,818**	**10,788**	**10,791**

Table 2: Nucleotide identity values for concatenated coding regions (N, P, M, G, L genes) of GBLV in comparison with all other identified lyssavirus species. See Annex Figure 1 for details of full genomes used. Concatenated sequences were aligned using ClustalW and a distance matrix was calculated as implemented in MegAlign.

Species	RABV	LBV	MOKV	DUVV	EBLV-1	EBLV-2	ABLV	ARAV	KHUV	IRKV	WCBV	SHIBV	BBLV	**IKOV**	**LLEBV**
RABV															
LBV	67.0-67.7														
MOKV	66.9-67.3	73.3-74.1													
DUVV	71.2-71.8	67.4-67.6	67.0-67.1												
EBLV-1	71.7-72.4	68.1-68.5	67.3-67.7	76.1											
EBLV-2	72.7-73.8	67.3-68.1	67.8-68.0	73.1-73.3	74.2-74.4										
ABLV	73.2-73.8	67.0-67.4	66.4-66.8	71.2	72.2-72.3	73.9									
ARAV	72.9-73.2	68.2-68.3	67.7-68.1	73.4-73.5	75.4-75.5	76.9	73.6								
KHUV	72.9-73.4	67.5-68.0	67.1-67.3	73.5-73.6	74.7	78.7-78.9	74.5	77.5							
IRKV	71.5-72.3	67.9-68.5	67.7-68.3	74.3-74.4	76.3-76.5	73.9	71.6	73.6-74.2	73.9-74.3						
WCBV	64.8-65.5	65.8-66.0	65.3-65.5	65.8	65.5-65.7	65.5	65.2	65.7	65.4	65.2					
SHIBV	67.2-67.7	73.8-75.1	71.9-72.0	67.7-67.8	68.2-68.3	68.1	67.2	68.1	67.9	68.7	66.4				
BBLV	72.7-73.6	67.5-67.9	67.6-68.1	73.1	74.2-74.3	78.2	74.3	76.3	78.4	73.6	65.1	68.7			
IKOV	61.9-62.5	62.7-62.9	62.3-62.5	62.5-62.6	62.6-62.7	62.8	62.3	62.6	62.4	62.4	63.2	63.5	62.5		
LLEBV	61.3-61.4	63.4-64.5	63.0-63.6	63.1	63.5	62.7-62.9	62.9-63.5	63.0	63.6	63.9	64.6	64.8	63.2	71.1	
GBLV	76.2-76.6	68.1-68.3	67.6-68.1	72.1	73.7-73.8	75.1-75.2	75.8-77.1	75.3	75.8	72.7	65.6	68.1	75.4	62.9	63.9

Table 3: Reaction pattern of a panel of 10 anti-nucleocapsid monoclonal antibodies with selected lyssaviruses. Note that MSA6.3 was no longer available for testing.

anti-NC mAb	RABV	LBV	MOKV	DUVV	EBLV-1	EBLV-2	ABLV	BBLV	GBLV	LLEBV
W239.17	+++	+++	+++	+++	+++	+++	+++	+++	+++	-
W187.5	+++	-	-	-	-	-	+++	-	-	-
W187.11.2	+++	-	-	-	-	-	+++	+++	+++	-
MW187.6.1	+++	+++	+++	+++	-	-	+++	+++	+++	+
MSA6.3	-	-	+++	-	+++	+++	-	+++	n.a.	n.a.
LBV7.36	-	+++	-	-	-	+++	-	-	-	-
DUV6.15.19	-	-	-	+++	+++	-	-	-	+	-
S62.1.2	-	-	-	-	+++	+++	-	-	-	-
P 41	-	-	-	-	-	-	-	-	-	-
Z144.88	-	-	-	-	-	-	-	-	-	-

Table 4: Cross neutralization of sera from rabies vaccinated humans and animals against GBLV in comparison with RABV. International units (IU) are given for RABV by comparison to a standard control (not applicable to GBLV). Reciprocal titres over 10,000 are considered exceptionally high, and less than 20 would be considered below detectable threshold and therefore effectively negative.

Sample ID	IU/ml	Reciprocal titre	
		CVS	GBLV
HUMAN-1	40.49	1448.15	512.00
HUMAN-2	364.42	23170.48	23170.48
Dog1	830.70	23170.48	23170.48
Dog2	4.50	362.04	181.02
Dog3	53.29	4096.00	1024.00
Dog4	1.97	64.00	90.51
Dog5	92.30	5792.62	1024.00
Dog6	53.29	4096.00	1024.00
Dog7	0.13	8.00	8.00
Dog8	17.79	724.08	256.00
Dog9	7.79	362.04	362.04
Dog10	1.97	90.51	45.25

4.2.7: Article 21 published in Genome Announcements

Complete Genome Sequence of Lleida Bat Lyssavirus

DOI:10.1128/genomeA.01427-16

Summary: Lleida bat lyssavirus (LLEBV) was discovered by our collaborators in Spain from a bent-winged bat (*Miniopterus schreibersii*) found in the city of Lleida, Spain. The finding was reported in EID in 2013. Since then, the Spanish team had been unable to culture the virus, and did not have NGS capabilities to obtain full genome sequence. The brain material was received in our laboratory, then extracted and depleted using our methodology. The de novo assembly pipeline was used, removing the host reads using a mouse genome dataset, then using IVA to assemble the remaining reads. The de novo assembly resulted in 2 non-overlapping contigs which contained the 5 coding genes, but missing the G/L intergenic region. Primers were designed to span the G/L region and the other intergenic regions to confirm the sequence.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

VIRUSES

Complete Genome Sequence of Lleida Bat Lyssavirus

Denise A. Marston,[a,c] Richard J. Ellis,[b] Emma L. Wise,[a] Nidia Aréchiga-Ceballos,[d,e] Conrad M. Freuling,[f] Ashley C. Banyard,[a] Lorraine M. McElhinney,[a,g] Xavier de Lamballerie,[c] Thomas Müller,[f] Anthony R. Fooks,[a,g,h] Juan E. Echevarría[d,i]

Wildlife Zoonoses & Vector-Borne Diseases Research Group, Animal and Plant Health Agency (APHA), Surrey, United Kingdom[a]; Surveillance and Laboratory Services Department, APHA, Surrey, United Kingdom[b]; UMR_D 190, Emergence de Pathologies Virales, Aix Marseille University, Marseille, France[c]; Instituto de Salud Carlos III, Madrid, Spain[d]; Laboratorio de Rabia, Instituto de Diagnóstico y Referencia Epidemiológicos, Ciudad de México, Mexico[e]; Friedrich Loeffler Institute (FLI), Institute of Molecular Virology and Cell Biology, Greifswald-Insel Riems, Germany[f]; Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom[g]; Institute for Infection and Immunity, St. George's Hospital Medical School, University of London, London, United Kingdom[h]; Ciber de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain[i]

ABSTRACT All lyssaviruses (family *Rhabdoviridae*) cause the disease rabies, an acute progressive encephalitis for which, once symptoms occur, there is no effective cure. Using next-generation sequencing, the full-genome sequence for a novel lyssavirus, Lleida bat lyssavirus (LLEBV), from the original brain of a common bent-winged bat has been confirmed.

Received 24 October 2016 **Accepted** 8 November 2016 **Published** 12 January 2017

Citation Marston DA, Ellis RJ, Wise EL, Aréchiga-Ceballos N, Freuling CM, Banyard AC, McElhinney LM, de Lamballerie X, Müller T, Fooks AR, Echevarría JE. 2017. Complete genome sequence of Lleida bat lyssavirus. Genome Announc 5:e01427-16. https://doi.org/10.1128/genomeA.01427-16.

Address correspondence to Anthony R. Fooks, Tony.Fooks@apha.gsi.gov.uk.

The lyssavirus genome consists of a single-stranded negative-sense RNA of approximately 12 kb. The International Committee for the Taxonomy of Viruses recognizes 14 distinct lyssavirus species, with Lleida bat lyssavirus (LLEBV) and Gannoruwa bat lyssavirus (GBLV) awaiting classification (1, 2).

LLEBV was detected in a common bent-winged bat (*Miniopterus schreibersii*) found in Lleida, Spain, during July 2011 (1). Initial analysis of the partial nucleoprotein (N) sequence confirmed that this virus was most closely related to members of the *Lyssavirus* genus, grouping with West Caucasian bat lyssavirus and Ikoma lyssavirus sequences. Full-genome sequencing of the original brain material was undertaken to genetically characterize this virus. Generation of the full-genome sequence was particularly significant, as initial attempts at virus isolation failed. Viral RNA was extracted using TRIzol, and host genomic DNA and rRNA were depleted as described previously (3). Double-stranded cDNA (ds-cDNA) was synthesized using random hexamers and a cDNA synthesis system (Roche). The ds-cDNA was purified using AMPure XP magnetic beads (Beckman Coulter, Inc.), and 1 ng was used with the Nextera XT DNA sample preparation kit (Illumina), according to the manufacturer's instructions (omitting the bead normalization step) and sequenced using an Illumina MiSeq with 2 × 150-bp paired-end reads.

Mapping the sequencing data (6,651,896 reads) to the most closely related lyssavirus genome (Ikoma lyssavirus [IKOV]) failed to generate a consensus sequence. Therefore, the reads were processed to remove host genome by mapping to a mouse reference (sufficient homology to bat) using BWA version 0.7.5a-r405 (4). The remaining unmapped reads were assembled using IVA (5). Two virus-specific contiguous sequences (contigs) were obtained, which were confirmed by mapping the original data with BWA and consensus calling with a modified SAMtools script (6). The total number of assembled viral reads was 3,610 (0.25% of the host-depleted reads). The two contigs covered all coding regions, but a section of the G-L region was absent, and the genomic ends were atypical lengths. The G-L intergenic region and genomic ends were con-

firmed by Sanger sequencing of PCR amplicons, obtained using primers designed from the next-generation sequencing (NGS) consensus sequences, and paired with conserved primers that bind the conserved lyssavirus genomic termini (7). The sequence obtained by combining the two NGS contigs and three PCR amplicon sequences was used to map the original reads and determine the consensus sequence as described above. The total number of assembled viral reads for LLEBV was 12,629 (0.19% of total reads). The entire genome, apart from the last two nucleotides, was covered by NGS reads.

The genetic organization of the LLEBV genome is similar to that of other lyssaviruses, with a complete genome size of 11,931 nucleotides (nt). The open reading frame (ORF) lengths are as follows: 3′ untranslated region (UTR), 70 nt; N-ORF, 1,353 nt; N-P intergenic region, 68 nt; P-ORF, 870 nt; P-M intergenic region, 74 nt; M-ORF, 609 nt; M-G intergenic region, 198 nt; G-ORF, 1,578 nt; G-L intergenic region, 608 nt; L-ORF, 6,381 nt; and 5′ UTR, 122 nt. These data will contribute to our understanding of lyssavirus diversity and evolution and further our knowledge of vaccine-induced immunity and protection.

Accession number(s). The complete genomic sequence of LLEBV has been deposited in GenBank under accession number KY006983.

ACKNOWLEDGMENTS

This project received funding from Defra, the Scottish Government, and the Welsh Government through grants SE0427 and SE0431 and from the EU H2020-funded Research Infrastructure Grant "European Virus Archive Global (EVAg)" (H2O20—grant agreement no. 653316).

REFERENCES

1. Aréchiga Ceballos N, Vázquez Morón S, Berciano JM, Nicolás O, Aznar López C, Juste J, Rodríguez Nevado C, Aguilar Setién A, Echevarría JE. 2013. Novel lyssavirus in bat, Spain. Emerg Infect Dis 19:793–795. https://doi.org/10.3201/eid1905.121071.
2. Gunawardena PS, Marston DA, Ellis RJ, Wise EL, Karawita AC, Breed AC, McElhinney LM, Johnson N, Banyard AC, Fooks AR. 2016. Lyssavirus in Indian flying foxes, Sri Lanka. Emerg Infect Dis 22:1456–1459. https://doi.org/10.3201/eid2208.151986.
3. Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR. 2013. Next generation sequencing of viral RNA genomes. BMC Genomics 14:444. https://doi.org/10.1186/1471-2164-14-444.
4. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.
5. Hunt M, Gall A, Ong SH, Brener J, Ferns B, Goulder P, Nastouli E, Keane JA, Kellam P, Otto TD. 2015. IVA: accurate *de novo* assembly of RNA virus genomes. Bioinformatics 31:2374–2376. https://doi.org/10.1093/bioinformatics/btv120.
6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
7. Kuzmin IV, Wu X, Tordo N, Rupprecht CE. 2008. Complete genomes of Aravan, Khujand, Irkut and west Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions. Virus Res 136:81–90. https://doi.org/10.1016/j.virusres.2008.04.021.

4.2.8: Article 22 published on ICTV website

LLEBV ICTV classification document

https://talk.ictvonline.org/files/proposals/animal_dsrna_and_ssrna_viruses/m/animal_rna_minus_ec_approved/6950

Summary: The complete genome sequence, phylogenetic analysis, and antibody neutralisation data were combined in a new species ICTV document which has been approved and is awaiting ratification in early 2018.

This form should be used for all taxonomic proposals. Please complete all those modules that are applicable.
For guidance, see the notes written in blue and the separate document "Help with completing a taxonomic proposal"

Please try to keep related proposals within a single document.

Part 1: **TITLE, AUTHORS, etc**

Code assigned:		(to be completed by ICTV officers)
Short title: One new species (*Lleida bat lyssavirus*) in the genus *Lyssavirus*		
Modules attached (Modules 1, 4 and either 2 or 3 are required.)	**1** ☒ **2** ☒ **3** ☐ **4** ☒	

Author(s):

Denise A. Marston, Juan Echevarria, Nidia Aréchiga, Sonia Vázquez, Thomas Mueller, Conrad Freuling, Ashley Banyard and Anthony R Fooks

Corresponding author with e-mail address:

Denise.Marston@apha.gsi.gov.uk

List the ICTV study group(s) that have seen this proposal:

A list of study groups and contacts is provided at http://www.ictvonline.org/subcommittees.asp . If in doubt, contact the appropriate subcommittee chair (there are six virus subcommittees: animal DNA and retroviruses, animal ssRNA-, animal ssRNA+, fungal and protist, plant, bacterial and archaeal)	***Rhabdoviridae* Study Group**

ICTV Study Group comments (if any) and response of the proposer:

Date first submitted to ICTV:
Date of this revision (if different to above):

ICTV-EC comments and response of the proposer:

Part 2: **PROPOSED TAXONOMY**

Present the proposed new taxonomy on accompanying spreadsheet

Name of accompanying spreadsheet: TP_Template_Excel_module_2017.v2_LLEBV

Please display the taxonomic changes you are proposing on the accompanying spreadsheet module 2017_TP_Template_Excel_module. Submit both this and the spreadsheet to the appropriate ICTV Subcommittee Chair.

Part 4: APPENDIX: supporting material
additional material in support of this proposal

References:

ARECHIGA CEBALLOS, N., VAZQUEZ MORON, S., BERCIANO, J. M., NICOLAS, O., AZNAR LOPEZ, C., JUSTE, J., RODRIGUEZ NEVADO, C., AGUILAR SETIEN, A. & ECHEVARRIA, J. E. (2013) Novel lyssavirus in bat, Spain. Emerg Infect Dis 19, 793-795

DIETZGEN, R., CALISHER, CH, KURATH, G, KUZMIN, IV, RODRIGUEZ, LL, STONE, DM, TESH, RB, TORDO, N, WALKER, PJ, WETZEL, T AND WHITfiELD, A E (2011) Rhabdoviridae. In Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Ed A. KING, ADAMS, MJ, CARSTENS, EB AND LEFKOWITZ, EJ, San Diego: Elsevier. pp 654-681

HORTON, D. L., BANYARD, A. C., MARSTON, D. A., WISE, E., SELDEN, D., NUNEZ, A., HICKS, D., LEMBO, T., CLEAVELAND, S., PEEL, A. J., KUZMIN, I. V., RUPPRECHT, C. E. & FOOKS, A. R. (2014) Antigenic and genetic characterization of a divergent African virus, Ikoma lyssavirus. J Gen Virol 95, 1025-1032

MARSTON, D. A., ELLIS, R. J., WISE, E. L., ARECHIGA-CEBALLOS, N., FREULING, C. M., BANYARD, A. C., MCELHINNEY, L. M., DE LAMBALLERIE, X., MULLER, T., FOOKS, A. R. & ECHEVARRIA, J. E. (2017) Complete Genome Sequence of Lleida Bat Lyssavirus. Genome Announc 5

SCHNEIDER, L. G., BARNARD, B. J. H., SCHNEIDER, H. P., SCHNEIDER, L. G., ØDEGAARD, Ø. A., MUELLER, J., SELIMOV, M., SCHNEIDER, L. G., COX, J. H., WANDELER, A. I., BLANCOU, J. & MEYER, S. (1985) Application of Monoclonal Antibodies for Epidemiological Investigations and Oral Vaccination Studies. In Rabies in the Tropics. Eds E. KUWERT, C. MÉRIEUX, H. KOPROWSKI, K. BÖGEL. Berlin, Heidelberg, Springer Berlin Heidelberg. pp 47-59

Annex:

Please explain the reasons for the taxonomic changes you are proposing and provide evidence to support them. The following information should be provided, where relevant:

- **Species demarcation criteria**: Explain how new species differ from others in the genus and demonstrate that these differences meet the criteria previously established for demarcating between species. If no criteria have previously been established, and if there will now be more than one species in the genus, please state the demarcation criteria you are proposing.
- **Higher taxa**:
  - There is no formal requirement to state demarcation criteria when proposing new genera or other higher taxa. However, a similar concept should apply in pursuit of a rational and consistent virus taxonomy.
  - Please indicate the **origin of names** assigned to new taxa at genus level and above.
  - For each new genus a **type species** must be designated to represent it. Please explain your choice.
- **Supporting evidence**: The use of Figures and Tables is strongly recommended (note that copying from publications will require permission from the copyright holder). For phylogenetic analysis, try to provide a tree where branch length is related to genetic distance.

In general, demarcation criteria for species in the genus *Lyssavirus* include (Dietzgen et al., 2011):

1. Genetic distances, with the threshold of 80–82% nucleotide identity for the complete N gene, that provides a better quantitative resolution compared to other genes, or 80–81% nucleotide identity for concatenated coding regions of N+P+M+G+L genes. Globally, all isolates belonging to the same species have higher identity values than the threshold, except the viruses currently included into the LBV species. For that reason, some authors suggested that LBV be subdivided into several genotypes. However, as these LBV representatives are segregated into a monophyletic cluster in the majority of phylogenetic reconstructions, in the absence of other sufficient demarcation characters there is currently no possibility to subdivide LBV into several viral species.

2. Topology and consistency of phylogenetic trees, obtained with various evolutionary models.

3. Antigenic patterns in reactions with anti-nucleocapsid monoclonal antibodies (preceded by serologic cross-reactivity and definition of lyssavirus serotypes, using polyclonal antisera).

4. Whenever available, additional characters, such as ecological properties, host and geographic range, pathological features are recruited.

Lleida bat virus
Isolated from a bent-winged bat (*Miniopterus schreibersii*) discovered in July 2011 in the city of Lleida, Spain (Arechiga Ceballos et al., 2013).

1. Genetic distances, with the threshold of 80–81% nucleotide identity for concatenated coding regions of N+P+M+G+L genes to demarcate species.

- The complete genome has been reported in Marston et al., 2017 (Genbank KY006983) of LLEBV consists of 11,931 nucleotides and includes 5 genes: nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and RNA-dependent RNA polymerase (L) (see Annex Table 1 for comparison of gene and intergenic region sizes for all lyssaviruses)
- LLEBV demonstrates a range of identity to other lyssaviruses (concatenated genes range from 62.9-71.1%, EBLV-2 and IKOV respectively). The most closely related lyssavirus is IKOV with 71.1% identity to LLEBV which is below the threshold of within species identity suggesting that LLEBV is a virus belonging to a new lyssavirus species (see Annex Table 2).

2. Topology and consistency of phylogenetic trees, obtained with various evolutionary models.

- Phylogenetically LLEBV belongs to the *Lyssavirus* genus (Annex Figure 1). Within the genus it is placed ancestrally to all other lyssaviruses apart from IKOV outside of phylogroups I and II using neighbour-joining, and maximum likelihood reconstructions with Kimura-2 parameter and GTR+G+I models respectively. Regardless of the model or reconstruction used, the resulting tree topologies did not alter (Annex Figure 1).

3. Antigenic patterns in reactions with anti-nucleocapsid monoclonal antibodies (preceded by serologic cross-reactivity and definition of lyssavirus serotypes, using polyclonal antisera).

- Cross reactivity with a number of phylogroup I anti-nucleocapsid monoclonal antibodies clearly distinguished LLEBV from other lyssaviruses with no cross reactivity apart from low cross reactivity with MW187.6.1. (Schneider et al., 1985; Annex Table 3)
- Cross-neutralizing assays using a panel of vaccinated human and dog sera with proven high serum neutralizing antibodies to a laboratory adapted RABV strain (CVS) indicate that vaccination using a RABV vaccine strain shows little/no cross-neutralisation against LLEBV (methodology described in Horton et al., 2014), (Annex Table 4).

4. Additional characters, such as ecological properties, host and geographic range, pathological features are recruited.

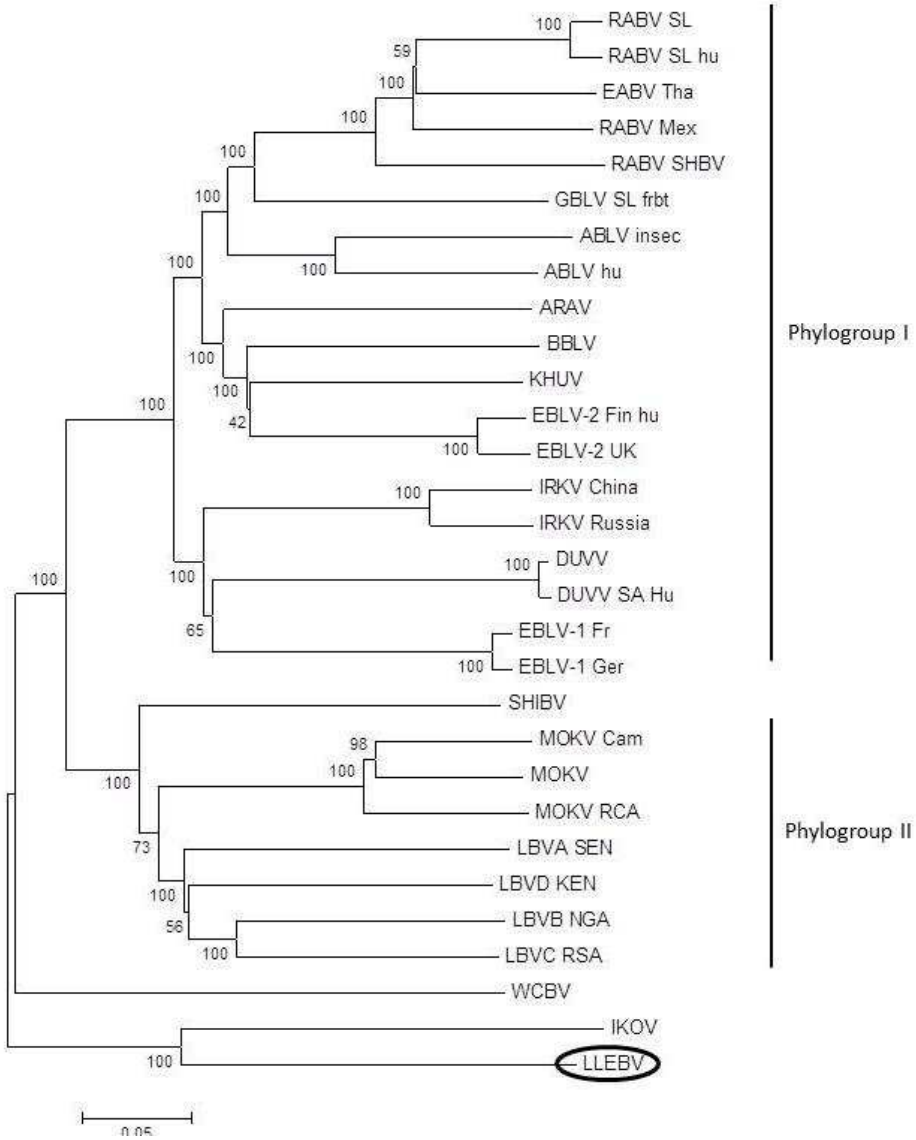
- First isolation of a lyssavirus from a *Miniopterus schreibersii* within Spain.
- Until the discovery of Lleida bat virus (LLEBV), the only known circulating lyssavirus in Spain was EBLV-1, which circulates within populations of *Eptesicus sp.* (*E. serotinus* and *E. isabellinus*)
- LLBV is significantly different to WCBV, which was isolated from the same bat species, *Miniopterus schreibersii*
- IKOV was identified in Tanzania and LLEBV in Spain, although the distribution of these viruses is unknown as only single isolations.
- Pathogenic to laboratory mice via intracranial and peripheral inoculation causing acute progressive fatal encephalitis (rabies).
- During infection LLEBV forms intracytoplasmic inclusions, detected by staining with FITC-conjugated anti-nucleocapsid monoclonal antibodies (Fujirebio).

Taken together, these data suggest that, LLEBV represents a new species in the *Lyssavirus* genus '*Lleida bat lyssavirus'*.

Annex Figures and Tables:

Annex Figure 1: Phylogenetic reconstructions of members of the *Lyssavirus* genus using concatenated gene sequences unless stated (A) neighbour-joining using Kimura-2 parameter implemented in MEGA 6, B) maximum likelihood using GTR G+I evolutional model implemented in MEGA 6. Significant bootstrap values are shown. Full genomes of BBLV (Bokeloh bat lyssavirus, JF311903), RABV (rabies virus, HQ450386; EU293111; AB635373, AB589299), DUVV (Duvenhage virus, EU293120; EU293119), EBLV-1 (European bat lyssavirus 1, EU293109, EF157976), IKOV (Ikoma virus, JX193798), ABLV (Australian bat lyssavirus, NC_003243; AF081020), KHUV (Khujand virus, EF614261), IRKV (Irkut virus, FJ905105, JX193798), EBLV-2 (European bat lyssavirus 2, JX129233, KF155004), ARAV (Aravan virus, EF614259), LBV (Lagos bat virus, EU293108; EU293110, GU170202), MOKV (Mokola virus, EU293118, EU293117, NC_006439), SHIBV (Shimoni bat virus, GU170201), GBLV (Gannoruwa bat lyssavirus (KU244266) and WCBV (West Caucasian bat virus, EF614258) were derived from NCBI Genbank.

A)


RABV SL
RABV SL hu
EABV Tha
RABV Mex
RABV SHBV
GBLV SL frbt
ABLV insec
ABLV hu
ARAV
BBLV
KHUV
EBLV-2 Fin hu
EBLV-2 UK
IRKV China
IRKV Russia
DUVV
DUVV SA Hu
EBLV-1 Fr
EBLV-1 Ger
SHIBV
MOKV Cam
MOKV
MOKV RCA
LBVA SEN
LBVD KEN
LBVB NGA
LBVC RSA
WCBV
IKOV
LLEBV
Phylogroup I
Phylogroup II

B)

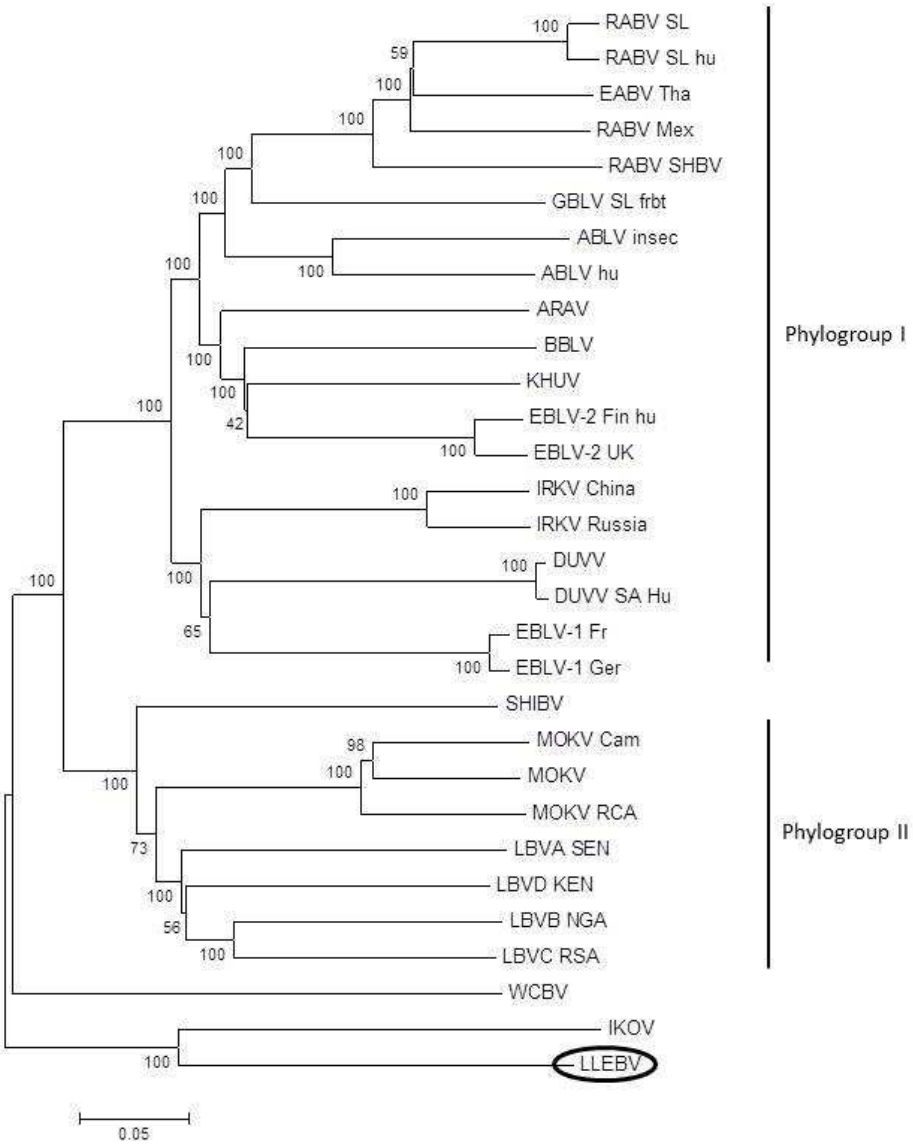

RABV SL
RABV SL hu
EABV Tha
RABV Mex
RABV SHBV
GBLV SL frbt
ABLV insec
ABLV hu
ARAV
BBLV
KHUV
EBLV-2 Fin hu
EBLV-2 UK
IRKV China
IRKV Russia
DUVV
DUVV SA Hu
EBLV-1 Fr
EBLV-1 Ger
SHIBV
MOKV Cam
MOKV
MOKV RCA
LBVA SEN
LBVD KEN
LBVB NGA
LBVC RSA
WCBV
IKOV
LLEBV
Phylogroup I
Phylogroup II
0.05

Table 1: Protein coding, non-coding and genome lengths for all lyssaviruses with LLEBV highlighted in grey.

	Phylogroup I										Phylogroup II			Phylogroup III		
	RABV	ABLV	GBLV	EBLV-1	EBLV-2	BBLV	DUVV	ARAV	KHUV	IRKV	SHIBV	LBV	MOKV	WCBV	IKOV	**LLEBV**
3' UTR*	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70
N protein	**1353**	**1353**	**1353**	**1356**	**1356**	**1356**	**1356**	**1356**	**1356**	**1356**	**1353**	**1353**	**1353**	**1353**	**1353**	**1353**
N–P	90-1	94	92	90	101	93	90	85	95	92	98	101-103	100-102	64	66	68
P protein	**894**	**894**	**894**	**897**	**894**	**894**	**897**	**894**	**894**	**897**	**918**	**918**	**912**	**894**	**870**	**870**
P–M	87-89	89	88	83	88	86	83	85	72	83	76	73-76	80-81	133	74	74
M protein	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**	**609**
M–G	211-5	207-209	212	211	210 (205)	210	191	210	208	214	205	204	203-204	206	209	198
G protein	**1575**	**1578-81**	**1581**	**1575**	**1575**	**1575**	**1602**	**1581**	**1581**	**1575**	**1569**	**1569**	**1569**	**1578**	**1575**	**1578**
G–L	516-522	508-509	505	560	512	496	562-563	514	504	569	613	574-588	546-563	862	569	608
L protein	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6384**	**6381**	**6381**
5' UTR	130-133	131	131	131	131	129	130-131	130	130	131	150	143-146	112-114	125	126	122
Genome	**11,923-8**	**11,918**	**11,919**	**11,966**	**11,930**	**11,902**	**11,975-6**	**11,918**	**11,903**	**11,980**	**12,045**	**12003-16**	**11,940-57**	**12,278**	**11,902**	**11,931**
Concat	**10,815**	**10,821**	**10,821**	**10,821**	**10,818**	**10,818**	**10,848**	**10,824**	**10,824**	**10,821**	**10,833**	**10,833**	**10,827**	**10,818**	**10,788**	**10,791**

Table 2: Nucleotide identity values for concatenated coding regions (N, P, M, G, L genes) of LLEBV in comparison with all other identified lyssavirus species. See Annex Figure 1for details of full genomes used. Concatenated sequences were aligned using ClustalW and a distance matrix was calculated as implemented in MegAlign.

Species	RABV	LBV	MOKV	DUVV	EBLV-1	EBLV-2	ABLV	ARAV	KHUV	IRKV	WCBV	SHIBV	BBLV	IKOV	GBLV
RABV															
LBV	67.0-67.7														
MOKV	66.9-67.3	73.3-74.1													
DUVV	71.2-71.8	67.4-67.6	67.0-67.1												
EBVL-1	71.7-72.4	68.1-68.5	67.3-67.7	76.1											
EBLV-2	72.7-73.8	67.3-68.1	67.8-68.0	73.1-73.3	74.2-74.4										
ABLV	73.2-73.8	67.0-67.4	66.4-66.8	71.2	72.2-72.3	73.9									
ARAV	72.9-73.2	68.2-68.3	67.7-68.1	73.4-73.5	75.4-75.5	76.9	73.6								
KHUV	72.9-73.4	67.5-68.0	67.1-67.3	73.5-73.6	74.7	78.7-78.9	74.5	77.5							
IRKV	71.5-72.3	67.9-68.5	67.7-68.3	74.3-74.4	76.3-76.5	73.9	71.6	73.6-74.2	73.9-74.3						
WCBV	64.8-65.5	65.8-66.0	65.3-65.5	65.8	65.5-65.7	65.5	65.2	65.7	65.4	65.2					
SHIBV	67.2-67.7	73.8-75.1	71.9-72.0	67.7-67.8	68.2-68.3	68.1	67.2	68.1	67.9	68.7	66.4				
BBLV	72.7-73.6	67.5-67.9	67.6-68.1	73.1	74.2-74.3	78.2	74.3	76.3	78.4	73.6	65.1	68.7			
IKOV	61.9-62.5	62.7-62.9	62.3-62.5	62.5-62.6	62.6-62.7	62.8	62.3	62.6	62.4	62.4	63.2	63.5	62.5		
GBLV	76.2-76.6	68.1-68.3	67.6-68.1	72.1	73.7-73.8	75.1-75.2	75.8-77.1	75.3	75.8	72.7	65.6	68.1	75.4	62.9	
LLEBV	61.3-61.4	63.4-64.5	63.0-63.6	63.1	63.5	62.7-62.9	62.9-63.5	63.0	63.6	63.9	64.6	64.8	63.2	71.1	63.9

Table 3: Reaction pattern of a panel of 10 anti-nucleocapsid monoclonal antibodies with selected lyssaviruses. Note that MSA6.3 was no longer available for testing.

anti-NC mAb	RABV	LBV	MOKV	DUVV	EBLV-1	EBLV-2	ABLV	BBLV	LLEBV	GBLV
W239.17	+++	+++	+++	+++	+++	+++	+++	+++	-	+++
W187.5	+++	-	-	-	-	-	+++	-	-	-
W187.11.2	+++	-	-	-	-	-	+++	+++	-	+++
MW187.6.1	+++	+++	+++	+++	-	-	+++	+++	+	+++
MSA6.3	-	-	+++	-	+++	+++	-	+++	n.a.	n.a.
LBV7.36	-	+++	-	-	-	+++	-	-	-	-
DUV6.15.19	-	-	-	+++	+++	-	-	-	-	+
S62.1.2	-	-	-	-	+++	+++	-	-	-	-
P 41	-	-	-	-	-	-	-	-	-	-
Z144.88	-	-	-	-	-	-	-	-	-	-

Table 4: Cross neutralization of sera from rabies vaccinated humans and animals against LLEBV in comparison with RABV, International units (IU) are given for RABV by comparison to a standard control (not applicable to LLEBV). Reciprocal titres over 10,000 are considered exceptionally high, and less than 20 would be considered below detectable threshold and therefore effectively negative.

Sample ID	CVS IU/ml	Reciprocal titre	
		CVS	LLEBV
HUMAN-1	40.49	1448.15	5.66
HUMAN-2	364.42	23170.48	5.66
Dog1	830.70	23170.48	5.66
Dog2	4.50	362.04	5.66
Dog3	53.29	4096.00	5.66
Dog4	1.97	64.00	5.66
Dog5	92.30	5792.62	5.66
Dog6	53.29	4096.00	5.66
Dog7	0.13	8.00	5.66
Dog8	17.79	724.08	5.66
Dog9	7.79	362.04	5.66
Dog10	1.97	90.51	5.66

4.3: Other viruses

In addition to sequencing lyssaviruses, our laboratory has also been involved in the discovery and characterisation of two novel viruses. The following articles describe the process involved in elucidating the characteristics of these novel viruses.

4.3.1: Article 23 published in Genome Announcements

Complete Genome Sequence of Issyk-Kul Virus

DOI:10.1128/genomeA.00662-15

Summary: ISKV is a nairovirus, isolated originally in Krgyzstan. Partial sequence of this virus had been available in Genbank, but full genome sequence was not. This genome sequence was obtained in collaboration with Dr Barry Atkinson at PHE (Porton down, UK). The virus was held at PHE and propagated in a suckling mouse model. RNA was extracted using TRIzol at PHE and sent to APHA for depletion and NGS sequencing. Initial NGS was performed on the 454 platform; de novo assembly was performed after removal of reads mapped to the mouse genome. Three contigs were obtained; matching the lengths expected of the three segments (S, M and L) based on comparison with other nairovirus segments lengths. I undertook blast analysis of the nucleotide sequences and the translated amino acid sequences showed homology to nairoviruses but identity was low (~50% homology to at nucleotide level and ~33% amino acid level). The depth of coverage was low and the ends were not represented, therefore the RNA was prepared for the Illumina platform, which had just been installed (summer 2014). The illumina data obtained was assembled using APHA's de novo pipeline after removing host reads and was compared to the 454 sequences, with much improved depth of coverage and improved genomic termini, although there were still issues with mapping of the reads to these complementary end sequences. In the meantime a Russian group had submitted ISKV full genome sequences to Genbank, however, the genomic termini for all three segments lacked the conserved terminal regions expected. Primers were designed and used by Dr Atkinson to obtain the genomic termini for ISKV at PHE. Once the end sequences for each segment were obtained, they were used to obtain full length references sequences for remapping. Remapping both Illumina and 454 data resulted in reads confirming the correct end sequence. Interestingly, the 454 data resolved the issues that were originally encountered with the S segment genomic termini, whereas the M and L segment ends were confirmed using Illumina. The genomic terminal of all three segments were shown to be incorrect in the Russian published genome sequence, the coding sequences matched the Russian coding sequences with only 5 mismatches identified.

genomeA Journals.ASM.org

Complete Genomic Sequence of Issyk-Kul Virus

Barry Atkinson,[a] Denise A. Marston,[b] Richard J. Ellis,[c] Anthony R. Fooks,[b,d,e] Roger Hewson[a,f]

Microbiology Services Division, Public Health England, Wilts, United Kingdom[a]; Wildlife Zoonoses & Vector-Borne Diseases Research Group, Animal and Plant Health Agency (APHA), Surrey, United Kingdom[b]; Specialist Scientific Support Department, APHA, Surrey, United Kingdom[c]; National Consortium for Zoonosis Research, University of Liverpool, Liverpool, United Kingdom[d]; Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, United Kingdom[e]; National Institute for Health Research, Health Protection Research Unit in Emerging and Zoonotic Infections, Liverpool, United Kingdom[f]

Issyk-Kul virus (ISKV) is an ungrouped virus tentatively assigned to the *Bunyaviridae* family and is associated with an acute febrile illness in several central Asian countries. Using next-generation sequencing technologies, we report here the full-genome sequence for this novel unclassified arboviral pathogen circulating in central Asia.

Received 16 May 2015 **Accepted** 1 June 2015 **Published** 2 July 2015

Citation Atkinson B, Marston DA, Ellis RJ, Fooks AR, Hewson R. 2015. Complete genomic sequence of Issyk-Kul virus. Genome Announc 3(4):e00662-15. doi:10.1128/genomeA.00662-15.

Address correspondence to Barry Atkinson, barry.atkinson@phe.gov.uk.

Issyk-Kul virus (ISKV) was first isolated in 1970 from a *Nyctalus noctula* bat trapped near Lake Issyk-Kul, Kyrgyzstan; the virus was subsequently identified in Tajikistan and Kazakhstan, with sporadic outbreaks of human disease reported in all 3 countries (1, 2). Clinical symptoms include fever (39 to 41°C), headache, myalgia, and nausea. Fatal outcomes are uncommon, although convalescence may take up to 6 weeks (1, 3). ISKV is likely to have a reservoir in both bats and ticks, with transmission to humans being associated with tick bites, exposure to bat urine/feces, or a possible involvement of mites (4). The virus was tentatively assigned to the *Bunyaviridae* family based on electron microscopy; however, to date, no confirmatory data have been published in English.

The LEIV-315K strain of ISKV was propagated in a suckling mouse model. TRIzol-extracted viral RNA was depleted of host genomic DNA using RNase-free DNase (Qiagen, United Kingdom), and host rRNA was depleted using Terminator 5′-phosphate-dependent exonuclease (Epicentre Biotechnologies), as described previously (5, 6). The depleted RNA was quantified using RiboGreen (Life Technologies). Double-stranded (ds) cDNA was synthesized, using a random-primed method, from 50 ng of depleted RNA using the cDNA synthesis system (Roche), according to the manufacturer's instructions. ds-cDNA was purified using AMPure XP magnetic beads (Beckman Coulter), with 1 ng used as input for the Nextera XT DNA sample preparation kit (Illumina). The sequencing library was prepared according to the manufacturer's instructions and sequenced on an Illumina MiSeq instrument with 2 × 150-bp paired-end reads, according to standard Illumina protocols. Sequencing data (27,029,078 reads) were processed to remove host genome by mapping to the mouse reference using BWA version 0.7.5a-r405 (7) and then by *de novo* assembly of unmapped reads (700,000 reads; 2.59%) using Velvet version 1.2.03 (8), with a *k*-mer value of 141. Of the 12 contigs obtained, 9 were identified as host using BLAST analysis and therefore removed from further analysis. The remaining 3 contigs (125,963 reads; 0.46%) were identified as viral with near-complete coverage of S, M, and L segments. Initial data identified the virus as belonging to the *Nairovirus* genus; unresolved terminal regions were confirmed using bespoke pan-nairovirus terminal region primers in combination with internal primers based on our next-generation sequencing (NGS) data. PCR amplicons were Sanger sequenced and aligned with the NGS *de novo* assembled sequences, resulting in complete S, M, and L segments. The NGS data were remapped against these reference sequences using previously described methods (9). A total of 249,869 reads were mapped (0.92% of total reads) as follows: S segment, 47,082 reads; M segment, 103,245 reads; and L segment, 99,542 reads, with average read depths of 3,522×, 2,620×, and 1,099×, respectively. The sequence data for ISKV described here align with independent ISKV virus genome data published during our investigations (GenBank accession numbers KF892055 to KF892057), with only 5 point mutations across all 3 segments. Importantly, our results confirm the terminal regions for all 3 segments that were not resolved by the initial submissions. The characterization of the genome will help improve our understanding of this human disease and provide data for the development of molecular diagnostics.

Nucleotide sequence accession numbers. The complete genomic sequence of ISKV has been deposited in GenBank under the accession numbers KR709219 to KR709221.

ACKNOWLEDGMENTS

We acknowledge the assistance of Dmitri L'vov for depositing the original isolate in 1984, the Centers for Disease Control and Prevention (Fort Collins, CO) for supplying the isolate of ISKV for investigative research, and staff at the Biological Investigations Group at Public Health England (Porton, United Kingdom) for propagating the original stock of ISKV in the suckling mouse model in order to obtain viable material (study 5108).

The sequencing was funded by EU FP7-funded Research Infrastructure Grant European Virus Archive (no. 19 228292).

The views expressed are those of the authors and not necessarily those of the funding body.

REFERENCES

1. **L'vov DK.** 1987. Natural foci of arboviruses in the USSR, p 153–196. *In* Ėdanov VM (ed), Virology reviews, Soviet medical reviews, section E. Harwood Academic Publishers, New York, NY.

2. **L'vov DK, Kostiukov MA, Daniiarov OA, Tukhtaev TM, Sherikov BK.** 1984. Outbreak of arbovirus infection in the Tadzhik SSR due to the Issyk-Kul virus (Issyk-Kul fever). Vopr Virusol **29:**89–92. (In Russian.)
3. **Pak TP.** 1981. Clinical-epidemiological characteristics of arbovirus infections in Tadzhikistan. Sborn nauch Tr Inst Virus Im Ivanov Akad Med Nauk SSSR 101–107. (In Russian.)
4. **Vargina SG, Kuchuk LA, Gershtein VI, Karas FR.** 1982. Transmission of Issyk Kul virus by *Argas vespertilionis* ticks in experiment. Sborn nauch Tr Inst Virus Im Ivanov Akad Med Nauk SSSR 123–127. (In Russian.)
5. **Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR.** 2013. Next generation sequencing of viral RNA genomes. BMC Genomics **14:**444. http://dx.doi.org/10.1186/1471-2164-14-444.
6. **Marston DA, Ellis RJ, Horton DL, Kuzmin IV, Wise EL, McElhinney LM, Banyard AC, Ngeleja C, Keyyu J, Cleaveland S, Lembo T, Rupprecht CE, Fooks AR.** 2012. Complete genome sequence of Ikoma lyssavirus. J Virol **86:**10242–10243. http://dx.doi.org/10.1128/JVI.01628-12.
7. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26:**589–595. http://dx.doi.org/10.1093/bioinformatics/btp698.
8. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
9. **Marston DA, Vázquez-Morón S, Ellis RJ, Wise EL, McElhinney LM, de Lamballerie X, Fooks AR, Echevarría JE.** 2015. Complete genomic sequence of European Bat lyssavirus 1, isolated from *Eptesicus isabellinus* in Spain. Genome Announc **3**(1):e01518-14. http://dx.doi.org/10.1128/genomeA.01518-14.

4.3.2: Article 24 published in Journal of General Virology

Discovery of hantavirus circulating among *Rattus rattus* in French Mayotte island, Indian Ocean

DOI:10.1099/jgv.0.000440

Summary: A novel Hantavirus species was discovered circulating on Mayotte Island in the Indian Ocean. RNA was sent to us for full genome sequencing. The RNA was depleted for gDNA and rRNA using our methodology. The de novo assembly pipeline was used, removing the host reads using the *Rattus rattus* reference genome, mapping with BWA. The unassembled non-host reads assembled using Velvet. The assembled non-host contigs were used in a BLAST analysis to identify the most closely related reference sequence (Anjozorbe virus – ANJV). The mapping assembly pipeline was then used to map the non-host reads against ANJV. The equivalent of an entire MiSeq run was required to generate sufficient viral reads to cover the three segments. S and M consensus sequences were resolved, but L segment had 409 reads and contained a number of gaps. I designed seven primer pairs to cover the gaps by Sanger sequencing. The updated consensus sequence was used to map the reads (no extra reads were mapped). There was an error in the consensus sequence of the L ORF resulting in a stop codon. The reads spanning this region were analysed and a 2bp and 4 bp indel were identified within a 13 bp region. A PCR was designed to span this region and the addition of the 6bps, confirmed by Sanger sequencing, corrected the stop codon.

Short Communication

Discovery of hantavirus circulating among *Rattus rattus* in French Mayotte island, Indian Ocean

Claudia Filippone,[1]† Guillaume Castel,[2]† Séverine Murri,[3] Frédérik Beaulieux,[4] Myriam Ermonval,[1] Corinne Jallet,[1] Emma L. Wise,[5] Richard J. Ellis,[5] Denise A. Marston,[5] Lorraine M. McElhinney,[5,6] Anthony R. Fooks,[5,6] Amélie Desvars,[7]‡ Lénaïg Halos,[8] Gwenaël Vourc'h,[7] Philippe Marianneau[3] and Noël Tordo[1]

Correspondence
Noël Tordo
ntordo@pasteur.fr

[1]Institut Pasteur, Unité des Stratégies Antivirales, F-75015, Paris, France

[2]INRA, UMR 1062 CBGP, F-34988 Montferrier-sur-Lez, France and Institut de Biologie Computationnelle, 34095 Montpellier, France

[3]Anses-Laboratoire de Lyon, Unité de Virologie, Lyon, France

[4]Centre de Recherche Clinique, Hôpital Croix Rousse, Lyon, France

[5]Wildlife Zoonoses and Vector-borne Diseases Research Group, Animal and Plant Health Agency, Woodham Lane, New Haw, Surrey, KT15 3NB, UK

[6]Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, UK

[7]UR346 Animal Epidemiology, INRA, Saint Genès Champanelle, France

[8]Mérial, Lyon, France

Received 5 November 2015
Accepted 27 February 2016

Hantaviruses are emerging zoonotic viruses that cause human diseases. In this study, sera from 642 mammals from La Réunion and Mayotte islands (Indian Ocean) were screened for the presence of hantaviruses by molecular analysis. None of the mammals from La Réunion island was positive, but hantavirus genomic RNA was discovered in 29/160 (18 %) *Rattus rattus* from Mayotte island. The nucleoprotein coding region was sequenced from the liver and spleen of all positive individuals allowing epidemiological and intra-strain variability analyses. Phylogenetic analysis based on complete coding genomic sequences showed that this Murinae-associated hantavirus is a new variant of Thailand virus. Further studies are needed to investigate hantaviruses in rodent hosts and in Haemorrhagic Fever with Renal Syndrome (HFRS) human cases.

Viruses belonging to the genus *Hantavirus*, family *Bunyaviridae*, are negative-sense tri-segmented RNA viruses, with L, M and S segments encoding the RNA-dependent RNA polymerase, glycoproteins Gc and Gn, and nucleoprotein, respectively. Hantaviruses are known to circulate in small mammals (rodents, shrews, moles and bats) in Europe, Asia, America and, more recently discovered, Africa. Infection of humans occurs through inhalation of rodents' aerosolized excreta and can cause two severe pathologies: Haemorrhagic Fever with Renal Syndrome (HFRS) in Asia and Europe, and Hantavirus Cardiopulmonary Syndrome (HCPS), in the Americas (Kruger *et al.*, 2015).

An extensive capture of almost 4000 wild and domestic mammals was conducted in La Réunion, Maurice and Mayotte islands between 2006 and 2007, during Chikungunya virus (CHIKV) outbreaks. Although a low seroprevalence against CHIKV was detected in several non-human primates and rats (Vourc'h *et al.*, 2014), the sample collection was used for virus hunting. Here, we report on an investigation of hantaviruses in 642 small wild mammals captured in La Réunion (193 *Rattus rattus*, 44 *Rattus norvegicus*, 67 *Mus musculus*, 133 *Suncus murinus*, 45 *Tenrec ecaudatus*) and Mayotte (160 *Rattus rattus*) islands.

†These authors contributed equally to this work.

‡Present address: Research Institute of Wildlife Ecology, University of Veterinary Medicine, Savoyenstrasse 1, A-1160 Vienna, Austria.

The GenBank/EMBL/DDBJ accession numbers for the sequences determined in this study are KT719320–KT719370 (N gene coding sequence), KT779091–KT779093 and KU587796 (S, M and L complete coding sequence).

One supplementary figure is available with the online Supplementary Material.

Pools of five individual sera were prepared for the collections from both La Réunion and Mayotte islands.

RNA was extracted from the pools using the MagAttract Viral RNA M48 kit and BioRobot M48 (Qiagen). The following reverse transcription (RT)-PCRs were performed for hantavirus screening: (i) consensus RT-PCR using degenerate primers targeting the L gene (Klempa *et al.*, 2006) for all samples; and (ii) specific RT-PCRs for insectivore hantaviruses targeting both the L and S segments (Kang *et al.*, 2009) for insectivore samples. Sera from positive pooled RNA were subsequently screened independently by RT-PCR. Although none of the samples from La Réunion was positive for hantavirus RNA, we observed hantavirus L gene amplicons for 18 % (29/160) of *Rattus rattus* samples from Mayotte island. Rat species identification was confirmed by a RT-PCR targeting the cytochrome *c* oxidase gene and using the RodentSEA Identification Tool (http://www.ceropath.org/barcoding_tool/rodentsea) for three representative animals (Fig. S1, available in the online Supplementary Material).

The characteristics of the 29 positive individuals are described in Table 1. Thirteen positive individuals (44.8 %) were males and 16 (55.2 %) were females, reflecting the gender proportion within all captured animals

Table 1. Description of hantavirus positive *Rattus rattus* in Mayotte island

Epidemiological information, geo-coordinates of captures and molecular analysis (RT-PCR; sequencing) are indicated for each individual.

						RT-PCR (L)‡			RT-PCR (S)/ sequences§		
***R. rattus* ID**	**Sex**	**Age***	**Capture area†**	**Latitude**	**Longitude**	**Serum**	**Spleen**	**Liver**	**Spleen**	**Liver**	**GenBank accession numbers**
463	M	a	SE – 1	−12.98468	45.18128	+	+	+	+	+	KT719351–KT719352
468	M	a	SE – 1	−12.98468	45.18128	+	−	+	+	+	KT719353–KT719354
469	M	j	SE – 1	−12.98468	45.18128	+	+	+	+	+	KT719355–KT719356
470	F	a	SE – 1	−12.98468	45.18128	+	+	+	+	+	KT719357–KT719358
90	F	a	SE – 2	−12.97011	45.16034	+	+	+	+	+	KT719366–KT719367
412	F	a	SE – 4	−12.91093	45.1904	+	+	+	+	+	KT719329–KT719330
422	M	a	SE – 4	−12.91093	45.1904	+	+	+	+	+	KT719331–KT719332
423	M	a	SE – 4	−12.91093	45.1904	+	+	+	+	+	KT719333–KT719334
96	M	a	SE – 5	−12.92727	45.177	+	+	+	−	+	KT719368
99	M	a	SE – 5	−12.92727	45.177	+	+	+	+	+	KT719369–KT719370
102	F	a	SE – 5	−12.92727	45.177	+	−	+	−	+	KT719320
425	F	a	SE – 5	−12.92727	45.177	+	+	+	+	−	KT719335
426	F	a	SE – 5	−12.92727	45.177	+	+	+	−	+	KT719336
428	F	a	SE – 5	−12.92727	45.177	+	+	+	+	+	KT719337–KT719338
429	F	NA	SE – 5	−12.92727	45.177	+	+	+	+	−	KT719339
376	M	a	W – 6	−12.83119	45.13716	+	+	+	+	+	KT719321–KT719322
379	M	a	W – 6	−12.83119	45.13716	+	+	+	+	+	KT719323–KT719324
382	M	a	W – 6	−12.83119	45.13716	+	+	−	+‖	+‖	KT719325–KT719326
408	F	a	W – 7	−12.87164	45.11745	+	+	+	+	+	KT719327–KT719328
441	M	a	N – 8	−12.70017	45.12168	+	+	+	+	+	KT719340–KT719341
444	F	a	N – 8	−12.70017	45.12168	+	+	−	+	−	KT719342
445	M	a	N – 8	−12.70017	45.12168	+	+	+	+	+	KT719343–KT719344
449	F	a	N – 8	−12.70017	45.12168	+	+	+	+	+	KT719345–KT719346
456	F	a	N – 8	−12.70017	45.12168	+	+	+	+	+	KT719347–KT719348
457	F	a	N – 8	−12.70017	45.12168	+	+	+	+¶	+	KT719349–KT719350
76	M	a	W – 9	−12.76478	45.10652	+	+	−	+	−	KT719363
82	F	a	W – 9	−12.76478	45.10652	+	+	+	+	+	KT719364–KT719365
493	F	a	W – 9	−12.76478	45.10652	+	+	+	+¶	+	KT719359–KT719360
494	F	a	W – 9	−12.76478	45.10652	+	+	+	+	+	KT719361–KT719362

*a, Adult; j, juvenile; NA, information not available.

†SE, South-east; N, north; W, west.

‡Klempa *et al.* (2006).

§GenBank accession numbers KT719320–KT719370.

‖Different sequences from spleen and liver of the same individual. E382: C/T spleen/liver at nt 945 (corresponding to nt 957 of the coding region) of the S segment. This polymorphism is a silent mutation.

¶Only half the length of the sequence obtained (562 nt vs 1005 nt).

(44 and 56 %, respectively). No significant association was observed between gender and infection (P=0.923), therefore male and female black rats were equally susceptible to hantavirus infection (13/70 : 18.6 % and 16/89 : 18.0 %, respectively). Of the 28 individuals with known age, 27 (96.4 %) were adults, which was significantly higher (P<0.05) than the proportion of adults within the whole sampled population (58.8 %). Despite the limited number of individuals, our results support the hypothesis that hantavirus infection is not acquired vertically or during the neonatal stage, as demonstrated previously experimentally (Taruishi *et al.*, 2008).

A precise location of each capture according to geo-coordinate (latitude and longitude) values (Table 1) allowed the differentiation of nine geographical areas across Mayotte island according to the virus circulation (Fig. 1). Five areas were in the south-east: areas 1 (prevalence 11.8 %), 2 (16.7 %), 3 (0 %), 4 (23.1 %) and 5 (26.9 %); three were in the west of the island: areas 6 (13.6 %), 7 (5.6 %) and 9 (17.4 %); and one was in the north: area 8 (35.3 %). We have noticed two main prevalence foci in the north (area 8) and south-east (area 5) of the island. However, the relatively low number of samples per area did not allow statistical comparison tests to be applied to the regional prevalence data. Additional samples and further ecological analysis are required to fully elucidate hantavirus transmission and circulation in Mayotte island.

To explore further the hantavirus genetic diversity in *Rattus rattus* from Mayotte island, RNA was extracted (QIAamp Viral RNA Extraction kit and QIAcube; Qiagen) from homogenized (TissueLyser; Qiagen) liver and spleen from the 29 positive animals. The majority of the spleen (27/29=93.1 %) and liver (26/29=90.0 %) samples tested positive using the nested RT-PCR targeting the L segment (Table 1). Both degenerate and specific primers were designed along the S segment by multiple alignments of the Murinae-associated hantavirus sequences (primers available on request) to obtain the nucleoprotein coding sequence from at least one organ per animal. Sequences from both organs (22/26) were identical to each other, except for one individual (E382) where a C/T polymorphism was observed (spleen/liver) at position 957 of the coding region on the S segment (Table 1), this mutation being silent at the amino acid level. The determined partial N gene coding sequences (1005 nt) within the S segment (nt 13–1017), were used. The method used was the maximum-likelihood method with PhyML v.3.0 (Guindon *et al.*, 2010) under the GTR (general time reversible) substitution model with a γ-distribution model among site rate heterogeneity and a proportion of invariant sites

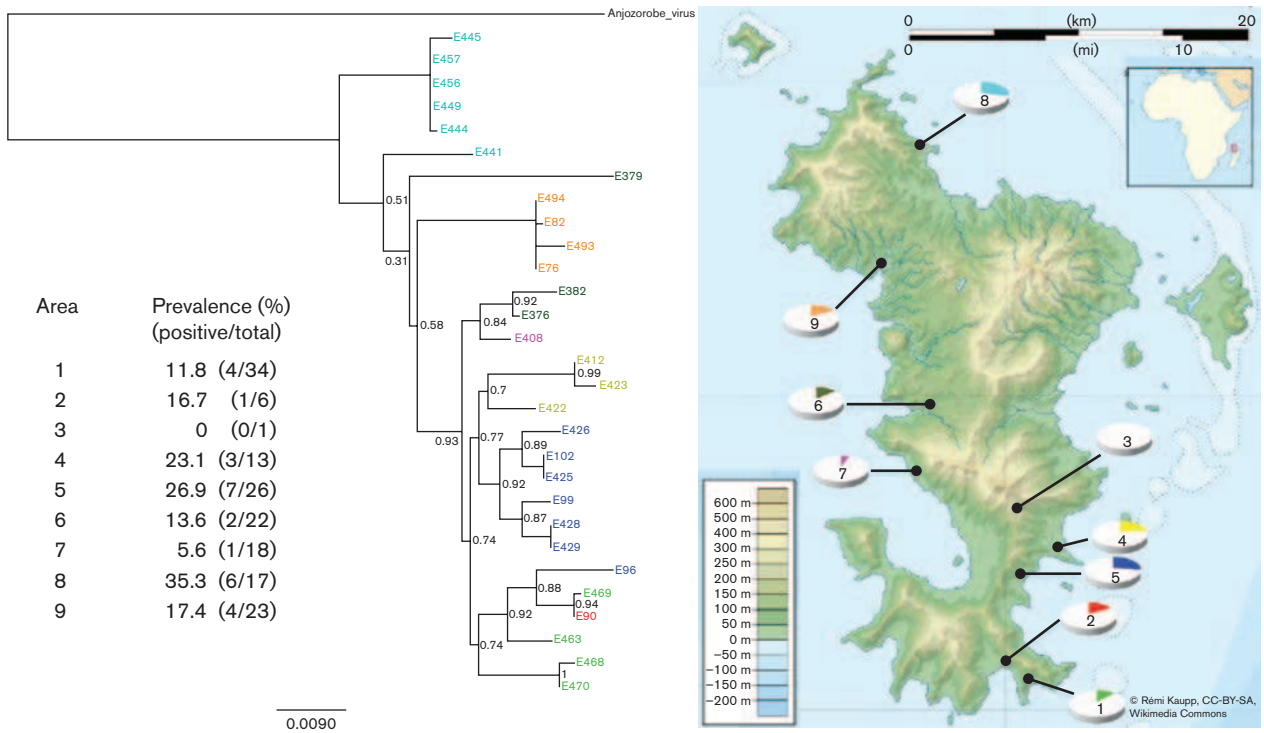


Fig. 1. Hantavirus prevalence and phylogeography in *Rattus rattus* from Mayotte island. Prevalence and intra-strain variability of MAYOV hantavirus according to localization of the captured *R. rattus*. Phylogenetic analysis was done using partial S segment sequences (nt 13–1017). Anjozorobe virus, previously detected in Madagascar (GenBank accession no. KC490916) was used as an outgroup. Bar, mean number of nucleotide substitutions per site. Colour codes correspond between the map and the tree.

(GTR + G + I), as determined by MEGA v.6.0 (Tamura *et al.*, 2013) and with a statistical approximate likelihood ratio test of branch support. Fig. 1 shows the phylogenetic pattern merged with the virus distribution across the island. The genetic cluster of most of the hantavirus sequences appeared to follow a geographical trajectory with a probable gradient of the virus distribution from north to south. However, individuals E96 and E379 did not follow this pattern, branching out of their group.

More exhaustive genetic information for this newly observed hantavirus was obtained from a representative individual (E469) by Next-Generation Sequencing (NGS) (Illumina MiSeq; Illumina). RNA was extracted and depleted of genomic DNA and ribosomal RNA using a protocol optimized for RNA viruses (Marston *et al.*, 2013). Double-stranded cDNA and subsequent library preparation for Illumina sequencing were undertaken as described previously (Marston *et al.*, 2013). Host sequences were removed by mapping to the *Rattus rattus* reference genome (BWA-mem v.0.7.5). BLAST analysis of the assembled contigs of non-host reads (Velvet v.1.2.10) was used to select the most similar reference sequence [Anjozorobe virus (ANJV)]. Iterative mapping and intermediate consensus alignment was used to generate the final consensus sequence. RT-PCR targeting with Sanger sequencing was undertaken to fill missing gaps in the sequence. Complete protein coding sequences (S, M and L) of the newly identified virus isolate, named Mayotte virus (MAYOV), were obtained. Sequences from the S and M segments were used for phylogenetic analysis (Fig. 2a, b), following the methodology described above. This analysis showed that MAYOV clusters within the Thailand hantavirus clade closer to the isolates circulating in the Indian Ocean, South-East Asia (Hugot *et al.*, 2006; Plyusnina *et al.*, 2009; Johansson *et al.*, 2010) and Madagascar (Reynes *et al.*, 2014), rather than the Murinae-associated hantaviruses circulating in continental Africa (Klempa *et al.*, 2006, 2012; Witkowski *et al.*, 2014). The data suggested that both

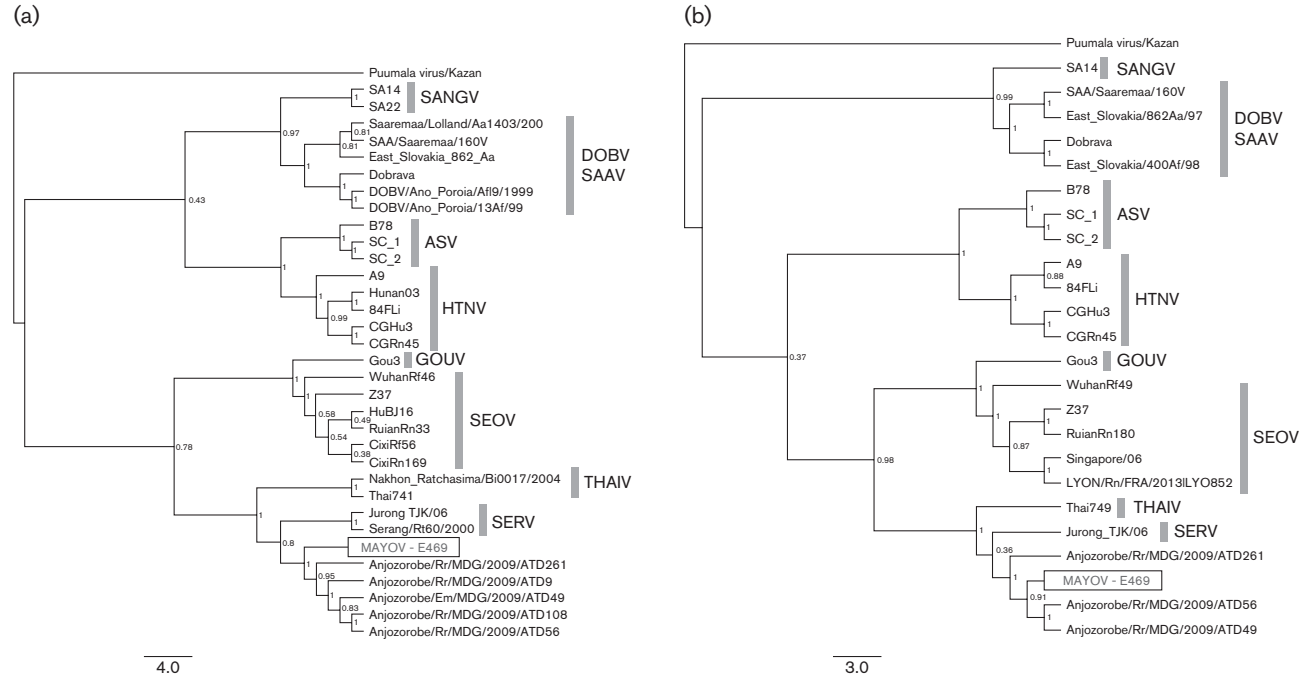


Fig. 2. Phylogenetic analysis of Murinae-associated hantaviruses including MAYOV. Phylogenetic analysis was performed for the S (a) and M (b) segments among the Murinae-associated hantaviruses. MAYOV sequences were obtained by next-generation sequencing (S, 142 reads; M, 306 reads) and Sanger sequencing. The sequences used for phylogenetic analysis were obtained from GenBank: Sangassou virus (SANGV) (N: JQ082303, JQ082300; Gc-Gn: JQ082301); Dobrava virus (DOBV)/Saarema virus (SAAV) (N: AJ269550, AJ009773, AJ616854, L41916, AJ410619, NC005233; Gc-Gn: AY168578, AJ009774, L33685, AY168577); Amur/Soochong virus (ASV) (N: AB127997, AY675349, AY675350; Gc-Gn: AB127994, AY675353, DQ056293); Hantaan virus (HTNV) (N: AF329390, AF366568, JN712306, EU363809, EU092221; Gc-Gn: AF035831, AF366569, EU363818, EU092225); Gou virus (GOUV) (N: AF184988; Gc-Gn: AF145977); Seoul virus (SEOV) (N: JQ665919, AF187082, FJ803202, FJ803207, GQ279380, FJ803215; Gc-Gn: JQ665895, AF187081, GU592931, GQ274942, KF387724); Thailand virus (THAIV) (N: AM397664, AB186420; Gc-Gn: L08756); Serang virus (SERV) (N: AM998808, GQ274941; Gc-Gn: GQ274939); Anjozorobe (N: KC490914-KC490918; Gc-Gn: KC490919-KC490921). The Kazan strain of Puumala virus (Z84204, Z84205) was used as an outgroup for both trees. Bars, mean number of nucleotide substitutions per site.

MAYOV and Anjozorobe virus, the closest variant within the Thailand hantavirus clade, may have resulted from a westward expansion of an ancestral South-East Asian hantavirus introduced with *R. rattus* (Cheke & Hume, 2008; Tollenaere *et al.*, 2010). Pairwise sequence identities between their coding regions, calculated with the BLASTN tool available online at http://blast.ncbi.nlm.nih.gov/Blast.cgi, were 91/99 % (nucleotide/amino acid identity) for the S segment, 90–91/97–98 % (nucleotide/amino acid identity) for the M segment (for ATD49 and AT261 Anjozorobe virus strains, respectively) and 90/99 % (nucleotide/amino acid identity) for the L segment. It is likely that exchanges between the Indian Ocean islands may have occurred as indicated by the phylogenetic trees in Fig. 2: Fig. 2(a) indicates that MAYOV and Anjozorobe virus have a common ancestor from which MAYOV diverged first, whereas Fig. 2(b) is less definitive and shows only the common ancestor of these two viral strains.

Conclusions

This study was performed in the framework of a campaign aimed at elucidating potential reservoirs of CHIKV, which induced several outbreaks in the Indian Ocean between 2005 and 2007. While screening for hantaviruses, we demonstrated the prevalence (18 %) of a newly identified isolate in *R. rattus* on Mayotte island.

Interestingly, antibodies against hantaviruses were reported previously in rats from the Eastern Horn of Africa (Rodier *et al.*, 1993). One hypothesis is that hantaviruses were introduced by the shipping route to East Africa through infected *R. rattus*, originating from South-East Asia, via the Middle East during Arabian trade in the 10th century, and then shipped in the same reservoir host to the Indian Ocean islands (Rollin *et al.*, 1986; Brouat *et al.*, 2014).

MAYOV could have been acquired earlier by *R. rattus* through a spillover infection event from other hantavirus rodent reservoirs such as *Bandicota indica* for Thailand virus and *Rattus tanezumi* for Jurong/Serang virus in South-East Asia (Thailand, Indonesia) (Hugot *et al.*, 2006; Plyusnina *et al.*, 2009; Johansson *et al.*, 2010). The poor intrinsic genetic variability of MAYOV in *R. rattus* reflects a limited evolution and suggests a relatively recent colonization of Mayotte island. Further studies are required to explore hantavirus and rodent genetic diversity to understand better the origin of MAYOV and its adaptation to local hosts. Identifying the presence of hantavirus in Mayotte island is of great importance to evaluate its distribution in the tropical Indian Ocean region. A more extensive surveillance of rodents and /or other reservoirs is required to fully understand the circulation of these zoonotic viruses and to assess their potential risk of transmission to humans and subsequent occurrence of HFRS.

Acknowledgements

This study was supported by the EU FP7 programmes: EDENext (no. 261504), EMPERIE (no. 223498), ANTIGONE (no. 278976) and the Research Infrastructure Grant European Virus Archive (no. 19 228292). We thank Sandra Lacote (Virology Unit, Anses-Laboratoire, Lyon, France) for helping during RNA extraction. A. R. F. was supported by the Research and Policy for Infectious Disease Dynamics Program (RAPIDD), Science and Technology Directorate, US Department of Homeland Security and Fogarty International Center, US National Institutes of Health. Sampling was supported by the French Research National Agency ChikAni (ANR no. 06SEST06). We thank all the people who have participated in the collection of data in the field. We thank Dr Alexander Plyusnin for critical reading of the manuscript.

References

Brouat, C., Tollenaere, C., Estoup, A., Loiseau, A., Sommer, S., Soanandrasana, R., Rahalison, L., Rajerison, M., Piry, S. & other authors (2014). Invasion genetics of a human commensal rodent: the black rat *Rattus rattus* in Madagascar. *Mol Ecol* **23**, 4153–4167.

Cheke, A. & Hume, J. (2008). *Lost Land of the Dodo: The Ecological History of Mauritius, Réunion & Rodrigues*. London: T. & A.D. Poyser.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321.

Hugot, J. P., Plyusnina, A., Herbreteau, V., Nemirov, K., Laakkonen, J., Lundkvist, A., Supputamongkol, Y., Henttonen, H. & Plyusnin, A. (2006). Genetic analysis of Thailand hantavirus in *Bandicota indica* trapped in Thailand. *Virol J* **3**, 72.

Johansson, P., Yap, G., Low, H. T., Siew, C. C., Kek, R., Ng, L. C. & Bucht, G. (2010). Molecular characterization of two hantavirus strains from different *Rattus* species in Singapore. *Virol J* **7**, 15.

Kang, H. J., Bennett, S. N., Sumibcay, L., Arai, S., Hope, A. G., Mocz, G., Song, J. W., Cook, J. A. & Yanagihara, R. (2009). Evolutionary insights from a genetically divergent hantavirus harbored by the European common mole (*Talpa europaea*). *PLoS One* **4**, e6149.

Klempa, B., Fichet-Calvet, E., Lecompte, E., Auste, B., Aniskin, V., Meisel, H., Denys, C., Koivogui, L., ter Meulen, J. & Krüger, D. H. (2006). Hantavirus in African wood mouse, Guinea. *Emerg Infect Dis* **12**, 838–840.

Klempa, B., Witkowski, P. T., Popugaeva, E., Auste, B., Koivogui, L., Fichet-Calvet, E., Strecker, T., Ter Meulen, J. & Krüger, D. H. (2012). Sangassou virus, the first hantavirus isolate from Africa, displays genetic and functional properties distinct from those of other Murinae-associated hantaviruses. *J Virol* **86**, 3819–3827.

Kruger, D. H., Figueiredo, L. T., Song, J. W. & Klempa, B. (2015). Hantaviruses – globally emerging pathogens. *J Clin Virol* **64**, 128–136.

Marston, D. A., McElhinney, L. M., Ellis, R. J., Horton, D. L., Wise, E. L., Leech, S. L., David, D., de Lamballerie, X. & Fooks, A. R. (2013). Next generation sequencing of viral RNA genomes. *BMC Genomics* **14**, 444.

Plyusnina, A., Ibrahim, I. N. & Plyusnin, A. (2009). A newly recognized hantavirus in the Asian house rat (*Rattus tanezumi*) in Indonesia. *J Gen Virol* **90**, 205–209.

Reynes, J. M., Razafindralambo, N. K., Lacoste, V., Olive, M. M., Barivelo, T. A., Soarimalala, V., Heraud, J. M. & Lavergne, A. (2014). Anjozorobe hantavirus, a new genetic variant of Thailand virus detected in rodents from Madagascar. *Vector Borne Zoonotic Dis* **14**, 212–219.

Rodier, G., Soliman, A. K., Bouloumié, J. & Kremer, D. (1993). Presence of antibodies to Hantavirus in rat and human populations of Djibouti. *Trans R Soc Trop Med Hyg* **87**, 160–161.

Rollin, P. E., Mathiot, C., Nawrocka, E., Ravaoalimalala, V. E., Coulanges, P., Sureau, P. & McCormick, J. B. (1986). [Hemorrhagic

fever with renal syndrome in Madagascar. First seroepidemiologic survey of rat populations]. *Arch Inst Pasteur Madagascar* **52**, 181–186 (in French).

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.

Taruishi, M., Yoshimatsu, K., Hatsuse, R., Okumura, M., Nakamura, I. & Arikawa, J. (2008). Lack of vertical transmission of Hantaan virus from persistently infected dam to progeny in laboratory mice. *Arch Virol* **153**, 1605–1609.

Tollenaere, C., Brouat, C., Duplantier, J. M., Rahalison, L., Rahelinirina, S., Pascal, M., Moné, H., Mouahid, G., Leirs, H. & Cosson, J.-F. (2010). Phylogeography of the introduced species *Rattus rattus* in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. *J Biogeogr* **37**, 398–410.

Vourc'h, G., Halos, L., Desvars, A., Boué, F., Pascal, M., Lecollinet, S., Zientara, S., Duval, T., Nzonza, A. & Brémont, M. (2014). Chikungunya antibodies detected in non-human primates and rats in three Indian Ocean islands after the 2006 ChikV outbreak. *Vet Res* **45**, 52.

Witkowski, P. T., Klempa, B., Ithete, N. L., Auste, B., Mfune, J. K., Hoveka, J., Matthee, S., Preiser, W. & Kruger, D. H. (2014). Hantaviruses in Africa. *Virus Res* **187**, 34–42.

4.4: Conclusions

These data presented within this chapter represents a substantial contribution to increasing knowledge of lyssaviruses. Three novel lyssaviruses have been identified and complete genome sequence obtained using true de novo assembly. IKOV, the first lyssavirus detected during this PhD has been fully characterised and accepted as a new species *Ikoma lyssavirus*. LLEBV, more recently identified by collaborators in Spain, is most closely related to IKOV, but also proposed to ICTV as a new species in its own right. Equally GBLV, identified in a fruit bat colony in Sri Lanka has been proposed as a new species. ICTV have clear demarcation criteria to differentiate lyssavirus species which include geographical and host information. The GBLV and LLEBV proposals have been ratified by the Rhabdovirus working group and have been formally agreed at the annual ICTV meeting. They are expected to be ratified early 2018. Although ICTV do not stipulate virus isolation as a criterion for acceptance, many virologists insist that a virus isolate should be available, not least so that full characterisation including in vitro and in vivo data can be obtained. Virus isolation for both IKOV and LLEBV was not straight forward. This was hindered by the lack of original clinical material available. Fortunately in both cases the whole carcase was stored at -80oC without thawing, unlike the brain samples which were freeze/thawed multiple times. Multiple attempts to propagate virus from the brain samples were unsuccessful. Finally virus was propagated from spinal cord from the LLEBV infected bat and a previously unthawed brain sample from the IKOV infected African civet.

Increasing the knowledge of lyssaviruses, particularly divergent viruses, is vital to understand the evolution and adaptation of these viruses. Article 23 ‘The Global Phylogeography of Lyssaviruses - Challenging the 'Out of Africa' Hypothesis’, utilises the whole genome sequences obtained in this chapter to investigate the origins of bat lyssaviruses.

Our success in elucidating novel virus genomes has facilitated a number of collaborations including with PHE and Pasteur Institute and gave opportunity for us to test our methodologies on other virus samples. The frustration that a complete genome sequence for ISKV was published during our own analysis was eased when we realised that the genomic termini were incorrect on the published sequence and a good example of the problems underlying a sequence database which is reliant on the submitters to self-check the data.

Obtaining complete genome sequences from clinical samples where the animal is asymptomatic can be challenging and Article 24 is a good example of this. Hantavirus is present at such low levels that RT-PCR detection often requires a nested or hemi-nested second round amplification and NGS on clinical sample only results in low proportion of reads. Despite this severe limitation a complete

genome sequence was obtained for MAYOV and the same methodology has been used for UK hantavirus clinical samples.

Armed with this powerful de novo pipeline, complete genome sequences of novel viruses are becoming more commonplace, and along with the mapping pipeline, complete genome sequences will become the goal of virus researchers globally.

Chapter 5: Utilising Whole Genome Sequences for Phylogeography

Chapter 5: Utilising Whole Genome Sequences for Phylogeography

5.1: Introduction

One of the most useful applications of the whole genome sequences obtained from lyssaviruses is to use them to analyse the origin and movement of lyssaviruses in a local area or population. Before WGS was widely available this analysis was performed using partial or complete gene sequence. Although this approach is valid, often in a cohort of RABV samples from the same location, sequences were 100% identical to each other resulting in a lack of high resolution. As we have developed a method to obtained complete genome sequences directly from clinical material, we applied this methodology to sequence whole genomes from a number of cohorts to infer phylogenetic and geographic relationships.

5.2: Using WGS to infer evolutionary relationships using phylogenetic reconstructions

Phylogenetic trees can be created using many different methods, from the simplistic 'Neighbour Joining' (NJ) to more computationally complex 'Maximum Likelihood' (ML) or 'Bayesian inference' (BI). Regardless of the method employed the topology of the trees should remain the same. It is computationally impossible to create every tree topology for ML and BI, therefore each program uses heuristic approaches to find the most likely tree(s). The majority of phylogenetic programs require the input sequences to be aligned for a meaningful output. Aligning genomic sequences is the single biggest assumption in phylogenetic analysis. Manual correction of alignments must be undertaken before implementing a phylogenetic analysis. This is particularly critical when comparing complete genome sequences (in comparison to discrete coding regions) as the lengths vary and there is high genetic diversity in the intergenic regions.

Both ML and BI evaluate the match between the aligned sequences and the trees, using the nucleotide substitution rate and evolutionary models specified, but they do so in complementary ways. ML starts with a 'good fit tree' then searches for improvements by breaking and reattaching branches. BI programs start with a random tree which is modified each iteration according to the prior distributions. These Markov chain Monte Carlo (MCMC) algorithms quickly find more likely trees then spends the majority of time modifying these 'likely trees'. The 10% burn in parameter is important to

implement, as this removes the initial 'poor trees' from the analysis. The 'confidence' parameters differ between ML and BI. ML uses additional testing – usually bootstrap values – to indicate the support for branches y resampling and replacing sites to create 'Pseudoreplicates'. This is typically done between 100-1000 times and the bootstrap value presented on each node represents the number of pseudoreplicate trees where the clade was reconstructed and can be viewed in a similar way to a percentage score, with 100 indicating that every pseudoreplicate reconstructed the clade the same way indicating a well-supported node, whereas 60 indicates a less-supported node. BI methods use 'posterior probabilities', these are applied 'post-hoc', determining how many trees (excluding 10% burn in) support the clades in the final tree and are usually high, therefore cannot be compared to bootstrap values (Cummings et al., 2003).

In addition to the above (providing rooted, time measured phylogenies using relaxed or strict molecular clock models), BI analyses can be applied to answer many other important questions by using a series of prior probability distributions (or priors). The analysis will use the information provided (for example year of isolation, or location, or host reservoir) to obtain a probability distribution which can be used to refine repeat analysis or used to make conclusions, using posterior probabilities to infer confidence.

Below are manuscripts describing the analysis of WGS using ML and / or BI methodologies.

5.2.1: Article 25 published in PLOS Neglected Tropical Diseases

Diversity and Epidemiology of Mokola Virus

DOI:10.1371/journal.pntd.0002511

Summary: This manuscript was is a good example of a successful international collaboration between a number of OIE reference laboratories. Historically, unusual lyssaviruses such as Mokola virus (MOKV) were distributed amongst rabies reference laboratories to aid diagnostic assay development and research. Therefore each laboratory held in their repositories the same MOKV isolates. A coordinated plan to sequence the 5 genes encoded was drawn up using primers and sanger sequencing, but due to the diversity observed amongst MOKV isolates progress was extremely slow and required many iterations of primer re-optimisation to obtain the gene sequences. Therefore we chose to obtain complete genome sequences for the isolates held at APHA using NGS.

APHA contributed 6 full length MOKV genomes to this manuscript (see table 2). The sequences were obtained using our depletion methodology and the 454 platform. De novo assembly was performed after removal of reads mapped to the mouse genome using the GS de novo assembly software. The assembled contig sequences were aligned with EU293118 (a previously published MOKV genome) and any other specific sequences previously obtained for the isolate in question. A single consensus sequence was obtained for each virus containing the 5 ORFs, but lacking the genomic UTRs. The genomic termini sequences were obtained using the method detailed in section 2.2.2 and assembled the sequences in SeqMan to obtain a reference sequence which was used to map the 454 reads and obtain a final sequence for each virus.

Table 2: Mokola virus reports from 1968-2010.

Highlighted viruses sequenced at APHA

Geographical location	Year	Origin	Lab No.	Accession numbers
Ibadan, Nigeria	1968	Shrew (*Crocidura spp*)	RV4	KF155005

Umhlanga Rocks, KwaZulu Natal Province, South Africa	1970	Feline	700/70	FJ465416 (N), AF049118 (P), GQ472989 (M), GQ473001 (G)
Yaounde, Cameron	1974	Shrew (*Crocidura spp*)	RV39	EU293117[#^]
Bangui, Central African Republic	1981	Rodent (*Lophuromys sikapusi*)	RV40	EU293118
Bulawayo, Zimbabwe	1981	Feline	13270	KC218932 (N), GQ500114 (P), GQ472990 (M), GQ473002 (G)
Bulawayo, Zimbabwe	1981	Feline	12341	FJ465417 (N), GQ861350 (P), GQ472991 (M), GQ473003 (G)
Bulawayo, Zimbabwe	1981	Feline	S59448	NC_006429
Bulawayo, Zimbabwe	1981	Feline	12574	FJ465418 (N), GQ861352 (P), GQ472992 (M), GQ473004 (G)
Bulawayo, Zimbabwe	1982	Feline	Zim82/ RV1035	KF155006
Addis Adaba, Ethiopia	1989-1990	Feline	RV610	AY333111 (N)[#]
Selous, Zimbabwe	1993	Feline	21846/ RV1017	KF155007
Mdantsane, Eastern Cape Province, South Africa	1995	Feline	543/95	FJ465415 (N), GQ500116 (P), GQ472994 (M), GQ500110 (G)
East London, Eastern Cape Province, South Africa	1996	Feline	112/96/ RV1021	KF155008

Yellow Sands, Eastern Cape Province, South Africa	1996	Feline (vaccinated)	322/96	FJ465414 (N), GQ861353 (P), GQ472996 (M), GQ500111 (G)
Pinetown, KwaZulu Natal Province, South Africa	1997	Feline (vaccinated)	252/97	JN944637 (N), AF369376 (P), GQ472997 (M), GQ500112 (G)
Pinetown, KwaZulu Natal Province, South Africa	1997	Feline (vaccinated)	229/97	FJ465413 (N), AF369377 (P), GQ472998 (M), GQ500113 (G)
Pietermaritzburg, KwaZulu Natal Province, South Africa	1998	Feline (vaccinated)	071/98	FJ465410 (N), AF369378 (P), GQ473000 (M), GQ500108 (G)
East London, Eastern Cape Province, South Africa	2006	Feline (vaccinated)	173/06	FJ465412 (N), GQ861351 (P), GQ472999 (M), HQ266624 (G)
Grahamstown, Eastern Cape Province, South Africa	2008	Feline (vaccinated)	226/08	KC218934 (N), KC218935 (P), KC218936 (M), KC218937 (G)

[†] Indicates that the existence of the isolate is not known and no full gene sequences were available in Genbank, therefore the isolate was not included for analysis.

[#] these sequences were likely the same isolate, therefore not included for MCC analysis

^ already published but the isolate held at APHA was also full genome sequenced

OPEN ACCESS Freely available online

PLOS | NEGLECTED TROPICAL DISEASES

Diversity and Epidemiology of Mokola Virus

Joe Kgaladi[1], Nicolette Wright[1], Jessica Coertse[1], Wanda Markotter[1], Denise Marston[2], Anthony R. Fooks[2,3], Conrad M. Freuling[4], Thomas F. Müller[4]*, Claude T. Sabeta[5], Louis H. Nel[1]

1 Department of Microbiology and Plant Pathology, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa, **2** Wildlife Zoonoses and Vector-borne Diseases Research Group, OIE Rabies Reference Laboratory/WHO Collaborating Centre for the Characterization of Rabies and Rabies-related Viruses), Department of Virology, Animal Health Veterinary Laboratories Agency (Weybridge), Addlestone, Surrey, United Kingdom, **3** University of Liverpool, Department of Clinical Infection, Microbiology and Immunology, Liverpool, United Kingdom, **4** OIE Rabies Reference Laboratory/WHO Collaborating Centre for Rabies Surveillance and Research, Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Institute of Molecular Biology, Greifswald-Insel Riems, Germany, **5** OIE Rabies Reference Laboratory, Agricultural Research Council, Onderstepoort Veterinary Institute, Pretoria, South Africa

Abstract

Mokola virus (MOKV) appears to be exclusive to Africa. Although the first isolates were from Nigeria and other Congo basin countries, all reports over the past 20 years have been from southern Africa. Previous phylogenetic studies analyzed few isolates or used partial gene sequence for analysis since limited sequence information is available for MOKV and the isolates were distributed among various laboratories. The complete nucleoprotein, phosphoprotein, matrix and glycoprotein genes of 18 MOKV isolates in various laboratories were sequenced either using partial or full genome sequencing using pyrosequencing and a phylogenetic analysis was undertaken. The results indicated that MOKV isolates from the Republic of South Africa, Zimbabwe, Central African Republic and Nigeria clustered according to geographic origin irrespective of the genes used for phylogenetic analysis, similar to that observed with Lagos bat virus. A Bayesian Markov-Chain-Monte-Carlo-(MCMC) analysis revealed the age of the most recent common ancestor (MRCA) of MOKV to be between 279 and 2034 years depending on the genes used. Generally, all MOKV isolates showed a similar pattern at the amino acid sites considered influential for viral properties.

Citation: Kgaladi J, Wright N, Coertse J, Markotter W, Marston D, et al. (2013) Diversity and Epidemiology of Mokola Virus. PLoS Negl Trop Dis 7(10): e2511. doi:10.1371/journal.pntd.0002511

Editor: Charles E. Rupprecht, The Global Alliance for Rabies Control, United States of America

Received June 20, 2013; **Accepted** September 18, 2013; **Published** October 24, 2013

Funding: We thank the South African National Research Foundation (NRF) and the South African Polioemylitis Research Foundation (PRF) for financial support. This work was also supported by grants from the Department for Environment, Food and Rural Affairs (SE0423 and SE0427), the Department of Science and Technology (04/17/c215), BMBF (01KI1016A), the Research and Policy for Infectious Disease Dynamics (RAPIDD) programme of the Science and Technology Directorate, US Department of Homeland Security, at the Fogarty International Center, National Institutes of Health, and by the EU FP7–funded Research Infrastructure Grant "European Virus Archive" (no. 19 228292). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Thomas.Mueller@fli.bund.de

Introduction

The lyssavirus genus consists of twelve species recognized by ICTV [1] of which five [(Rabies virus (RABV), Lagos bat virus (LBV), Mokola virus (MOKV), Duvenhage virus (DUVV), and Shimoni bat virus (SHIBV)] have been isolated in Africa [2]. LBV, MOKV, DUVV and SHIBV occur exclusively in Africa. SHIBV was recently isolated from *Hipposideros vittatus* (formerly known as *H. commersoni*) [3]. Another proposed lyssavirus species is Ikoma lyssavirus (IKOV) isolated from an African civet in Tanzania [4].

The first isolations of MOKV were made in 1968 and 1969 from organ pools of shrews (*Crocidura flavescens manni*) in Ibadan, Nigeria [5,6,7]. The only isolations from humans were in 1968 and 1971 from two girls from Nigeria [6,8,9]. However, there were no classical signs of rabies in either of these cases. Whilst the 1968 isolation was made from the cerebrospinal fluid of a girl who presented with fever and convulsions but fully recovered with no neurological damage, the 1971 isolate was from the brain of a girl who died of a poliomyelitis-like encephalitic disease. A further isolation was made in 1974 from a shrew (*Crocidura* spp.) in Yaounde, Cameroon [10]. The only isolation from a rodent (*Lophuromys sikapusi*) was in 1981, from Bangui, Central African Republic [11]. MOKV was also isolated from other animal species including companion animals. A survey on lyssaviruses undertaken in Zimbabwe between 1981 and 1984 revealed six isolations of MOKV from domestic animals, namely a dog and cats that had been previously vaccinated against rabies and unvaccinated cats [12,13]. In 1989 MOKV was isolated from a cat in Addis Ababa, Ethiopia [14]. No further isolation of MOKV was made in Zimbabwe until 1993, when the virus was again isolated from a domestic cat [15]. In the Republic of South Africa, the first isolation was made in 1970 from a domestic cat in Umhlanga Rocks, Kwa-Zulu Natal Province (KZN) [16]. At the time the isolate was assumed to be RABV and the isolate was only identified retrospectively using antigenic typing with monoclonal antibodies during the discovery of MOKV in Zimbabwe in the 1980s [17]. Twenty five years later, in 1995, MOKV was isolated from a domestic cat in South Africa, this time from Mdantsane in the Eastern Cape Province (EC) [17]. Two more isolations followed in 1996, one each in KZN and EC and both from domestic cats of which one was vaccinated against rabies [18,19]. In 1997 and 1998 three more isolations were made from rabies-vaccinated cats in KZN [18,19]. Following several years in which MOKV was not encountered, two isolations were from rabies-vaccinated domestic cats in 2006 and 2008 from the EC province

Author Summary

According to the World Health Organization, rabies is considered both a neglected zoonotic and tropical disease. Among all the lyssavirus species known to exist today, Mokola virus is unique and appears to be exclusive to Africa. In contrast to all other known virus species in the genus *Lyssavirus* of the family *Rhabdoviridae*, its reservoir host has not been identified yet. As only limited sequence information is available, this study significantly contributes to the understanding of the genetic diversity and relatedness of Mokola viruses. In a collective approach, the complete nucleoprotein, phosphoprotein, matrix, and glycoprotein genes of all Mokola viruses isolated to date were sequenced in various rabies laboratories across the world. A phylogenetic analysis was undertaken and the most recent common ancestor was determined. Subsequently, results were linked to epidemiological background data. We also conducted a comparative study of distinct antigenic sites considered influential for viral properties within those genes.

and these are the most recent known isolations of this virus [20,21]. From South Africa all isolations of MOKV were from a domestic cat. Viral RNA was detected by PCR from a domestic dog (in 2005) from Mpumalanga Province of South Africa, virus isolation was unsuccessful in this case [20]. A summary of all MOKV isolates and the approximate geographic location of their origin are presented in Table 1 and Fig. 1. Generally, MOKV infected domestic animals were not observed to be particularly aggressive, but displayed other rabies-like signs that included dehydration, unusual behavior, hypersensitivity, neurological disturbance and salivation [19]. Despite MOKV being isolated from a variety of mammal species, this species is the only lyssavirus never to have been isolated from bats.

Cross protection of WHO and OIE recommended rabies vaccines against various rabies-related lyssavirus species have been reported in a number of studies [22,23,24,25,26]. However, no rabies vaccine provided complete protection against MOKV [27,28,29,30,31,32]. More evidence that RABV derived vaccines do not protect against MOKV infection is shown by circumstantial evidence of the fatal infections of numerous domestic animals that had been vaccinated against RABV [13,18,19,20,21]. Given this scenario and the apparent obscurity of MOKV, we argue that much more information is needed to improve our scant understanding of the epidemiology, disease dynamics and the ecology of this virus.

Some phylogenetic studies have been undertaken on MOKV [18,20,21,33,34], but these studies were invariably performed on a smaller number of isolates and limited to partial gene sequences. Despite these limitations, these studies provided some evidence of the existence of different virus clusters, delineated according to geographical incidence. Generally, the genetic variance was shown to be inversely related to the spatial distribution of isolates. For example, South African MOKV isolates were shown to be closely related, but distinguishable based on province and as a cluster more distant from those made in a neighboring country, Zimbabwe [21,33,34]. Such patterns of genetic diversity may indicate extended periods of isolated evolution, as have been reported for terrestrial rabies virus variants [35]. An exception appeared to be a grouping that included one isolate from Cameroon and one from Ethiopia.

The study reported here involved a multi-disciplinary collaborative effort amongst various laboratories in order to generate for the first time a comprehensive dataset of all the known MOKV isolates available. We have shown that most, but not all of the viruses mentioned in literature could be tracked and that some contamination or misnaming occurred. Given a final cohort of eighteen MOKV isolates, the objective of the study was to sequence full nucleoprotein (N), phosphoprotein (P), matrix (M) and glycoprotein (G) genes. The estimation of viral lineage divergence times and subsequent application of a molecular clock is dependent on an accurate estimation of the rate of nucleotide substitution. Bayesian techniques using the Markov Chain Monte Carlo (MCMC) methods have been successfully applied to estimate the evolutionary rate and divergence times from dated sequences of RABVs [36,37,38,39,40]. This study applied a relaxed molecular clock to N-, M-, P- and G-gene datasets to obtain estimates of the time to the most recent common ancestor (MRCA) and rate of evolution for MOKV. The subsequent analysis allowed for study of the phylogeny and diversity within this African lyssavirus species.

Materials and Methods

Virus isolates

MOKV included in this study were comprised of archived isolates. Information on the geographic location, year of isolation, species origin and references of those MOKV isolates is presented in Table 1. The isolates were either passaged several times (passage number unknown) in suckling mice or in tissue culture, or both. Total RNA was extracted from the samples using the TRIzol® method (Invitrogen) according to the manufacturer's instructions.

Primer design, RT-PCR and sequencing

The N, P, M and G genes were sequenced using different primer combinations and cycling conditions available from the authors upon request. All PCR products were analyzed by agarose gel electrophoresis and subsequently purified (Wizard PCR Preps DNA Purification System; Promega). The purified PCR products were sequenced with BigDye Termination Cycle Sequencing Ready Reaction Kit 3.1 (Applied Biosystems) according to the manufacturer's protocol and analyzed on an ABI Prism 3100 DNA sequencer (Applied Biosystems). Within the duration of this project next generation sequencing technology became available and was applied on a selection of samples. Complete genome sequence was obtained directly from brain tissue for four MOKV isolates (RV4, RV1017, RV1021 and RV1035) (Marston, unpublished). Briefly, TRIzol (Invitrogen) extracted viral RNA was depleted of host genomic DNA using RNase-free DNAse (Qiagen, UK) and host ribosomal RNA was depleted using Terminator 5′-Phosphate-Dependent Exonuclease (Epicentre Biotechnologies). The RNA was fragmented, a random-primed cDNA library was made and run using the Roche 454 GS FLX System. The sequencing data were assembled in the GS de novo assembly software (Roche). The de novo assembled contigs for each isolate were individually aligned using Seqman (DNAStar) using reference sequence EU293117 and/or specific isolate sequences where available. The resulting consensus sequences were used in GS Reference Mapper (Roche) to obtain further sequence reads from the original raw data for each isolate. All four complete genome sequences were obtained, apart from the extremities of the genome (UTRs). The UTRs were inferred from the previously determined MOKV UTR sequences by using RT-PCR primers situated at the beginning and end of the genome (Marston, unpublished).

Table 1. Mokola virus reports from 1968–2012.

Geographical location	Year	Origin	Lab reference number	Reference	Accession numbers
Ibadan, Nigeria	1968	Shrew (*Crocidura spp*)	RV4	[7,8]	KF155005
Ibadan, Nigeria	1968	Human		[7,10]	?†
Ibadan, Nigeria	1969	Shrew (*Crocidura spp*)		[7]	?†
Umhlanga Rocks, KwaZulu Natal Province, South Africa	1970 (Identified in the 1980's)	Feline	700/70	[17]	FJ465416 (N), AF049118 (P), GQ472989 (M), GQ473001 (G)
Ibadan, Nigeria	1971	Human		[7,9]	?†
Yaounde, Cameron	1974	Shrew (*Crocidura spp*)	RV39	[11]	#EU293117
Bangui, Central African Republic	1981	Rodent (*Lophuromys sikapusi*)	RV40	[12]	EU293118
Bulawayo, Zimbabwe	1982	Feline	13270	[13]	KC218932 (N), GQ500114 (P), GQ472990 (M), GQ473002 (G)
Bulawayo, Zimbabwe	1981	Feline	12341	[13]	FJ465417 (N), GQ861350 (P), GQ472991 (M), GQ473003 (G)
Bulawayo, Zimbabwe	1981	Feline		[51]	NC_006429
Bulawayo, Zimbabwe	1981	Feline	12574	[13]	FJ465418 (N), GQ861352 (P), GQ472994 (M), GQ473004 (G)
Bulawayo, Zimbabwe	1982	Feline	Zim82/RV1035	[13,31]	KF155006
Bulawayo, Zimbabwe	1981	Canine (vaccinated)		[13,31]	?†
Addis Adaba, Ethiopia	1989–1990	Feline	RV610	[15]	#AY333111 (N)
Selous, Zimbabwe	1993	Feline	21846/RV1017	[16]	KC218933 (N), GQ500115 (P), GQ472993 (M), GQ500109 (G)
Mdantsane, Eastern Cape Province, South Africa	1995	Feline	543/95	[18]	FJ465415 (N), GQ500116 (P), GQ472992 (M), GQ500110 (G)
East London, Eastern Cape Province, South Africa	1996	Feline	112/96/RV1021	[19,20]	KF155008
Yellow Sands, Eastern Cape Province, South Africa	1996	Feline (vaccinated)	322/96	[19,20]	FJ465414 (N), GQ861353 (P), GQ472996 (M), GQ500111 (G)
Pinetown, KwaZulu Natal Province, South Africa	1997	Feline (vaccinated)	252/97	[19,20]	JN944637 (N), AF369376 (P), GQ472997 (M), GQ500112 (G)
Pinetown, KwaZulu Natal Province, South Africa	1997	Feline (vaccinated)	229/97	[19,20]	FJ465413 (N), AF369375 (P), GQ472998 (M), GQ500113 (G)
Pietermaritzburg, KwaZulu Natal Province, South Africa	1998	Feline (vaccinated)	071/98	[19,20]	FJ465410 (N), AF369378 (P), GQ473000 (M), GQ500108 (G)
Nkomazi, Mpumalanga Province, South Africa	2005	Canine	404/05	[21]	?†
East London, Eastern Cape Province, South Africa	2006	Feline (vaccinated)	173/06	[21]	FJ465412 (N), GQ861351 (P), GQ472999 (M), HQ266624 (G)
Grahamstown, Eastern Cape Province, South Africa	2008	Feline (vaccinated)	226/08	[21]	KC218934 (N), KC218935 (P), KC218936 (M), KC218937 (G)

†Indicates that the existence of the isolate is not known and no full gene sequences were available in Genbank or
#the sequences were likely the same isolate and therefore were not included for MCC analysis.
doi:10.1371/journal.pntd.0002511.t001

Phylogenetic analyses

Nucleotide sequences were assembled and edited using Vector NTI 9.1.0 (Invitrogen). Multiple sequence alignments were generated using ClustalX and exported in FASTA format. Phylogenetic and evolutionary analyses were conducted using Mega 5.05 [41] for a variety of data sets, i.e. the N, P, M and G gene nucleotide sequences as well as the concatenated sequence. The p-distances between MOKV N gene nucleotide and amino acids sequences were also calculated.

The Maximum Clade Credibility (MCC) phylogenetic tree, estimates of the rate of molecular evolution (substitutions per site

Figure 1. Map of Africa indicating approximate locations of MOKV isolations. The number next to the dots indicates the number of isolates isolated in the same (or in close locations such that the difference cannot be seen in the figure) location.
doi:10.1371/journal.pntd.0002511.g001

per year) and the most recent common ancestor (MRCA) for MOKV were inferred using a Bayesian Markov Chain Monte Carlo (MCMC) method in the BEAST package (BEAST and associated programmes are available via http://beast.bio.ed.ac.uk/) [42]. For this analysis, an input file for BEAST was generated using the BEAUti programme. For MCMC analysis of the concatenated gene sequence dataset (N, P, M and G genes) partitioning into genes was implemented The analysis utilized the general time reversible model with gamma distribution and proportion of invariable sites (GTR+G+I) with site heterogeneity [43] and population histories were constructed using the Bayesian skyline plot [44]. The relaxed (uncorrelated lognormal) molecular clock was chosen as demographic model. The statistical uncertainty in the data for each parameter estimate is reflected by the value of the 95% highest posterior density (HPD). For each estimate, duplicate BEAST runs were performed to test the reproducibility of the analysis. The BEAST output was assessed using the TRACER programme. For each analysis, a chain length of >30 million steps resulted in an effective sampling size (ESS>200 unless noted), with 10% burn-in removed. Trees and parameters were recorded every 10 000 steps. The trees obtained from BEAST were used as input for the TREEANNOTATOR programme to find the MCC tree. Phylogenetic trees were edited for publication using FigTree (version 1.3.1; http://tree.bio.ed.ac.uk/software/figtree/) [42]. Posterior probability values represent the degree of support for each node on the tree.

Results

We have generated a comprehensive dataset of all available isolates of MOKV, however, we were unable to trace some, as indicated in Table 1. Of the 24 reported detections of MOKV over the past 50 years, only 18 isolates could be included in this study. We were unable to track virus isolates for four cases reported in literature. Three of these were historical cases from Nigeria, the existence of which are now uncertain viz. two human isolates and a further isolate from a shrew [6,8,9]. The fourth was a dog-associated case reported in recent times from South Africa

[20], for which an isolate was never produced. Isolates RV39 (Cameroon, 1974) and RV610 (Ethiopia, 1990) were excluded from the MCC phylogenetic analyses as sequence data indicated that these isolates, in our hands, were likely to be the same virus (Fig. 2). A small number of nucleotide differences on some genes were believed to be due to mutations introduced from multiple cell culture passages over years of laboratory maintenance.

A number of domains on the lyssavirus genome have been implicated in the varying degrees of virulence between virus isolates of a lyssavirus species, as well as between virus isolates of different species of the *Lyssavirus* genus. A comparison of these amino acid positions is provided in Table 2. A similar pattern of amino acid substitutions on these positions was observed for the majority of MOKV isolates with specific differences observed on AA 144 (P-gene), AA 81 (M-gene) and AA 194, 198, 268, 352 and 330 (G-gene).

The genetic relationships between the different MOKV isolates was determined by construction of a MCC tree using the concatenated full coding regions of the N, P, M and G gene sequences as well as the individual genes (supplementary material). The MOKV isolates analyzed in this study formed a cluster supported by bootstrap values >70% when nucleotide sequences from the concatenated N, P, M and G genes (Fig. 3), N (supplementary material, Fig. S1), P (supplementary material, Fig. S2), M (supplementary material, Fig. S3) and G gene (supplementary material, Fig. S4) were used. The same tree topology was observed for both nucleotide and amino acids

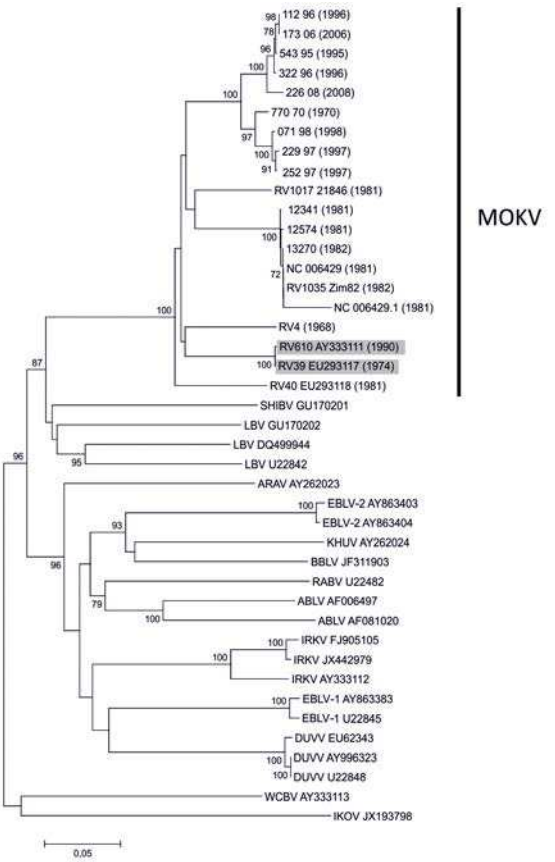


Figure 2. Evolutionary relationships of lyssaviruses inferred using the Neighbor-Joining method (500 replicates) using the Kimura 2-parameter method.
doi:10.1371/journal.pntd.0002511.g002

sequence analysis (data not shown). The MCC trees indicated that isolates grouped according to geographic location. Phylogenetic analysis of MOKV isolates from South Africa and Zimbabwe demonstrated geographic clustering consistent with previous findings [18,34]. The South African isolates formed two clusters consisting of KZN and EC provinces respectively. The Zimbabwean isolates from the 1980s (all from Bulawayo) formed a single cluster, distinct from the single 1993 isolate (from Selous). The same grouping was demonstrated for these isolates (South African and Zimbabwean) irrespective of the gene used for phylogenetic analysis.

The Central African Republic and Nigeria isolates formed independent clusters irrespective of the gene used for analyses.

The P-distance comparison between different MOKV isolates was performed using the N gene nucleotide and amino acid sequences (Table 3). Comparison of the nucleotide sequences indicated the difference between the MOKV isolates to be between 0 and 15% (85% nt seq identity), with the highest value (15%) being between U22843 (Zimbabwe) and RV4 (Nigeria). The highest nucleotide difference between South African isolates was 5.7% (226/08 and 229/97) while for Zimbabwean isolates it was 12.3% (between U22843 and RV1017/21846). Collectively, the nucleotide difference between South African isolates and Zimbabwean isolates was 14.4% (U22843 and 226/08). When comparing amino acid differences between MOKV isolates the same trend was observed, with MOKV isolates displaying an overall intragenotypic amino acid variation of 6.4%.

In order to investigate the evolutionary relationship of MOKV, a MCMC analysis was used to estimate the rate of nucleotide substitution calculated in substitutions/site/year as well as the time of the most recent common ancestor (MRCA) of MOKV (Table 4). When analyzing the N and G gene datasets, the mean nucleotide substitution rate was estimated to be 2.172×10^{-4} (N) and 2.123×10^{-4} (G). This is in agreement with previously published nucleotide substitution rate estimates (N gene: 1.1×10^{-4} to 3.8×10^{-4} substitutions per site per year, G gene: 5.56×10^{-4} to 1.286×10^{-3} substitutions per site per year) [40,45,46,47,48]. The age of the MRCA of MOKV was estimated to be 591 years old (95% HPD 294–1005 years) or 657 years old (95%HPD 279–1174 years), respectively. Analyses of the M-gene and P-gene yielded less robust estimates (1883 years and 1703 years respectively) (supplementary material, Fig. S2 and S3) with 95% HPD ranges much wider that the estimates for the N and G gene datasets. This is possibly due to the more variable nature of the M- and P-genes. Estimates based on the concatenated sequences (N, P, M and G-gene coding regions) yielded estimates within the ranges of the other genes (1157 years old, 95%HPD 413–2034 years) (Fig. 3).

The successful use of NGS on four of the isolates (RV4, RV1017, RV1021 and RV1035) enabled full genome consensus sequences to be obtained without the use of specific primers. In comparison to the work involved to obtain gene specific sequences for N, P, M and G on each of the MOKV isolates this approach was relatively simple and time efficient. Of the total number of reads obtained from the brain RNA preparations, between 0.25 and 1% were viral equating to between 294 and 1006 reads.

Discussion

This study was aimed at producing further insights into the phylogeny and diversity within a unique African lyssavirus species, MOKV. It was our objective to include all MOKV's encountered in history, but the existence or identity of several reported viruses and/or isolates could not be corroborated (Table 1). These included 3 virus isolates that were reported from Nigeria, the

Table 2. Comparison of amino acid (AA) sites of MOKV isolates influential for viral properties.

protein	region	Function/Effect	most common motive	deviations (isolate)	reference
N	AA 273	evasion of retinoic acid-inducible gene I mediated innate immunity and pathogenicity	F		[60]
	AA 394				
P	AA 144-148	P protein binding to the LC8 dynein light chain	IQIQT	VQIQT (229-97; 071-98; 252-97; 770-70)	
M	AA 22-25	important for pathogenicity of a related rhabdovirus,VSV	A		[61]
	AA 35-38	efficient virion release and pathogenicity	PPEYVPL		[62]
	AA 77	important in disruption of the mitochondrion and induction of apoptosis	K		
	AA 81		N	S (071-98; 770-70; 229-97; 252-97)	[63]
	AA 95	Val to Ala at position 95 results in increased apoptosis	V		[64]
G	AA 194	Asn – Lys increased viral spread, internalization & pathogenicity	S		[65]
	AA 198	mutation of Arg/Lys 198 resultsin reduced pathogenicity	K	Q (RV4; EU293118)	[66]
	AA 242	important for pathogenicity of the Nishigahara strain (Ile 268 most important residue)	S		[67]
	AA 255		N		
	AA 268		I	V (RV1017)	
	AA 318	p75NTR receptor binding	L		[68]
	AA 352		M	L (226-08)	
	AA 330-333	Arg/Lys 330 responsible for virulence in mice	KRVD	NRVD (RV1017)	[69,70]
		Double mutation R/K 333 and R/K 330 further reduces virulence			[71]

doi:10.1371/journal.pntd.0002511.t002

existence of which is now doubtful [6,8,9] and an isolate reported recently from South Africa [20]. It was also unfortunate that isolates from Cameroon and Ethiopia had to be excluded from this study, as these viruses, in our hands, were likely of the same original stock.

Nevertheless, a panel of 18 MOKV isolated over a period of nearly 50 years (Table 1) could be included and thus represents the most comprehensive phylogenetic analysis of full N, P, M and G genes of the MOKV species.

The monophyletic grouping of isolates from the KZN province of South Africa, isolated over a period of 28 years, indicates the continual presence and stability of the same viral lineage in this geographical domain. This KZN group could be distinguished from the other South African MOKV group, from the EC province, but the time point of divergence is rather recent with the MCRA for these two MOKV groups in the order of 150 years. The sequence diversity observed also seems to determine biological properties of the isolates.

Parallel experimental infection studies in mice showed that the pathogenicity of MOKV (isolates 12341, 252/97, Table 1) had been underestimated, although specific markers could not be determined [49]. Our analysis using a more comprehensive set of sequences corroborated these results (Table 2), but the relevance need to be confirmed by further studies.

Previously, when lyssaviruses were still classified according to genotypes, it was proposed that a new genotype is defined by >80% nucleotide differences and >92–93% amino acid differences [50,51]. Although this classification is no longer used, the p-distance analysis in this study indicated the MOKV isolates fall within the defined ranges. The MOKV isolates also displayed less sequence divergence than that seen among LBV isolates (20.9% nucleotide sequence difference and 6.7% amino acid sequence difference between LBV isolates). However, the delineation between LBV and MOKV using maximum clade credibility appears not as robust as when using M-gene where LBV isolates rather cluster with MOKV (supplementary material, Fig. S3).

MRCA estimates of MOKV utilizing different genes ranged widely from 591 to 1883 years (HPD 214–4318 years). Although these dates for the MOKV MRCA correspond with the timeframe estimated by Bourhy et al. [47] for the emergence of RABV associated with non-flying mammals (749 years ago, 95% HPD 363–1215 years), it must be noted that the small sample size of MOKV could also influence the robustness of the results. Also, purifying selection can mask the ancient age of viruses that appear to have recent origins as shown for other RNA-viruses [52,53], thus making it difficult to objectively model the evolutionary history of MOKV.

Use of NGS technologies to obtain four of the MOKV genomes directly from RNA preparations without amplification using specific primers was highly successful. Unlike the approach taken for the other isolates where often primers had to be designed for each specific isolate due to the high divergence seen in the sequences between the MOKV viruses, for the NGS approach

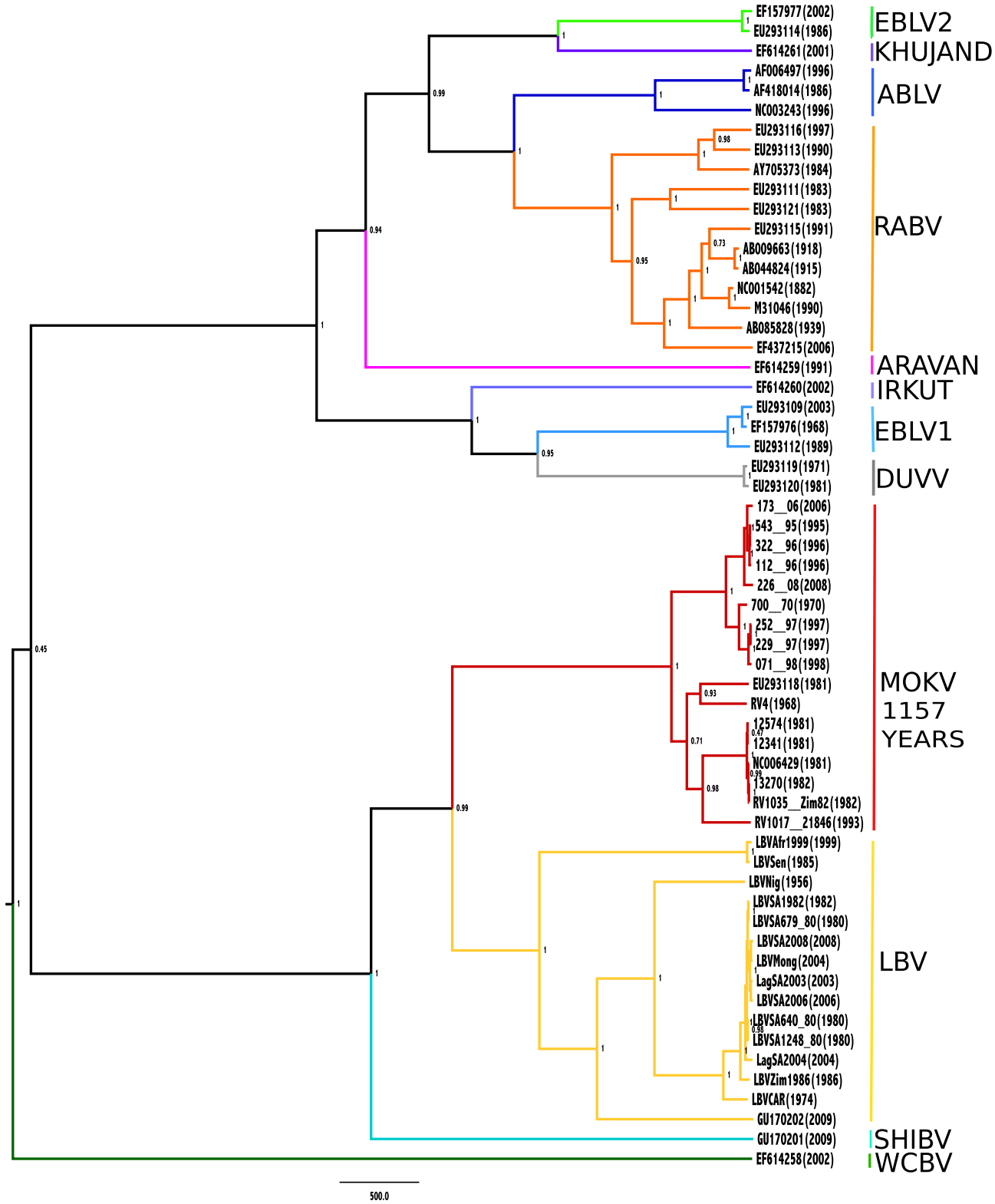


Figure 3. MCC phylogenetic tree based on the concatenated nucleotide sequence of the complete N, P, M and G gene of MOKV isolates and representative lyssavirus isolates. A table indicating the details of the isolates used in the analysis is provided in the supplementary material (Table S1).
doi:10.1371/journal.pntd.0002511.g003

the same methodology, i.e. random priming was applied to all isolates, regardless of their divergence. Although a high percentage of non-target sequences were produced, full coding sequences were obtained from all four MOKV isolates which was then put forward to the individual gene analyses. Given the inherent accuracy of 99.9% of Roche 454 and a sequence depth between

Table 3. Percentage difference of the nucleotides (A) and amino acids (B) of the N protein of MOKV isolates.

A	LBVSA2008	071/98	252/97	229/97	700/70	112/96	173/06	543/95	226/08	322/96	U22843	12341	12574	NC_006429	13270	RV1017/21864	EU293118	RV4	RV1035/Zim82
LBVSA2008																			
071/98	0.224																		
252/97	0.224	0.004																	
229/97	0.227	0.007	0.003																
700/70	0.226	0.021	0.023	0.023															
112/96	0.223	0.048	0.049	0.050	0.036														
173/06	0.224	0.050	0.050	0.052	0.037	0.010													
543/95	0.224	0.045	0.047	0.049	0.033	0.003	0.007												
226/08	0.224	0.053	0.056	0.057	0.041	0.018	0.019	0.015											
322/96	0.224	0.046	0.048	0.050	0.033	0.005	0.010	0.002	0.017										
U22843	0.239	0.136	0.138	0.138	0.135	0.139	0.142	0.141	0.144	0.139									
12341	0.213	0.136	0.110	0.110	0.107	0.112	0.115	0.113	0.116	0.111	0.032								
12574	0.214	0.110	0.111	0.111	0.109	0.115	0.116	0.115	0.118	0.113	0.033	0.003							
NC_006429	0.215	0.110	0.111	0.111	0.109	0.113	0.116	0.115	0.118	0.113	0.030	0.001	0.003						
13270	0.215	0.110	0.111	0.111	0.109	0.113	0.116	0.115	0.118	0.113	0.030	0.001	0.003	0.000					
RV1017/21846	0.226	0.104	0.105	0.107	0.096	0.108	0.107	0.107	0.108	0.106	0.123	0.099	0.099	0.099	0.099				
EU293118	0.220	0.117	0.116	0.119	0.116	0.118	0.119	0.118	0.118	0.117	0.145	0.118	0.119	0.119	0.119	0.111			
RV4	0.224	0.110	0.111	0.113	0.107	0.113	0.112	0.112	0.117	0.111	0.150	0.124	0.124	0.124	0.124	0.112	0.111		
RV1035/Zim82	0.215	0.110	0.111	0.111	0.109	0.113	0.116	0.115	0.118	0.113	0.030	0.001	0.003	0.000	0.000	0.099	0.119	0.124	

B	LBVSA2008	071/98	252/97	229/97	700/70	112/96	173/06	543/95	226/08	322/96	U22843	12341	12574	NC_006429	13270	RV1017/21864	EU293118	RV4	RV1035/Zim82
LBVSA2008																			
071/98	0.107																		
252/97	0.107	0.000																	
229/97	0.109	0.002	0.002																
700/70	0.104	0.007	0.007	0.009															
112/96	0.109	0.013	0.013	0.016	0.007														
173/06	0.109	0.013	0.013	0.016	0.007	0.000													
543/95	0.109	0.013	0.013	0.016	0.007	0.000	0.000												
226/08	0.113	0.018	0.018	0.020	0.011	0.004	0.004	0.004											
322/96	0.109	0.013	0.013	0.016	0.007	0.000	0.000	0.000	0.004										
U22843	0.147	0.058	0.058	0.060	0.053	0.060	0.060	0.060	0.064	0.060									
12341	0.102	0.013	0.013	0.016	0.009	0.016	0.016	0.016	0.020	0.016	0.044								
12574	0.102	0.013	0.013	0.016	0.009	0.016	0.016	0.016	0.020	0.016	0.044	0.000							
NC_006429	0.102	0.013	0.013	0.016	0.009	0.016	0.016	0.016	0.020	0.016	0.044	0.000	0.000						
13270	0.102	0.013	0.013	0.016	0.009	0.016	0.016	0.016	0.020	0.016	0.044	0.000	0.000	0.000					

Table 3. Cont.

B	LBVSA2008	071/98	252/97	229/97	700/70	112/96	173/06	543/95	226/08	322/96	U22843	12341	12574	NC_006429	13270	RV1017/21864	EU293118	RV4	RV1035/Zim82
RV1017/21846	0.107	0.013	0.013	0.016	0.011	0.018	0.018	0.018	0.022	0.018	0.053	0.009	0.009	0.009	0.009				
EU293118	0.107	0.011	0.011	0.013	0.013	0.016	0.016	0.016	0.020	0.016	0.056	0.011	0.011	0.011	0.011	0.011			
RV4	0.107	0.020	0.020	0.022	0.018	0.020	0.020	0.020	0.024	0.020	0.064	0.020	0.020	0.200	0.200	0.020	0.018		
RV1035/Zim82	0.102	0.013	0.013	0.016	0.009	0.016	0.016	0.016	0.020	0.016	0.044	0.000	0.000	0.000	0.000	0.009	0.011	0.020	

doi:10.1371/journal.pntd.0002511.t003

Table 4. Results from molecular clock analysis for all genes.

Gene analyzed	Age of MOKV MRCA	Nucleotide substitution rate (substitutions/site/year)
N gene	591 yrs (294–1005 yrs)	$2.516E^{-4}$ ($1.2009E^{-4}$ to $3.8862E^{-4}$)
P gene	1703 yrs (214–3907 yrs)	$1.114E^{-4}$ ($2.0828E^{-5}$ to $2.5012E^{-4}$)
M gene	1883 yrs (392–4318 yrs)	$7.963E^{-5}$ ($1.4543E^{-5}$ to $1.698E^{-4}$)
G gene	657 yrs (279–1174 yrs)	$2.123E^{-4}$ ($5.5067E^{-5}$ to $3.7878E^{-4}$)

Ninety five percent HPD values are indicated in brackets.

doi:10.1371/journal.pntd.0002511.t004

294 and 1006 reads per position the inclusion of sequencing errors are highly unlikely.

Twenty three MOKV isolations and one PCR-based identification have been made to date from six African countries in a period of more than 40 years. Since these African countries are from regions in Africa that are far apart, it is likely that MOKV is present in many others African countries and spread over vast portions of the continent. Moreover, over the past almost 20 years MOKV has only been isolated from South Africa (Fig. 1). Since it is known that MOKV is not only limited to South Africa, the lack of isolation from elsewhere is reflective of the non-existence of appropriate surveillance, including for rabies virus, across Africa. Limited diagnostic capabilities (e.g. typing or sequencing of rabies cases/specimens/isolates) across the continent, remains a key factor. Such enhanced surveillance would likely result in the discovery of more isolates and therefore, a higher diversity of MOKV and would thus improve our understanding of MOKV incidence and circulation. Since rabies vaccines do not offer protection against MOKV, a case can be made for the relative importance of a better understanding of the ecology of MOKV [32]. One of the limiting factors in studying MOKV is the fact that the reservoir species for this virus is not known. Although shrews have been implicated, it remains speculative. Lyssaviruses have a strong association with bats and it seems peculiar that MOKV may be the only exception in this regard - among all the other members of the genus. Indeed, virus neutralizing antibodies (VNA), neutralizing both LBV and MOKV have been detected in sera from frugivorous bats (*Rousettus aegyptiacus* and *Eidolon helvum*) [54,55,56]. However, belonging to the same phylogroup II, LBV and MOKV have been reported to cross react in serological assays [7,24,45,57]. Since there have been repeated reports of LBV isolations from fruit bats [58,59], the neutralizing activity of bat sera to MOKV apparently does not confirm the circulation of MOKV in those bat species. However, it cannot be excluded that other yet unidentified African bats may act as reservoir for MOKV. On the other hand, consistent encounters of MOKV in domestic cats and small mammalian species invite speculation along the lines of a prey-to-predator transmission pathway. For MOKV, the estimated MRCA from our study coincides and provides support for the timeframe suggested for the emergence of terrestrial rabies [47]. It is possible that MOKV remained stable in an extant African host environment, while RABV evolution was vastly accelerated given a plethora of host opportunities and global distribution.

Supporting Information

Figure S1 MCC phylogenetic tree based on the entire nucleotide sequence of the N gene of MOKV isolates and representative lyssavirus isolates.
(EPS)

Figure S2 MCC phylogenetic tree based on the entire nucleotide sequences of the P gene of MOKV isolates and representative lyssavirus isolates.
(EPS)

Figure S3 MCC phylogenetic tree based on the entire nucleotide sequences of the M gene of MOKV isolates and representative lyssavirus isolates.
(EPS)

Figure S4 MCC phylogenetic tree based on the entire nucleotide sequences of the G gene of MOKV isolates and representative lyssavirus isolates.
(EPS)

Table S1 Additional lyssavirus sequences used in phylogenetic analyses (see figure 2–3 and figure S1-S4).
(DOCX)

Acknowledgments

We thank Jacobeth Miyen, Ernest Ngoepe, Susann Schares and Jeannette Kliemt for excellent technical assistance, and three anonymous reviewers for their helpful comments.

Author Contributions

Conceived and designed the experiments: LHN TFM ARF CTS. Performed the experiments: JK NW JC CMF DM WM CTS. Analyzed the data: JK NW JC CMF DM WM CTS. Contributed reagents/materials/analysis tools: JK NW JC WM DM ARF CFM TFM CTS LHN. Wrote the paper: JK CMF DM LHN TFM ARF CTS WM.

References

1. Dietzgen RG, Calisher CH, Kurath G, Kuzmin IV, Rodriguez LL, et al(2012) Family Rhabdoviridae. In: A.M.Q. . King, M.J. . Adams, E.B. . Carstens and E.J. . Lefkowitz (Eds), Virus taxonomy : classification and nomenclature of viruses - Ninth Report of the International Committee on Taxonomy of Viruses, pp. 686–713. Elsevier, San Diego.
2. Banyard AC, Hayman D, Johnson N, McElhinney LM, Fooks AR (2011) Bats and Lyssaviruses. Adv Virus Res 79: 239–289.
3. Kuzmin IV, Mayer AE, Niezgoda M, Markotter W, Agwanda B et al. (2010) Shimoni bat virus, a new representative of the lyssavirus genus. Virus Res 149: 197–210.
4. Marston DA, Horton DL, Ngeleja C, Hampson K, McElhinney LM et al. (2012) Ikoma lyssavirus, Highly Divergent Novel Lyssavirus in an African Civet. Emerg Infect Dis 18: 664–667.
5. Aitken THG, Kowalski RW, Beaty BJ, Buckley SM, Wright JD, et al. (1984) Arthropod studies with rabies-related Mokola virus. Am J Trop Med Hyg 33: 945–952.
6. Kemp GE, Causey OR, Moore DL, Odelola A, Fabiyi A (1972) Mokola virus: Further studies on IbAn 27377, a new rabies-related etiological agent of zoonosis in Nigeria. Am J Trop Med Hyg 21: 356–359.
7. Shope RE, Murphy FA, Harrison AK, Causey OR, Kemp GE et al. (1970) Two African viruses serologically and morphologically related to rabies virus. J Virol 6: 690–692.
8. Familusi JB, Osunkoya BO, Moore DL, Kemp GE, Fabiyi A (1972) A fatal human infection with Mokola virus. Am J Trop Med Hyg 21: 959–963.
9. Familusi JB, Moore DL (1972) Isolation of a rabies-related virus from the cerebrospinal fluid of a child with 'aseptic meningitis. Afr J Med Sci 3: 93–96.
10. Le Gonidec G, Rickenbach A, Robin Y, Heme G (1978) Isolation of a strain of Mokola virus in Cameroon. Annu Microbiol 129: 245–249.
11. Saluzzo JF, Rollin PE, Dauguet C, Digoutte JP, Georges AJ et al. (1984) Premier isolment du virus Mokola a partir D'un rongeur (Lophuromys sikapusi). Ann Inst Pasteur Virol 135: 57–66.
12. Foggin CM (1982) Atypical rabies virus in cats and a dog in Zimbabwe. Vet Rec 110: 338.
13. Foggin CM (1988) Rabies and rabies-related viruses in Zimbabwe: Historical, virological and ecological aspects. PhD thesis. Harare: University of Zimbabwe.
14. Mebatsion T, Cox JH, and Frost JW (1992) Isolation and characterisation of 115 street rabies virus isolates from Ethiopia by using monoclonal antibodies: identification of 2 isolates of Mokola and Lagos Bat viruses. J Infect Dis 166: 972–977
15. Bingham J, Javangwe S, Sabeta CT, Wandeler AI, Nel LH (2001) Report of isolations of unusual lyssaviruses (rabies and Mokola virus) identified retrospectively from Zimbabwe. J S Afr Vet Assoc 72: 92–94.
16. Schneider LG, Barnard BJH, Schneider HP (1985) Application of monoclonal antibodies for epidemiological investigations and oral vaccination studies I-African viruses. In: "Rabies in the Tropics" pp. 47–59. E. Kuwert & C. Merieux. Berlin: Springer Verlag.
17. Meredith CD, Nel LH, Von Teichman BF (1996) Further isolation of Mokola virus in South Africa. Vet Rec 138: 119–120.
18. Nel LH, Jacobs J, Jafta J, von Teichman B, Bingham J (2000) New cases of Mokola virus in South Africa. Virus Genes 20: 103–106.
19. Von Teichman BF, De Koker WC, Bosch SJ, Bishop GC, Meredith CD et al. (1998) Mokola virus infection: description of recent South African cases and a review of the virus epidemiology. J S Afr Vet Assoc 69: 169–171.
20. Sabeta CT, Makotter W, Mohale DK, Shumba W, Wandeler AI et al. (2007) Mokola virus in domestic mammals, South Africa. Emerg Infect Dis 13: 1371–1373.

21. Sabeta C, Blumberg L, Miyen J, Mohale D, Shumba W et al. (2010) Mokola virus involved in a human contact (South Africa) FEMS Immunol Med Microbiol 58: 85–90
22. Brookes SM, Parsons G, Johnson N, McElhinney LM, Fooks AR (2005) Rabies human diploid cell vaccine elicits cross-neutralising and cross-protecting immune responses against European and Australian bat lyssaviruses. Vaccine 23: 4101–4109.
23. Hooper PT, Lunt RA, Gould AR, Samaratunga H, Hyatt AD et al. (1997) A new lyssavirus - the first endemic rabies-related virus recognized in Australia. Bull Institut Pasteur 95: 209–218.
24. Jallet C, Jacob Y, Bahloul C, Drings A, Desmezieres E et al.(1999) Chimeric lyssavirus glycoprotein genes with increased immunological potential. J Virol 73: 225–233.
25. Lafon M, Bourhy H, Sureau P (1988) Immunity against the European bat rabies (Duvenhage) virus induced by rabies vaccines: an experimental study in mice. Vaccine 6: 362–368.
26. Malerczyk C, Selhorst T, Tordo N, Moore S, Müller T (2009) Antibodies induced by vaccination with purified chick embryo cell culture vaccine (PCECV) cross-neutralize non-classical bat lyssavirus strains. Vaccine 27: 5320–5325.
27. Bahloul C, Jacob Y, Tordo N, Perrin P (1998) DNA-based immunization for exploring the enlargement of immunological cross-reactivity against the lyssaviruses. Vaccine 16: 417–425.
28. Fekadu M, Shaddock JK, Sanderlin DW, Smith JS (1988) Efficacy of rabies vaccines against Duvenhage virus isolated from European house bats (Eptesicus serotinus), classic rabies and rabies-related viruses. Vaccine 6: 533–539.
29. Fooks AR (2004) The challenge of emerging lyssaviruses. Expert Rev Vaccines 4: 89–92.
30. Le Mercier P, Jacob Y, Tordo N (1997) The complete Mokola virus genome sequence: structure of the RNA-dependent RNA polymerase. J Gen Virol 78: 1571–1576.
31. Nel LH, Niezgoda M, Hanlon CA, Morril PA, Yager PA et al. (2003) A comparison of DNA vaccines for the rabies-related virus, Mokola. Vaccine 21: 2598–2606.
32. Nel LH (2005) Vaccines for lyssaviruses other than rabies. Expert Rev Vaccines 4: 533–540.
33. Sabeta CT (2002) Aspects of the molecular epidemiology of rabies in Zimbabwe and South Africa. PhD thesis. Pretoria: University of Pretoria, South Africa.
34. Van Zyl N (2009) Molecular epidemiology of African mongoose rabies and Mokola virus. MSc thesis. Pretoria: University of Pretoria, South Africa.
35. Van Zyl N, Markotter W, Nel LH (2010) Evolutionary history of African mongoose rabies. Virus Res 150: 93–102.
36. Hayman DT, Johnson N, Horton DL, Hedge J, Wakeley PR et al. (2011) Evolutionary history of rabies in Ghana. PLoS Negl Trop Dis 5: e1001. doi:10.1371/journal.pntd.0001001
37. Hughes GJ, Orciari LA, Rupprecht CE (2005) Evolutionary timescale of rabies virus adaptation to North American bats inferred from the substitution rate of the nucleoprotein Gene. J Gen Virol 86: 1467–1474.
38. McElhinney LM, Marston D, Freuling CM, Cragg WJ, Stankov S et al. (2011) Molecular diversity and evolutionary history of rabies virus strains circulating in the Balkans. J GenVirol 92: 2171–2180.
39. Ming P, Yan J, Rayner S, Meng S, Xu G et al. (2010) A history estimate and evolutionary analysis of rabies virus variants in China. J Gen Virol 91: 759–764
40. Talbi C, Holmes EC, de Benedictis P, Faye O, Nakoune E et al. (2009) Evolutionary history and dynamics of dog rabies virus in western and central Africa. J Gen Virol 90: 783–791.
41. Tamura K, Peterson D, Peterson N, Stecher G, Nei M et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
42. Drummond AJ, Suchard MA, Xie D, Rambaunt A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29: 1969–1973.
43. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phyloegenetics and dating with confidence. PLoS Biol 4: e88. doi:10.1371/journal.pbio.0040088
44. Drummond AJ, Nicholis GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161: 1307–1320.
45. Badrane H, Bahloul C, Perrin P, Tordo N (2001) Evidence of two phylogroup with distinct pathogenicity and immunogenicity. J Virol 75: 3268–3276.
46. Badrane H, Tordo N (2001) Host switching in the Lyssavirus history from the Chiroptera to Carnivora orders. J Virol 17: 8096–8014.
47. Bourhy H, Reynes J, Dunham EJ, Dacheux L, Larrous F et al. (2008) The origin and phylogeography of dog rabies virus. J Gen Virol 89: 2673–2681
48. Holmes EC, Woelk CH, Kassis R, Bourhy H (2002) Genetic constraints and the adaptive evolution of rabies virus in nature. Virology 292: 247–257.
49. Kgaladi J, Nel LH, Markotter W (2013) Comparison of pathogenic domains of rabies and African rabiesrelated lyssaviruses and pathogenicity observed in mice. Onderstepoort J Vet Res 80: doi: 10.4102/ojvr.v80i1.511
50. Bourhy H, Kissi B, Tordo N (1993) Molecular diversity of the lyssavirus genus. Virology 194: 70–81.
51. Kissi B, Tordo N, Bourhy H (1995) Genetic polymorphism in the rabies virus nucleoprotein gene. Virology 209: 526–537.
52. Wertheim JO, Chu DKW, Peiris JSM, Pond SLK, Poon LLM (2013) A Case for the Ancient Origin of Coronaviruses. J Virol 87: 7039–7045.
53. Wertheim JO, Pond SLK (2011) Purifying Selection Can Obscure the Ancient Age of Viral Lineages. Mol Biol Evol 28: 3355–3365.
54. Dzikwi AA, Kuzmin IV, Umoh JU, Kwaga JKP, Ahmad AA et al. (2010) Evidence of Lagos bat virus circulation among Nigerian fruit bats. J Wildl Dis 46: 267–271.
55. Kuzmin IV, Niezgoda M, Franka R, Agwanda B, Markotter W et al. (2008) Lagos bat virus in Kenya. J Clin Microbiol 46: 1451–1461.
56. Wright E, Hayman DT, Vaughan A, Temperton NJ, Wood JL et al. (2010) Virus neutralising activity of African fruit bat (Eidolon helvum) sera against emerging lyssaviruses. Virology 408: 183–189.
57. Hanlon CA, Kuzmin IV, Blanton JD, Weldon WC, Manangan JS et al. (2005) Efficacy of rabies biologics against new lyssaviruses from Eurasia. Virus Res 111: 44–54.
58. Hayman DT, Fooks AR, Horton D, Suu-Ire R, Breed AC et al. (2008) High-level antibody seroprevalence to Lagos bat virus detected in megachiroptera from West Africa. Emerg Infect Dis 14: 926–928.
59. Hayman DT, Fooks AR, Rowcliffe M, Restif O, Baker K et al. (2012) Endemic Lagos bat virus infection in Eidolon helvum. Epidemiol Infect 140: 2163–2171.
60. Masatania T, Ito N, Shimizua K, Ito Y, Nakagawaa K et al. (2011) Amino acids at positions 273 and 394 in rabies virus nucleoprotein are important for both evasion of host RIG-I mediated antiviral response and pathogenicity. Virus Res 155: 168–174.
61. Irie T, Carnero E, Okumura A, García-Sastre A, Harty RN (2007) Modifications of the PSAP region of the matrix protein lead to attenuation of vesicular stomatitis virus in vitro and in vivo', J Gen Virol 88: 2559–2567.
62. Wirblich C, Tan GS, Papaneri A, Godlewski PJ, Orenstein JM et al. (2008) PPEY motif within the rabies virus (RV) matrix protein is essential for efficient virion release and RV pathogenicity. J Virol 82: 9730–9738.
63. Gholami A, Kassis R, Real E, Delmas O, Guadagnini S et al. (2008) Mitochondrial dysfunction in lyssavirus-induced apoptosis', J Virol 82: 4774–4784.
64. Mita T, Shimizu K, Ito N, Yamada K, Ito Y et al. (2008) Amino acid at position 95 of the matrix protein is a cytopathic determinant of rabies virus. Virus Res 137: 33–39.
65. Faber M, Faber ML, Papaneri A, Bette M, Weihe E et al. (2005) A single amino acid change in rabies virus glycoprotein increases virus spread and enhances virus pathogenicity. J Virol 79: 14141–14148
66. Lentz TL, Wilson PT, Hawrot E, Speicher DW (1984) Amino acid sequence similarity between rabies virus glycoprotein and snake venom curaremimetic neurotoxins. Science 226: 847–848.
67. Takayama-Ito M, Ito N, Yamadaa K, Sugiyama M, Minamoto N (2006) Multiple amino acids in the glycoprotein of rabies virus are responsible for pathogenicity in adult mice. Virus Res 115: 169–175.
68. Langevin C, Tuffereau C (2002) Mutations conferring resistance to neutralization by a soluble form of the neurotrophin receptor (p75NTR) map outside of the known antigenic sites of the rabies virus glycoprotein. J Virol 76: 10756–10765.
69. Tuffereau C, Leblois H, Bénéjean J, Coulon P, Lafay F et al. (1989) Arginine or lysine in position 333 of ERA and CVS glycoprotein is necessary for rabies virulence in adult mice. Virology 172: 206–212.
70. Dietzschold B, Wunner WH, Wiktor TJ, Lopes AD, Lafon M et al., 1983, 'Characterization of an antigenic determinant of the glycoprotein that correlates with pathogenicity of rabies virus. Proc Natl Acad Sci U S A 80: 70–74.
71. Coulon P, Ternaux JP, Flamand A, Tuffereau C (1998) An avirulent mutant of rabies virus is unable to infect motoneurons in vivo and in vitro. J Virol 7: 273–8

5.2.2: Article 26: published in PLOS Neglected Tropical Diseases

The Global Phylogeography of Lyssaviruses - Challenging the 'Out of Africa' Hypothesis

DOI:10.1371/journal.pntd.0005266

Summary: Understanding the evolution of lyssaviruses is fundamental to unravelling the processes occurring during viral emergence and spread. This manuscript utilizes whole genome sequences of lyssaviruses isolated from bats to ask the question 'where did lyssaviruses originate?' The manuscript complements the impressive WGS analysis of over 200 RABV isolates (bat and meso-carnivore hosts) which tackled the question of how RABV has evolved in multiple host reservoirs (Troupin et al., 2016). The genome sequences obtained for IKOV, GBLV and LLEBV (Section 3.2: Articles 16, 19 and 21) are unique representatives for each species and enabled a complete analysis of bat lyssaviruses identified to date.

These analyses were undertaking using Bayesian Inferences testing the hypothesis using a probabilistic phylogeographical approach, initially on N gene sequences. The resulting Maximum Clade Credibility (MCC) tree was confirmed using complete genome sequences and the ancestral nodes were confirmed using three different approaches. The data presented suggests that the MRCA of the contemporary lyssaviruses originated in the Palearctic region. This is perhaps a controversial discovery as Africa contains the most diverse range of lyssavirus species, including representative from each phylogroup, therefore a more intuitive hypothesis would be that lyssaviruses evolved 'out of Africa'(Rupprecht et al., 2017). However, more lyssavirus species have been isolated from the Palearctic region than any other region, including representatives of Phylogroup I and III. The data suggests that there have been 3 independent transmission events to the Afrotropical region representing the three major lyssavirus phylogroups. It should be noted that IKOV - the African representative in phylogroup III has only been detected once – (in an African civet in Tanzania), therefore the distribution of IKOV is unknown and possibly covers the Palearctic region.

PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

The Global Phylogeography of Lyssaviruses - Challenging the 'Out of Africa' Hypothesis

David T. S. Hayman[1]*, Anthony R. Fooks[2,3], Denise A. Marston[2], Juan C. Garcia-R[1]

1 Molecular Epidemiology and Public Health Laboratory, Hopkirk Research Institute, Massey University, Palmerston North, New Zealand, **2** Wildlife Zoonoses and Vector-borne Diseases Research Group, Animal and Plant Health Agency (APHA), Weybridge-London, United Kingdom, **3** Department of Clinical Infection, Microbiology & Immunology, Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom

* d.t.s.hayman@massey.ac.nz, davidtshayman@gmail.com

Citation: Hayman DTS, Fooks AR, Marston DA, Garcia-R JC (2016) The Global Phylogeography of Lyssaviruses - Challenging the 'Out of Africa' Hypothesis. PLoS Negl Trop Dis 10(12): e0005266. doi:10.1371/journal.pntd.0005266

Editor: Remi Charrel, Aix Marseille University, Institute of Research for Development, and EHESP School of Public Health, FRANCE

Received: July 24, 2016

Accepted: December 15, 2016

Published: December 30, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files (all accession numbers are in S2-S6 Figs and S1 Table).

Funding: DTSH acknowledges funding from the Royal Society of New Zealand Marsden Fund (MAU1503) and the Research and Policy for Infectious Disease Dynamics program of the Science and Technology Directorate, Department of Homeland Security, Fogarty International Center, National Institutes of Health; JCGR and DTSH from

Abstract

Rabies virus kills tens of thousands of people globally each year, especially in resource-limited countries. Yet, there are genetically- and antigenically-related lyssaviruses, all capable of causing the disease rabies, circulating globally among bats without causing conspicuous disease outbreaks. The species richness and greater genetic diversity of African lyssaviruses, along with the lack of antibody cross-reactivity among them, has led to the hypothesis that Africa is the origin of lyssaviruses. This hypothesis was tested using a probabilistic phylogeographical approach. The nucleoprotein gene sequences from 153 representatives of 16 lyssavirus species, collected between 1956 and 2015, were used to develop a phylogenetic tree which incorporated relevant geographic and temporal data relating to the viruses. In addition, complete genome sequences from all 16 (putative) species were analysed. The most probable ancestral distribution for the internal nodes was inferred using three different approaches and was confirmed by analysis of complete genomes. These results support a Palearctic origin for lyssaviruses (posterior probability = 0.85), challenging the `outof Africa' hypothesis, and suggest three independent transmission events to the Afrotropical region, representing the three phylogroups that form the three major lyssavirus clades.

Author Summary

Rabies virus kills tens of thousands of people globally each year and causes indescribable misery and family disturbance, especially in developing countries. Yet in much of the world there are related viruses, called lyssaviruses, which circulate among bats without causing conspicuous outbreaks. The greater diversity of African lyssaviruses has led to the hypothesis that Africa is the origin of these viruses. To test this hypothesis, the genetic data from 153 representative viruses from 16 available lyssavirus species from across the world dated between 1956 and 2015 were analysed. Statistical models were used to reconstruct the historical processes that lead to the contemporary distribution of these viruses.

the New Zealand Ministry of Health; and ARF was financially supported by the UK Department for Environment, Food and Rural Affairs (Defra), Scottish Government and Welsh Government by grants SEV3500 and SE0431, The Research and Policy for Infectious Disease Dynamics program of the Science and Technology Directorate, Department of Homeland Security, Fogarty International Center, National Institutes of Health and from the EU H2020-funded Research Infrastructure Grant 'European Virus Archive Global (EVAg)' [H2020 - grant agreement n°653316]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Our results support a Palearctic origin for lyssaviruses, not Afrotropic, and suggest three independent transmission events to Africa from the Palearctic region.

Introduction

Determining the evolutionary history of viruses is fundamental to our understanding of the patterns and processes occurring during viral emergence and spread. Emergence and spread of viral diseases is a permanent threat in animal and public health and special attention has been given to fast-evolving RNA viruses due to the high mortality rates recorded worldwide. The family *Rhabdoviridae* contains a diverse variety of RNA viruses that replicate in vertebrates, invertebrates and plants. The vast majority of rhabdoviruses have invertebrate vectors that play a role in the transmission to plants, fishes or mammals. The lyssaviruses, which cause the disease rabies, are unique within these negative-sense, single-stranded RNA viruses because they do not require arthropod vectors and are well-adapted to their mammalian hosts [1].

The prototypic virus within the genus *Lyssavirus* is rabies virus (RABV) [2]. RABV has a global distribution (with the exception of Australasia, Antarctica, and some islands). The principal reservoir of RABV is the domestic dog (*Canis familiaris*), although RABV is enzootic in a number of wildlife Carnivora, including fox, mongoose, raccoon, and skunk populations. RABV variants undergo genetic adaptation to each particular host, resulting in new clades or biotypes relating to the local fauna [3–11]. Well-established wildlife Carnivora reservoirs for RABV are apparently absent in South America and Australasia. However, lyssaviruses are present in both of these regions and throughout the rest of the world in bat hosts (Chiroptera).

All lyssaviruses have been isolated from bats, with the exception of Mokola virus (MOKV) and Ikoma lyssavirus (IKOV), and phylogenetic analyses suggest all lyssaviruses have bat origins [12–16]. RABV has only been identified in bats in the Americas and is the only lyssavirus detected circulating in American bat species. This observation is in direct contrast to the rest of the world, where RABV has not been detected in bat populations. The greatest diversity of lyssaviruses occurs in bats in Eurasia and on the African continent. The species are divergent enough that sera raised against specific virus species often do not neutralize other virus species and are divided into phylogroups (PG) [14, 17, 18].

Rabies virus is a member of phylogroup 1 (PG1). European bat lyssavirus-1 (EBLV-1) and 2 (EBLV-2), Bokeloh bat lyssavirus (BBLV), Irkut (IRKV), Aravan (ARAV) and Khujand (KHUV) viruses all belong to PG1. These PG1 viruses have all been isolated from bats in Eurasia, as has West Caucasian Bat Virus (WCBV) from PG3. Lleida bat lyssavirus (LLEBV) is a tentative new species yet to be recognised by the International Committee on Taxonomy of Viruses (ICTV). LLEBV has been identified from a bat in Spain and is most closely related to the African Ikoma lyssavirus (IKOV) genetically in PG3 [19–21]. Gannoruwa bat lyssavirus (GBLV), also awaiting ICTV recognition as a new species, has recently been isolated from fruit bats in Sri Lanka, and phylogenetic analysis indicated that it is most closely related to RABV (PG1) [22]. In Australasia the only lyssavirus identified is Australian bat lyssavirus (ABLV), also a PG1 virus, which circulates within the Australian bat populations. In Africa, however, all three phylogroups are represented: Duvenhage virus (DUVV) from PG1, Lagos bat virus (LBV), MOKV and Shimoni bat virus (SHIBV) from PG2 [13], and IKOV [23] from PG3.

The observation that the greatest genetic diversity of lyssaviruses is in Africa has led to the hypothesis that Africa is the continent where lyssaviruses originated, most likely from an African bat reservoir [24]. This hypothesis is founded on sound observations, though does not

address the lack of known reservoir(s) for MOKV, and was proposed despite a lack of understanding of the ecology of MOKV and many other lyssavirus species. Here, the hypothesis that lyssaviruses had their origins in Africa was tested by using a probabilistic phylogeographical approach, which provides additional insights into the historical biogeography of lyssaviruses.

Methods

We used 153 nucleoprotein (N) gene sequences from the 14 recognized and 2 putative lyssavirus species (LLEBV, GBLV) in these analyses (final dataset, see below), along with complete genome sequences from all 16 (putative) species when appropriate for confirmatory analyses. Lyssavirus sequences were mostly derived from bats, with exception of MOKV and IKOV (see Introduction). Using bat-derived sequences was particularly important for RABV because the evolutionary history was less likely to be confounded by the global spread of RABV by human movement of terrestrial carnivores and the post-war RABV epidemics in wildlife [25–27]. Sequences represent serially sampled data; the earliest sequence from 1956 and the last from 2015, including the most recently available sequences from GBLV and LLEBV, spanning a 59-year period.

The dataset was aligned in ClustalX2.1 [28] and inspected by eye. Bayesian Markov Chain Monte Carlo (MCMC) implemented in BEAST software v.1.8.3 [29] was used for phylogenetic analysis and estimation of divergence times. A codon partition strategy was implemented with a general time reversible (GTR) model of substitution with gamma distributed variation in rates amongst sites and a proportion of sites assumed to be invariant according to the Akaike criterion in Modeltest [30]. The lower number of substitutions per site in EBLV-1 (S1 Fig) [31] compared across the tree can potentially cause problems for the estimation of the posterior probabilities and other parameters. To mitigate this issue, the number of EBLV-1 sequences was reduced in the final dataset (S1 Table). Divergence times were estimated using a strict clock model in BEAST assuming an underlying coalescent process with a constant population size. To be more conservative in our estimates of the divergence times and assuming that purifying selection has removed deleterious mutations from rate estimates in short timeframes [32–34], a 2.3 x 10^{-4} substitution rate estimated by Bourhy and colleagues [35] was used. An uncorrelated lognormal relaxed clock was also considered because it assumes that the substitution rate along branches is not correlated.

Parameter effective sample sizes were visualized in Tracer v.1.5 (http://tree.bio.ed.ac.uk/software/tracer/). The results from two independent runs with 5x10^{7} MCMC length of chain were combined. The first 10% of maximum clade credibility (MCC) trees were discarded as burn-in in TreeAnnotator v1.8.3 [36]. Final trees were visualized in FigTree v.1.4.2 [37]. The fit of each analysis (strict and lognormal) were evaluated with Bayes Factors in Tracer v.1.5 (http://tree.bio.ed.ac.uk/software/tracer/).

Ancestral distributions were first performed in BEAST as part of the analysis for the estimations of the divergence times. The analysis assumed the forward and reverse rates to be symmetrical (Mk1). For comparison, our sampling was reduced to contain a single complete genome representative of each taxon, including GBLV and LLEBV, and we performed a RAxML analysis [38] to obtain a Maximum Likelihood phylogenetic tree. This phylogenetic tree was used to infer the most probable ancestral distribution for the internal nodes with Likelihood and Parsimony approaches and restriction of equal probability for all state changes with the Mk1 model using Mesquite v3.04 [39]. For ancestral state analyses viruses were categorized by the terrestrial ecozone from which they were isolated [35]. The major ecozones were Afrotropical, Palearctic, Oriental and Nearctic and Neotropic (both Americas) (Fig 1). Note that we use the term Oriental for the often named Indomalaya region but do not partition

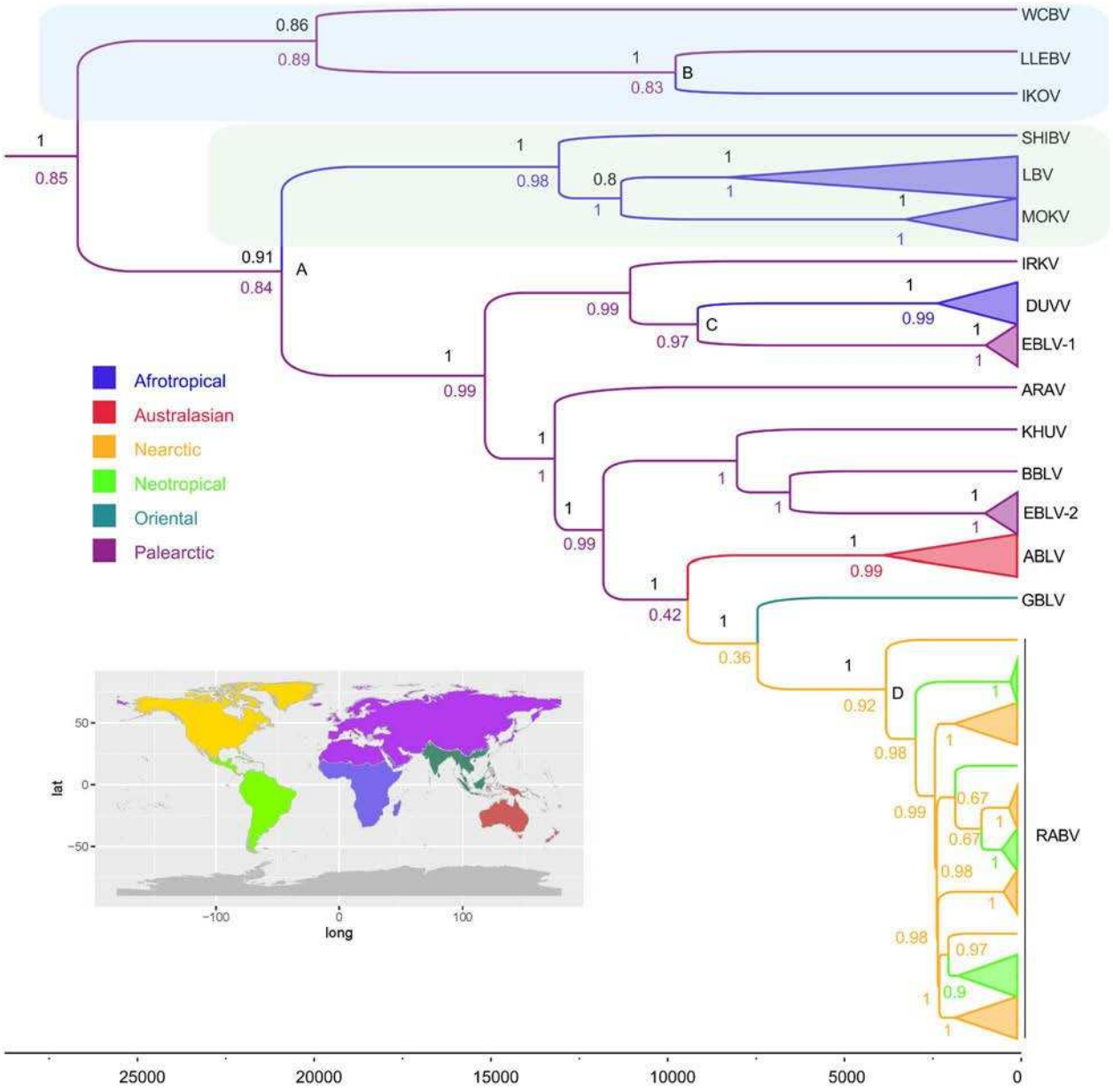


Fig 1. Evolutionary relationships between lyssaviruses. The time-scaled phylogeny was generated from 153 nucleoprotein gene sequences and inferred with a Lognormal relaxed-clock Bayesian analysis using BEAST. Branch colours correspond to ecozones shown on the inset map. Support values corresponding to Bayesian posterior probabilities (above branches) and states probabilities from the different assigned ecozones (below branches) are indicated for key nodes. The time scale in years is shown. Phylogroups 3 (green, top) and 2 (blue) are shaded and key nodes discussed in the text labelled A-D. Virus names are *Mokola virus* (MOKV); *Australian bat lyssavirus* (ABLV); *European bat lyssavirus-1* (EBLV-1); *European bat lyssavirus-2* (EBLV-2), *Irkut* (IRKV), *Aravan* (ARAV), *Khujand* (KHUV); *West Caucasian Bat Virus* (WCBV); *Lagos bat virus* (LBV); *Duvenhage virus* (DUVV); *Shimoni bat virus* (SHIBV); *Bokeloh bat lyssavirus* (BBLV); *Ikoma virus* (IKOV); *Lleida virus* (LLEBV); *Gannoruwa bat lyssavirus* (GBLV); *Rabies virus* (RABV).

doi:10.1371/journal.pntd.0005266.g001

the Palearctic into Palearctic, Saharo-Arabian and Sino-Japanese regions [40] because no viruses were available from these latter two regions. Because RABV can be identified in more than one ecozone, its distribution category was treated as uncertain for the Likelihood analysis whilst polymorphic for the Parsimony analysis. Summary trees are presented in the main text

and detailed trees with additional information, such as with all tip labels, are presented as figures in the supplementary information.

Results

Using a panel of lyssavirus N genes with global distributions there was strong support for the overall tree topology (Fig 1). The topology was confirmed through complete genome analysis (Fig 2). The uncorrelated relaxed clock (-25893.0, Fig 1) outperformed the strict clock model (-25999.2, S2 Fig). The results were supportive of a Palearctic origin for the lyssaviruses (posterior probability (PP) = 0.85, Fig 1, PP = 0.86 S2 Fig, strict clock). The results suggest there were three independent transmission events from the Palearctic to the African region, one each from the three putative phylogroups (Fig 1). One event led to the presence of IKOV in Africa (node B). Another event led to the distinct PG2 virus clade (SHIBV, LBV and MOKV) having their current African distribution (node A). It has been proposed that EBLV-1 had its origins in Africa, being closely related to DUVV, whereas our analysis suggested there was greater support for DUVV being a subsequent introduction into the Afrotropical region from the Palearctic (PP = 0.96, node C). Likelihood and Parsimony analyses of individual viral species genomes both provided support for these results (Fig 2). Our analysis supports an easterly spread of lyssaviruses to Australasia, the Oriental realm and the Americas. The inclusion of GBLV into our dataset supports previous findings that GBLV shares a most recent common ancestors (MRCA) with RABV (PP = 1), but support for this ancestral state node was the weakest (PP = 0.36). ABLV (previously the most closely related lyssavirus species to RABV) shares a MRCA with GBLV, although support for this ancestral state is also weak (PP = 0.42) (Fig 1).

Though not the aim of our analysis, we also estimate the time to the MRCA (tMRCA). The uncertainty around these estimates is large; however they suggest that these events probably took place tens of thousands of years ago. The three median tMRCA for the branching events relating to the Palearctic to Afrotropical region clades are 20820 years ago (95% highest posterior density, HPD, 3995 to 166820) for the PG1/PG2 ancestors (node A), 9676 year ago (1102 to 83408 95% HPD) for IKOV and LLEBV MRCA (node B) and 9048 years ago (1405 to 77181 95% HPD) for the DUVV and EBLV-1 MRCA (node C). The results also suggest that once RABV entered the Americas there was widespread dispersal of RABV between the Neotropical and Nearctic regions (Fig 1). Estimated tMRCA for the earliest RABV in our dataset is 3726 years ago (593 to 31478 95% HPD, node D).

Discussion

Our analyses support a Palearctic origin for the lyssaviruses (PP = 0.85). This is despite the high diversity of lyssaviruses found in the Africotropical region. The support for most state probabilities is high, suggesting there is strong geographic and temporal structure to lyssavirus evolution as previously demonstrated by RABV in non-volant Carnivora species [41, 42]. Our estimation of the temporal origins of extant lyssaviruses varies depending on whether we used a fixed rate of evolution or estimated the rate using a relaxed clock. This does not, however, affect the phylogeographic inferences from this analysis because of the accurate estimation of evolutionary relationships among species (Figs 1 and 2). These analyses reject the hypothesis that lyssaviruses emerged as viruses of bats in Africa and suggest three distinct emergence events from the Palearctic region into the Afrotropical region (Fig 1).

There is topological support for the current phylogroups (PG1-3), each with monophyletic origins (Fig 1). A PG 3, with lower support, contains species IKOV, LLEBV and WCBV. PG 2 has a single Afrotropical ecozone range composed by SHIBV, MOKV and LBV. PG 1 divides into two major lineages, one which contains IRKV, EBLV-1 and DUVV, and another lineage

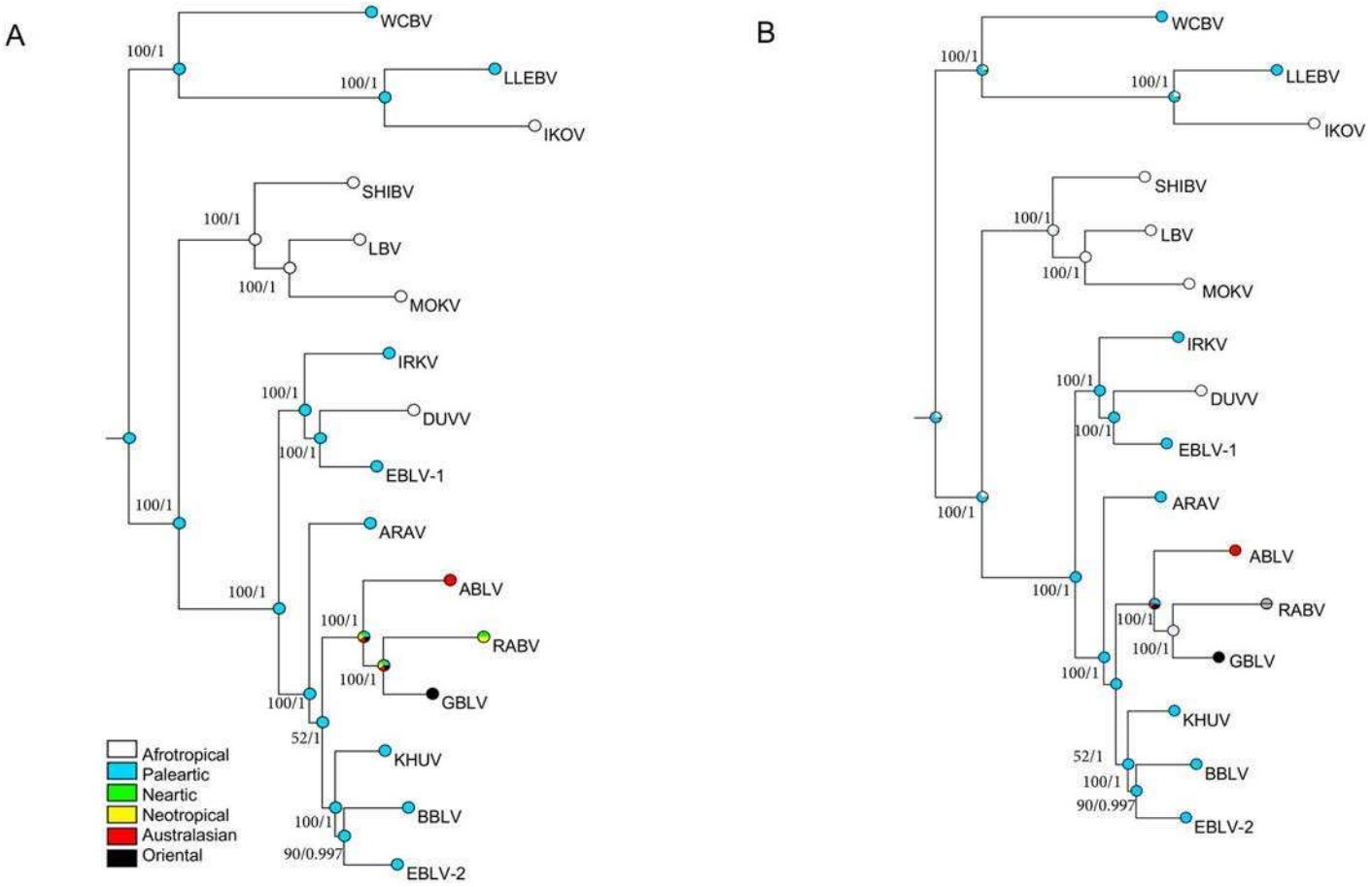


Fig 2. Ancestral state reconstruction using complete genomes of the 16 *Lyssavirus* species. The reconstruction was based on the Maximum Likelihood tree and using the parsimony (A) and likelihood (B) models in Mesquite v.3.0.4. Coloured pie-charts represent proportions generated from the different assigned states of the character (see colour legends). Grey terminal pie-chart in likelihood analysis indicated a polymorphic state that was coded as uncertain in the data matrix for RABV species. Support values are indicated above branches and correspond to bootstrap and posterior probabilities, respectively. Virus names are detailed in Fig 1.

doi:10.1371/journal.pntd.0005266.g002

that includes the Palearctic (ARAV, BBLV, KHUV and EBLV-2), Australian (ABLV), Oriental (GBLV) and American (RABV) lyssavirus species. A possible mechanism for the distribution of these viruses may be provided by considering the hosts of WCBV and LLEBV (PG3) and potential hosts of EBLV-1 and DUVV (PG1). Each of these viruses has been detected in the same bat genus, *Miniopterus*, thus these bats may have played a role in the inter-continental spread of lyssaviruses. This bat genus occurs in the Afrotropical and Palearctic ecozones, though species such as *Miniopterus schreibersi* previously thought to be distributed across both ecozones are now recognized to have more complex taxonomies [43]. However, the Palearctic species, *Miniopterus schreibersi*, from which WCBV was isolated in the Russian Caucasus and LLEBV and EBLV-1 were identified in Spain, occurs in both North Africa and Eurasia. WCBV is yet to be detected in Africa, but specific antibodies against WCBV have been detected in African *Miniopterus* in Kenya, providing possible evidence of a bat lyssavirus species circulating in both Africa and Europe [14, 44, 45]. There is no evidence of cross-neutralization between WCBV and IKOV (most closely related lyssavirus to WCBV), suggesting the antibodies detected in the *Miniopterus* spp in Kenya were WCBV specific [46]. The genetically close relationship between EBLV-1 and DUVV has been used as evidence of viruses from Africa

entering Europe [47–49]. However, in our analysis, a greater support for transmission of EBLV-1 from the Palearctic to the Afrotropical zone was observed. The observation that both EBLV-1 and DUVV share a common ancestor with IRKV (Palearctic) strengthens this finding. The isolation of SHIBV from *Hipposideros vittatus* in Kenya [13] is the first lyssavirus isolation from a bat of the genus *Hipposideros*. *Hipposideros* has a broad distribution in the Old World from tropical Africa through to China, although like *Miniopterus* each species has a more specific distribution. Further sampling within this taxon will determine if SHIBV has crossed from the Palearctic to Afrotropical ecozones, as may be the case from the limited data available from WCBV [13] and our analyses (Fig 1).

Our results were confirmed through complete genome analyses with single representatives from each species. Despite this strong support for our conclusion that lyssaviruses have spread from the Palearctic region, it should be recognized that two of the most divergent viruses, WCBV and LLEBV, isolated from Russia [14] and Spain [20] respectively, may influence these findings [50, 51] and the ancestors of the *Lyssavirus* genus themselves may have originated from outside the Palearctic region, indeed even within the Afrotropical ecozone. The relationships of RNA viruses can be influenced by the different processes, such as through high evolutionary rates and subsequent purifying selection [52, 53]. We observed that EBLV-1 had reduced nucleotide substitutions per site compared to other viruses. The earliest EBLV-1 sequence is from 1968 and the sequences are taken from a relatively large geographic area (ranging from Ukraine to Spain), however EBLV-1 has a restricted host range. EBLV-1 has only ever been detected in the serotine bat (*Eptesicus serotinus*) outside Spain. In Spain, it has been reported from six European insectivorous bat species through active survey, although the majority were *E. serotinus* [47, 54]. Therefore, future studies should determine if host restriction is reducing the rate of nucleotide substitution and by what mechanism. The data from Spain may be evidence of this virus expanding into new niches and undergoing a bottleneck that reduced genetic diversity during the process. However, of the relatively well studied bat lyssaviruses, EBLV-2 has even greater known host restriction to *M. daubentonii*, and narrower geographic range, so alternative mechanisms may be responsible for the reduced substitution rate we estimated. As more full length genomes from more viruses become available, especially from bats, better inference can be made regarding evolutionary relationships for lyssaviruses [22]. In particular, we suggest that future studies aim to discover more viruses from Africa, Asia and the eastern Palearctic region.

Other significant dispersal events observed from this dataset are from the Palearctic to Australasia and between the Nearctic and Neotropical ecozones in both directions. These analyses provide support for the spread of lyssaviruses, in an easterly direction, from the Palearctic to Australasia and subsequent colonization into the Americas. Australia is free of RABV within its wildlife population with only occasional cases of imported human rabies [55–58], however, isolation of ABLV from sick bats [59], people [60–63], and through surveillance in bats [64–66] suggest ABLV is well established. Indeed, two distinct ABLV lineages apparently circulate, one now isolated from all four species of Australian Pteropodidae and another from an insectivorous bat species, *Saccolaimus flaviventris* [59, 64, 65, 67]. The isolation of GBLV from *Pteropus medius* in Sri Lanka in 2015 was significant because it is more closely related to RABV than any of the other Old World lyssaviruses currently identified. The host reservoir of GBLV (*P. medius*) is interesting for a number of reasons when considering spread of lyssaviruses between ecozones. Firstly, ABLV has been isolated from all four *Pteropus* spp in Australia. Secondly, members of the *Pteropus* genus are present throughout Asia and Australasia, providing a possible mechanism for transmission between the two ecozones. Thirdly, serosurveillance of bat populations in Asia has detected lyssavirus-specific antibodies, yet no virus had been

isolated [16]. However, the ABLV host *S. flaviventris* has also been reported in Papua New Guinea [68] suggesting other pathways may exist.

Our analyses of the RABV data provide strong support for transmission between the American ecozones (Fig 1). This may be due to two, not mutually exclusive mechanisms. One is that despite recent evidence of host phylogeny constraining inter-species virus transmission with the USA [69], RABV has been isolated from over 23 bat species in the USA alone. Host signatures for species variants exist, but adaptation is not sufficient to prevent cross-species transmission [69]. Therefore, mutations in the RABV genome may have led to reduced host restriction and RABV being able to spread more easily between bats in both North and South American locations. An additional mechanism for this finding may be the hosts themselves. There are highly sociable, numerous, and/or migratory bats which occur throughout the Americas, such as *Tadarida brasiliensis*, *Lasiurus cinereus*, and *Eptesicus fuscus*. These migratory and widespread bat species of the Americas may have rapidly disseminated RABV between ecozones, enabling the promiscuous RABV to rapidly exploit unoccupied niches. Furthermore, the presence of *L. cinereus* in both Hawaii and British Columbia (http://www.iucnredlist.org/details/11345/0) demonstrates how a migratory species may be a potential vector of RABV or a RABV ancestor from the Palearctic to the Americas.

Our molecular clock analyses provide support for the hypothesis that RABV was circulating in the bat populations of the Americas before the arrival of Europeans in the late 15th Century. Previously it has been claimed that Spanish conquistadors reported attacks by bats on humans and that native Americans knew that cauterisation may prevent disease development [70]. The median tMRCA for GBLV and RABV is >7,000 years ago. The median tMRCA for the first internal branch in RABV is 3726 years ago (593 to 31478 95% HPD). We are cautious when interpreting the ages of RNA viruses using molecular clock analysis because of the impacts of purifying selection on RNA viruses [52, 71], however purifying selection should push our dates further into the past rather than bring the estimates forward in time. Thus, we are confident that our analysis provides support for the probable presence of RABV in bats, and likely in the Americas, before the arrival of Europeans in 1492CE, because our lowest 95% HPD RABV tMRCA date is 1423CE and the median estimate is 3726 years ago in 1790 BCE.

The origins of RABV in dogs is debated and phylogenetic analyses have questioned the reports of RABV from ancient Greek references [70]. Our analyses demonstrate that lyssaviruses were almost certainly circulating in Palearctic bats at this time. Similarly, whether estimating evolutionary rates using relaxed clocks as in other analyses [35] or fixing the evolutionary rate, there appear to be extant RABV circulating in bats in the Americas at this time. Thus, it may be possible that the reports from 23rd Century BC are due to an "extant" RABV if spillover had occurred from bats to terrestrial carnivores already. Future analyses of all extant lyssaviruses accounting for purifying selection may help elucidate these relationships further [52].

Lastly, extant Chiroptera bats likely originated in Asia/Europe and young clades are found in the Americas [72]. This biogeographic reconstruction reflects a similar pattern as the one found in the *Lyssavirus* genus. However, the time of divergence for extant Chiroptera is in Millions of years and there is no information to suggest co-speciation between bats and lyssavirus reflects a possible ancient origin of these viruses, as has been found in other groups (e.g. coronaviruses [71] and papillomaviruses [73]).

In conclusion, our analyses provide support for the monophyletic, Palearctic origins of lyssaviruses with dispersal from there to the rest of the world. And while three dispersal events have been from the Palearctic to the Afrotropical regions, arguably the dispersal events that led to the greatest impact on animal and human health are those eastward, where the RABV species appears to have evolved and dispersed globally from, leading to 23,000–93,000 human

deaths a year [74]. Understanding why this lyssavirus, but not others, has emerged globally will provide insights into the processes that drive viral emergence.

Supporting Information

S1 Table. Partial nucleoprotein (N) or complete (C) lyssavirus genome sequences from GenBank used in the analysis. Virus names are as Fig 1.
(DOCX)

S1 Fig. Phylogeny of a panel of lyssaviruses showing the slow rate of nucleotide substitutions for European bat lyssavirus-1 (green) compared to the other viruses (black).
(PDF)

S2 Fig. Evolutionary relationships between lyssaviruses based on the nucleoprotein (N) gene with a strict clock analysis using a fixed evolutionary rate of $2.3x10^{-4}$ (according to [35]) in BEAST. Support values corresponding to Bayesian posterior probabilities are indicated. Tips are labelled with the following information: Ecozone, follow by the_species name, country where it was isolated, GenBank accession number and year of isolation. Virus names and other details are as Fig 1.
(PDF)

S3 Fig. Evolutionary relationships between lyssaviruses. The phylogenetic tree was generated from 153 nucleoprotein gene sequences and inferred with a Lognormal relaxed-clock Bayesian analysis using BEAST. Support values corresponding to Bayesian state probabilities from the different assigned ecozones are indicated. Tips are labelled with the following information: Ecozone, follow by the species name, country where it was isolated, GenBank accession number and year of isolation. Virus names and other details are as Fig 1.
(PDF)

S4 Fig. Evolutionary relationships between lyssaviruses. The phylogenetic tree was generated from 153 nucleoprotein gene sequences and inferred with a Lognormal relaxed-clock Bayesian analysis using BEAST showing divergence times in years with 95% credible intervals. Branch colours correspond to ecozones shown on the map. The time scale is in years. Virus names are as Fig 1.
(PDF)

S5 Fig. Ancestral state reconstruction using complete genomes of the 16 *Lyssavirus* species based on the Maximum Likelihood tree using likelihood models in Mesquite v.3.0.4. Coloured pie-charts represent proportions generated from the different assigned states of the character (see colour legends). The grey terminal pie-chart indicated a polymorphic state that was coded as uncertain in the data matrix for RABV species. Support values are indicated above branches and correspond to bootstrap and posterior probabilities, respectively. Virus names are as Fig 1.
(PDF)

S6 Fig. Ancestral state reconstruction using complete genomes of the 16 *Lyssavirus* species based on the Maximum Likelihood tree and using the parsimony models in Mesquite v.3.0.4. Coloured pie-charts represent proportions generated from the different assigned states of the character (see colour legends). Support values are indicated above branches and correspond to bootstrap and posterior probabilities, respectively. Virus names are as Fig 1.
(PDF)

Acknowledgments

Thank you to Dr. Jessica Hedge, University of Oxford, for early useful discussions and two anonymous reviewers for their thoughtful reviews.

Author Contributions

Conceptualization: DTSH.

Data curation: JCGR DAM.

Formal analysis: DTSH JCGR.

Funding acquisition: DTSH ARF.

Investigation: DTSH JCGR.

Methodology: DTSH JCGR.

Resources: ARF DAM.

Software: DTSH JCGR DAM.

Supervision: DTSH.

Validation: ARF DAM.

Visualization: DTSH JCGR.

Writing – original draft: DTSH.

Writing – review & editing: DTSH JCGR DAM ARF.

References

1. Bourhy H, Cowley JA, Larrous F, Holmes EC, Walker PJ. Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene. Journal of General Virology. 2005; 86(Pt 10):2849±58. doi: 10.1099/vir.0.81128-0 PMID: 16186241
2. Afonso CL, Amarasinghe GK, Banyai K, Bao Y, Basler CF, Bavari S, et al. Taxonomy of the order Mononegavirales: update 2016. Archives of Virology. 2016; 161(8):2351±60. doi: 10.1007/s00705-016-2880-1 PMID: 27216929
3. Real LA, Russell C, Waller L, Smith D, Childs J. Spatial dynamics and molecular ecology of North American rabies. Journal of Heredity. 2005; 96(3):253±60. doi: 10.1093/jhered/esi031 PMID: 15677743
4. Hass CC, Dragoo JW. Rabies in hooded and striped skunks in Arizona. Journal of Wildlife Diseases. 2006; 42(4):825±9. doi: 10.7589/0090-3558-42.4.825 PMID: 17255450
5. Velasco-Villa A, Orciari LA, Souza V, Juarez-Islas V, Gomez-Sierra M, Castillo A, et al. Molecular epizootiology of rabies associated with terrestrial carnivores in Mexico. Virus Research. 2005; 111(1):13±27. doi: 10.1016/j.virusres.2005.03.007 PMID: 15896399
6. Rupprecht CE, Smith JS, Fekadu M, Childs JE. The ascension of wildlife rabies: a cause for public health concern or intervention? Emerging Infectious Diseases. 1995; 1(4):107±14. doi: 10.3201/eid0104.950401 PMID: 8903179
7. von Teichman BF, Thomson GR, Meredith CD, Nel LH. Molecular epidemiology of rabies virus in South Africa: evidence for two distinct virus groups. The Journal of General Virology. 1995; 76 (Pt 1)(1):73±82.
8. Nel LH, Thomson GR, Von Teichman BF. Molecular epidemiology of rabies virus in South Africa. Onderstepoort Journal of Veterinary Research. 1993; 60(4):301±6. PMID: 7777315
9. Swanepoel R, Barnard BJ, Meredith CD, Bishop GC, Bruckner GK, Foggin CM, et al. Rabies in southern Africa. Onderstepoort Journal of Veterinary Research. 1993; 60(4):325±46. PMID: 7777317
10. Davis PL, Rambaut A, Bourhy H, Holmes EC. The evolutionary dynamics of canid and mongoose rabies virus in Southern Africa. Archives of Virology. 2007; 152(7):1251±8. doi: 10.1007/s00705-007-0962-9 PMID: 17401615

11. Childs J, Real AL. Epidemiology. In: Jackson AC, Wunner WH, editors. Rabies. 2nd ed. London: Elsevier; 2007. p. 123±200.

12. Badrane H, Tordo N. Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders. Journal of Virology. 2001; 75(17):8096±104. doi: 10.1128/JVI.75.17.8096-8104.2001 PMID: 11483755

13. Kuzmin IV, Mayer AE, Niezgoda M, Markotter W, Agwanda B, Breiman RF, et al. Shimoni bat virus, a new representative of the Lyssavirus genus. Virus Research. 2010; 149(2):197±210. doi: 10.1016/j.virusres.2010.01.018 PMID: 20138934

14. Kuzmin IV, Hughes GJ, Botvinkin AD, Orciari LA, Rupprecht CE. Phylogenetic relationships of Irkut and West Caucasian bat viruses within the Lyssavirus genus and suggested quantitative criteria based on the N gene sequence for lyssavirus genotype definition. Virus Research. 2005; 111(1):28±43. doi: 10.1016/j.virusres.2005.03.008 PMID: 15896400

15. Kuzmin IV, Orciari LA, Arai YT, Smith JS, Hanlon CA, Kameoka Y, et al. Bat lyssaviruses (Aravan and Khujand) from Central Asia: phylogenetic relationships according to N, P and G gene sequences. Virus Research. 2003; 97(2):65±79. PMID: 14602198

16. Banyard AC, Hayman D, Johnson N, McElhinney L, Fooks AR. Bats and lyssaviruses. Advances in Virus Research. 2011; 79:239±89. doi: 10.1016/B978-0-12-387040-7.00012-3 PMID: 21601050

17. Fooks A. The challenge of new and emerging lyssaviruses. Expert Review of Vaccines. 2004; 3 (4):333±6. doi: 10.1586/14760584.3.4.333 PMID: 15270628

18. Horton DL, McElhinney LM, Marston DA, Wood JLN, Russell CA, Lewis N, et al. Quantifying Antigenic Relationships among the Lyssaviruses. Journal of Virology. 2010; 84(22):11841±8. doi: 10.1128/JVI.01153-10 PMID: 20826698

19. Freuling CM, Beer M, Conraths FJ, Finke S, Hoffmann B, Keller B, et al. Novel lyssavirus in Natterer's bat, Germany. Emerging Infectious Diseases. 2011; 17(8):1519±22. doi: 10.3201/eid1708.110201 PMID: 21801640

20. Ceballos NA, Moron SV, Berciano JM, Nicolas O, Lopez CA, Juste J, et al. Novel Lyssavirus in Bat, Spain. Emerging Infectious Diseases. 2013; 19(5):793±5. doi: 10.3201/eid1905.121071 PMID: 23648051

21. Marston DA, Ellis RJ, Wise EL, Aréchiga-Ceballos N, Freuling CM, Banyard AC, et al. Complete genomic sequence of Lleida bat lyssavirus. Genome Announcements 2017; 5(2):e01427±16.

22. Gunawardena PS, Marston DA, Ellis RJ, Wise EL, Karawita AC, Breed AC, et al. Lyssavirus in Indian Flying Foxes, Sri Lanka. Emerging Infectious Diseases. 2016; 22(8):1456±9. doi: 10.3201/eid2208.151986 PMID: 27434858

23. Marston DA, Ellis RJ, Horton DL, Kuzmin IV, Wise EL, McElhinney LM, et al. Complete genome sequence of Ikoma lyssavirus. Journal of Virology. 2012; 86(18):10242±3. doi: 10.1128/JVI.01628-12 PMID: 22923801

24. Nel LH, Rupprecht CE. Emergence of lyssaviruses in the Old World: the case of Africa. Current Topics in Microbiology and Immunology. 2007; 315:161±93. PMID: 17848065

25. Velasco-Villa A, Reeder SA, Orciari LA, Yager PA, Franka R, Blanton JD, et al. Enzootic rabies elimination from dogs and reemergence in wild terrestrial carnivores, United States. Emerging Infectious Diseases. 2008; 14(12):1849±54. doi: 10.3201/eid1412.080876 PMID: 19046506

26. Childs JE, Curns AT, Dey ME, Real AL, Rupprecht CE, Krebs JW. Rabies epizootics among raccoons vary along a North-South gradient in the Eastern United States. Vector-borne and Zoonotic Diseases. 2001; 1(4):253±67. doi: 10.1089/15303660160025895 PMID: 12653126

27. Flamand A, Coulon P, Lafay F, Kappeler A, Artois M, Aubert M, et al. Eradication of rabies in Europe. Nature. 1992; 360(6400):115±6. doi: 10.1038/360115a0 PMID: 1436089

28. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23(21):2947±8. doi: 10.1093/bioinformatics/btm404 PMID: 17846036

29. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology. 2007; 7:214. doi: 10.1186/1471-2148-7-214 PMID: 17996036

30. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics. 1998; 14 (9):817±8. PMID: 9918953

31. Davis PL, Holmes EC, Larrous F, Van der Poel WH, Tjornehoj K, Alonso WJ, et al. Phylogeography, population dynamics, and molecular evolution of European bat lyssaviruses. Journal of Virology. 2005; 79(16):10487±97. doi: 10.1128/JVI.79.16.10487-10497.2005 PMID: 16051841

32. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. Journal of Virology. 2010; 84(19):9733±48. doi: 10.1128/JVI.00694-10 PMID: 20660197

33. Holmes EC. Error thresholds and the constraints to RNA virus evolution. Trends in Microbiology. 2003; 11(12):543±6. PMID: 14659685

34. Holmes EC. Molecular clocks and the puzzle of RNA virus origins. Journal of Virology. 2003; 77 (7):3893±7. doi: 10.1128/JVI.77.7.3893-3897.2003 PMID: 12634349

35. Bourhy H, Reynes JM, Dunham EJ, Dacheux L, Larrous F, Huong VT, et al. The origin and phylogeography of dog rabies virus. The Journal of General Virology. 2008; 89(Pt 11):2673±81. doi: 10.1099/vir.0.2008/003913-0 PMID: 18931062

36. Rambaut A, Drummond A. TreeAnnotator v1. 7.5. 2013.

37. Rambaut A. FigTree v. 1.4. Molecular evolution, phylogenetics and epidemiology. Edinburgh, UK: University of Edinburgh, Institute of Evolutionary Biology. 2012.

38. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22(21):2688±90. doi: 10.1093/bioinformatics/btl446 PMID: 16928733

39. Maddison W, Maddison DR. Mesquite: a modular system for evolutionary analysis. 2001.

40. Holt BG, Lessard J-P, Borregaard MK, Fritz SA, Araújo MB, Dimitrov D, et al. An update of Wallace's zoogeographic regions of the world. Science. 2013; 339(6115):74±8. doi: 10.1126/science.1228282 PMID: 23258408

41. Biek R, Henderson JC, Waller LA, Rupprecht CE, Real LA. A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. Proceedings of the National Academy of Sciences U S A. 2007; 104(19):7993±8.

42. Real LA, Henderson JC, Biek R, Snaman J, Jack TL, Childs JE, et al. Unifying the spatial population dynamics and molecular evolution of epidemic rabies virus. Proceedings of the National Academy of Sciences U S A. 2005; 102(34):12107±11.

43. Miller-Butterworth CM, Eick G, Jacobs DS, Schoeman MC, Harley EH. Genetic and phenotypic differences between South African long-fingered bats, with a global miniopterine phylogeny. Journal of Mammalogy. 2005; 86(6):1121±35.

44. Kuzmin IV, Niezgoda M, Franka R, Agwanda B, Markotter W, Beagley JC, et al. Possible emergence of West Caucasian bat virus in Africa. Emerging Infectious Diseases. 2008; 14(12):1887±9. doi: 10.3201/eid1412.080750 PMID: 19046512

45. Botvinkin AD, Poleschuk EM, Kuzmin IV, Borisova TI, Gazaryan SV, Yager P, et al. Novel lyssaviruses isolated from bats in Russia. Emerging Infectious Diseases. 2003; 9(12):1623±5. doi: 10.3201/eid0912.030374 PMID: 14720408

46. Horton DL, Banyard AC, Marston DA, Wise E, Selden D, Nunez A, et al. Antigenic and genetic characterization of a divergent African virus, Ikoma lyssavirus. Journal of General Virology. 2014; 95(Pt 5):1025±32. doi: 10.1099/vir.0.061952-0 PMID: 24496827

47. Serra-Cobo J, Amengual B, Abellan C, Bourhy H. European bat lyssavirus infection in Spanish bat populations. Emerging Infectious Diseases. 2002; 8(4):413±20. doi: 10.3201/eid0804.010263 PMID: 11971777

48. Schneider LG, Cox JH. Bat lyssaviruses in Europe. Current Topics in Microbiology and Immunology. 1994; 187:207±18. PMID: 7859491

49. Amengual B, Whitby JE, King A, Cobo JS, Bourhy H. Evolution of European bat lyssaviruses. The Journal of general virology. 1997; 78 (Pt 9):2319±28.

50. Crisp MD, Cook LG. Do early branching lineages signify ancestral traits? Trends in Ecology and Evolution. 2005; 20(3):122±8. doi: 10.1016/j.tree.2004.11.010 PMID: 16701355

51. Losos JB. CommentariesÐUn certainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. Animal Behaviour. 1999; 58(6):1319±24. doi: 10.1006/anbe.1999.1261 PMID: 10600155

52. Wertheim JO, Pond SLK. Purifying Selection Can Obscure the Ancient Age of Viral Lineages. Molecular Biology and Evolution. 2011; 28(12):3355±65. doi: 10.1093/molbev/msr170 PMID: 21705379

53. Revell LJ, Harmon LJ, Collar DC. Phylogenetic signal, evolutionary process, and rate. Systematic Biology. 2008; 57(4):591±601. doi: 10.1080/10635150802302427 PMID: 18709597

54. Echevarria JE, Avellon A, Juste J, Vera M, Ibanez C. Screening of active lyssavirus infection in wild bat populations by viral RNA detection on oropharyngeal swabs. Journal of Clinical Microbiology. 2001; 39 (10):3678±83. doi: 10.1128/JCM.39.10.3678-3683.2001 PMID: 11574590

55. McColl KA, Gould AR, Selleck PW, Hooper PT, Westbury HA, Smith JS. Polymerase chain reaction and other laboratory techniques in the diagnosis of long incubation rabies in Australia. Australian Veterinary Journal. 1993; 70(3):84±9. PMID: 8476363

56. Faoagali JL, De Buse P, Strutton GM, Samaratunga H. A case of rabies. Medical Journal of Australia. 1988; 149(11±12):702±7. PMID: 3200197

57. Johnson N, Fooks A, McColl K. Human rabies case with long incubation, Australia. Emerging Infectious Diseases. 2008; 14(12):1950±1. doi: 10.3201/eid1412.080944 PMID: 19046531

58. McColl KA, Chamberlain T, Lunt RA, Newberry KM, Middleton D, Westbury HA. Pathogenesis studies with Australian bat lyssavirus in grey-headed flying foxes (Pteropus poliocephalus). Australian Veterinary Journal. 2002; 80(10):636±41. PMID: 12465817

59. Fraser GC, Hooper PT, Lunt RA, Gould AR, Gleeson LJ, Hyatt AD, et al. Encephalitis caused by a Lyssavirus in fruit bats in Australia. Emerging Infectious Diseases. 1996; 2(4):327±31. doi: 10.3201/eid0204.960408 PMID: 8969249

60. Warrilow D. Australian bat lyssavirus: a recently discovered new rhabdovirus. Current Topics in Microbiology and Immunology. 2005; 292:25±44. PMID: 15981466

61. Warrilow D, Smith IL, Harrower B, Smith GA. Sequence analysis of an isolate from a fatal human infection of Australian bat lyssavirus. Virology. 2002; 297(1):109±19. PMID: 12083841

62. Hanna JN, Carney IK, Smith GA, Tannenberg AE, Deverill JE, Botha JA, et al. Australian bat lyssavirus infection: a second human case, with a long incubation period. Medical Journal of Australia. 2000; 172 (12):597±9. PMID: 10914106

63. Allworth AM, Murray K, Morgan J. A human case of encephalitis due to a lyssavirus recently identified in fruit bats. Communicable Diseases Intelligence. 1996; 20:325.

64. Gould AR, Kattenbelt JA, Gumley SG, Lunt RA. Characterisation of an Australian bat lyssavirus variant isolated from an insectivorous bat. Virus Research. 2002; 89(1):1±28. PMID: 12367747

65. Gould AR, Hyatt AD, Lunt R, Kattenbelt JA, Hengstberger S, Blacksell SD. Characterisation of a novel lyssavirus isolated from Pteropid bats in Australia. Virus Research. 1998; 54(2):165±87. PMID: 9696125

66. Hooper PT, Lunt RA, Gould AR, Samaratunga H, Hyatt AD, Gleeson LJ, et al. A new lyssavirusÐthe first endemic rabies-related virus recognized in Australia. B I Pasteur. 1997; 95(4):209±18.

67. Guyatt KJ, Twin J, Davis P, Holmes EC, Smith GA, Smith IL, et al. A molecular epidemiological study of Australian bat lyssavirus. The Journal of General Virology. 2003; 84(Pt 2):485±96. doi: 10.1099/vir.0.18652-0 PMID: 12560583

68. Flannery TF. Mammals of New Guinea: Reed; 1995.

69. Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science. 2010; 329 (5992):676±9. doi: 10.1126/science.1188836 PMID: 20689015

70. Baer G. The History of Rabies. In: Jackson AC, Wunner WH, editors. Rabies. 2nd ed. London: Academic Press; 2007.

71. Wertheim JO, Chu DK, Peiris JS, Kosakovsky Pond SL, Poon LL. A case for the ancient origin of coronaviruses. Journal of Virology. 2013; 87(12):7039±45. doi: 10.1128/JVI.03273-12 PMID: 23596293

72. Teeling EC, Springer MS, Madsen O, Bates P, O'Brien S J, Murphy WJ. A molecular phylogeny for bats illuminates biogeography and the fossil record. Science. 2005; 307(5709):580±4. doi: 10.1126/science.1105113 PMID: 15681385

73. Rector A, Lemey P, Tachezy R, Mostmans S, Ghim SJ, Van Doorslaer K, et al. Ancient papillomavirus-host co-speciation in Felidae. Genome Biology. 2007; 8(4):R57. doi: 10.1186/gb-2007-8-4-r57 PMID: 17430578

74. Knobel DL, Cleaveland S, Coleman PG, Fevre EM, Meltzer MI, Miranda ME, et al. Re-evaluating the burden of rabies in Africa and Asia. Bulletin of the World Health Organisation. 2005; 83(5):360±8.

5.2.3: Article 27 published in PLOS Neglected Tropical Diseases

Spatio-temporal analysis of the genetic diversity of arctic Rabies viruses and their reservoir hosts in Greenland

DOI:10.1371/journal.pntd.0004779

Summary: This paper is a result of collaborations between multiple Institutes, coordinated by FLI (Germany). It represents the most thorough analysis of the RABV lineages circulating in Greenland (and the Arctic as a whole) combining archives with a collective dataset of RABVs from 1977 to 2014. 79 full genome sequences were obtained, of these 7 were provided by APHA. The samples were prepared and sequenced on the Illumina platform and the raw data sent to FLI for assembly. The fox (host) data was used to identify patterns between the host and virus evolution, although none were observed, the data lead to a better understanding of the evolution, dynamics and geographical spread of arctic rabies in Greenland.

PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Spatio-temporal Analysis of the Genetic Diversity of Arctic Rabies Viruses and Their Reservoir Hosts in Greenland

Dennis Hanke[1], Conrad M. Freuling[2], Susanne Fischer[3], Karsten Hueffer[4], Kris Hundertmark[5], Susan Nadin-Davis[6], Denise Marston[7], Anthony R. Fooks[7,8], Anette Bøtner[9], Thomas C. Mettenleiter[2], Martin Beer[1], Thomas B. Rasmussen[9], Thomas F. Müller[2]*, Dirk Höper[1]

1 Friedrich-Loeffler-Institut (FLI), Institute of Diagnostic Virology, Greifswald-Insel Riems, Germany, **2** FLI, Institute of Molecular Virology and Cell Biology, Greifswald-Insel Riems, Germany, **3** FLI, Institute of Epidemiology, Greifswald-Insel Riems, Germany, **4** Department of Veterinary Medicine, University of Alaska, Fairbanks, Alaska, United States of America, **5** Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska, United States of America, **6** Animal Health Microbiology Research, Canadian Food Inspection Agency (CFIA), Centre of Expertise for Rabies, Ottawa Laboratory, Ottawa, Ontario, Canada, **7** Animal and Plant Health Agency (APHA), Wildlife Zoonoses and Vector-borne Diseases Research Group, Addlestone, Surrey, United Kingdom, **8** University of Liverpool, Department of Clinical Infection, Microbiology and Immunology, Liverpool, United Kingdom, **9** DTU National Veterinary Institute, Technical University of Denmark, Lindholm, Kalvehave, Denmark

* thomas.mueller@fli.bund.de

Citation: Hanke D, Freuling CM, Fischer S, Hueffer K, Hundertmark K, Nadin-Davis S, et al. (2016) Spatio-temporal Analysis of the Genetic Diversity of Arctic Rabies Viruses and Their Reservoir Hosts in Greenland. PLoS Negl Trop Dis 10(7): e0004779. doi:10.1371/journal.pntd.0004779

Editor: Sergio Recuenco, Centers for Disease Control and Prevention, UNITED STATES

Received: January 19, 2016

Accepted: May 24, 2016

Published: July 26, 2016

Data Availability Statement: All sequence files are available from the INSDC database (accession number(s) AY352480; DQ010123; DQ010129; KU198460—KU198479; LM645015—LM645056; KX036361—KX036367; LT598538—LT598543).

Funding: This study was supported by BMBF grant 01KI1016A "Lyssaviruses - a potential re-emerging public health threat", by an intramural collaborative research grant of the Friedrich-Loeffler-Institut, Riems, Germany, and financing of the Research and Policy for Infectious Disease Dynamics (RAPIDD) program of the Science & Technology Directorate,

Abstract

There has been limited knowledge on spatio-temporal epidemiology of zoonotic arctic fox rabies among countries bordering the Arctic, in particular Greenland. Previous molecular epidemiological studies have suggested the occurrence of one particular arctic rabies virus (RABV) lineage (arctic-3), but have been limited by a low number of available samples preventing in-depth high resolution phylogenetic analysis of RABVs at that time. However, an improved knowledge of the evolution, at a molecular level, of the circulating RABVs and a better understanding of the historical perspective of the disease in Greenland is necessary for better direct control measures on the island. These issues have been addressed by investigating the spatio-temporal genetic diversity of arctic RABVs and their reservoir host, the arctic fox, in Greenland using both full and partial genome sequences. Using a unique set of 79 arctic RABV full genome sequences from Greenland, Canada, USA (Alaska) and Russia obtained between 1977 and 2014, a description of the historic context in relation to the genetic diversity of currently circulating RABV in Greenland and neighboring Canadian Northern territories has been provided. The phylogenetic analysis confirmed delineation into four major arctic RABV lineages (arctic 1–4) with viruses from Greenland exclusively grouping into the circumpolar arctic-3 lineage. High resolution analysis enabled distinction of seven geographically distinct subclades (3.I – 3.VII) with two subclades containing viruses from both Greenland and Canada. By combining analysis of full length RABV genome sequences and host derived sequences encoding mitochondrial proteins obtained simultaneously from brain tissues of 49 arctic foxes, the interaction of viruses and their hosts was explored in detail. Such an approach can serve as a blueprint for analysis of

U.S. Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. We thank the Department of Natural Resources of Newfoundland and Labrador, Canada, for funds that helped support this work. The work undertaken at APHA was funded by the UK Department for Environment, Food and Rural Affairs grant SE0427 "Lyssavirus Heterogeneity and Evolution: Evaluating risks of adaption and cross-species transmission". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

infectious disease dynamics and virus-host interdependencies. The results showed a fine-scale spatial population structure in Greenland arctic foxes based on mitochondrial sequences, but provided no evidence for independent isolated evolutionary development of RABV in different arctic fox lineages. These data are invaluable to support future initiatives for arctic fox rabies control and elimination in Greenland.

Author Summary

Next to dog-mediated rabies, wildlife rabies continues to pose a public health problem, particularly in the northern hemisphere. Control of this zoonosis at the animal source has been proven the most efficient route to reduction of human rabies burden. Successful elimination of red fox-mediated rabies in Western Europe and parts of North America has demonstrated the viability of wildlife rabies control strategies. In some regions, the epidemiology of wildlife rabies is well understood; this is not the case for arctic rabies, particularly in Greenland. Previous molecular epidemiological studies demonstrated the occurrence of one particular arctic rabies virus (RABV) lineage (arctic-3) but were limited by low sample numbers and limited sequence length so as to preclude generation of high resolution phylogenetic analysis. Here, a unique set comprised of 79 complete genome sequences of RABVs from Greenland, Canada, USA (Alaska) and Russia collected over the past four decades was analysed. The use of next generation sequencing (NGS) allowed simultaneous determination of host derived sequences encoding mitochondrial proteins from the same brain tissue of 49 arctic foxes. These sequence data combined with geographical and temporal information permit the study of the genetic diversity and evolution of circulating RABVs in Greenland against the background of reservoir host genetics. The results reveal the existence of a single arctic RABV lineage (arctic-3) in Greenland, which has evolved into multiple distinct variants. These analyses provide an improved knowledge of the evolution of the circulating viruses at the molecular level and a better understanding of the historical perspective of the disease in Greenland compared to other parts of the Arctic. This knowledge will support policy on rabies control in mammalian wildlife reservoirs.

Introduction

Rabies, an ancient disease known for millennia, is caused by lyssaviruses of the Rhabdoviridae family [1]. The prototypical rabies virus (RABV) has a global distribution and the domestic dog is the host reservoir responsible for the vast majority of the estimated 60,000 human rabies cases annually [2]. Other RABV host reservoirs in terrestrial wildlife are primarily meso-carnivores. In Arctic regions, RABV is believed to be maintained by the arctic fox (*Vulpes lagopus*) [3], which has a circumpolar distribution and has uniquely adapted to the extreme climatic and ecologic conditions of this northern environment [4]. The distribution and group size of arctic foxes are strongly influenced by the distribution and density of prey [5]. Notably, arctic foxes have considerably variable home ranges (5–120 km^2) than any other meso-carnivore RABV reservoir host [6–8] and can roam over large areas and migrate over extremely long distances [4,8]. From an epidemiological point of view, this may be an important factor for the spread of RABV in northern Polar regions where rabies-like diseases have been described for about 150 years [9], specifically among sledge dogs in Greenland as early as 1859. However,

confirmation of the existence of rabies in Greenland was only provided 100 years later, when Jenkins and Wamberg [10] demonstrated the presence of RABV in dogs and arctic foxes. The disease is considered endemic among the arctic fox population of Greenland [11,12]. A recent epidemiological study of arctic fox rabies in Greenland between 1969 and 2011 revealed that the disease flared up every 5–10 years on average, whereby most rabid foxes were reported from southern Greenland [13].

Historically, some properties of arctic RABVs were regarded as "atypical" [14]. However, early genetic virus characterizations based on the nucleoprotein (N) gene clearly identified it as RABV but as a separate virus lineage designated as "arctic" [15]. This lineage circulates throughout the circumpolar region including northern regions of North America, Europe, and Asia. More detailed phylogenetic analyses revealed that the arctic RABV variant can be further delineated into at least four distinct groups [16–18]. The arctic-1 lineage, recovered from southern Ontario, Canada, in the late 20th and early 21st centuries, represented the remnants of an epidemic that spread from northern Canada in the mid-1900s; reports of this strain are now rare due to the rabies control program carried out by provincial authorities. The arctic-4 lineage has only ever been recovered from regions of Alaska and viruses of the arctic-2 lineage appear to be restricted to Siberia, the Russian Far East, and Alaska. In contrast, lineage arctic-3 has a circumpolar distribution [17]. Arctic rabies was also detected on the European Svalbard Islands with the prevailing RABV lineage having a closer phylogenetic relationship to those occurring in the polar regions of Russia [19]. Viruses closely related to those of the arctic clade have been designated as 'arctic-like' or 'arctic-related' but have a broad distribution in central, east, and southeast Asia [18,20].

Published phylogenetic analysis demonstrated that RABV isolates from Greenland belong to the arctic-3 lineage [16,17]. These studies did not allow for a more comprehensive evolutionary analysis as the datasets were restricted in terms of the number of samples, time, geographic origin, and sequence length. In this present study, a comprehensive panel of 58 RABVs from Greenland between 1990 and 2014 was analyzed. Additionally, 24 arctic RABVs from Canada/Alaska and Russia were also sequenced and added to this dataset to provide some context to the situation in Greenland. The principal objectives were (i) to infer the viral phylogenetic relationships in space and time based on complete genome sequences and (ii) to gain more insights into the contribution of the host population to the spatial spread of individual RABVs in Greenland. In particular, evidence based on sequence analysis for any links between phylogenetic clusters of arctic RABV and the arctic fox population in Greenland was sought.

Materials and Methods

Origin of viruses

All samples in this study were either taken from officially implemented passive rabies surveillance programs or already existing collections at the (i) DTU National Veterinary Institute, Technical University of Denmark, Denmark, (ii) Canadian Food Inspection Agency (CFIA), Canada, (iii) Friedrich-Loeffler-Institut (FLI), Germany, and (iv) Animal and Plant Health Agency (APHA), UK. Samples (Table 1) comprised original clinical brain samples submitted for passive surveillance from arctic foxes (*Vulpes lagopus*), red foxes (*Vulpes vulpes*), dogs (*Canis lupus familiaris*), cats (*Felis silvestris catus*), sheep (*Ovis aries*), and a long-tailed ground squirrel (*Citellus undulates*) collected between 1977 and 2014 in Greenland, Northern Canada, Alaska and Russia that tested positive in the direct fluorescent antibody test (FAT, [21]). Because this is a multi-center study, RNA extraction, library preparation and sequencing were done using slightly different protocols. The 3 protocols are briefly outlined in the following paragraphs and the respective protocol is denoted for each sample in Table 1. In no case was

Table 1. Details of RABV samples investigated in this study.

Sample name for this study[1]	Year	Month[1]	Host	Country	Location[1]	Area code[1]	Gauss-Krueger coordinates[1]		Arctic fox mitochondrial sequences available [1,2]	RABV INSDC accession[2]	Sequencing Protocol
							N	W			
Gra21.05-GRL-1-AF-2005	2005	12	Arctic fox	Greenland	Siorapaluk/Qaanaq	1	77.47	70.46	yes	LM645017	1
Gra07.06-GRL-1-AF-2006	2006	3	Arctic fox	Greenland	Savissivik	1	76.01	65.08	yes	LM645015	1
Gra14.06-GRL-1-AF-2006	2006	4	Arctic fox	Greenland	Pituffik	1	76.32	68.45	yes	LM645016	1
Gra03.13-GRL-1-AF-2013	2013	3	Arctic fox	Greenland	Qaasuitsup/Thule	1	76.53	68.70	yes	LT598541	1
Gra23.06-GRL-2-AF-2006	2006	11	Arctic fox	Greenland	Torsukattak/Qeqertaq	2	70.03	51.27	yes	LM645018	1
Gra03.10-GRL-2-AF-2010	2010	6	Arctic fox	Greenland	Nuussuaq/Upernavik	2	74.07	57.04	yes	LM645019	1
Gra01.13-GRL-2-AF-2013	2013	3	Arctic fox	Greenland	Upernavik/Kullorsuaq	2	74.58	57.22	yes	LT598539	1
Gra02.13-GRL-2-AF-2013	2013	3	Arctic fox	Greenland	Upernavik/Aappilattoq	2	60.13	44.30	yes	LT598538	1
Gra03.06-GRL-3-AF-2006	2006	2	Arctic fox	Greenland	Qasigiannguit	3	68.49	51.05	yes	LM645025	1
Gra04.06-GRL-3-AF-2006	2006	2	Arctic fox	Greenland	Niaqornaarsuk	3	68.18	53.27	yes	LM645021	1
Gra05.06-GRL-3-AF-2006	2006	2	Arctic fox	Greenland	Aasiaat	3	68.49	51.05	yes	LM645026	1
Gra08.06-GRL-3-AF-2006	2006	3	Arctic fox	Greenland	Aasiaat	3	68.42	52.52	yes	LM645023	1
Gra09.06-GRL-3-AF-2006	2006	3	Arctic fox	Greenland	Iginniarfik	3	68.85	53.10	yes	LM645028	1
Gra10.06-GRL-3-AF-2006	2006	3	Arctic fox	Greenland	Aasiaat	3	68.42	52.52	yes	LM645024	1
Gra24.06-GRL-3-D-2006	2006	11	Dog	Greenland	Kangaatsiaq	3	68.18	53.27	ND	LM645022	1
Gra25.06-GRL-3-AF-2006	2006	11	Arctic fox	Greenland	Qasigiannguit	3	68.49	51.05	yes	LM645027	1
Gra02.07-GRL-3-AF-2007	2007	1	Arctic fox	Greenland	Kangaatsiaq	3	68.18	53.27	yes	LM645020	1
Gra01.14-GRL-3-AF-2014	2014	10	Arctic fox	Greenland	Kangaatsiaq/ Ikerasaarsuk	3	68.14	53.44	yes	LT598543	1
Gra16.06-GRL-4-AF-2006	2006	5	Arctic fox	Greenland	Kangerlussuaq	4	67.04	50.41	yes	LM645031	1
Gra09.07-GRL-4-AF-2007	2007	4	Arctic fox	Greenland	Sisimiut	4	66.55	53.40	yes	LM645029	1
Gra10.07-GRL-4-AF-2007	2007	5	Arctic fox	Greenland	Sisimiut	4	66.55	53.40	yes	LM645030	1
Gra02.14-GRL-4-AF-2014	2014	9	Arctic fox	Greenland	Kangerlussuaq	4	67.01	50.70	yes	LT598540	1
Gra03.14-GRL-4-AF-2014	2014	10	Arctic fox	Greenland	Kangerlussuaq	4	67.01	50.70	yes	LT598542	1
Gra02.10-GRL-5-AF-2010	2010	5	Arctic fox	Greenland	Qoqqut/Nuuk	5	64.16	50.54	yes	LM645033	1
Gra05.10-GRL-5-AF-2010	2010	6	Arctic fox	Greenland	Kobbefjord/Nuuk	5	64.10	51.44	yes	LM645032	1
Gra10.08-GRL-6-AF-2008	2008	11	Arctic fox	Greenland	Paamiut	6	62.00	49.43	yes	LM645037	1
Gra07.09-GRL-6-AF-2009	2009	5	Arctic fox	Greenland	Godthåbsfjorden/Nuuk	6	60.55	48.15	yes	LM645034	1
Gra13.09-GRL-6-AF-2009	2009	12	Arctic fox	Greenland	Paamiut	6	62.00	49.43	yes	LM645038	1
Gra06.10-GRL-6-AF-2010	2010	9	Arctic fox	Greenland	Paamiut	6	62.00	49.43	yes	LM645036	1
Gra07.10-GRL-6-AF-2010	2010	9	Arctic fox	Greenland	Kangilinnguit/Ivittuut	6	61.14	48.60	yes	LM645035	1
Gra18.05-GRL-7-AF-2005	2005	11	Arctic fox	Greenland	Qaqortoq	7	60.43	46.03	yes	LM645048	1
Gra01.06-GRL-7-AF-2006	2006	1	Arctic fox	Greenland	Narsaq	7	60.54	46.03	yes	LM645051	1
Gra02.06-GRL-7-AF-2006	2006	2	Arctic fox	Greenland	Narsaq	7	60.54	46.03	yes	LM645052	1
Gra06.06-GRL-7-C-2006	2006	3	Cat	Greenland	Qaqortoq	7	60.43	46.03	ND	LM645046	1
Gra20.06-GRL-7-AF-2006	2006	9	Arctic fox	Greenland	Qaqortoq	7	60.43	46.03	no	NA	1
Gra21.06-GRL-7-AF-2006	2006	9	Arctic fox	Greenland	Narsarsuaq	7	61.08	45.26	yes	LM645055	1
Gra26.06-GRL-7-AF-2006	2006	11	Arctic fox	Greenland	Qaqortoq	7	60.43	46.03	no	NA	1
Gra04.07-GRL-7-AF-2007	2007	1	Arctic fox	Greenland	Igaliko	7	60.59	45.25	yes	LM645053	1
Gra05.07-GRL-7-AF-2007	2007	2	Arctic fox	Greenland	Qaqortoq	7	60.43	46.03	yes	LM645045	1
Gra15.07-GRL-7-S-2007	2007	10	Sheep	Greenland	Narsarsuaq	7	61.08	45.26	ND	LM645054	1
Gra3a.07-GRL-7-AF-2007	2007	1	Arctic fox	Greenland	Qaqortoq	7	60.43	46.03	yes	LM645049	1
Gra3b.07-GRL-7-AF-2007	2007	1	Arctic fox	Greenland	Qaqortoq	7	60.43	46.03	yes	LM645050	1

(*Continued*)

Table 1. (*Continued*)

Sample name for this study[1]	Year	Month[1]	Host	Country	Location[1]	Area code[1]	Gauss-Krueger coordinates[1]		Arctic fox mitochondrial sequences available [1,2]	RABV INSDC accession[2]	Sequencing Protocol
							N	W			
Gra02.08-GRL-7-AF-2008	2008	4	Arctic fox	Greenland	Ammassalik	7	65.43	37.35	yes	LM645056	1
Gra03.08-GRL-7-AF-2008	2008	4	Arctic fox	Greenland	Nanortalik	7	60.08	45.14	yes	LM645041	1
Gra13.08-GRL-7-AF-2008	2008	12	Arctic fox	Greenland	Qaqortoq	7	60.43	46.03	yes	LM645047	1
Gra01.09-GRL-7-S-2009	2009	1	Sheep	Greenland	Nanortalik	7	60.08	45.14	ND	LM645039	1
Gra02.09-GRL-7-S-2009	2009	1	Sheep	Greenland	Nanortalik	7	60.08	45.14	ND	LM645040	1
Gra03.09-GRL-7-AF-2009	2009	3	Arctic fox	Greenland	Nanortalik	7	60.08	45.14	yes	LM645042	1
Gra04.09-GRL-7-AF-2009	2009	4	Arctic fox	Greenland	Qaqortoq	7	60.43	46.03	yes	NA	1
Gra01.10-GRL-7-AF-2010	2010	1	Arctic fox	Greenland	Qaqortoq	7	60.43	46.03	yes	LM645043	1
Gra01.11-GRL-7-AF-2011	2011	3	Sheep	Greenland	Qaqortoq	7	60.43	46.03	ND	LM645044	1
13232-CAN-NT-AF-1977	1977	-	Arctic fox	Canada	Northwest Territories	NT	NA	NA	yes	LT598537	1
RV1391-GRL-1-F-1990	1990	3	Arctic fox	Greenland	Thule	1	NA	NA	no	KX036361	3
RV1415-GRL-4-F-2001	2001	12	Arctic fox	Greenland	Kangerlussuaq	4	NA	NA	no	KX036362	3
RV1416-GRL-5-AF-2002	2002	1	Arctic fox	Greenland	Nuuk	5	NA	NA	yes	KX036363	3
RV1417-GRL-3-AF-2002	2002	1	Arctic fox	Greenland	Kangaasiaq	3	NA	NA	yes	KX036364	3
RV1418-GRL-4-AF-2002	2002	1	Arctic fox	Greenland	Kangerlussuaq	4	NA	NA	yes	KX036365	3
RV1419-GRL-3-AF-2002	2002	10	Arctic fox	Greenland	Ilulissat	3	NA	NA	yes	KX036366	3
RV1420-GRL-4-AF-2002	2002	10	Arctic fox	Greenland	Kangerlussuaq	4	NA	NA	yes	KX036367	3
07V1483RFX-USA-AK-RF-2007	2007	NA	Red fox	USA	NA	AK	NA	NA	ND	KU198471	2
89V809AFX-USA-AK-AF-1989	1989	NA	Arctic fox	USA	Barrow	AK	NA	NA	no	KU198460	2
90V814AFX-USA-AK-AF-1990	1990	NA	Arctic fox	USA	Kwigillingok	AK	NA	NA	no	KU198462	2
90V820RFX-USA-AK-RF-1990	1990	NA	Red fox	USA	Cold Bay	AK	NA	NA	ND	KU198463	2
12N0215RFX-CAN-NL-RF-2012	2012	NA	Red fox	Canada	St. John's Co	NL	NA	NA	ND	KU198473	2
12N0280RFX-CAN-NL-RF-2012	2012	NA	Red fox	Canada	St. John's Co	NL	NA	NA	ND	KU198474	2
14N0601RFX-CAN-NL-RF-2014	2014	NA	Red fox	Canada	Nain	NL	NA	NA	ND	KU198479	2
12L1020AFX-CAN-NT-AF-2012	2012	NA	Arctic fox	Canada	Holman	NT	NA	NA	no	KU198472	2
13L0040AFX-CAN-NT-AF-2013	2013	NA	Arctic fox	Canada	Holman	NT	NA	NA	no	KU198475	2
13L0094AFX-CAN-NT-AF-2013	2013	NA	Arctic fox	Canada	Paulatuk	NT	NA	NA	no	KU198476	2
90L2968AFX-CAN-NT-AF-1990	1990	NA	Arctic fox	Canada	Grise Fiord	NT	NA	NA	no	KU198461	2
91L0085AFX-CAN-NT-AF-1991	1991	NA	Arctic fox	Canada	Cambridge Bay	NT	NA	NA	no	KU198464	2
93N1395AFX-CAN-NT-AF-1993	1993	NA	Arctic fox	Canada	Resolute Bay	NT	NA	NA	no	KU198466	2
97L1793AFX-CAN-NT-AF-1997	1997	NA	Arctic fox	Canada	Arctic Bay	NT	NA	NA	no	KU198467	2
13N0473AFX-CAN-NU-AF-2013	2013	NA	Arctic fox	Canada	Resolute Bay	NU	NA	NA	no	KU198477	2
13N0643AFX-CAN-NU-AF-2013	2013	NA	Arctic fox	Canada	Grise Fiord	NU	NA	NA	no	KU198478	2
00N3340RFX-CAN-ON-RF-2000	2000	NA	Red fox	Canada	Walkerton	ON	NA	NA	ND	KU198468	2
01N10254RFX-CAN-ON-RF-2001	2001	NA	Red fox	Canada	Shelburne	ON	NA	NA	ND	KU198469	2
92N7894RFX-CAN-ON-RF-1992	1992	NA	Red fox	Canada	Swastika	ON	NA	NA	ND	KU198465	2
02N2980RFX-CAN-QC-RF-2002	2002	NA	Red fox	Canada	Salluit	QC	NA	NA	ND	KU198470	2
RV53-USA-F-1988	1988	NA	Fox	USA	NA	NA	NA	NA	ND	DQ010123	3
RV250-RUS-Squirrel-1983	1983	NA	*Citellus undulatus*	Russia	Tuvia	NA	NA	NA	ND	AY352480	3
RV1336-RUS-AF-1996	1996	NA	Arctic fox	Russia	Yakutia	NA	NA	NA	no	DQ010129	3

[1] NA, not available

[2] ND, not determined

doi:10.1371/journal.pntd.0004779.t001

5'- or 3'-RACE performed to confirm the genome termini. The accessions for all RABV genome sequences generated in this study are given in Table 1.

Sequencing protocol 1

RNA extraction. RNA was extracted from approximately 20 mg of brain tissue. To this end, the material was frozen in liquid nitrogen, homogenized using the Mikro-Dismembrator S (Sartorius, Göttingen, Germany), and the homogenate was suspended in 2 ml Buffer AL (Qiagen, Hilden, Germany) pre-heated to 56°C. The resulting suspension was mixed with 3 volumes of TRIzol LS Reagent (Life Technologies, Carlsbad, California, USA) and 0.2 volumes chloroform (Carl Roth GmbH + Co. KG, Karlsruhe, Germany) and after addition of 1 volume of 100% ethanol to the aqueous phase RNA was extracted using the RNeasy Mini Kit (Qiagen) as per the manufacturer's instructions including an on-column DNase I digestion. If necessary, extracted RNA was concentrated with Agencourt RNA Clean XP magnetic beads (Beckman Coulter, Fullerton, USA). RNA quantity was determined using the Nanodrop ND1000 UV spectrophotometer (Peqlab, Erlangen, Germany).

cDNA-synthesis, library preparation and sequencing. cDNA was generated from total RNA with the cDNA synthesis system kit (Roche, Mannheim, Germany) and random hexamer primers (Roche) according to the Genome Sequencer RNA rapid library preparation manual (Roche). The resulting cDNA was fragmented to a target size of 300 bp using a Covaris M220 instrument (Covaris, Brighton, United Kingdom) and subsequently transformed to barcoded sequencing libraries using Illumina compatible adapters (Bioo Scientific Corp., Austin, USA) on a SPRI-TE library system (Beckman Coulter) with SPRIworks Fragment Library Cartridge II (Beckman Coulter) without size selection. After manual size selection with Agencourt AMPure XP magnetic beads (Beckman Coulter) for a target peak size of 350 bp, library quality was assessed using a Bioanalyzer 2100 instrument (Agilent Technologies, Böblingen, Germany) with a High Sensitivity DNA kit (Agilent Technologies). Finally, the libraries were quantified with the KAPA Library Quantification Kit for Illumina (Kapa Biosystems, Cape Town, South Africa) on a CFX96 Real-Time System (Bio-Rad Laboratories, Hercules, USA) and sequenced using the Illumina MiSeq instrument with MiSeq reagent kit v2 (Illumina, San Diego, USA) in 2 x 250 bp mode.

Sequence assembly. For sequence assembly, reads representing the respective sequences were selected by mapping the complete data set to reference sequences using the Genome Sequencer software suite (v2.6; Roche). Subsequently, the sorted reads were used for a de novo assembly using newbler (Roche) and the complete raw data sets were mapped along the resulting consensus sequences in order to identify potential sequencing errors. Thereafter, the sequences were visually inspected using Geneious (v6.1.7; Biomatters, Auckland, New Zealand).

Sequencing protocol 2

RNA extraction. RNA was extracted from approximately 100 mg of brain tissue using TRIzol reagent as per the supplier's instructions (Life Technologies).

cDNA-synthesis, library preparation and sequencing. A protocol for efficient RT-PCR amplification of the entire viral genome as a small number of overlapping amplicons was employed (Primers and protocols are available upon request). Amplicons derived from a single sample were pooled in equimolar concentrations and then used to generate a sequencing library using a Nextera XT kit as per the manufacturer's directions (Illumina). Libraries were sequenced on an Illumina Miseq instrument using a MiSeq reagent kit v2 in 2 x 250 bp mode.

Sequence assembly. Sequence reads were assembled with the Lasergene software (v11; DNASTAR, Madison, Wisconsin USA). The paired end fastq files for each sample were assembled using a template-based method. Base coverage >200 was obtained throughout the genome with the exception of a small stretch of bases at the genomic termini for which coverage was very limited, regions which could not be unambiguously sequenced in any event due to their use as targets for amplification primers. As these regions tend to be highly conserved in all rabies viruses this was not considered to be a significant limitation. Complete assembled genomes were exported in fasta format for subsequent alignment and phylogenetic analysis as detailed below.

Sequencing protocol 3

RNA extraction, cDNA-synthesis, library preparation and sequencing. RNA from brain material stored at -80°C was prepared for Next Generation Sequencing on the MiSeq platform (Illumina). Briefly, TRIzol (Life Technologies) extracted viral RNA was depleted of host genomic DNA and rRNA as described previously [22]. Double stranded (ds) cDNA was synthesized from 50 ng RNA, using a random cDNA synthesis system (Roche), according to the manufacturers' instructions. The ds cDNA was purified using Ampure XP magnetic beads (Beckman Coulter) and 1ng used for the Nextera XT DNA sample preparation kit (Illumina). A sequencing library was prepared according to the manufacturers' instructions and sequenced on an Illumina MiSeq with 2 x 150 bp paired end reads following standard Illumina protocols.

Sequence assembly. The total reads were mapped to an appropriate reference sequence in Burrows-Wheeler Alignment Tool (BWA, http://bio-bwa.sourceforge.net) using a script to generate intermediate consensus sequences in which any indels relative to the original reference sequence were appropriately called, then visualized in Tablet [23] as described previously [24].

Sequence determination of arctic fox mitochondrial genes

For the determination of the genetic diversity of the reservoir host, mitochondrial (mtDNA) reference genes of arctic foxes, i.e. ATP6, ATP8, COX1, COX2, COX3, CYTB, ND1, ND2, ND3, ND4, ND4L and NDS [25] were selected from the International Nucleotide Sequence Database (INSDC) databases. Additionally, the mitochondrial D-Loop sequence as suggested before [26] was also used for mapping. Briefly, raw reads from sequencing were mapped along the reference genes and all reads identified as fox mitochondrial sequences were assembled de novo. Subsequently, the resulting consensus sequence was used as reference to map all reads of the dataset in order to identify potential sequencing errors. The resulting sequences were inspected visually in Geneious (v6.1.7; Biomatters).

Evolutionary analyses

All obtained complete RABV genome sequences were aligned using ClustalW [27] (http://www.clustal.org) as implemented in Geneious (Biomatters), the sequences were trimmed to equal length and labelled with the collection year. Subsequently, phylogenetic analyses were performed using the Bayesian Markov Chain Monte Carlo (MCMC) simulation in the BEAST (Bayesian Evolutionary Analysis Sampling Trees) package v1.8.2 [28]. Selection of the evolutionary model using IQ-Tree (v1.1.0, [29] proposed use of the General Time Reversible model with rate heterogeneity (GTR + G). This model was used for MCMC simulation together with a relaxed molecular clock model and Bayesian Skyline population for 100,000,000 iterations, sampling every 10,000 states to give effective sample sizes. Maximum clade credibility trees (MCC) were annotated using TreeAnnotator (v1.8.2), 10% of the trees were removed as burn-

in. The resulting final trees were visualized using FigTree (v1.4.2; http://tree.bio.ed.ac.uk/software/figtree/).

For phylogenetic analysis of the RABV Nucleoprotein (N) gene sequences, the dataset was extended with additional sequences from the INSDC databases (Table 2). All calculations were performed as described above for the full genome sequences except that for MCMC simulation the transitional model with rate heterogeneity (TIM + G) model was used.

From 55 of 58 Greenland samples, the substitutions per site and year were determined for the whole genome and all five protein coding nucleotide-sequences (N-, P-, M-, G- and L-

Table 2. Details of additional sequences used for phylogenetic analyses of complete N gene.

Isolate name project	Location	Year	Host	INSDC accession no.	References
AY352458.1-RUS.Chabarovsk-RD-1980	Russia/ Chabarovsk	1980	Raccon dog	AY352458.1	[60]
AY352459.1-RUS.Chita-SF-1977	Southern Siberia	1977	Corsac fox	AY352459.1	[60]
AY352462.1-RUS.Norilsk-H-1998	Russia/Norilsk	1998	Human (ex wolf)	AY352462.1	[60]
AY352474.1-RUS.Pskov-RF-1990	Russia/Pskov	1990	Red fox	AY352474.1	[60]
AY352487.1-RUS-SA-1986	Russia/Yakutia	1986	Arctic fox	AY352487.1	[60]
AY352491.1-KAZ-RF-1988	Kazakhstan	1988	Red fox	AY352491.1	[60]
AY352498.1-USA-AK-D-1988	Alaska	1988	Dog	AY352498.1	[60]
AY352500.1-USA-AK-RF-1988	Alaska	1988	Red fox	AY352500.1	[60]
AY730596.1-KOR-D-2004	South Korea	2004	Dog	AY730596.1	[61]
AY956319.1-IND-H-2004	India	2004	Human	AY956319.1	-
EF611828.1-RUS-SA-1987	Russia/Yakutia	1987	Arctic fox	EF611828.1	[17]
EF611830.1-RUS-SA-1987	Russia/Yakutia	1987	Arctic fox	EF611830.1	[17]
EF611843.1-USA-AK-RF-2007	Alaska	2007	Red fox	EF611843.1	[17]
EF611848.1-USA-AK-RF-2008	Alaska	2008	Red fox	EF611848.1	[17]
EF611851.1-USA-AK-AF-2006	Alaska	2006	Arctic fox	EF611851.1	[17]
EF611853.1-USA-AK-D-2006	Alaska	2006	Dog	EF611853.1	[17]
EF611855.1-USA-AK-D-2006	Alaska	2006	Dog	EF611855.1	[17]
EF611868.1-SouthernSiberia-D-1980	Southern Siberia	1980	Dog	EF611868.1	[17]
JX944565.1-NPL-H-2003	Nepal	2003	Human	JX944565.1	[20]
JX944601.1-NPL-M-2010	Nepal	2010	Mongoose	JX944601.1	[20]
JX987745.1-GRL-D-1980	Greenland	1980	Dog	JX987745.1	[20]
L20675.2-CAN-ON-AF-1991	Canada/ Ontario	1991	Arctic fox	L20675.2	[62]
U03769.1-CAN.Arctic-D-1993	Canada/Arctic	1993	Dog	U03769.1	[63]
U03770.1-CAN.HudsonBay-D-1992	Canada/ Hudson Bay	1992	Dog	U03770.1	[63]
U11735.1-CAN-ON-RF-1993	Canada/ Ontario	1993	Red fox	U11735.1	[63]
U22474.1-France-RF-1991	France	1991	Red fox	U22474.1	[15]
U22477.1-MEX-D-1991	Mexico	1991	Dog	U22477.1	[15]
U22654.1-GRL-AF-1981	Greenland	1981	Arctic fox	U22654.1	[15]
U42705.1-F.R.Yugoslavia-cattle-1981	F.R.Yugoslavia	1981	Cattle	U42705.1	-
U42432.1-EST-RD-1991	Estonia	1991	Raccoon dog	U43432.1	-

doi:10.1371/journal.pntd.0004779.t002

Gene), respectively. Sequences were aligned and best fitting evolutionary models selected as described above. The best model for the N and P genes was Kimura 3-parameter (K81, [30], for the G and L genes Kimura 3-parameter with unequal frequencies (K81uf), for the M gene Kimura 2-parameter (K2P, [31], and for the whole genome Kimura 3-parameter with unequal frequencies and proportion of invariable sites (K81uf + I). Beast analysis was performed as described above except for the N gene for which 2,000,000,000 iterations and samples every 200,000 states were needed.

To infer the evolutionary and phylogenetic relationships of foxes, the sequences were aligned using ClustalW and subsequently phylogenetic trees were calculated using the maximum likelihood (ML) method as implemented in MEGA (v5.2; [32]). For these calculations, the best fitting evolutionary model, Tamura and Nei 1993 with rate heterogeneity (TN93 + G), was selected using MEGA's model test function and 1000 bootstrap replicates were calculated. The resulting phylogenetic trees were visualized in MEGA.

Geographical mapping

Approximate sampling locations in Greenland, Canada and the USA (Table 1, Fig 1) were visualized using ArcGis 10.0 (ESRI) at the highest spatial resolution available.

Results

To enable in-depth phylogenetic analysis, viral sequences encoding all genes were generated for all arctic RABV samples (S1 Table, Fig 1). For Greenland fox samples, complete coding sequences of 12 mitochondrial genes were determined from the available host sequence data in order to relate the viral phylogeny to fox populations and to detect potentially existing spatial fox distribution patterns.

Greenland RABV segregation into distinct arctic-3 subclades

To classify the sequenced viruses globally into pre-defined lineages, only complete N gene sequences (n = 109, Tables 1 and 2) were used, since to date no full-genome sequences are available for the Arctic rabies strains. All sequenced Greenland RABV (Table 1) clustered within the previously established circumpolar arctic-3 lineage (Fig 2). This lineage appears to be rather heterogeneous. Older arctic-3 viruses from Greenland (obtained in 1980/81) are clearly separated from recent Greenland RABVs.

For higher resolution of the heterogeneous arctic-3 lineage the analysis was repeated using full-genome sequences obtained for 79 RABV samples from Greenland, Canada, USA (Alaska) and Russia between 1977 and 2014 (Table 1). This high resolution analysis confirmed previous delineation of the arctic RABV strain into four major lineages (arctic 1–4, Figs 1 and 3), whereby RABV from Greenland exclusively grouped into arctic-3. Furthermore, it enabled the distinction of seven subclades and seven outliers (1977–2013) within arctic-3 (Fig 3), with the most recent common ancestor (MRCA) of the analyzed viruses (Fig 3) occurring approximately 82 years ago (95% HPD values, 73–92 years). The MRCA for the Greenland samples within this dataset dates back circa 35 years (95% HPD values, 32–37), whilst the divergence into subclades occurred between 15 years (subclade 3.VII; 95% HPD values, 14–17 years) and 6 years (subclade 3.IV; 95% HPD values, 5–8 years) ago.

Of these seven subclades, only arctic-3.IV and 3.V contained viruses from both Greenland and Canada (2006 to 2014). Greenland RABV from arctic-3.IV -3.VI (n = 10) originated from north-western and western Greenland (regions 1–4, Fig 1), collected between 2005 and 2014. Greenland specific subclades 3.II, 3.III, and 3.VII consisted of 24 RABV from the western coastline (regions 2–6) obtained between 2001 and 2014. In contrast, all but one RABV within

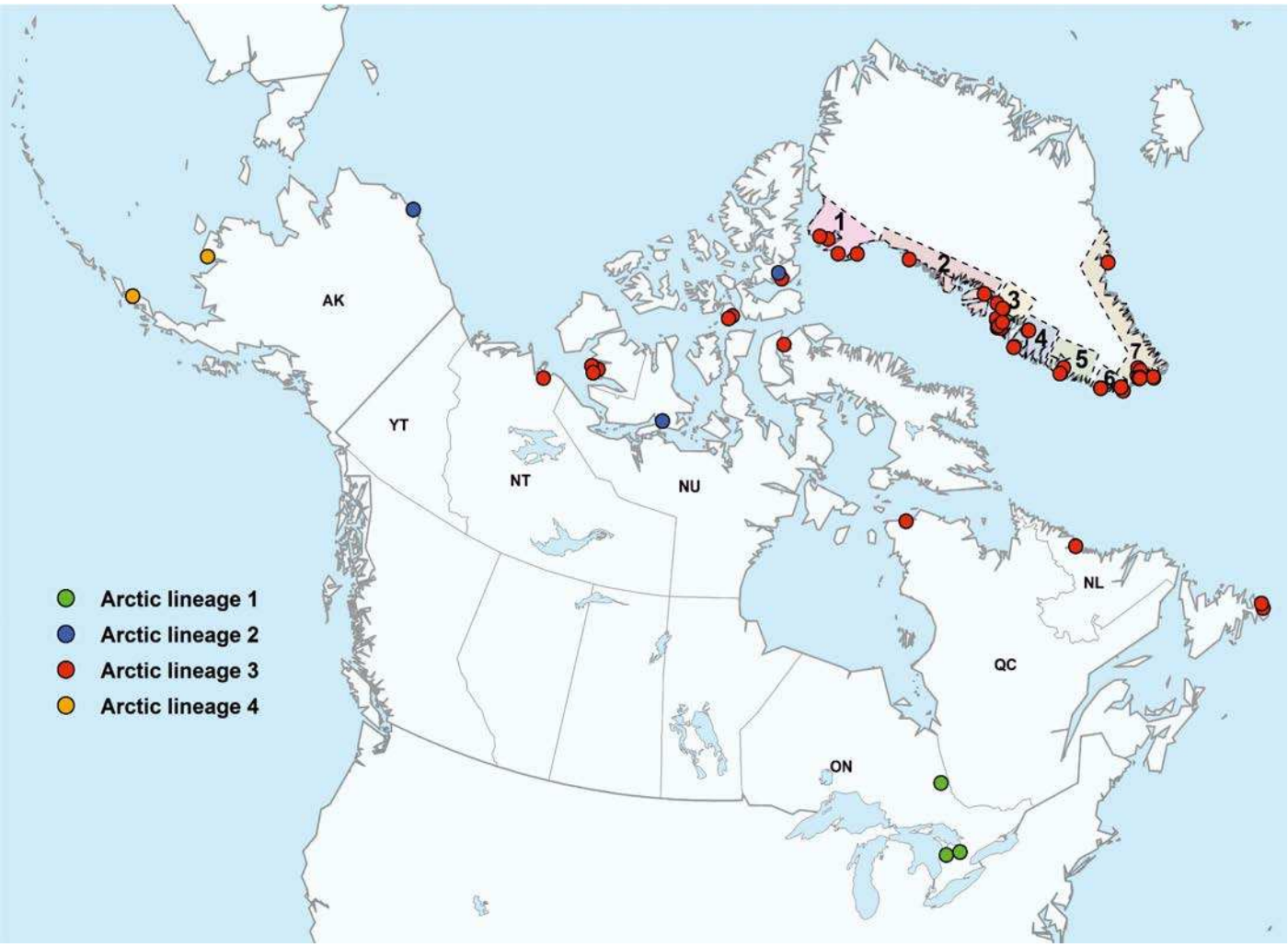


Fig 1. Map illustrating the origin of samples from Greenland, Canada and Alaska and their assignment to arctic lineages (Green = arctic lineage 1; Blue = 2; Red = 3; Yellow = 4). Numbers in Greenland represent regions, while for Canada and Alaska postal abbreviation of the province/territory/states is indicated.

doi:10.1371/journal.pntd.0004779.g001

subclade 3.I (2005–2011) originated from southern Greenland (region 7), with Gra10.08-GRL-6-AF-2008 being the only exception collected in region 6 (Figs 1 and 3).

Partial sequencing overestimates substitution rates

Both the individual gene sequences and the full-genome alignment were used to estimate the mean nucleotide substitution rate utilizing BEAST [28]. For the N gene sequences (Tables 1 and 2, Fig 2) an estimate of 2.5 E-4 substitutions per site per year (95% high posterior density (HPD), 1.9 E-4–3.1 E-4) was obtained. When only N gene sequences from Greenland (N = 55) were considered, the estimate was 3.1 E-4 substitutions per site per year, a value in the range of previous studies. Among the different genes, lowest and highest substitution rates were observed in the L gene and the P gene, respectively (Table 3). The substitution rate of 2.5 E-4 per site per year (95% HPD values, 2.1 E-4–3.0 E-4) for the full genome sequence is similar to the value observed for the L gene.

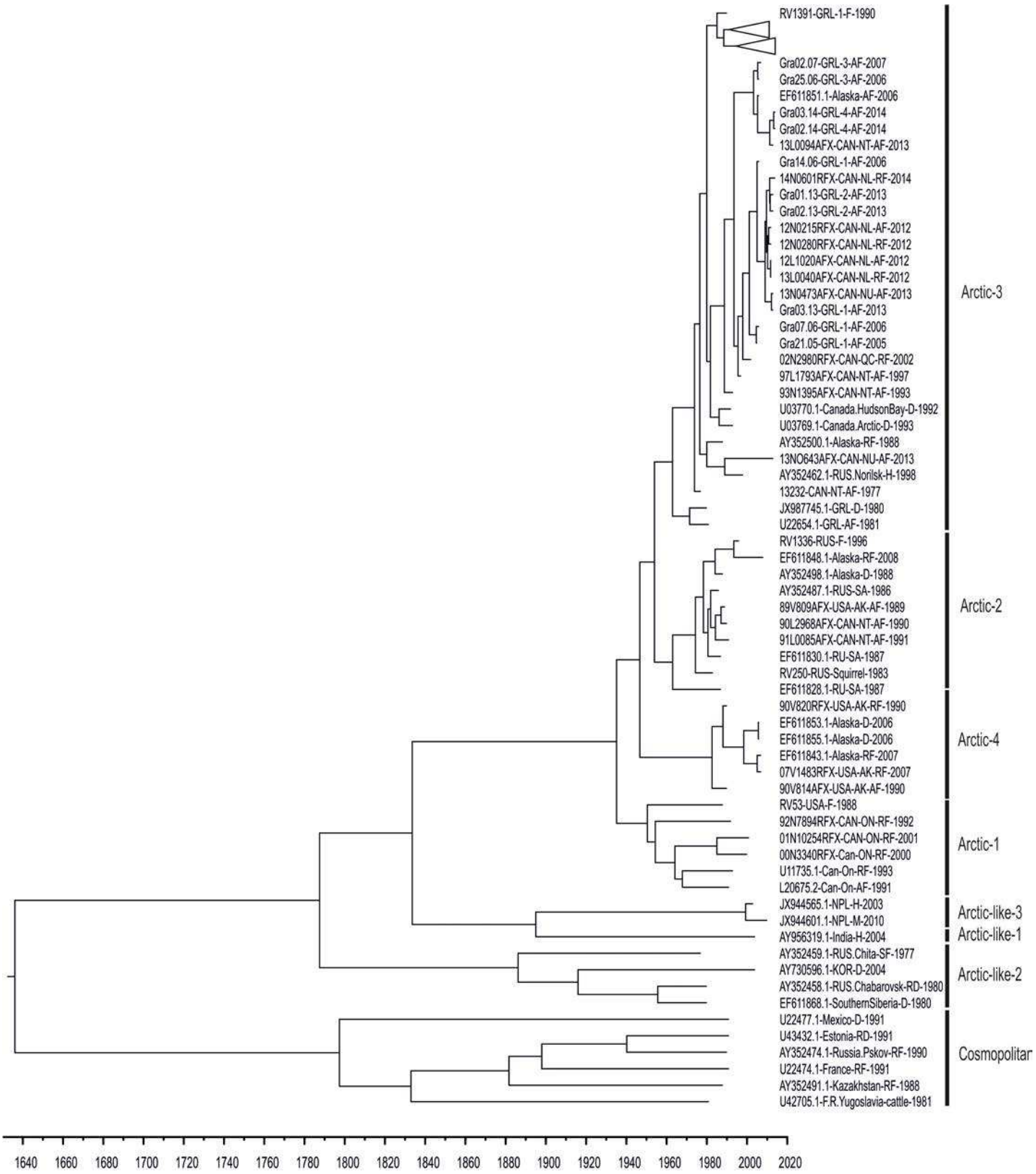


Fig 2. Time resolved phylogenetic tree of the RABV complete N gene (maximum clade credibility (MCC) calculated with BEAST from a total of 109 RABV sequences (Tables 1 and 2). RABV genetic lineages are labelled to pre-defined clusters according to a previous study [17].

doi:10.1371/journal.pntd.0004779.g002

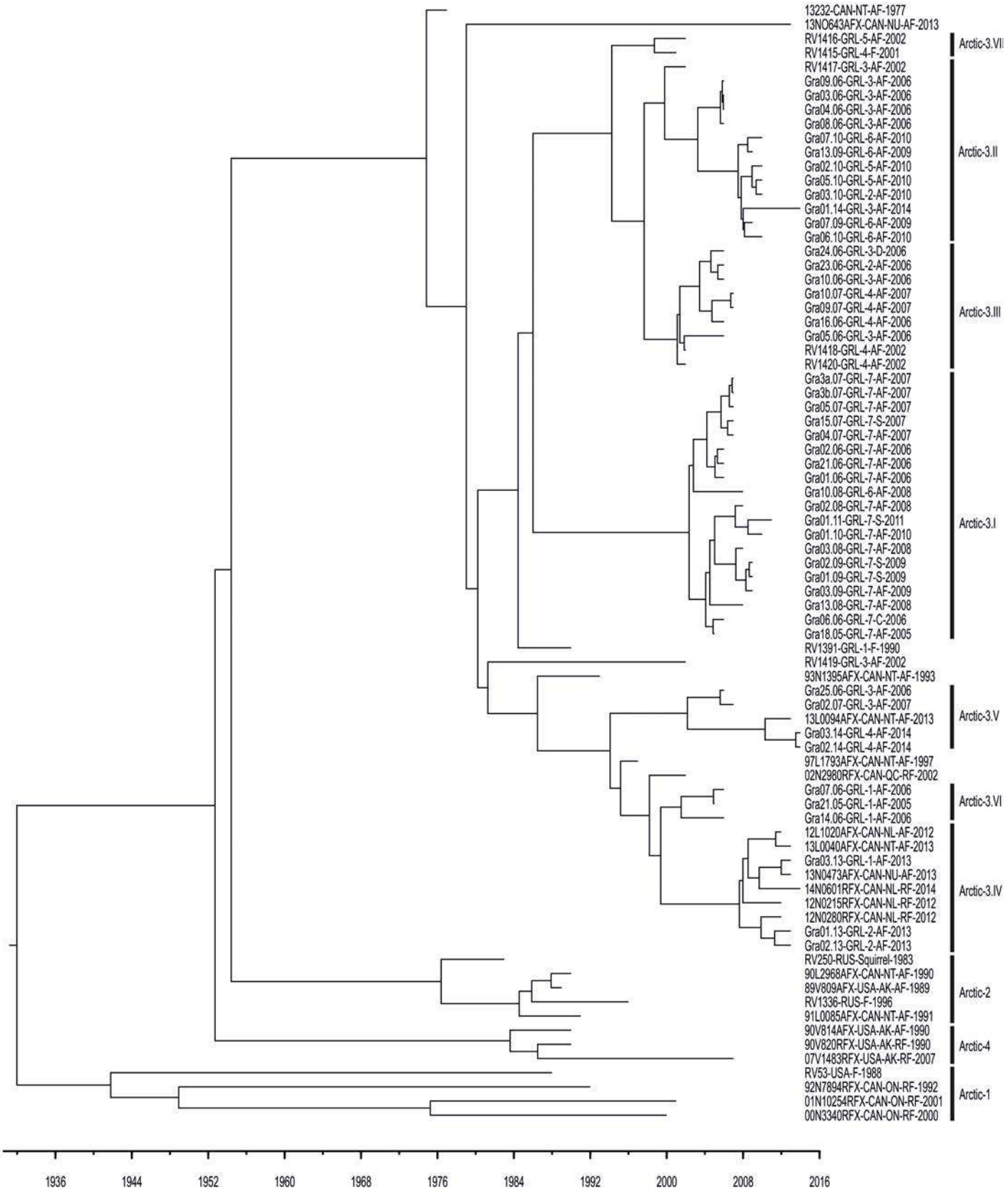

13232-CAN-NT-AF-1977
13NO643AFX-CAN-NU-AF-2013
RV1416-GRL-5-AF-2002
RV1415-GRL-4-F-2001
Arctic-3.VII
RV1417-GRL-3-AF-2002
Gra09.06-GRL-3-AF-2006
Gra03.06-GRL-3-AF-2006
Gra04.06-GRL-3-AF-2006
Gra08.06-GRL-3-AF-2006
Gra07.10-GRL-6-AF-2010
Gra13.09-GRL-6-AF-2009
Gra02.10-GRL-5-AF-2010
Gra05.10-GRL-5-AF-2010
Gra03.10-GRL-2-AF-2010
Gra01.14-GRL-3-AF-2014
Gra07.09-GRL-6-AF-2009
Gra06.10-GRL-6-AF-2010
Arctic-3.II
Gra24.06-GRL-3-D-2006
Gra23.06-GRL-2-AF-2006
Gra10.06-GRL-3-AF-2006
Gra10.07-GRL-4-AF-2007
Gra09.07-GRL-4-AF-2007
Gra16.06-GRL-4-AF-2006
Gra05.06-GRL-3-AF-2006
RV1418-GRL-4-AF-2002
RV1420-GRL-4-AF-2002
Arctic-3.III
Gra3a.07-GRL-7-AF-2007
Gra3b.07-GRL-7-AF-2007
Gra05.07-GRL-7-AF-2007
Gra15.07-GRL-7-S-2007
Gra04.07-GRL-7-AF-2007
Gra02.06-GRL-7-AF-2006
Gra21.06-GRL-7-AF-2006
Gra01.06-GRL-7-AF-2006
Gra10.08-GRL-6-AF-2008
Gra02.08-GRL-7-AF-2008
Gra01.11-GRL-7-S-2011
Gra01.10-GRL-7-AF-2010
Gra03.08-GRL-7-AF-2008
Gra02.09-GRL-7-S-2009
Gra01.09-GRL-7-S-2009
Gra03.09-GRL-7-AF-2009
Gra13.08-GRL-7-AF-2008
Gra06.06-GRL-7-C-2006
Gra18.05-GRL-7-AF-2005
Arctic-3.I
RV1391-GRL-1-F-1990
RV1419-GRL-3-AF-2002
93N1395AFX-CAN-NT-AF-1993
Gra25.06-GRL-3-AF-2006
Gra02.07-GRL-3-AF-2007
13L0094AFX-CAN-NT-AF-2013
Gra03.14-GRL-4-AF-2014
Gra02.14-GRL-4-AF-2014
Arctic-3.V
97L1793AFX-CAN-NT-AF-1997
02N2980RFX-CAN-QC-RF-2002
Gra07.06-GRL-1-AF-2006
Gra21.05-GRL-1-AF-2005
Gra14.06-GRL-1-AF-2006
Arctic-3.VI
12L1020AFX-CAN-NL-AF-2012
13L0040AFX-CAN-NT-AF-2013
Gra03.13-GRL-1-AF-2013
13N0473AFX-CAN-NU-AF-2013
14N0601RFX-CAN-NL-RF-2014
12N0215RFX-CAN-NL-RF-2012
12N0280RFX-CAN-NL-RF-2012
Gra01.13-GRL-2-AF-2013
Gra02.13-GRL-2-AF-2013
Arctic-3.IV
RV250-RUS-Squirrel-1983
90L2968AFX-CAN-NT-AF-1990
89V809AFX-USA-AK-AF-1989
RV1336-RUS-F-1996
91L0085AFX-CAN-NT-AF-1991
Arctic-2
90V814AFX-USA-AK-AF-1990
90V820RFX-USA-AK-RF-1990
07V1483RFX-USA-AK-RF-2007
Arctic-4
RV53-USA-F-1988
92N7894RFX-CAN-ON-RF-1992
01N10254RFX-CAN-ON-RF-2001
00N3340RFX-CAN-ON-RF-2000
Arctic-1
1936
1944
1952
1960
1968
1976
1984
1992
2000
2008
2016

Fig 3. Maximum clade credibility (MCC) phylogenetic tree using complete genome sequences from 79 RABV samples sequenced in this study (Table 1). The minimum identity of sequences to be grouped into a single subclade within arctic lineage 3 was set to 99.5%.

doi:10.1371/journal.pntd.0004779.g003

Greenland arctic foxes segregate into three major genetic clusters

Direct untargeted NGS determined by protocols 1 and 3 from original rabies positive brain samples from Greenland (n = 48) and Canada (n = 1) yielded a substantial amount of host sequences. This provided the unique opportunity to analyze the genetic structure of the arctic fox population. Analysis of 12 mitochondrial gene sequences revealed three major maternal lineages and additional outliers (Fig 4). Genetic identity within the main clusters was between 99.7% and 100% (Cluster 1), 99.8% and 100% (Cluster 2), and 99.9% (Cluster 3); while identity between clusters was between 99.4% and 99.6%. For all clusters there was a distinct, partly overlapping (cluster 1 and 2) geographical distribution. Twenty-one animals of cluster 1 were detected in North-western and Western Greenland (regions 1–6), while three foxes originated from the far distant Southern coast (region 7). Interestingly, a fox sample from the

Table 3. Comparison of substitution rates per site of RABVs.

References	Number of sequences	Whole genome	Substitution per site (95% HPD)					Comment
			N -gene	P—gene	M- gene	G—gene	L—gene	
This study	55	2.5	3.1 E-4	6.0 E-4	4.4 E-4	3.3 E-4	2.4 E-4	Whole genome sequences of this study (Greenland only)
		(2.1 E-4–3.0 E-4)	(1.8 E-4–4.4 E-4)	(4.2 E-4–7.9 E-4)	(2.7 E-4–6.4 E-4)	(2.2–4.5 E-4)	(1.9–2.9 E-4)	
	79	2.2 E-4						Whole genome sequences of this study
		(2.0 E-4–2.5 E-4)						
	109		2.5 E-4					N gene sequences (Fig 2)
			(1.9 E-4–3.1 E-4)					
[17]	32		1.23 E-4					Arctic/Arctic-like RABV
			(6.8 E-5–1.83 E-4)					
	17		3.89 E-4					European red fox
			(5.1 E-5–6.7 E-4)					
	60		1.48 E-4					Combination of arctic/arctic-like, European red fox and 11 further isolates
			(8.1 E-5–2.23 E-4)					
[64]	76				4.6 E-4			Chinese street RABV
					(2.5 E-4–6.6 E-4)			
[65]	N = 151		2.3 E-4			3.9 E-4		Global genetic diversity of RABV for which the time(year) of sampling was available
	G = 74		(1.1 E-4–3.6 E-4)			(1.2 E-4–6.5 E-4)		
[20]	67		3.8 E-4					Arctic—related RABV
			(2.3 E-4–5.4 E-4)					
[66]	N = 80		5.27 E-4			4.1 E-4		N = 1350 bp
	G = 71		(± 0.23 E-4)			(± 0.3 E-4)		G = 690 bp
[18]	212		3.817 E-4					Arctic RABV
			(2.816 E-4–4.825 E-4)					N = 500 bp

doi:10.1371/journal.pntd.0004779.t003

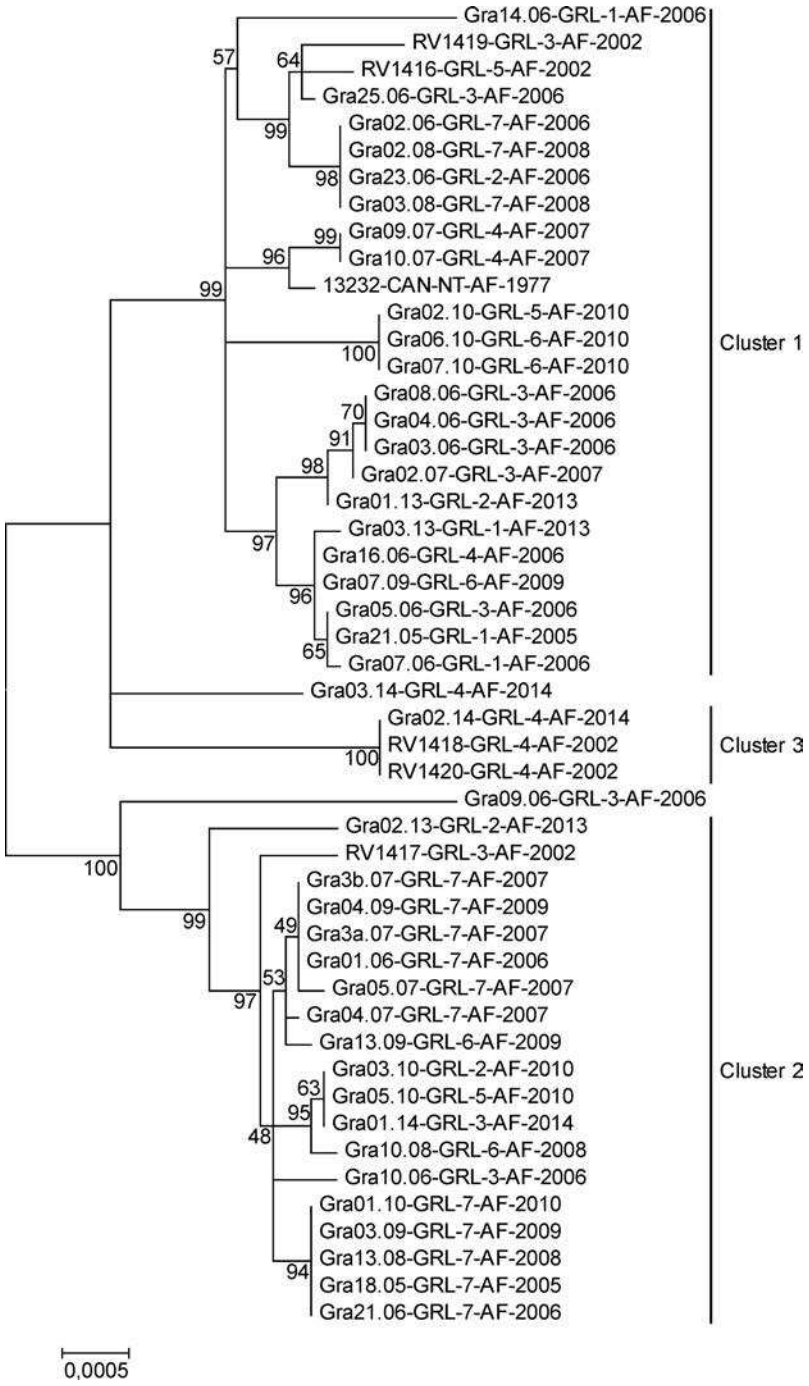


Fig 4. Phylogram of 49 arctic foxes generated from concatenated sequences of 12 mitochondrial protein coding sequences (ATP6, ATP8, COX1, COX2, COX3, CYTB, ND1, ND2, ND3, ND4, ND4L and NDS). Bootstrap values higher than 40% are indicated. The minimum identity of sequences to be grouped into a fox cluster was set to 99.7%.

doi:10.1371/journal.pntd.0004779.g004

Northwestern Territories of Canada obtained in 1977 was almost identical with two foxes collected in 2007 in Western Greenland (Fig 1, Table 1). Half of the arctic foxes from cluster 2 (n = 19) originated from southern Greenland (region 7), while the remaining foxes were

located in the western parts of the island (regions 2, 3, 5, and 6). The highly conserved cluster 3 comprised exclusively of arctic foxes from region 4, with samples from 12 years apart (Fig 4). The outliers originated from the western regions 2, 3, and 4 (Figs 1 and 4).

Occurrence of RABV variants in different fox clusters

Combinatorial analysis of RABV subclades, fox mitochondrial genes and the geographic origin revealed no association (Fig 5). The majority of samples were assigned to fox cluster 1 which was detected in all regions and comprised all RABV subclades. Fox cluster 2 was restricted mainly to the southern regions and therefore mostly infected with RABV subclades arctic 3.I and 3.II. Fox cluster 3 in contrast is only found in region 4, but represents 2 different RABV subclades (Figs 3 and 5).

Discussion

For a long time, there has been limited knowledge on the epidemiology of arctic rabies in Greenland [9–12]. It was only recently that the temporal occurrence, spatial distribution, and spread of arctic rabies in Greenland was investigated based on historical observations [13], however, this study did not provide any context to the RABV circulating on the island. While previous molecular epidemiological studies on RABV from Greenland using partial N gene sequences revealed that they belong to the arctic-3 lineage [16,17], unfortunately, the number of samples and the sequence length examined prevented further in-depth phylogenetic analyses. However, with the advent of NGS the determination of whole genome sequences for molecular epidemiological studies of RABV has become more efficient and comprehensive as illustrated recently for skunk variant of RABV from California [33].

Here, NGS derived sequence analysis is simultaneously applied to both arctic RABV strains and arctic fox genetics. A large number of RABV full genome sequences and mitochondrial genes were utilized for the phylogenetic analysis of arctic RABVs and their respective arctic fox host species.

This dataset comprises a comprehensive number of RABV samples of the arctic strain (n = 79), particularly from Greenland (n = 55), thereby providing an improved knowledge of the molecular evolution of the circulating viruses and a more comprehensive understanding of the historical perspective of the disease across the Arctic. One limitation of this and previous studies is that collection of samples within Greenland, an island comprising 2.1 million km², relies on a passive surveillance system that results in uneven distribution of submissions, i.e. samples are only submitted from areas where close fox-human encounters are possible and there is a lack of samples from central and eastern parts of Greenland (Fig 1). For instance, in sparsely populated east Greenland only 12 cases have been observed since the late 1960s until 2014 [13], suggesting either low level of infection or underreporting [17,34]. Whilst in previous studies only selected RABV were considered [16–18], here nearly all viruses from rabies cases detected between 2005 and 2014 were included, together with selected samples from an earlier study [16] that were available for re-analysis using NGS.

While in Alaska lineage 4, in Southern Canada lineage 1 and in Russia mainly RABV of arctic lineage 2 circulate, interestingly, both historic and recent RABV from Greenland exclusively fall within arctic-3 lineage (Fig 1). At present, based on our sample set of complete genome sequences, at least seven distinct subclades (arctic 3.I–3.VII) within this lineage have been circulating in Greenland (Fig 3). This delineation, however, contrasts results of a previous study, in which many of the Greenland samples (1990–2002) only clustered as one separate subgroup within arctic lineage 3 [18]. This can be explained by the limited length of the partial N gene sequence (500 bp) used which prevents a higher resolution of the tree. In fact, combining

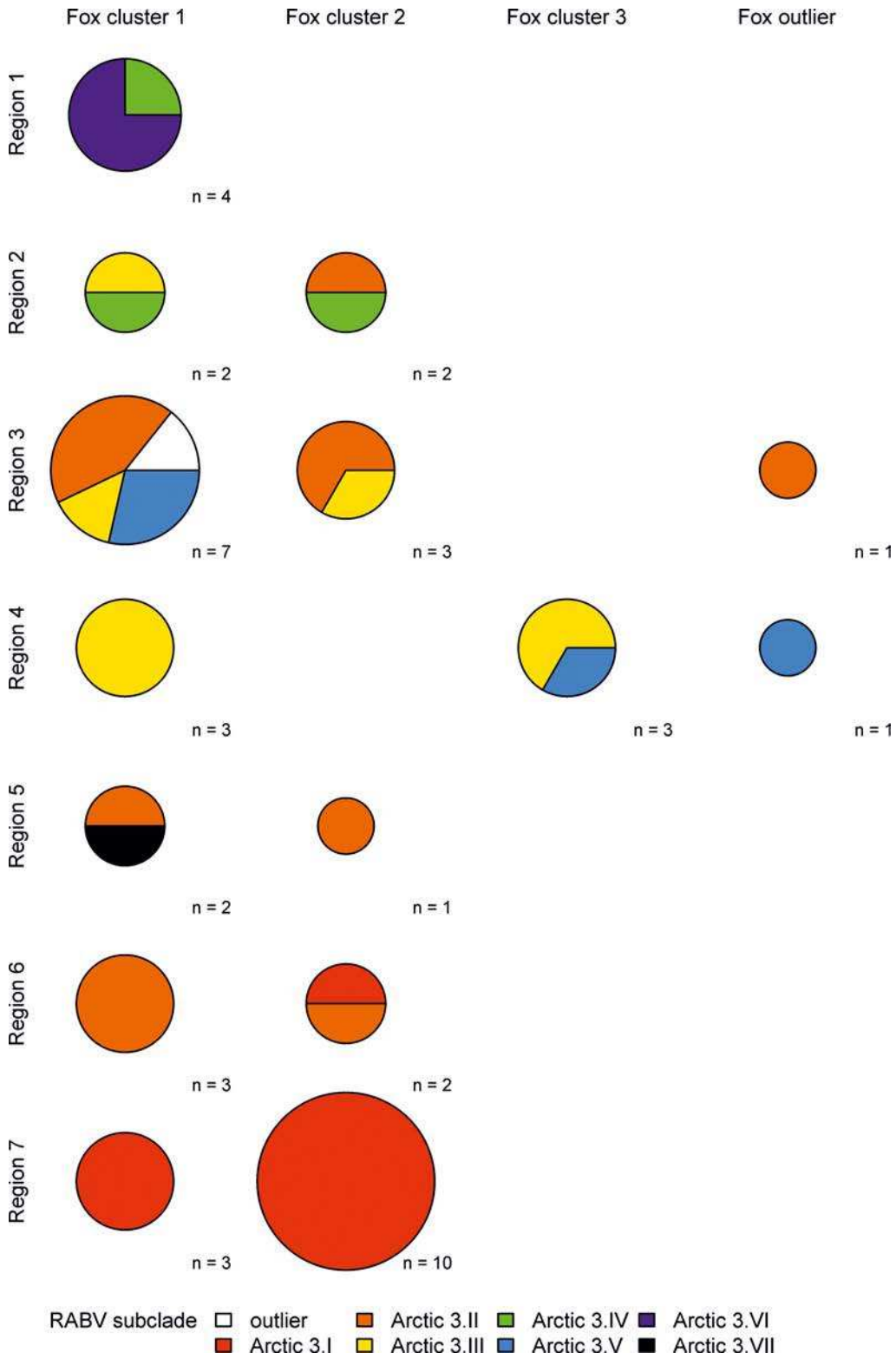


Fig 5. Comparison of landscape genetics of Greenland arctic foxes (Figs 1 and 4) and their RABV subclades (Fig 3). The size of circles represents the number of observations.

doi:10.1371/journal.pntd.0004779.g005

all previous partial N gene sequences (S2 Table) [16–18] with our dataset resulted in a rather limited alignment of 163 nucleotides in length which is too short to support phylogenetic analysis (S1 Fig). These limitations confound attempts to compare the studies and put them into a larger perspective, both temporally and spatially. Therefore, in this study this issue was addressed as best as possible by obtaining full length genome sequences from a subset of a previous study [16].

From a geographical perspective, the arctic-3 subclades defined here showed certain localized distributions (Fig 5). Considering the large home ranges of arctic foxes in connection with long distance migration of more than 1000 km [35–37], a larger geographical overlap would have been expected. However, this holds true only for subclade 3.II, the oldest cluster with viruses collected between 2002 and 2014. Phylogenetic analyses showed that all RABV from Greenland analyzed in this study have derived from Canadian incursions (Fig 3) with the oldest sample from 1977 having the most basal position in the tree. In contrast, an incursion and further spread of arctic RABV from Svalbard across the Greenland Sea into eastern and northeastern island as suggested previously [11] is not evident in this present study. In fact, Svalbard RABVs were shown to be more closely related to arctic RABVs from Russia [17,19].

The fact that subclades arctic-3.IV and 3.V are the only subclades to be identified on the North American mainland while all other subclades are restricted to Greenland's western and southern parts, respectively, may represent evidence for a more recent exchange of viruses (Figs 1 and 3). In fact, these subclades 3.IV and 3.V were found in regions of Greenland where the distance to neighboring Ellesmere Island, Canada, is shortest. Here, pack ice that frequently bridges these two land masses may facilitate the spread of the disease [16,18]. Hence, the Smith Sound, the uninhabited sea passage between Greenland and Canada's northernmost islands may play an important role in the exchange of arctic RABVs in this part of the Arctic [38]. This is supported by the close genetic relation between RABV Gra03.13-GRL-1-AF-2013 and 13N0473AFX-CAN-NU-AF-2013, with nearly identical genomes (single substitution), despite the fact that they are geographically separated by a distance of 750 km (Thule Air Base to Resolute Bay).

In Greenland, viruses with a high nucleotide substitution rate may have evolved into younger subclades that have a narrower geographical distribution and are only identified on the island (Figs 3 and 5). For instance, arctic subclade 3.I is only found in the southern parts (Region 6,7; Fig 5). Interestingly, the most ancestral RABV still circulating (subclade arctic 3. II) has the widest geographic spread across the entire western part of Greenland. Such spread has also been demonstrated epidemiologically [13]. The dynamic observed here is further demonstrated by the observation that older subclades previously found in Greenland (Fig 2), [15–18] seem to have disappeared.

It is interesting that arctic lineage 2 (Figs 1 and 3) had not (yet) been detected in Greenland, particularly considering that other arctic lineages appear to have spread [18]. As a case in point, one sample from 1990 from Grise Fjord, which has close proximity to Greenland, was arctic lineage 2, while a sample from the same place twenty-three years later was arctic 3 (Figs 1 and 3, Table 1). As regards the other arctic lineages, lineage 4 seems to be restricted to Alaska [18,39] (Fig 1), while circulation of arctic lineage 1 in Southern Ontario, Canada is highly restricted due to oral rabies vaccination [40,41].

The observed genetic dynamic within arctic RABVs is also demonstrated by the nucleotide substitution rates inferred from this dataset at the full genome level. Although in comparison with partial sequence analysis of other RABVs this appears low (Table 3), a substitution rate of 2.5 E-4 per site and year still indicates a substantial evolutionary dynamic. Discrepancies in the substitution rates and in the resulting MRCA as discussed before [18], are a result of the respective partial sequence used for calculation (Table 3). Thus, not all RABV genes are equal for

evolutionary analyses, as previously suggested [42,43]. Taken together, use of complete genome sequences should result in a more accurate substitution rate, closely reflecting the actual virus evolution and genetic dynamic.

The use of unbiased NGS offered the unique opportunity to simultaneously obtain reservoir host derived RNA sequences from the same sample for population analysis. Similar to previous studies [26,44] we initially used D-Loop sequences for this analysis. However, this did not allow a clear distinction because of the high genetic identities observed. Instead, for increased resolution, the 12 protein-coding mitochondrial genes enabled delineation of Greenland's fox population into three main genetic clusters as shown recently [39]. Still, the high genetic identity both within and between the three main genetic fox clusters ranging between 99.4% and 100% may support assumptions that the genetic diversity of island arctic foxes compared to main land populations as a result of colonization is low [37,45]. The detection of three almost genetically identical arctic fox samples from the Northwestern Territories of Canada and Western Greenland (Fig 1) three decades apart may be evidence for high gene flow among arctic fox subpopulations, contributing to low genetic differentiation at least in the mitochondrial genome and further corroborate this hypothesis. In contrast, the results from this present study of mitochondrial gene analysis are in agreement with previous observations showing a fine-scale spatial population structure in Alaskan arctic foxes [39]. Future more detailed comparative haplotype analysis of arctic foxes including nuclear loci as well as fox genetic data from Canada's northern territories and Alaska should corroborate this fact.

While previously, mtDNA structure in arctic and red foxes from Alaska did not correspond to RABV variant structure in either species, microsatellite analyses identified 3 and 4 groups of arctic foxes, closely matching the distribution of rabies virus variants in the state [39]. Although the Greenland RABV arctic-3 subclades were not evenly distributed among the different mitochondrial fox clusters (Fig 5), these data provide no evidence for independent isolated evolutionary development of RABV in different arctic fox lineages but rather resemble geographic separation.

In our study we focused on mitochondrial DNA (mtDNA) alone for characterization of the population genetics of arctic foxes. Analysis of mtDNA reveals a longer-term view of population structure than do fast evolving nuclear markers (e.g., microsatellites) and analysis of multiple mitochondrial genes, as in this study, yields finer resolution of relationship than studies focusing on a single gene. Nonetheless, any combination of mitochondrial genes is effectively considered a single locus. Recent studies included multiple markers of nuclear genes i.e. microsatellites to investigate host genetics in carnivores, sometimes resulting in conflicting results when compared to parallel investigated mtDNA [46–50]. However, a meta-analysis showed that mtDNA was robust in determining patterns of population history and yielded similar results to microsatellites [51]. Finally, the lack of a standardized approach for using both types of marker genes and the limited information on nuclear genes to those that are expressed as mRNA and are thus part of the dataset precluded further analyses. For instance, in a recent study the transcriptome was used for evolutionary studies of Arctic and red foxes [52]. While this information should be available it is unclear whether it would be sufficient for intra-species genetic studies.

Conclusion

By combining full length RABV genome sequence analysis and host derived sequences the interaction of viruses and their hosts was exemplarily demonstrated and may serve as a model approach for analysis of real-world understanding of infectious disease dynamics and virus-host interdependencies using a landscape genetics approach as suggested for dog mediated

rabies [53]. In this study, although no interdependencies based on mtDNA were identified, nevertheless this approach led to a better understanding of the evolution, dynamics and geographical spread of arctic rabies in Greenland.

A high degree of genetic identity both of RABVs and arctic foxes from Canada and Greenland suggests the movement of infected animals between the two landmasses. The overall diversity of arctic RABV in Greenland was very limited and only by analyzing the entire genome, a high resolution of the genetic evolution was possible, providing real-time insights into viral evolution. These results may be useful for future control strategies of arctic fox rabies. In contrast to previous statements [13], given the unique geographical location of Greenland, the expected reduction of connectivity by pack-ice due to climate change [54,55] and the geographic separation of individual host and virus genetic subclades despite long distance movement [36], the idea of arctic rabies control using oral rabies vaccination (ORV) in selected coastal areas appears feasible [56,57]. While preliminary field trials in Newfoundland (Canada) [58] and even northern Greenland [59] demonstrated in principle that ORV could be undertaken in remote northern regions, a targeted vaccination strategy would have to be developed before an elimination program could be implemented.

Supporting Information

S1 Table. Overview of raw sequence data using NGS generated in this study.
(PDF)

S2 Table. Reference sequences used in S1 Fig.
(PDF)

S1 Fig. Phylogram combining all previous partial N gene sequences (S2 Table) [16–18] with our dataset based on a rather limited alignment of 163 nucleotides in length.
(TIFF)

Acknowledgments

The authors would like to thank all representatives of responsible staff of local veterinary services in Greenland and Canadas Northern territories for obtaining rabies samples from coastal regions. Without their help this study would have been impossible. The Canadian portion of this study was inspired by the efforts of Dr. H. Whitney (former Chief Veterinary Officer for the province of Newfoundland and Labrador) to improve knowledge of the spread of fox rabies in the north. We are especially thankful for the excellent technical assistance provided by Jeannette Kliemt and Patrick Wysocki at FLI, by Helle Rasmussen at DTU, and by Emilie Falardeau at CFIA. We also acknowledge the assistance of Dr C. Fehlner-Gardiner and staff from the CFIA's Centre of Expertise for Rabies for providing rabies diagnostic and typing services. The authors also wish to thank Dr Karen Mansfield (APHA) for providing rabies virus RNA from Greenland samples recovered in 2006.

Author Contributions

Conceived and designed the experiments: TFM CMF TCM TBR ARF. Performed the experiments: DHa DHo SND DM. Analyzed the data: DHa DHo SF KHue KHun SND DM. Contributed reagents/materials/analysis tools: AB TBR DM ARF. Wrote the paper: TFM CMF MB DHo TCM.

References

1. Dietzgen RG, Calisher CH, Kurath G, Kuzmin IV, Rodriguez LL, et al. Family Rhabdoviridae. In: King AM, Adams MJ, Carstens EB, Lefkowitz EJ, editors. Virus taxonomy: classification and nomenclature of viruses—Ninth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier; 2012. pp. 686–713.
2. World Health Organisation. Expert Consultation on Rabies, Second Report. World Health Organ Tech Rep Ser. 2013; 982: 1–150. PMID: 24069724
3. World Health Organization. Report of A WHO/NVI Workshop on arctic rabies. Upsala, Sweden, 24–27 April 1990. 1–26 p.
4. Noren K, Carmichael L, Dalen L, Hersteinsson P, Samelius G, et al. Arctic fox Vulpes lagopus population structure: circumpolar patterns and processes. Oikos. 2011; 120: 873–885.
5. Macdonald DW. The Ecology of Carnivore Social-Behavior. Nature. 1983; 301: 379–384.
6. Eberhardt LE, Hanson WC, Bengtson JL, Garrott RA, Hanson EE. Arctic Fox Home Range Characteristics in an Oil-Development Area. J Wildl Manage. 1982; 46: 183–190.
7. Anthony RM. Home ranges and movements of arctic fox (Alopex lagopus) in Western Alaska. Arctic. 1997; 50: 147–157.
8. Frafjord K, Prestrud P. Home Range and Movements of Arctic Foxes Alopex-Lagopus in Svalbard. Polar Biol. 1992; 12: 519–526.
9. Crandell RA. Arctic Fox Rabies. In: Baer GM, editor. The Natural History of Rabies. New York: Academic Pr.; 1975. pp. 23–40.
10. Jenkins M, Wamberg K. Rabies discovered in Greenland. J Am Vet Med Assoc. 1960; 137: 183–185. PMID: 14407072
11. Leisner K. Rabies in Greenland, 1975–2001. Rabies Bulletin Europe. 2002; 26: 10–14.
12. Mork T, Prestrud P. Arctic Rabies—A Review. Acta Vet Scand. 2004; 45: 1–9. PMID: 15535081
13. Raundrup K, Moshoj CM, Wennerberg SE, Kapel CMO. Spatiotemporal distribution of rabies in Arctic foxes in Greenland. Eur J Wildl Res. 2015; 61: 457–465.
14. Crandell RA. Laboratory Investigation of Arctic Strains of Rabies Virus. Acta Pathol Microbiol Scand. 1965; 63: 587–596. PMID: 14324666
15. Kissi B, Tordo N, Bourhy H. Genetic Polymorphism in the Rabies Virus Nucleoprotein Gene. Virology 1995; 209: 526–537. PMID: 7778285
16. Mansfield KL, Racloz V, McElhinney LM, Marston DA, Johnson N, et al. Molecular epidemiological study of Arctic rabies virus isolates from Greenland and comparison with isolates from throughout the Arctic and Baltic regions. Virus Res. 2006; 116: 1–10. PMID: 16198016
17. Kuzmin IV, Hughes GJ, Botvinkin AD, Gribencha SG, Rupprecht CE. Arctic and Arctic-like rabies viruses: distribution, phylogeny and evolutionary history. Epidemiol Infect. 2008; 136: 509–519. PMID: 17599781
18. Nadin-Davis SA, Sheen M, Wandeler AI. Recent emergence of the Arctic rabies virus lineage. Virus Res. 2012; 163: 352–362. doi: 10.1016/j.virusres.2011.10.026 PMID: 22100340
19. Johnson N, Dicker A, Mork T, Marston DA, Fooks AR, et al. Phylogenetic comparison of rabies viruses from disease outbreaks on the Svalbard Islands. Vector Borne Zoonotic Dis. 2007; 7: 457–460. PMID: 17767407
20. Pant GR, Lavenir R, Wong FY, Certoma A, Larrous F, et al. Recent emergence and spread of an Arctic-related phylogenetic lineage of rabies virus in Nepal. PLoS Negl Trop Dis. 2013; 7: e2560. doi: 10.1371/journal.pntd.0002560 PMID: 24278494
21. Dean DJ, Abelseth MK, Athanasiu P. The fluorescence antibody test In: Meslin FX, Kaplan MM, Koprowski H, editors. Laboratory techniques in rabies. 4th ed. Geneva: World Health Organization; 1996. pp. 88–93.
22. Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, et al. Next generation sequencing of viral RNA genomes. BMC Genomics. 2013; 14: 444. doi: 10.1186/1471-2164-14-444 PMID: 23822119
23. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, et al. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform. 2013; 14: 193–202. doi: 10.1093/bib/bbs012 PMID: 22445902
24. Marston DA, Wise EL, Ellis RJ, McElhinney LM, Banyard AC, et al. Complete genomic sequence of rabies virus from an ethiopian wolf. Genome announc. 2015; 3: e00157–15. doi: 10.1128/genomeA.00157-15 PMID: 25814597
25. Delisle I, Strobeck C. A phylogeny of the Caniformia (order Carnivora)based on 12 complete protein-coding mitochondrial genes. Mol Phylogenet Evol. 2005; 37: 192–201. PMID: 15964215

26. Dalen L, Fuglei E, Hersteinsson P, Kapel CMO, Roth JD, et al. Population history and genetic structure of a circumpolar species: the arctic fox. Biol J Linn Soc. 2005; 84: 79–89.

27. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22: 4673–4680. PMID: 7984417

28. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012; 29: 1969–1973. doi: 10.1093/molbev/mss075 PMID: 22367748

29. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015; 32: 268–274. doi: 10.1093/molbev/msu300 PMID: 25371430

30. Kimura M. Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci U S A. 1981; 78: 454–458. PMID: 6165991

31. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980; 16: 111–120. PMID: 7463489

32. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol. 2011; 28: 2731–2739. doi: 10.1093/molbev/msr121 PMID: 21546353

33. Borucki MK, Chen-Harris H, Lao V, Vanier G, Wadford DA, et al. Ultra-deep sequencing of intra-host rabies virus populations during cross-species transmission. PLoS Negl Trop Dis. 2013; 7: e2555. doi: 10.1371/journal.pntd.0002555 PMID: 24278493

34. Prestrud P, Krogsrud J, Gjertz I. The Occurrence of Rabies in the Svalbard Islands of Norway. J Wildl Dis. 1992; 28: 57–63. PMID: 1548803

35. Goltsman M, Kruchenkova EP, Sergeev S, Johnson PJ, Macdonald DW. Effects of food availability on dispersal and cub sex ratio in the Mednyi Arctic fox. Behav Ecol Sociobiol. 2005; 59: 198–206.

36. Tarroux A, Berteaux D, Bety J. Northern nomads: ability for extensive movements in adult arctic foxes. Polar Biol. 2010; 33: 1021–1026.

37. Noren K, Carmichael L, Fuglei E, Eide NE, Hersteinsson P, et al. Pulses of movement across the sea ice: population connectivity and temporal genetic structure in the arctic fox. Oecologia. 2011; 166: 973–984. doi: 10.1007/s00442-011-1939-7 PMID: 21344255

38. Charles E. Movements of Arctic Fox populations in the region of Baffin Bay and Smith Sound. Polar Rec. 1949; 5: 296–305.

39. Goldsmith EW, Renshaw B, Clement CJ, Himschoot EA, Hundertmark KJ, et al. Population Structure of Two Rabies Hosts Relative to the Known Distribution of Rabies Virus Variants in Alaska. Mol. Ecol. 2016; 25:675–688. doi: 10.1111/mec.13509 PMID: 26661691

40. MacInnes CD, Smith SM, Tinline RR, Ayers NR, Bachmann P, et al. Elimination of rabies from red foxes in eastern Ontario. J Wildl Dis. 2001; 37: 119–132. PMID: 11272485

41. Rosatte RC, Power MJ, Donovan D, Davies JC, Allan M, et al. Elimination of arctic variant rabies in red foxes, metropolitan Toronto. Emerg Infect Dis. 2007; 13: 25–27. PMID: 17370512

42. Johnson N, McElhinney LM, Smith J, Lowings P, Fooks AR. Phylogenetic comparison of the genus Lyssavirus using distal coding sequences of the glycoprotein and nucleoprotein genes. Arch Virol. 2002; 147: 2111–2123. PMID: 12417947

43. Wu X, Franka R, Velasco-Villa A, Rupprecht CE. Are all lyssavirus genes equal for phylogenetic analyses? Virus Res. 2007; 129: 91–103. PMID: 17681631

44. Aubry KB, Statham MJ, Sacks BN, Perrine JD, Wisely SM. Phylogeography of the North American red fox: vicariance in Pleistocene forest refugia. Mol Ecol. 2009; 18: 2668–2686. doi: 10.1111/j.1365-294X.2009.04222.x PMID: 19457180

45. Geffen E, Waidyaratne S, Dalen L, Angerbjorn A, Vila C, et al. Sea ice occurrence predicts genetic isolation in the Arctic fox. Mol Ecol. 2007; 16: 4241–4255. PMID: 17868292

46. Noren K, Angerbjorn A, Hersteinsson P. Population structure in an isolated Arctic fox, Vulpes lagopus, population: the impact of geographical barriers. Biol J Linn Soc. 2009; 97: 18–26.

47. Bardeleben C, Moore RL, Wayne RK. A molecular phylogeny of the Canidae based on six nuclear loci. Mol Phylogenet Evol. 2005; 37: 815–831. PMID: 16213754

48. Cohen TM, King R, Dolev A, Boldo A, Lichter-Peled A, et al. Genetic characterization of populations of the golden jackal and the red fox in Israel. Conserv Genet. 2013; 14: 55–63.

49. Dalen L, Kvaloy K, Linnell JDC, Elmhagen B, Strand O, et al. Population structure in a critically endangered arctic fox population: does genetics matter? Mol Ecol. 2006; 15: 2809–2819. PMID: 16911202

50. Sacks BN, Louie S. Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae. Mol Ecol Resour. 2008; 8: 35–49. doi: 10.1111/j.1471-8286.2007.01830.x PMID: 21585716

51. Zink RM, Barrowclough GF. Mitochondrial DNA under siege in avian phylogeography. Mol Ecol. 2008; 17: 2107–2121. doi: 10.1111/j.1365-294X.2008.03737.x PMID: 18397219

52. Kumar V, Kutschera VE, Nilsson MA, Janke A. Genetic signatures of adaptation revealed from transcriptome sequencing of Arctic and red foxes. BMC Genomics. 2015; 16: 585. doi: 10.1186/s12864-015-1724-9 PMID: 26250829

53. Brunker K, Hampson K, Horton DL, Biek R. Integrating the landscape epidemiology and genetics of RNA viruses: rabies in domestic dogs as a model. Parasitology. 2012; 139: 1899–1913. doi: 10.1017/S003118201200090X PMID: 22814380

54. Mellows A, Barnett R, Dalen L, Sandoval-Castellanos E, Linderholm A, et al. The impact of past climate change on genetic variation and population connectivity in the Icelandic arctic fox. Proc R Soc Lond B —Biol Sci. 2012; 279: 4568–4573.

55. Freuling C, Vos A, Johnson N, Mühle RU, Müller T. Rabies. In: Johnson N, editor. The role of animals in emerging viral diseases. 1st ed. San Diego: Academic Press. 2013; pp. 63–87.

56. Follmann EH, Ritter DG, Baer GM. Oral rabies vaccination of arctic foxes (Alopex lagopus) with an attenuated vaccine. Vaccine. 1992; 10: 305–308. PMID: 1574916

57. Follmann EH, Ritter DG, Hartbauer DW. Oral Vaccination of Captive Arctic Foxes with Lyophilized SAG2 Rabies Vaccine. J Wildl Dis. 2004; 40: 328–334. PMID: 15362836

58. Johnston DH, Fong DW. Epidemiology of arctic fox rabies. In: Bögel K, Meslin F-X, Kaplan MM, editors. Wildlife rabies control. Kent, UK: Wells Medical Ltd. 1992; pp. 45–49.

59. Hansen EH. Oral vaccination of arctic foxes in Greenland: A field trial. 7th Annual International Meeting: Advances Towards Rabies Control in the Americas. Atlanta, USA: CDC; 1996. pp. 7.

60. Kuzmin IV, Botvinkin AD, McElhinney LM, Smith JS, Orciari LA, et al. Molecular epidemiology of terrestrial rabies in the former Soviet Union. J Wildl Dis. 2004; 40: 617–631. PMID: 15650080

61. Park YJ, Shin MK, Kwon HM. Genetic characterization of rabies virus isolates in Korea. Virus Genes. 2005; 30: 341–347. PMID: 15830152

62. Nadin-Davis SA, Casey GA, Wandeler A. Identification of regional variants of the rabies virus within the Canadian province of Ontario. J Gen Virol. 1993; 74: 829–837. PMID: 8492088

63. Nadin-Davis SA, Casey GA, Wandeler AI. A molecular epidemiological study of rabies virus in central Ontario and western Quebec. J Gen Virol. 1994; 75: 2575–2583. PMID: 7931145

64. Wu H, Wang L, Tao X, Li H, Rayner S, et al. Genetic diversity and molecular evolution of the rabies virus matrix protein gene in China. Infect Genet Evol. 2013; 16: 248–253. doi: 10.1016/j.meegid.2013.02.009 PMID: 23453987

65. Bourhy H, Reynes JM, Dunham EJ, Dacheux L, Larrous F, et al. The origin and phylogeography of dog rabies virus. J Gen Virol. 2008; 89: 2673–2681. doi: 10.1099/vir.0.2008/003913-0 PMID: 18931062

66. Holmes EC, Woelk CH, Kassis R, Bourhy H. Genetic constraints and the adaptive evolution of rabies virus in nature. Virology. 2008; 292: 247–257.

5.2.4: Article 28 in preparation

Molecular Epidemiology and Evolution of European Bat Lyssavirus Type 2

Summary: Unlike EBLV-1, which has been detected during active surveillance in bat species other than Serotine bats (Schatz et al., 2013), the chiropteran host range of EBLV-2 appears to be restricted to just two bat species. Many factors affecting the behaviour of Daubenton's bats, may allow EBLV-2 to become established and rapidly spread throughout the population. This may account for the detection of EBLV-2, albeit at low levels, throughout the UK. Cross species transmission (CST) of bat lyssaviruses to other mammals has been documented but is extremely rare. Two CST events involving EBLV-2 from bats to humans have been reported, in Finland in 1985 and in the UK in 2002. Unlike for EBLV-1, EBLV-2 spillover into other wildlife or domestic species has not been documented.

Full genome sequences have been published or obtained in this study for 29 of the 33 clinical EBLV-2 cases recorded. Thirteen EBLV-2 cases have been reported in the UK, 12 in Daubenton's bats and a single case human case. In addition, EBLV-2 viral RNA was detected in a single oro-pharyngeal swab specimen collected from a healthy Daubenton's bat sampled as part of an active surveillance programme in Scotland in 2008. In 2015, Norway reported its first case of EBLV-2 in a Daubenton's bat. The bat was found by a member of the public in Oppland, in the southern part of Norway and has recently been published (see article 10).

The two Finnish EBLV-2 sequences and U89479 consistently formed a well-supported sub lineage (EBLV-2b) irrespective of the model with all others grouping in lineage EBLV2a, apart from the 1992 & 1993 Swiss EBLV-2 sequences that were not consistently associated with either sub-lineage when individual sequences were removed or additional lyssavirus sequences added. In general EBLV-2 sequences cluster geographically. The unique opportunity to study variation within a single roost site, Stokesay Castle, over a 7 year period using WGS identified the 3 genomes were 99.9% identical, with substitutions identified at only 4 positions. Furthermore, sub-consensus level read analysis was undertaken and identified minor variants which fixed over time.

Molecular Epidemiology and Evolution of European Bat Lyssavirus Type 2

Lorraine M. McElhinney[1,2#], Denise A. Marston[1*], Emma L. Wise[1], Conrad M. Freuling[3], Herve Bourhy[4], Reto Zanoni[5], Torfinn Moldal[6], Bart Kooi[7], Antonie Neubauer-Juric[8], Tiina Nokireki[9], Thomas Müller[3], Anthony R. Fooks,[1,2]*

[1] Wildlife Zoonoses and Vector Borne Disease Research Group, Animal and Plant Health Agency (APHA), Surrey, United Kingdom.
[2] Institute of Global Health, University of Liverpool, Liverpool, United Kingdom
[3] Institute of Molecular Virology and Cell Biology, Friedrich Loeffler Institute, (FLI), Greifswald-Insel Riems, Germany
[4] Institute Pasteur, París, France
[5] Institute of Virology and Immunology, University of Berne, Switzerland
[6] Norwegian Veterinary Institute, Oslo, Norway
[7] Wageningen University and Research Centre, Lelystad, The Netherlands.
[8] Bavarian Health and Food Safety Authority, Oberschleissheim, Germany.
[9] Finnish Food Safety Authority Evira, Helsinki, Finland

Key words: European Bat Lyssavirus Type 2, EBLV-2, Europe, evolution, epidemiology, Myotis, bat.

*Equal contribution

#Corresponding author: Dr Lorraine Marie McElhinney, Wildlife Zoonoses and Vector Borne Diseases Research Group, Animal & Plant Health Agency, Woodham Lane, New Haw Addlestone, Surrey KT15 3NB, UK. E-mail: Lorraine.McElhinney@apha.gsi.gov.uk

Abstract:
The first case of bat rabies in Europe was recorded in Hamburg, Germany, in 1954. Bat rabies cases in Europe are principally attributed to two lyssaviruses, namely European Bat 1 Lyssaviruses (EBLV-1) and European Bat 2 Lyssavirus (EBLV-2). Prior to the death of a bat worker in Finland in 1985, very few bat rabies cases were reported. Enhanced surveillance in the two subsequent years (1986-87), identified 263 cases (more than a fifth of all reported cases to date). Between 1977 and 2016, 1183 cases of bat rabies were reported, with the vast majority (>97%) being attributed to EBLV-1. In contrast, there have been only 38 suspected cases of EBLV-2, of which 32 have been confirmed by virus typing. Unlike EBLV-1, which has been detected in a range of bat species, the chiropteran host range of EBLV-2 is presently restricted to just two bat species; *M. daubentonii* and *M. dascyneme*.

The limited number of EBLV-2 cases in Europe prompted the establishment of a network of European reference laboratories to collate all available viruses and data. Despite the relatively low number of EBLV-2 cases, a large amount of anomalous data had been published in the scientific literature, which we have now reviewed and clarified. In this review, 29 EBLV-2 full genome sequences have been analysed to further our understanding of the diversity and molecular evolution of EBLV-2 in Europe.

In addition to the 29 EBLV-2 genomes, partial nucleoprotein (N) gene sequences were available from two oro-pharyngeal swabs collected from healthy *Myotis Daubentonii* during active surveillance programmes in Scotland and Switzerland. Phylogenetic analysis of 31 EBLV-2 partial N gene sequences defined both geographical and host specific relationships but no chronological clustering was evident. Analysis of the 29 complete EBLV-2 genome sequences, clearly defined geographical relationships with all EBLV-2 sequences clustering at the country level irrespective of the gene studied. Further geographical clustering was also observed at a local level. There is high levels of homogeneity within the EBLV-2 species with nucleotide identities ranging from 95.5-100% and amino acid identities between 98.7-10%, despite the widespread isolation of the isolates both geographically and chronologically. The mean substitution rate for EBLV-2 across the five concatenated genes was 1.65E-5 and evolutionary clock analysis confirms the slow evolution of EBLV-2 both between and within countries in Europe. This is further supported by the relatively high

sequence homogeneity across the genome of three EBLV-2 isolates obtained several years apart (2007, 2008 and 2014) from *Myotis Daubentonii* at the same site (Stokesay Castle, Shropshire UK).

Introduction

There are currently fourteen recognised species in the lyssavirus genus of the family *Rhabdoviridae* (ICTV 2015). These include, *Rabies lyssavirus* (RABV), *Lagos Bat Lyssavirus* (LBV), *Mokola Lyssavirus* (MOKV), *Duvenhage Lyssavirus* (DUVV), *European Bat Lyssavirus Type 1* (EBLV-1), *European Bat Lyssavirus Type 2* (EBLV-2), *Australian Bat Lyssavirus* (ABLV), *Aravan Lyssavirus* (ARAV), *Khujand Lyssavirus* (KHUV), *Irkut Lyssavirus* (IRKV), *West Caucasian Bat Lyssavirus* (WCBV) *Shimoni Bat Lyssavirus* (SHIBV), *Ikoma Lyssavirus* (IKOV) and *Bokeloh Bat Lyssavirus* (BBLV). A further two putative bat lyssaviruses await classification; Lleida lyssavirus (LLBV) (Arechiga-Ceballos et al., 2013) and *Gannoruwa bat lyssavirus* (GBLV) (Gunawardena et al., 2016). With the exception of MOKV and IKOV, all lyssaviruses have been isolated from bats.

Cases of bat rabies occur globally, although different bat lyssavirus species are present in different regions (Streicker et al., 2010, Banyard et al., 2011). Bat rabies cases in Europe are principally attributed to EBLV-1 and EBLV-2. However, three other lyssaviruses have been detected in European insectivorous bats, albeit in very few cases. WCBV was isolated in 2002 from a bent-winged bat (Miniopterus schreibersii) in the Caucasus Mountains and LLBV from the same bat species in Spain in 2011 (Botvinkin et al. 2003, Echevarria et al., 2013). Antigenic and phylogenetic profiling of BBLV, isolated from *Myotis nattererii* in Germany and France, confirms the presence of a fifth lyssavirus in European bats (Freuling et al., 2011, Picard-Meyer et al., 2013).

The first rabies infected insectivorous bat in Europe was recorded in Hamburg, Germany, in 1954 (Mohr, 1957, Schindler & Dennig, 1958**)**. The reports described a biting incident from a bat that was caught and tested but not formally identified. Negri bodies were detected following mouse inoculation tests but no tissue was retained for virus typing, so the bat and virus species involved are unknown. Prior to the death of a bat worker in Finland in 1985 (Lumio et al., 1986), very few bat rabies cases were reported and 263 cases, 23% of all subsequent reported cases, were identified through enhanced surveillance in the subsequent two years (1986 and 1987). With the advent of RT-PCR and molecular sequencing, the viruses were genetically typed as two separate lyssavirus species, EBLV-1 and -2 (Bourhy et al., 1992).

Between 1977 and 2016, 1183 infected bats were reported in total in Europe with an average of 34 infected bats reported to Rabies Bulletin Europe in the last decade. More than 80% of the reported cases were identified in Denmark (21%), Germany (27%) and the Netherlands (33%). Differences in surveillance systems, bat ecology, public awareness and available resources may account for the relatively low numbers of confirmed EBLV cases (Schatz et al., 2013). However, the apparent host specificity may also limit the prevalence of EBLVs when considering the wide distribution, high frequency and sympatric habitation of some European insectivorous bat species. Whilst lyssaviruses have the ability to cause clinical disease in bats, they can also circulate within healthy bat populations detectable only via active surveillance screening (Brookes et al., 2005; Amengual et al., 2007).

Evolutionary analyses, sub-clinical infections and relatively restricted virus–host relationships suggest that all lyssaviruses, including RABV, most likely originated in bats (Badrane & Tordo, 2001; Vos et al., 2007). Previous attempts to study the phylogeography, molecular epidemiology and evolutionary rates for EBLV-2 have been restricted to only a small number of virus isolates and partial genomic sequences (Davis et al., 2005, McElhinney et al., 2013, Jakava-Viljanen et al., 2015). The historic sharing of viruses between European rabies laboratories for confirmation and research was not standardised. This is particularly evident for the EBLV-2 viruses due to the small number involved. Unfortunately, differences in virus handling (multiple passages in vivo and in vitro), various sequencing approaches (direct sequencing from amplicons, sequencing from cloned products, deep sequencing) and improvements in the proof reading fidelity of molecular enzymes

have resulted in anomalous published sequences being submitted for the same isolate (McElhinney et al., 2013). To address this, a network of rabies reference laboratories involved in EBLV-2 surveillance and diagnosis was established, to review, collate and analyse all available EBLV-2 submissions.

Methods

A literature review was undertaken of all reported cases of EBLV-2, submission details, Rabies Bulletin Europe Reports, reference laboratory archives and published sequences. When available and where gaps existed, original material was sought to clarify anomalies in published data and obtain complete genome sequences (Table 1).

Next generation sequencing

Total RNA was depleted of host genomic DNA (gDNA) and ribosomal RNA (rRNA) following methods described previously (Marston et al., 2013; Brunker et al., 2015). Briefly, gDNA was depleted using the on-column DNase digestion protocol in RNeasy plus mini kit (Qiagen) following manufacturer's instructions, eluting in 30µl molecular grade water. Subsequently, rRNA was depleted, using Terminator 5'-phosphate-dependent exonuclease (Epicentre Biotechnologies). Briefly, 30µl of gDNA depleted RNA was mixed with 3µl of Buffer A, 0.5µl of RNAsin Ribonuclease inhibitor (20-40 U/µl) and incubated at 30^{o}C for 60 minutes. The depleted RNA was purified to remove the enzyme using the Rneasy plus mini kit as above, without the Dnase digestion, eluting in 30µl of molecular grade water. Double stranded cDNA (ds-cDNA) was synthesised using random hexamers and a cDNA synthesis kit (Roche) following manufacturer's instructions. The resulting ds-cDNA was purified using AMPure XP magnetic beads (Beckman Coulter), quantified using Quantifluor (Promega) and approximately 1ng of each sample was used in a 'tagmentation' reaction mix using a Nextera XT DNA sample preparation kit (Illumina) following manufacturer's instructions – without the bead normalization step. DNA libraries were quantified using Quantifluor (Promega). Individual libraries were pooled (approximately 8-12 samples per MiSeq run) and normalized to equimolar concentrations. Libraries were sequenced as 2 x 150bp paired-end reads on an Illumina MiSeq platform.

Mapping to reference sequence to obtain complete genome sequences

Reads were mapped using the Burrow-Wheeler Aligner (BWA version 0.7.5a-r405) and were visualized in Tablet (Milne et al., 2013). A modified SAMtools/vcfutils (Li et al., 2009) script was used to generate an intermediate consensus sequence in which any indels and SNPs relative to the original reference sequence were appropriately called. The intermediate consensus sequence was used as the reference for four subsequent iterations of mapping and consensus calling. Sequencing resulted in 100% coverage of the five genes for all samples. Consensus sequences were submitted to Genbank (please see Table for accession numbers)

Evolutionary Analysis

Sequences generated in this study were combined with published EBLV-2 sequences and aligned using ClustalX2 (version 1.2). Nucleotide and amino acid identities were obtained using MegAlign, v11.1.0 (Lasergene)

A phylogenetic analysis of all 29 EBLV-2 concatenated coding gene sequences (Table 1) was undertaken using Bayesian Markov Chain Monte Carlo simulation in the BEAST package v1.8.0 (Lemey et al., 2009). A GTR nucleotide substitution model with rate variation and a proportion of invariant sites (Gamma) were determined to best fit the data using Akaike Information Criterion (AIC) in MEGA 6.0 (Tamura et al., 2013). A relaxed molecular clock model with constant population prior were used in Markov Chain Monte Carlo (MCMC) simulation for 100,000,000 iterations, sampling every 10,000 states to give effective sample sizes of over 200. Molecular clock and population coalescent models were compared using a modified AIC (AICM) in Tracer v1.5, as described previously (Baele et al., 2013), and maximum clade credibility trees were annotated using TreeAnnotator (v1.8.4) after 10% of trees (burn-in) were discarded. The maximum clade credibility (MCC) trees were then visualised using Fig Tree (v1.4.3). Maximum likelihood phylogenetic

reconstructions were also undertaken on the EBLV-2 dataset (n=29) in MEGA6 for each individual gene and for partial N gene sequences (400 nucleotides).

Results and discussion

Review of reported cases

EBLV-2 was first isolated in 1985 from a Swiss bat biologist who had been working with bats in Finland, Switzerland and Malaysia. A year later, EBLV-2 was isolated in Switzerland from *Myotis daubentonii*. Since 1985 a total of 33 clinical rabies cases are believed to have been associated with EBLV-2 (Table 1). Full genome sequences have been previously published or obtained in this study for 29 EBLV-2. In addition, EBLV-2 RNA has been detected by RT-PCR in oro-pharyngeal swabs of two apparently healthy *M. daubentonii* bats during active surveillance (Scotland and Switzerland), yielding partial N gene sequences. Assuming that rabies in Daubenton's and Pond bats is restricted to EBLV-2 infection there were three further possible EBLV-2 cases in Denmark from single *Myotis dasycneme* (1986, 1987) and *M. daubentonii* (1986) bats (King et al., 1990). Also in Germany, bat rabies was diagnosed in two *M. daubentonii* in 1986, however in none of those cases is material available for confirmation of an EBLV-2 infection.

In the UK, thirteen cases of EBLV-2 have been reported in *M. daubentonii* and a single case in a male in Scotland (Fooks et al., 2003). In addition, EBLV-2 viral RNA was detected in a single oro-pharyngeal swab specimen collected from a healthy Daubenton's bat sampled as part of an active surveillance programme in Scotland in 2008 (McElhinney et al., 2013).

In Finland, the discovery of their first EBLV-2 in *M. daubentonii* (2009) provided evidence for the origin of the human infection in 1985 (Jakava-Viljanen et al., 2010). Prior to this latter case, passive surveillance had failed to detect EBLV-2 in Finland despite intensive efforts in 1986 and continued surveillance since (albeit for low submissions averaging 10/year). Subsequent active surveillance has failed to detect viral RNA in the oro-pharyngeal swabs of healthy Daubenton's bats. A second case of EBLV-2 in *M.daubentonii* has more recently been reported in Inkoo, Southern Finland (October 2016) but this case was unavailable for this study.

In Germany, despite extensive surveillance programmes, only five EBLV-2 cases have been confirmed as EBLV-2 (Freuling et al., 2008, Schatz et al., 2014), all of which are represented in this study.

Whilst EBLV-2 was detected in the Netherlands, passive surveillance has not yielded any rabies cases since 1997 (van der Poel et al., 2005). All five cases of EBLV-2 were identified in Myotis dasycneme. However, sequences are only available for four isolates (Table 1). No other M. dascyneme isolates have been identified elsewhere and no cases of EBLV-2 have been detected in any of the 111 *M. daubentonii* tested in the Netherlands to date (van der Poel., 2005).

In Switzerland, passive surveillance, between 1976 and 2009 revealed only three cases of EBLV-2, all in M.daubentonii and a single PCR positive oro-pharyngeal swab sample via active surveillance. The three Swiss EBLV-2 isolates identified through passive surveillance have previously been analysed (Amengual et al., 1997, Davis et al., 2005, Megali et al., 2010). In Megali et al., 2010, the 1992 and 1993 isolates have 100% identity in the 220bp region studied. However, the 1992 (RV594) and 1993 (RV621) Swiss isolates received at APHA in the early 1990s, differed in the 400bp sequence with 97% sequence identity (McElhinney et al., 2013). To add to the confusion, the 1992 isolate held at APHA (RV594) showed 100% identity to the published Swiss 2002 sequence (AY863408) (Figure 2). In addition, it is unclear if the partial N gene sequence published as U89479 represents an additional Swiss virus or sequencing error? Original virus isolates for the Swiss EBLV-2 held at APHA and University of Berne were compared using NGS and the correct genome sequences confirmed.

In 2015, Norway reported its first case of EBLV-2 in a Daubenton's bat. The bat was found by a member of the public in Oppland, in the southern part of Norway.

Phylogenetic analysis

The phylogenetic analysis confirmed the closer relationship of EBLV-2 sequences to the Phylogroup I lyssaviruses, in particular Khujand and BBLV, irrespective of the gene studied. Where available, a number of representative lineages for the lyssaviruses were included (Figure 2). The genetic diversity observed for the other lyssavirus species is in contrast to the highly homogeneous sequence identities observed for EBLV-2. However, despite this homogeneity, the EBLV-2 are resolved geographically for all analysis, with clear clustering of the EBLV-2 at the country level. Even within country resolution can be observed when sufficient numbers allow. The Scotland and Northern England EBLV-2 isolates can be seen to cluster separately to the southern England and Welsh isolates, although bootstrap support is not consistent (Figure 2 and 3). The human EBLV case (RV1333) in Scotland in 2002, resolved closest to the Northern England and Scotland cases. In the partial N400 analysis (not shown), human UK EBLV-2 was mst similar to the partial sequence obtained from a Daubenton's saliva sample during active surveillance (McElhinney et al., 2013). Although originating from disparate regions in the South and the North of Germany, the German EBLV-2 viruses cluster together and are most closely associated with the 2002 Swiss isolate (TW0118/02) (Figure 2). All of the Dutch EBLV-2 cluster separately from the other sequences possibly reflecting either geographical and/or host species relationships. The latter may be likely as the EBLV-2 isolates from the Netherlands are the only ones sampled from *Myotis dasycneme*. The Norwegian EBLV-2 is distinct but is most closely related to the Swiss EBLV-2.

The two available Finnish EBLV-2 sequences are clearly distinct from the other EBLV-2 studied and have previously been proposed as a separate lineage, however depending upon the gene studied, the bootstrap support is not always consistent. However, the relatively low number of available EBLV-2 viral sequences makes it difficult to confidently determine how this lyssavirus species can be further divided into monophyletic sub lineages. Although, the subtypes 2a and 2b of EBLV-2 were proposed in earlier studies (Amengual et al., 1997), it has more recently been reported that the highest nucleotide sequence divergence was between two isolates (AY863408 – Switzerland isolate and AY863406 – Finland human case), previously assigned as EBLV-2b (Marston et al., 2007). The two Finnish EBLV-2 sequences and U89479 consistently formed a well-supported sub lineage (EBLV-2b) irrespective of the model used (bootstrap support 100% for Neighbour Joining (p-values) and 99% for Maximum Likelihood (GTR) models). In contrast, the 1992 & 1993 Swiss EBLV-2 sequences were not consistently associated with either sub-lineage when individual sequences were removed or additional lyssavirus sequences added (data not shown).

Intra-genotypic comparisons were reported previously for the five viral proteins of EBLV-2 (Marston et al., 2007) and demonstrated the limited genetic diversity of these lyssaviruses compared to RABV at both the nucleotides and amino acid levels. In this analysis using concatenated gene sequences, the similarity of the EBLV-2 ranged from 95.5-100% for nucleotides and 98.7-100% for amino acids.

Inter-host variation

The three EBLV-2 isolates from Daubenton's bats found at Stokesay castle (Shropshire UK) were studied in greater detail as they represented the only cluster of EBLV-2 with three separate isolations at the same site in 2007, 2008 and 2014. The percentage of viral reads assembled from the total reads varied between isolates, ranging from 0.98% (RV2473) to 2.08% (RV2336). Depth of coverage (average and maximum) was variable and correlated with the number of viral reads assembled.

The three genomes were aligned and found to be 99.9% identical to one another. Positions at which single nucleotide polymorphisms (SNPs) occurred were identified and investigated to determine whether they denote a change in amino acid within a region of the genome coding for a viral protein (Figure 4). One such SNP was identified within the P gene, at position 2015. RV2336 and RV2974 code for lysine at this position, whereas the change in base from A to G in RV2473 results in the translation of arginine instead (supplementary table S2). This SNP is located within the

LC8 domain (cytoplasmic dynein light chain) at residue 164. The LC8 domain has been previously identified as a region at which P protein interacts with LC8, a cytoplasmic protein associated with axonal transport of RABV within neurons (Marston et al. 2007, Lo et al. 2001), however, the SNP is not within the key residues of 145-149, the motif to which LC8 binds. In addition, the substitution is considered conservative due to the similar biochemical properties of the amino acids.

Base substitutions were identified at three other positions, however these are within non-coding regions of the genome, therefore do not affect the viral proteins. This is not to say that mutations in non-coding regions are unimportant, as they may play a part in areas such as viral replication or structure.

Minor differences in genome sequence exist between the three isolates from Stokesay castle. To investigate these differences further, we looked at the raw sequence reads at the positions where SNPs were shown, to identify whether or not viral variants were observed that differed from the consensus sequence at each given position.

In the 2007 isolate (RV2336), at the positions at which SNPs are seen, no variants exceed 1% of the total viral reads, however a number of variants are seen, four of which result in an amino acid change, all of which are non-conservative changes (supplementary table xx). At position 2423, within the P-M intergenic regions, the 2008 isolate (RV2473) shows a variant (T) that occurs in 1.01% of the total viral reads. None of the RV2473 variants shown in supplementary table xx result in an amino acid change. Three positions within the 2014 isolate (RV2974) show variants that exceed 1% of the total viral reads; 1.01% of reads at position 1566 (P gene) are A, 2.26% of reads at position 1752 (P gene) are A, and 1.05% of reads at position 7355 (L gene) are T (supplementary table xx). Six variants observed result in amino acid changes; three of which are conservative, two of which are non-conservation and one of which results in a stop codon (position 9374).

To investigate whether the variants seen at >1% of the total reads were observable at lower proportions previously, or if they became fixed over time, pie charts were created to show the proportion of each base over time, at each of the four positions where a >1% minority variant is seen (figure 5). Position 1566 (P gene) is 99.7% C with a minority population of A in 2007. This changes dramatically by 2008, in which 100% of the reads are T at this position. In 2014, T remains the major variant, with two minor variants of A (>1%) and C (0.09%). Position 1752 (P gene) is 99.7% A with a minority population of G in 2007. In 2008 the minority population has disappeared and all reads show an A at this position. By 2014, the majority population has changed and 97.1% of the reads show a G, with a smaller remaining population of A (2.26%) and a minority population of T (0.6%). Position 2423 (P-M intergenic region) shows no variation in 2007; 100% of reads are C. In 2008, a minority T variant appears (1.05%), which has become fixed at this position by 2014 (99.84%). No C remains in 2014, however 0.16% of reads show A. Position 7355 (L gene) is 99.82% T in 2007, with a minority population of A. In 2008, 100% of reads have a C at this position, and by 2014 the majority population of C remains, with two minor variants of T (1.05%) and A (0.45%).

Intra-host variation

To investigate heterogeneity of the 3 Stokesay Castle isolates across the length of each genome, and not just at the positions at which SNPs were identified, each position in the genome was analysed to determine the number of reads showing each base A, T, C or G. The proportion of reads showing each base was calculated and plotted to give a graphical representation of heterogeneity across the genome for each of the three EBLV-2 isolates (supplementary figure xx). Data was filtered to remove positions that showed less than one hundred reads, and less than or equal to 1% heterogeneity. Across the length of the genome, RV2336 has 23 positions at which heterogeneity equals or exceeds 1%, RV2473 has 257 positions and RV2974 has 212. The similar level of heterogeneity between RV2473 and RV2974 gives us confidence that this data is not merely a reflection of the varying depth of coverage between the genomes; RV2473 has the lowest average depth of coverage (245 reads) whilst RV2974 has the highest (970 reads). There are three positions at which heterogeneity equals or exceeds 2% in RV2336, 50 positions in RV2473 and 86 positions in RV2974.

Evolutionary analysis of EBLV-2

The evolutionary analysis in this study supports the relatively low evolutionary rates for EBLV-2 compared to other RNA viruses (Jakava-Viljanen et al., 2015, Davis et al., 2005). The long TMRCAs observed in Figure 3 is suggestive of minimal and discrete evolution at the local level. Migration of infected bats, whilst possible does not appear to be supported in our dataset. Mean rate of evolution for the concatenated genes for the EBLV-2 dataset (n=29) was 1.7E-5 which is within the ranges observed previously (Davis et al., 2005,

Conclusions

Unlike EBLV-1, which has been detected during active surveillance in bat species other than Serotine bats (Schatz et al., 2013), the chiropteran host range of EBLV-2 appears to be restricted to just two bat species *Myotis Daubentonii* (n=27) and *Myotis dascyneme* (n=7) (Table 1). *The distribution of M. dasycneme* is confined to Central and Eastern Europe, whereas *M. daubentonii* is widely distributed throughout Europe (Stebbings and Griffith, 1986). *Myotis dascyneme* is considered to be absent in the UK. The genetic structure of *M. daubentonii* in Western Europe is relatively homogenous and data suggests more migratory movements between the UK and the mainland by Daubenton's bats than Serotine bats (Smith et al., 2011). Many factors affecting the behaviour of Daubenton's bats, may allow EBLV-2 to become established and rapidly spread throughout the population. This may account for the detection of EBLV-2, albeit at low levels, throughout the UK (Harris et al., 2009).

Cross species transmission (CST) of bat lyssaviruses to other mammals has been documented but is extremely rare (Dacheux et al., 2009). Two CST events involving EBLV-2 from bats to humans have been reported, in Finland in 1985 and in the UK in 2002 (Table 1). Unlike for EBLV-1, EBLV-2 spillover into other wildlife or domestic species has not been documented.

In recent years, multiple novel lyssaviruses have been isolated from bats: ABLV, ARAV, KHUV, IRKV, SHIBV, WCBV, BBLV, LLBV and GBLV. The genetic relationships of these novel viruses to known lyssaviruses, including RABV, have been studied with particular attention paid to the respective Phylogroups. The bat lyssaviruses ABLV, ARAV, BBLV, DUVV, EBLV-1, EBLV-2, GBLV, KHUV, IRKV and LLBV accompany the RABV bat variants within phylogroup I. The African bat lyssaviruses LBV and SHIBV are Phylogroup II viruses whilst the more divergent bat lyssaviruses WCBV and LLBV represent a third phylogroup (Kuzmin et al., 2010; Banyard et al., 2011). No phylogroup II viruses have been detected in indigenous European bat species.

All current human and animal vaccines are based on RABV. Predicting antigenic relatedness of the lyssaviruses to RABV, particularly in the glycoprotein gene, is important for predicting vaccine protection (Badrane et al., 2001, Wright et al., 2010; Horton et al., 2010). *In-vivo* protection experiments have shown vaccines offer the best protection against viruses in phylogroup I, limited protection against Phylogroup II lyssaviruses and no protection against Phylogroup III lyssaviruses (Brookes et al., 2005; Brookes et al., 2006; Hanlon et al., 2005). However, cross protection studies are recommended for newly identified novel lyssaviruses as sequence identity alone can be a poor predictor of antigenic relatedness, due to the disproportionate effect of some amino acid substitutions (Horton et al., 2010).

Passive bat surveillance, complemented by active surveillance where appropriate, will facilitate the further isolation of novel bat lyssaviruses. Such surveillance must be directed in collaboration with the relevant local bat conservation authorities using internationally recognised procedures. This will facilitate comparative analysis of these surveillance data (particularly with respect to serology) and the continued conservation of bats whilst establishing the cooperation of the national and international groups of bat specialists with respect to bat population, ecology and speciation. To assess the diversity of lyssaviruses in Europe, all isolates from bat rabies cases should be archived and further characterized. Molecular typing of wildlife rabies cases will facilitate the early detection of CST events which could give rise to the adaption and maintenance of bat lyssaviruses in non-flying mammals.

To ensure that a true representation of the epidemiology and transmission dynamics of bat lyssaviruses is determined and given the presence of five bat lyssavirus species and at least 52 species of bats (www.eurobats.org) in Europe, it is extremely important to identify both the bat species and the viral species involved in any suspected case of bat rabies. In the past, misrepresentation of the role of bats as reservoirs and vectors of lyssaviruses have on occasion resulted in attempts to exclude or cull bat colonies, despite their protected status in some countries. In addition, molecular typing of viral strains can exclude laboratory contamination in cases where new host species are implicated in EBLV transmission (Picard-Meyer et al., 2004). Bats which are difficult to speciate phenotypically can be identified by molecular typing (Harris et al., 2008).

Acknowledgements

This work was supported by grants from Defra, the Scottish Government and Welsh Government through grants SE0431 and SV3500, by the EU Horizon 2020 research and innovation programme under grant agreement No 653316, and the German Ministry for Education and Research (grant 01KI1016A). The authors acknowledge support from the Research and Policy for Infectious Disease Dynamics program of the Science and Technology Directorate, Department of Homeland Security USA. The authors would also like to acknowledge the excellent technical support of staff within the Wildlife Zoonoses and Vector Borne Diseases Research Group at APHA.

References

Arechiga Ceballos N, Vazquez Moron S, Berciano JM, Nicolas O, Aznar Lopez C, Juste J, Rodriguez Nevado C, Aguilar Setien A, Echevarria JE. 2013. Novel lyssavirus in bat, Spain. Emerging infectious diseases 19:793-795.

Badrane H, and Tordo N. Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders. J. Virol. 2001; 75:8096–8104.

Badrane H, Bahloul C, Perrin P, Tordo N. Evidence of two lyssavirus phylogroups with distinct pathogenicity and immunogenicity.. J. Virol., 2001; 75 pp. 3268–3276.

Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P (2013) Accurate Model Selection of Relaxed Molecular Clocks in Bayesian Phylogenetics. Mol Biol Evol 30: 239–243. Doi: 10.1093/molbev/mss243

Banyard AC, Johnson N, Voller K., Hicks D, Nunez A, Hartley M, Fooks AR. Repeated detection of European bat lyssavirus type 2 in dead bats found at a single roost site in the UK. Arch. Virol. 2009; 154:1847–1850.

Banyard AC, Hayman D., Johnson N, McElhinney LM, Fooks AR. Bats and Lyssaviruses. Adv. Virus Res. 2011; 79; 239 — 89.

Botvinkin et al 2003

Bourhy H, Kissi B, Lafon M, Sacramento D, Tordo N. Antigenic and molecular characterization of bat rabies virus in Europe. J. Clin. Microbiol. 1992; Sep;30(9):2419-26.

Brookes SM, Aegerter JN, Smith GC, Healy DM, Jolliffe T, Swift SM, Mackie I, Pritchard JS, Racey PA, Moore NP, Fooks AR. Prevalence of antibodies to European Bat Lyssavirus type-2 in Scottish bats. Emerg. Infect. Dis. 2005; 11(4); 572 – 578.

Brookes SM, Healy DM, Fooks AR. Ability Of Rabies Vaccine Strains To Elicit Cross-Neutralising Antibodies. Dev Biol (Basel) (eds: Dodet, B, Schudel A, Pastoret P.P, Lombard M) Basel, Karger, 2006; 125; 185-193.

Brunker K, Marston DA, Horton DL, Cleaveland S, Fooks AR, Kazwala R, Ngeleja C, Lembo T, Sambo M, Mtema ZJ.. Elucidating the phylodynamics of endemic rabies virus in eastern Africa using whole-genome sequencing. Virus Evolution 2015, 1:vev011.

Dacheux L, Larrous F, Mailles A, Boisseleau D, Delmas O, Biron C, Bouchier C, Capek I, Muller M, Ilari F, Lefranc T, Raff, F, Gouda M, Bourhy H. European bat lyssavirus transmission among cats, Europe. Emerg. Infect. Dis. 2009; 15:280–284.

Davis PL, Holmes EC, Larrous F, Van der Poel WH, Tjornehoj K, Alonso WJ, Bourhy H: Phylogeography, population dynamics, and molecular evolution of European bat lyssaviruses. J. Virol. 2005; 79:10487–10497.

Echevarría JE et al 2013

Fooks, AR, Brookes SM, Johnson N, McElhinney LM Hutson AM. European bat lyssaviruses: an emerging zoonosis. Epidemiol. Infect. 2003a; 131(3); 1029-1039.

Fooks AR, McElhinney L M, Pounder DJ, Finnegan CJ, Mansfield K, Johnson N, Brookes SM, Parsons G, White K, McIntyre PG, Nathwani D. Case report: Isolation of a European bat lyssavirus type 2a from a fatal human case of rabies encephalitis. J. Med. Virol. 2003b; 71:281–289.

Fooks, AR: The challenge of Emerging *Lyssaviruses.* Expert Rev. Vaccines 2004; 3(4):89-92.

Fooks AR, Selden D, Brookes SM, Johnson N, Marston DA, Jolliffe TA, Wakeley PR, McElhinney LM. Identification of a European bat lyssavirus type 2 in a Daubenton's bat found in Lancashire. Vet. Rec. 2004; Nov 6;155(19):606-7.

Fooks AR, McElhinney LM, Marston DA, Selden D, Jolliffe TA, Wakeley PR, Johnson N, Brookes SM. Identification of a European bat lyssavirus type 2 in a Daubenton's bat found in Staines, Surrey, UK. Vet. Rec. 2004 Oct 2;155(14):434-5.

Fooks AR, Marston D, Parsons G, Earl D, Dicker A, Brookes SM. Isolation of EBLV-2 in a Daubenton's bat (*Myotis daubentonii*) found in Oxfordshire. Vet. Rec. 2006 Oct 14;159(16):534-5.

Freuling CM, Beer M, Conraths FJ, Finke S, Hoffmann B, Keller B, Kliemt J, Mettenleiter TC, Mühlbach E, Teifke JP, Wohlsein P, Müller T. Novel lyssavirus in Natterer's bat, Germany. Emerg. Infect. Dis. 2011; Aug;17(8):1519-22.

Freuling C, Grossmann E, Conraths FJ, Schameitat A, Kliemt J, Auer E, Greiser-Wilke I, Muller T. (2008) First isolation of EBLV-2 in Germany. Vet. Microbiol. 131 (1-2), 26-34.

Freuling CM, Kliemt J, Schares S, Heidecke D, Driechciarz R, Schatz J, Müller T. Detection of European bat lyssavirus 2 (EBLV-2) in a Daubenton's bat (Myotis daubentonii) from Magdeburg, Germany. Berl Munch Tierarztl Wochenschr. 2012 May-Jun;125(5-6):255-8.

Grauballe PC, Baagøe HJ, Fekadu M, Westergaard JM, Zoffmann H. Bat rabies in Denmark. Lancet. 1987 Feb 14;1(8529):379-80.

Gunawardena PS, Marston DA, Ellis RJ, Wise EL, Karawita AC, Breed AC, McElhinney LM, Johnson N, Banyard AC, Fooks AR. 2016. Lyssavirus in Indian flying foxes, Sri Lanka. Emerg Infect Dis 22:1456 –1459.

Hanlon CA, Kuzmin IV, Blanton JD, Weldon WC, Manangan JS, Rupprecht CE. Efficacy of rabies biologics against new lyssaviruses from Eurasia. Virus Res. 2005; Jul;111(1):44-54.

Harris SL, Mansfield K, Marston DA, Johnson N, Pajamo K, O'Brien N, Black C, McElhinney LM, Fooks AR. Isolation of European bat lyssavirus type 2 from a Daubenton's bat (*Myotis daubentonii*) in Shropshire. Vet. Rec. 2007; Sep 15;161(11):384-6.

Harris SL, Johnson N, Brookes SM, Hutson AM, Fooks AR, Jones G. The application of genetic markers for EBLV surveillance in European bat species. Dev Biol (Basel). 2008;131:347-63.

Harris SL, Aegerter JN, Brookes SM, McElhinney LM, Jones G, Smith GC, Fooks AR. Targeted surveillance for European bat lyssaviruses in English bats (2003-06). J. Wildl. Dis. 2009 Oct;45(4):1030-41.

Horton DL, Voller K, Haxton B, Johnson N, Leech S, Goddard T, Wilson C, McElhinney LM, Fooks AR. European bat lyssavirus type 2 in a Daubenton's bat in Scotland. Vet. Rec. 2009; Sep 26;165(13):383-4.

Horton DL, McElhinney LM, Marston DA, Wood JL, Russell CA, Lewis N, Kuzmin IV, Fouchier RAM, Osterhaus ADME, Fooks AR, Smith DJ. Quantifying antigenic relationships among the lyssaviruses. J. Virol. 2010;84(22):11841–8.

ICTV official taxonomy: updates http://talk.Ictvonline.Org/media/p/1208.Aspx.

Jakava-Viljanen M, Lilley T, Kyheroinen EM, and Huovilainen A. First encounter of European bat lyssavirus type 2 (EBLV-2) in a bat in Finland. Epidemiol. Infect. 2010 138:1581–1585.

Jakava-Viljanen M, Nokireki T, Sironen T, Vapalahti O, Sihvonen L, Huovilainen A. Evolutionary trends of European bat lyssavirus type 2 including genetic characterization of Finnish strains of human and bat origin 24 years apart. Arch Virol. 2015 Jun;160(6):1489-98

Johnson N, Selden D, Parsons G, Healy D, Brookes SM, McElhinney LM, Hutson AM, Fooks AR. Isolation of a European bat lyssavirus type 2 from a Daubenton's bat in the United Kingdom. Vet. Rec. 2003;152:383–387.

Johnson N, Goddard TM, Goharriz H, Wise E, Jennings D, Selden D, Marston DA, Banyard AC, McElhinney LM, Fooks AR. Two EBLV-2 infected Daubenton's bats detected in the north of England. Vet Rec. 2016 Sep 24;179(12):311-2

King AA, Davies P, Lawrie A. The rabies viruses of bats. Vet. Microbiol. 1990; 23, 165-74

Kuzmin IV, Mayer AE, Niezgoda M, Markotter W, Agwanda B, Breiman RF, Rupprecht C.E. Shimoni bat virus, a new representative of the Lyssavirus genus. Virus Res. 2010;149 (2), pp. 197-210

Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. PloS Comput Biol 5: e1000520. Doi: 10.1371/journal.pcbi.1000520 PMID: 19779555

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S.. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009 25:2078-2079.

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

Lumio J, Hillbom M, Roine R, Ketonen L, Haltia M, Valle M, Neuvonen E, Lahdevirta J. Human rabies of bat origin in Europe. Lancet 1986; 1:378.

McElhinney LM, Marston DA, Leech S, Freuling CM, van der Poel WH, Echevarria J, Vázquez-Moron S, Horton DL, Müller T, Fooks AR. Molecular epidemiology of bat lyssaviruses in Europe Zoonoses Public Health. 2013 Feb;60(1):35-45. Doi: 10.1111/zph.12003.

Marston D, McElhinney LM, Johnson N, Müller T, Conzelmann KK, Tordo N, Fooks AR Comparative analysis of the full genome sequence for European Bat Lyssavirus type-1 and type-2 with other lyssaviruses and evidence for a conserved transcription termination and polyadenylation motif in the G-L 3'-nontranslated region. J. Gen. Virol. 2007; 88; 1302 -1314.

Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR. 2013. Next generation sequencing of viral RNA genomes. BMC genomics 14:444.

Megali A, Yannic G, Zahno ML, Brügger D, Bertoni G, Christe P, Zanoni R. Surveillance for European bat lyssavirus in Swiss bats. Arch. Virol. 2010; Oct;155(10):1655-62.

Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D.. Using Tablet for visual exploration of second-generation sequencing data. Briefings in bioinformatics 2013, 14:193-202.

Mohr W. Die tollwut. Med Klinik 1957; 52:1057–1060.

Nieuwenhuijs J, Haagsma J, Lina P. Epidemiology and control of rabies in bats in The Netherlands. Rev Sci Tech. 1992 Dec;11(4):1155-61.

Pajamo K, Harkess G, Goddard T, Marston D, McElhinney L, Johnson N, Fooks A. Isolation of European bat lyssavirus type 2 (EBLV-2) in a Daubenton's bat in the UK with a minimum incubation period of 9 Months. Rabies Bulletin Europe 2008-2.

Picard-Meyer E, Barrat J, Wasniewski M, Wandeler A, Nadin-Davis S, Lowings JP, Fooks AR, McElhinney L, Bruyère V, Cliquet F. Epidemiology of rabid bats in France, 1989 to 2002. Vet. Rec. 2004; Dec 11;155(24):774-7.

Picard-Meyer E, Servat A, Robardet E, Moinet M, Borel C, Cliquet F. Isolation of Bokeloh bat lyssavirus in Myotis nattereri in France. Arch Virol. 2013 Nov;158(11):2333-40

Schatz J, Fooks AR, McElhinney L, Horton D, Echevarria J, Vázquez-Moron S, Kooi EA, Rasmussen TB, Müller T, Freuling CM. Bat rabies surveillance in Europe. Zoonoses Public Health. 2013 Feb; 60(1):22-34

Schatz J, Freuling CM, Auer E, Goharriz H, Harbusch C, Johnson N, Kaipf I, Mettenleiter TC, Mühldorfer K, Mühle RU, Ohlendorf B, Pott-Dörfer B, Prüger J, Ali HS, Stiefel D, Teubner J, Ulrich RG, Wibbelt G, Müller T. Enhanced passive bat rabies surveillance in indigenous bat species from Germany—a retrospective study. PloS Negl Trop Dis. 2014 May 1;8(5):e2835.

Schindler, R., Denning H. K. 1958. Unter-suchungen über die Bedeutung der Fledermäusefür den gegenwärtigen Tollwut-Seuchenzug inDeutschland. Monatsschrift für Tierheilkunde,10: 169–177.

Stebbings RE, Griffith F.1986 Distribution and status of bats in Europe 1. Bats I. Title II. Griffith, F. Ill. Institute of Terrestrial Ecology 599.4 OL737. C5.

Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science. 2010; Aug 6;329(5992):676-9.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30: 2725–2729. Doi: 10.1093/molbev/mst197 PMID: 24132122

Van der Poel WHM, Van der Heide R, Verstraten ER, Takumi K, Lina PHC, Kramps JA: European bat lyssaviruses, The Netherlands. Emerg. Infect. D.is 2005; 11:1854-1859.

Vos A, Kaipf I, Denzinger A, Fooks AR, Johnson N, Müller T. European Bat Lyssaviruses – an ecological enigma. Acta Chiropterologica 2007; 9(1); 283-296.

Whitby JE, Heaton PM, Black EM, Wooldridge M, McElhinney LM, Johnstone P. First isolation of a rabies-related virus from a Daubenton's bat in the United Kingdom. Vet. Rec. 2000; 147, 385–388.

Wright E, Hayman DT, Vaughan A, Temperton NJ, Wood JL, Cunningham AA, Suu-Ire R, Weiss RA, Fooks AR. Virus neutralising activity of African fruit bat (Eidolon helvum) sera against emerging lyssaviruses. Virology 2010; 408(2); 183-9.

Figure 1 Distribution of cases of EBLV-2 in Europe

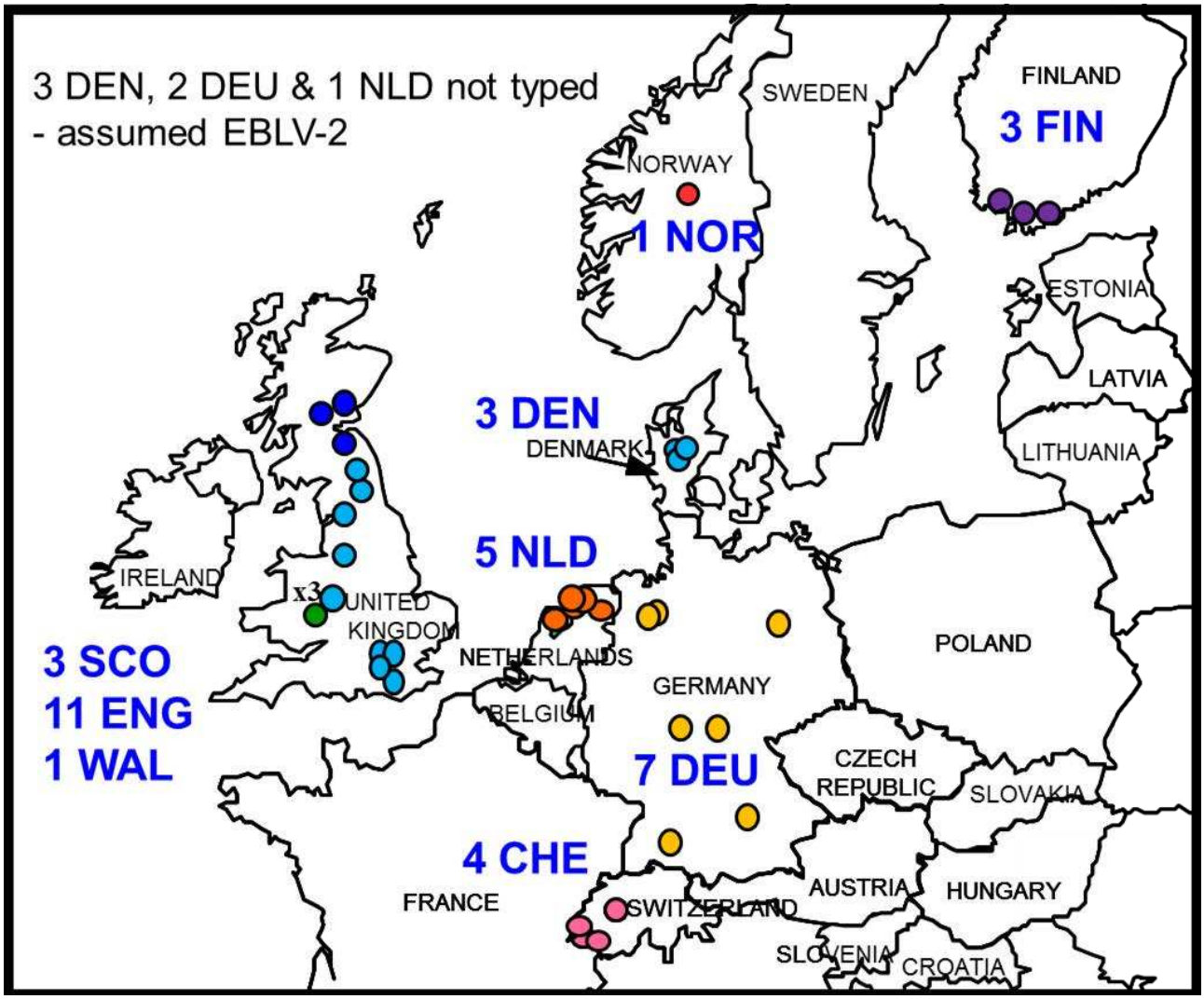


Table 1 Reported cases of EBLV-2 in Europe 1985-2016 and sequences used to generate phylogenetic trees.

Year	Submission Reference	APHA Virus Ref	PI Ref	Location	Country	Species	N400 data Accession No.	Genome Accession No.	Case/Sequence Ref
1985	FINMAN	RV8	9007FIN	Helsinki	Finland	Human	AY062091	JX129233 KY688151	Lumio et al., 1986
1986	??	na	na	Lower Saxony	Germany	*M.daub*	NA	NA	Rabies Bulletin Europe 86/3
1986	??	na	na	Lower Saxony	Germany	*M.daub*	NA	NA	Rabies Bulletin Europe 86/3
1986	?	na	na	Jutland	Denmark	*M.daub*	NA	NA	Grauballe et al., 1987
1986	?	na	na	Jutland	Denmark	*M. dasyc*	NA	NA	Grauballe et al., 1987
1987	?	na	na	Jutland	Denmark	*M. dasyc*	NA	NA	Rabies Bulletin Europe 87/4
1987	41729	RV30	9018HOL	Wommels	Netherlands	*M. dasyc*	EU293114	EU293114	Davis et al., 2005
1987	47072	RV29	9482HOL	Tjerkwerd	Netherlands	*M. dasyc*	U89480	KY688145	Rabies Bulletin Europe 87/4
1987	?	na	na	Tjerkgaast	Netherlands	*M. dasyc*	NA	NA	Nieuwenhuijs et al., 1992
1989	92666	RV228	94112HOL	Andijk	Netherlands	*M. dasyc*	AY062089	KY688146	Davis et al., 2005
1992	TW1814/92	RV594, RV2478	na	Plaffeien	Switzerland	*M. daub.*	AY212117	KY688133	Megali et al., 2010
1993	?	na	9375HOL	Roden	Netherlands	*M. dasyc*	AY863404	KY688152	Davis et al., 2005
1993	TW1392/93	RV621, RV2479	9337SWI	Versoix	Switzerland	*M. daub.*	AY212118	KY688140	Davis et al., 2005
1996	18/96	RV628	EBL2GB	Sussex	England	*M. daub.*	U89478	KY688136	Whitby et al., 2000
2002	105/02	RV1332	na	Lancashire	England	*M. daub.*	AY212120	KY688150	Johnson et al., 2003
2002	Human case	RV1333	na	Angus	Scotland	Human	EF157977	EF157977	Fooks et al., 2003
2002	TW0118/02	RV2480	02053SWI	Geneva	Switzerland	*M. daub.*	AY863408	KY688132	Davis et al., 2005
2004	603/04	RV1787	na	Surrey	England	*M. daub.*	JQ796807	KY688142	Fooks et al., 2004
2003	696/04	RV1788	na	Lancashire	England	*M. daub.*	JQ796808	KF155004	Fooks et al., 2004
2006	672/06	RV2159	na	Oxfordshire	England	*M. daub.*	JQ796809	KY688144	Fooks et al., 2006
2006	18856	RV2506	na	Magdeburg	Germany	*M. daub.*	JQ796805	KY688135	Freuling et al., 2012
2006	25538	na	na	Thuringia	Germany	*M.daub*	KF826115	KY688137	Schatz et al., 2014
2007	16618	RV2505	na	Bad Buchau	Germany	*M. daub.*	GU227648	KY688138	Freuling et al., 2008
2007	762/07	RV2336	na	Shropshire	England	*M. daub.*	JQ796810	KY688139	Harris et al., 2007
2008	166/08	RV2418	na	Surrey	England	*M. daub.*	JQ796811	KY688141	Pajamo et al., 2008
2008	1218/08	RV2473	na	Shropshire	England	*M. daub.*	JQ796812	KY688147	Banyard et al., 2009
2009	882/09	RV2482	na	West Lothian	Scotland	*M. daub.*	JQ796806	KY688143	Horton et al., 2009
2009	ra3278/09	na	na	Nr Turku	Finland	*M. daub.*	GU002399	JX129232	Jakava-Viljanen et al., 2010
2012	F500/LGL	na	na	Ingolstadt, Bavaria	Germany	*M. daub.*	NA	KY688149	Schatz et al., 2014

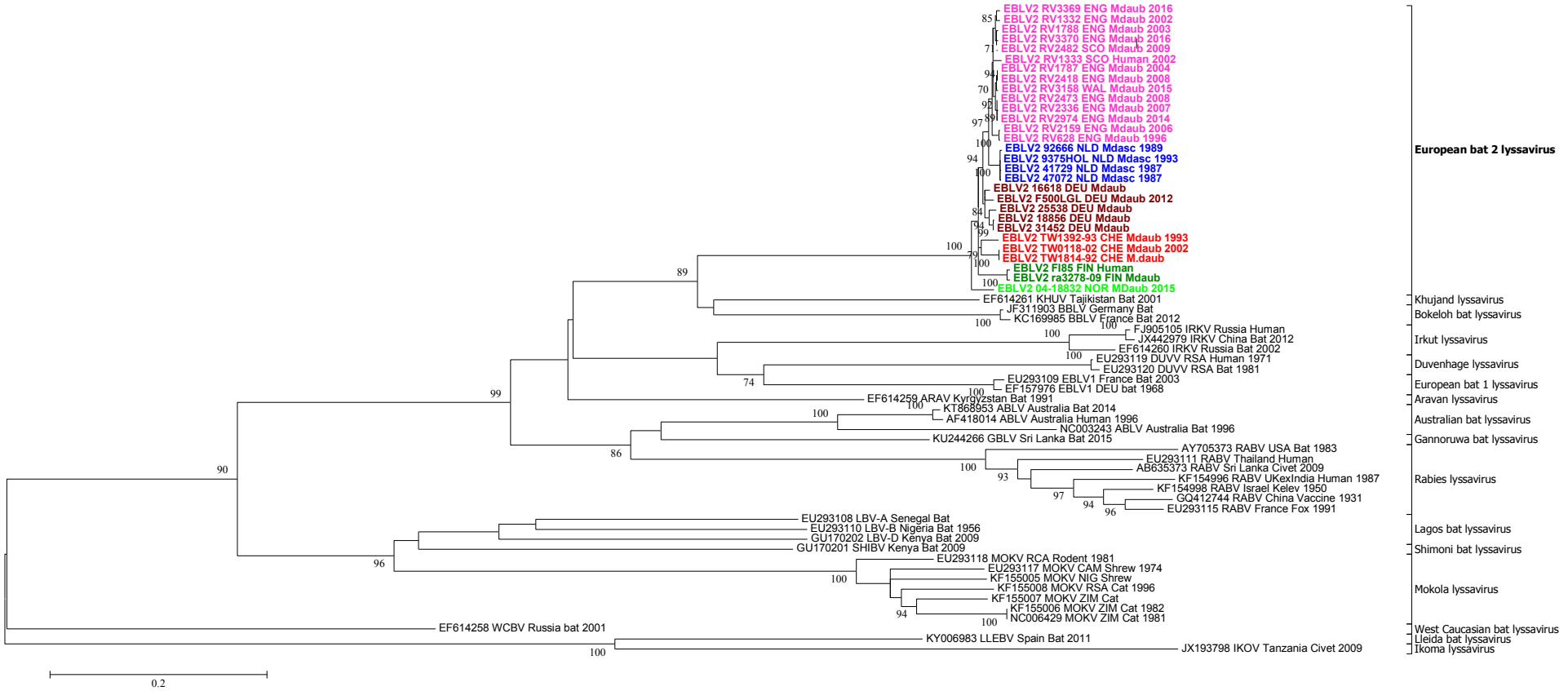
2013	31452	na	na	Gießen, Hesse	Germany	*M. daub.*	KF826149	KY688134	Schatz et al., 2014
2014	281/14	RV2974	na	Shropshire	England	*M. daub.*	NA	KY688148	This study
2015	421/15	RV3158		Newtown, Powys	Wales	*M. daub.*	NA	KY688153	This study
2015	2015-04-18832	na	na	Oppland	Norway	*M. daub.*	NA	KY688154	This study
2016	307/16	RV3369	na	North Yorkshire	England	*M. daub.*	NA	KY688156	This study
2016	432/16	RV3370	na	Northumberland	England	*M. daub.*	NA	KY688155	Johnson et al., 2016
2016	???	na	na	Inkoo, Uusimaa	Finland	*M. daub.*	NA	NA	This study
				Active Surveillance Detection of viral RNA in oro-pharyngeal swab only:					
2008	80063	M08/09	na	Perthshire	Scotland	*M. daub.*	JQ796804	NA	PCR only-McElhinney et al 2013
2009	70	na	na	Genthod	Switzerland	*M. daub.*	HM067110	NA	PCR only-Megali et al.,2010

Table 2 Details of non-EBLV-2 Lyssavirus sequences used in this study

	Species	Country	Host	Date	Acc No		Species	Country	Host	Date	Acc No
1	LBV A	Senegal	E helvum	1985	EU293108	19	IRKV	Russia	M leucigaster	2002	EF614260
2	LBV B	Nigeria	E helvum	1956	EU293110	20	IRKV	Russia	human	2007	FJ905105
3	LBV D	Kenya	R aegyptiacus	2009	GU170202	21	IRKV	China	M leucigaster	2012	JX442979
4	MOKV	RSA	Rodent	1981	EU293118	22	KHUV	Tajikistan	M mystacinus	2001	EF614261
5	MOKV	Cameroon	Shrew	1974	EU293117	23	WCBV	Russia	M schreibersii	2001	EF614258
6	MOKV	Zimbabwe	Cat	1981	NC_006429	24	SHIBV	Kenya	H commersoni	2009	GU170201
7	MOKV	Nigeria	Shrew	1968	KF155005	25	BBLV	Germany	M. naterreri	2010	JF311903
8	MOKV	Zimbabwe	Cat	1982	KF155006	26	BBLV	France	M. naterreri	2012	KC169985
9	MOKV	Zimbabwe	Cat	1993	KF155007	27	IKOV	Tanzania	African Civet	2009	JX193798
10	MOKV	RSA	Cat	1996	KF155008	28	GBLV	Sri Lanka	P.giganteus	2016	KU244266
11	DUVV	RSA	Miniopterus	1981	EU293120	29	Lleida	Spain	M. schreibersii	2016	KY006983
12	DUVV	RSA	Human	1971	EU293119	30	RABV	Sri Lanka	Civet	2009	AB635373
13	EBLV-1	Germany	E serotinus	1968	EF157976	31	RABV	USA	Bat	1983	AY705373
14	EBLV-1	France	E serotinus	2003	EU293109	32	RABV	Thailand	Human	1983	EU293111
15	ABLV	Australia	Pteropus	1996	NC_003243	33	RABV	France	Fox	1991	EU293115
16	ABLV	Australia	Human	1996	AF418014	34	RABV	China	Vaccine	1931	GQ412744
17	ABLV	Australia	P. alecto	2014	KT868953	35	RABV	UKexIndia	Human	1987	KF154996
18	ARAV	Kyrgyzstan	Myotis blythi	1991	EF614259	36	RABV	Israel	Kelev vaccine	1950	KF154998

Figure 2. Molecular phylogenetic analysis of (a) nucleoprotein sequences (1353 nucleotides) and (b) glycoprotein sequences (1569 nucleotides) from 29 EBLV-2 cases and 36 non EBLV-2 lyssavirus sequences. Sequences were analysed by the Maximum Likelihood method. Bootstrap values >70% are shown for each branch and analysis was performed using the GTR+G+I model in MEGA6 with 1000 bootstrap samples.

(a) N1353



(b) G1569

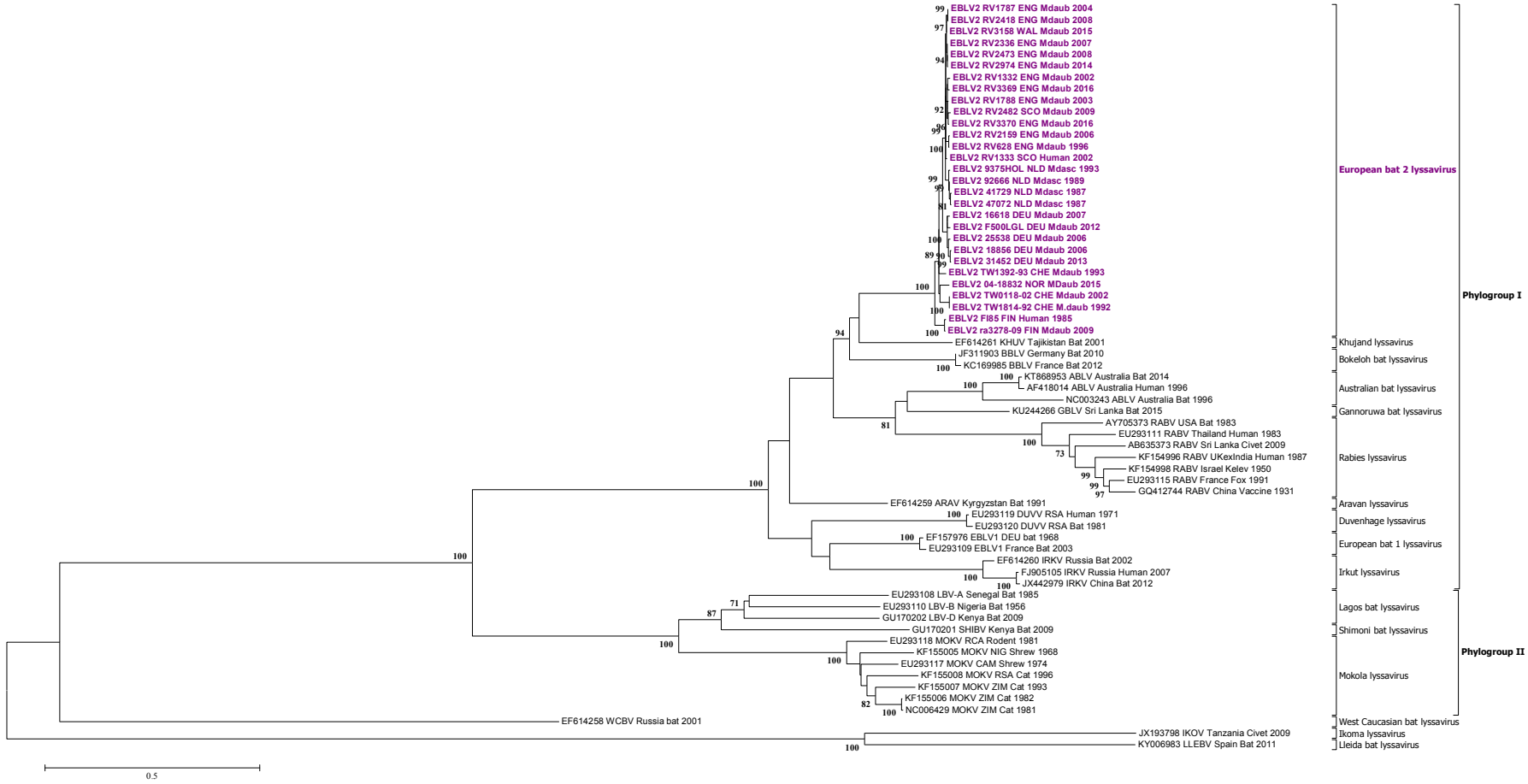
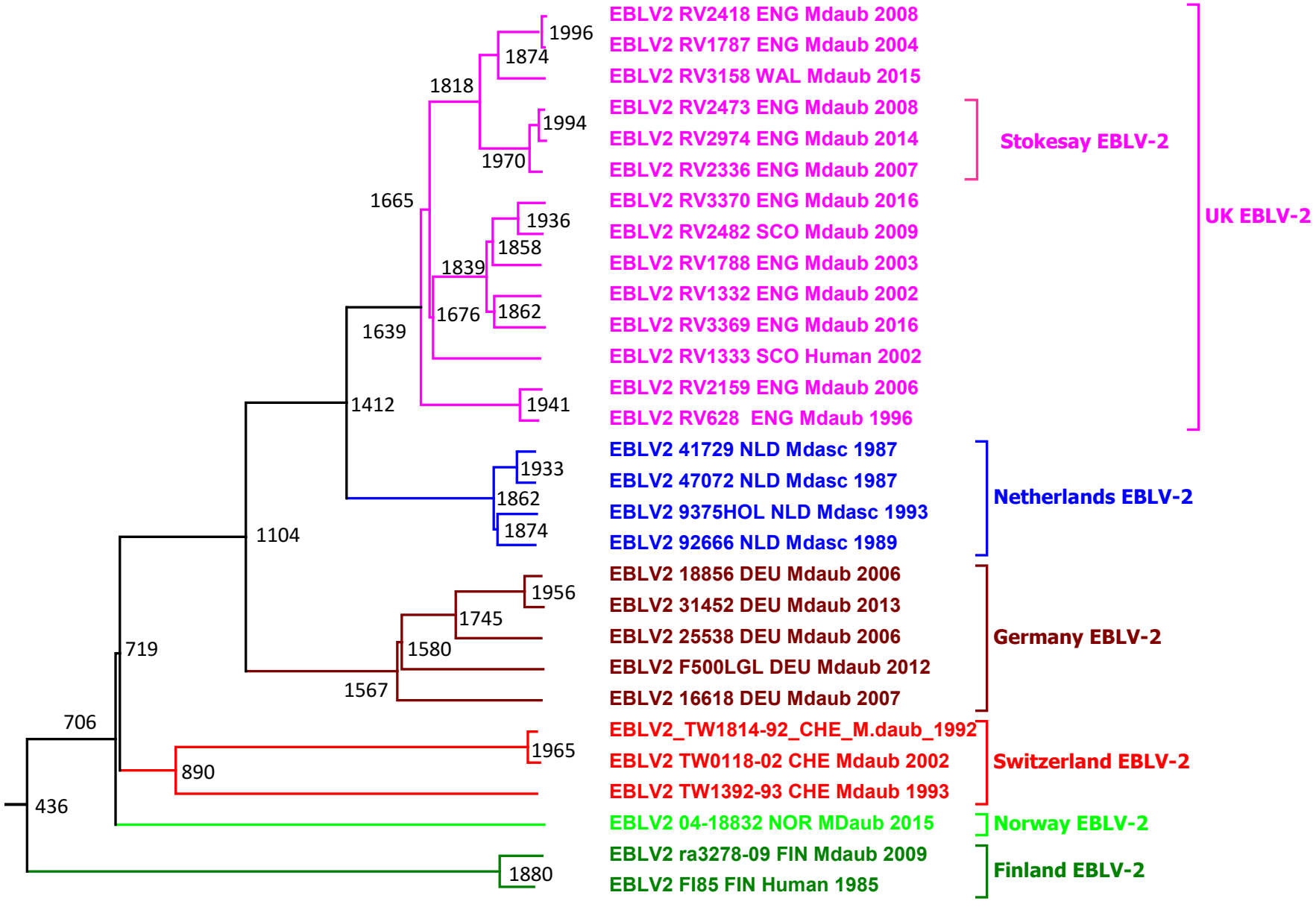

EBLV2 RV1787 ENG Mdaub 2004
EBLV2 RV2418 ENG Mdaub 2008
EBLV2 RV3158 WAL Mdaub 2015
EBLV2 RV2336 ENG Mdaub 2007
EBLV2 RV2473 ENG Mdaub 2008
EBLV2 RV2974 ENG Mdaub 2014
EBLV2 RV1332 ENG Mdaub 2002
EBLV2 RV3369 ENG Mdaub 2016
EBLV2 RV1788 ENG Mdaub 2003
EBLV2 RV2482 SCO Mdaub 2009
EBLV2 RV3370 ENG Mdaub 2016
EBLV2 RV2159 ENG Mdaub 2006
EBLV2 RV628 ENG Mdaub 1996
EBLV2 RV1333 SCO Human 2002
EBLV2 9375HOL NLD Mdasc 1993
EBLV2 92666 NLD Mdasc 1989
EBLV2 41729 NLD Mdasc 1987
EBLV2 47072 NLD Mdasc 1987
EBLV2 16618 DEU Mdaub 2007
EBLV2 F500LGL DEU Mdaub 2012
EBLV2 25538 DEU Mdaub 2006
EBLV2 18856 DEU Mdaub 2006
EBLV2 31452 DEU Mdaub 2013
EBLV2 TW1392-93 CHE Mdaub 1993
EBLV2 04-18832 NOR MDaub 2015
EBLV2 TW0118-02 CHE Mdaub 2002
EBLV2 TW1814-92 CHE M.daub 1992
EBLV2 FI85 FIN Human 1985
EBLV2 ra3278-09 FIN Mdaub 2009
EF614261 KHUV Tajikistan Bat 2001
JF311903 BBLV Germany Bat 2010
KC169985 BBLV France Bat 2012
KT868953 ABLV Australia Bat 2014
AF418014 ABLV Australia Human 1996
NC003243 ABLV Australia Bat 1996
KU244266 GBLV Sri Lanka Bat 2015
AY705373 RABV USA Bat 1983
EU293111 RABV Thailand Human 1983
AB635373 RABV Sri Lanka Civet 2009
KF154996 RABV UKexIndia Human 1987
KF154998 RABV Israel Kelev 1950
EU293115 RABV France Fox 1991
GQ412744 RABV China Vaccine 1931
EF614259 ARAV Kyrgyzstan Bat 1991
EU293119 DUVV RSA Human 1971
EU293120 DUVV RSA Bat 1981
EF157976 EBLV1 DEU bat 1968
EU293109 EBLV1 France Bat 2003
EF614260 IRKV Russia Bat 2002
FJ905105 IRKV Russia Human 2007
JX442979 IRKV China Bat 2012
EU293108 LBV-A Senegal Bat 1985
EU293110 LBV-B Nigeria Bat 1956
GU170202 LBV-D Kenya Bat 2009
GU170201 SHIBV Kenya Bat 2009
EU293118 MOKV RCA Rodent 1981
KF155005 MOKV NIG Shrew 1968
EU293117 MOKV CAM Shrew 1974
KF155008 MOKV RSA Cat 1996
KF155007 MOKV ZIM Cat 1993
KF155006 MOKV ZIM Cat 1982
NC006429 MOKV ZIM Cat 1981
EF614258 WCBV Russia bat 2001
JX193798 IKOV Tanzania Civet 2009
KY006983 LLEBV Spain Bat 2011
0.5
European bat 2 lyssavirus
Khujand lyssavirus
Bokeloh bat lyssavirus
Australian bat lyssavirus
Gannoruwa bat lyssavirus
Rabies lyssavirus
Aravan lyssavirus
Duvenhage lyssavirus
European bat 1 lyssavirus
Irkut lyssavirus
Lagos bat lyssavirus
Shimoni bat lyssavirus
Mokola lyssavirus
West Caucasian bat lyssavirus
Ikoma lyssavirus
Lleida bat lyssavirus
Phylogroup I
Phylogroup II

Figure 3 Bayesian maximum clade credibility (MCC) tree of 29 EBLV-2 concatenated gene sequences (10818 nucleotide sites comprising the nucleoprotein, matrix, phospoprotein, glycoprotein and polymerase genes) with branch lengths scaled in time by enforcing a relaxed molecular clock. The EBLV-2 strains are coloured to indicate the location of sampling. Mean substitution rate of the dataset was 1.7E-5.



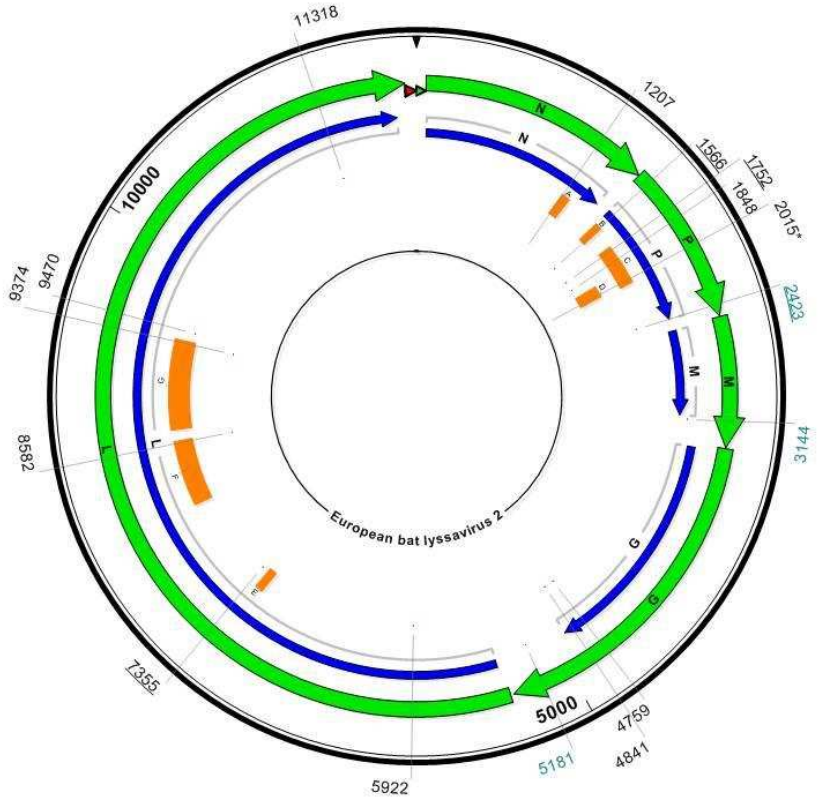
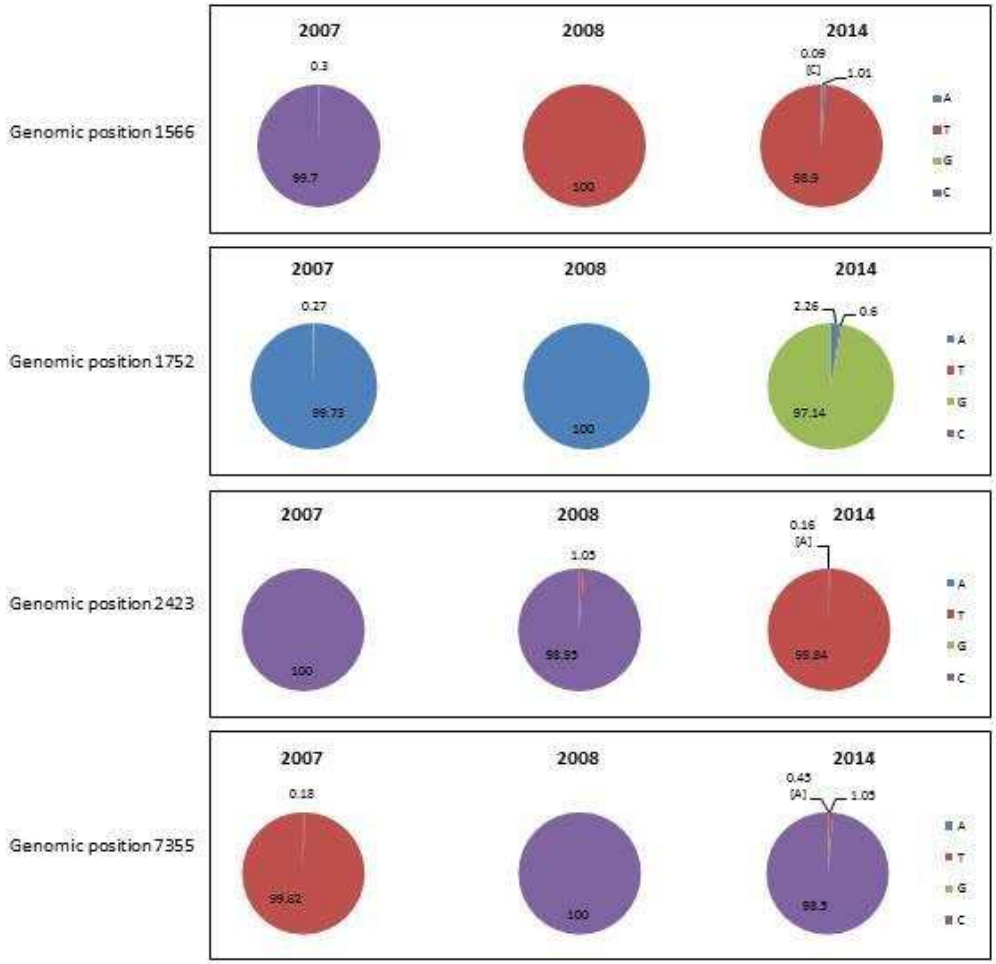


Figure 4. A circular map of the EBLV2 genome. Numbers indicate genomic positions at which heterogeneity is seen between the three genomes from Stokesay Castle. Black positions are located within coding regions, teal positions are in non-coding regions. Position 2015 (asterisk) has a base change corresponding to an amino acid substitution. Orange blocks indicate known domains within each protein where heterogeneity between the three genomes is observed. Underlined positions indicate those at which >1% heterogeneity is seen within the host (refer to figure 2). Domains key: A = T-cell epitopes, B = L binding region, C = soluble N binding region, D = LC8 protein binding region, E = Domain III, F = Domain IV G = Domain V

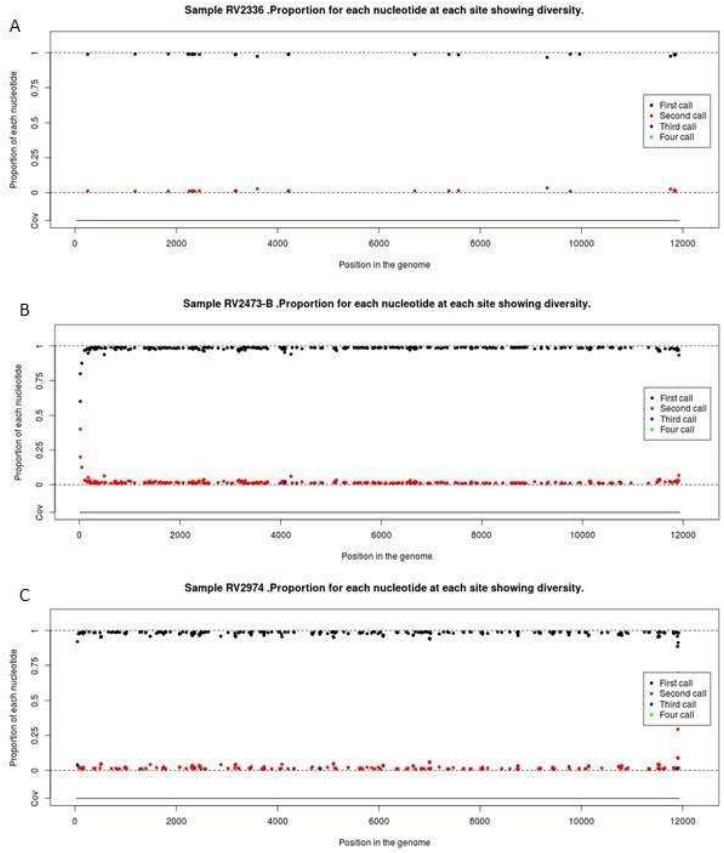


Supplementary Figure 1.

Supplementary Table 1. Viral variants identified at positions with SNPs. For each position the total number of reads is given, as well as the consensus base and whether variants from the consensus are seen. Where variants exist, the variant base is given as well as the percentage of viral reads that show this base. Variants that exceed 1% of the total viral reads are highlighted in yellow. * indicates an amino acid substitution. - indicates no variants at this position. Text in blue highlights the only SNP denoting an amino acid substitution within the consensus sequences.

			Base				
Genomic region	**Domain**	**Position**	**2007**	**2008**	**2014**	**Base position within amino acid**	**Amino acid change in coding region?**
N gene	T cell epitope	1207	A	G	G	3rd	No
P gene	L binding region	1566	C	T	T	3rd	No
P gene	N° binding region	1752	A	A	G	3rd	No
P gene	N° binding region	1848	G	A	A	3rd	No
P gene	LC8 binding region	2015	A	G	A	2nd	Yes (K-R-K)
P-M	None	2423	C	C	T	n/a	No
M-G	None	3144	T	C	C	n/a	No
G gene	None	4759	C	T	T	3rd	No
G gene	None	4814	C	C	T	3rd	No
G-L	None	5181	C	T	T	n/a	No
L gene	None	5922	C	T	T	1st	No
L gene	L domain III	7355	T	C	C	3rd	No
L gene	L domain IV	8582	A	A	G	3rd	No
L gene	L domain V	9374	T	C	C	3rd	No
L gene	None	9470	A	G	G	3rd	No
L gene	None	11318	T	C	T	3rd	No

Supplementary Figure 3. SNPs identified between the three EBLV2 genomes from Stokesay Castle. For each position, information on genomic region, domain within that region, position within amino acid and whether the SNP results in a substitution of a coding amino acid. Domain information from Marston *et al.* 2007. K = Lysine, R = Arginine. Colour highlighting indicates pattern of change in base at this position; red = different, same, same; green = same, same, different; blue = same, different, same.



Supplementary Figure x. Positions at which intra-host variation is seen within the three EBLV-2 isolates from Stokesay castle. A. 2007 (RV2336), B. 2008 (RV2473), C. 2014 (RV2974). Mapping quality 30, call quality 30, error 0.01

5.3: Using WGS to infer evolutionary relationships using Affinity Propagation

Affinity propagation is an algorithm that can be used to identify exemplars within data points and then form clusters around these exemplars. It considers simultaneously all data points as potential exemplars and exchanges messages between the data points until a good set of exemplars are identified and clusters are obtained (Frey and Dueck, 2007; Kiddle et al., 2010). AP can be used whenever there is a way to measure a number of each pair of data points so that similarity can be compared. AP has been used for a diverse range of applications including analysis of network traffic, analysing basketball statistics and clustering astrophysical objects. AP has also been used for more biologically relevant applications such as gene expression data.

The following manuscript has been prepared and submitted to PLoS NTD describing utilising APC to differentiate RABVs at the sub-species level.

5.3.1: Article 29 submitted to PLoS NTD

Affinity propagation as a novel tool for objective clustering of rabies virus sequences

Summary: Traditional phylogenetic analyses, in combination with virus metadata, suggests spatial-dependent distributions of RABV. However, these analyses were limited by the lack of objective criterions for cluster allocations in phylogenetic trees. Therefore, we additionally applied a mathematical approach called affinity propagation clustering (APC) to eliminate this limitation and to propose a way to classify full genome RABV sequences. Because of its computational power, APC has already been successfully applied for other bioinformatical purposes e.g. microarray and gene expression data but not for cluster analysis in sequence comparisons. To this end, we used the most extended panel of existing and newly obtained full genome RABV sequences to exemplarily demonstrate the application of APC for RABV clustering. On a global scale, APC determined the most supported number of clusters as four, i.e. New Word cluster, Arctic/Arctic-like, Cosmopolitan, and Asian as previously assigned by phylogenetic studies. We propose a combination of APC with established phylogenetic analyses to resolve phylogenetic relationships between objectively determined clusters and sequences. This workflow could help to substantiate a transparent cluster distribution, not only for RABV but also for other comparative sequence analyses.

5.4: Conclusions

These data presented within this chapter represents large multi-national collaborative studies. Each study incorporates complete genome sequences, obtained during this PhD using NGS technologies. Utilising complete genome sequences improves the resolution of the phylogenetic analyses ensuring the most comprehensive analyses. Challenging the presumed 'out of Africa' hypothesis for the origins of lyssaviruses has proven to be controversial, sparking debate amongst lyssavirus researchers through Twitter, Research Gate and email. The robustness of the analysis will ensure that the data is correctly critiqued and built on as more lyssaviruses and analyses will obtained in the future.

The future of lyssavirus phylogeography will undoubtedly be shaped by the increasing availability of genome sequences. The resolution afforded will provide the foundation for transmission analysis of highly related virus genomes which is investigated in detail in Chapter 6.

Chapter 6: Utilising WGS to investigate lyssavirus transmissions

Chapter 6: Utilising WGS to investigate lyssavirus transmissions

6.1: Introduction

Utilising WGS for broad-scale phylogenetics and phylogeography was discussed in detail in Chapter 5. This Chapter will focus on utilising WGS for fine-scale population dynamics and transmission. One of the major limitations in partial genome sequence analysis, which has dominated the phylogenetics field until recently, is that highly related virus sequences were unable to be resolved. In part, this was addressed within the lyssavirus field, by sequencing the G-L region of the genome (see Fig 6) which is the most variable region of the genome (Mollentze et al., 2014b; Sabeta et al., 2007; Talbi et al., 2010). Indeed, analysis with this region resolved many highly related viruses, enabling a finer scale analysis.

With this finer-resolution, researchers began to question whether transmission chains could be resolved, to identify 'who infected whom' in a disease outbreak or indeed, endemic rabies region. To address these questions, both epidemiological (spatiotemporal) and genetic data are required. Samples from the UK FMDV outbreaks have been analysed to reconstruct both the 2001 and 2007 outbreaks using Bayesian inference scheme combining genetic and epidemiological data. The models assume a single introduction, followed by widespread transmission. These analyses robustly inferred transmission chains and identified undetected premises providing future tools for real-time tracing of epidemics (Morelli et al., 2012). In an endemic scenario, with multiple sources of infection, for example, rabies in South Africa, the framework was altered to accommodate the epidemiological landscape. Using G-L data from approximately 200 RABV isolates direct transmission was resolved to a fine level, while introductions and transmissions from an unsampled source was resolved at a moderate level (Mollentze et al., 2014b).

In a comprehensive study recently published, partial genomes from five publicly available viral genetic data sets associated with distinct epidemics were utilised: (1) RABV in North American skunks, (2) RABV in North American raccoons, (3) RABV in North African domestic dogs, (4) RABV in vampire bat populations in eastern Argentina and (5) RABV in vampire bats in eastern Brazil. The data sets all contain acceptably long nucleotide sequences (800–3,000 nt) that are spatially distributed over broad geographic ranges and are associated with precise information on the dates and locations of sampling, and on host species (Dellicour et al., 2017). The analysis focused on spatial dynamics of the different RABVs and estimated the rate of viral movements. They concluded that dog-mediated rabies was spread more rapidly than in wildlife populations, due to human mediated factors. Interestingly, they

did not identify any environmental factors with strong confidence, although it is likely that the most applicable environment factors for bats and terrestrial mammals are different, therefore in appropriate environmental factors were considered for bats.

Incomplete sampling will always be the major challenge in the resolution of transmission chains, however, utilising WGS with samples that have accurate epidemiological data with ever improving computer modelling will provide a unique insight into the transmission of pathogens such as RABV.

The articles in this chapter are among the first endemic RABV transmission studies utilising WGS.

6.2: Article 30 published in Virus Evolution

Elucidating the phylodynamics of endemic rabies virus in eastern Africa using whole-genome sequencing

DOI:10.1093/ve/vev011

Summary: In this study, canine RABV in Tanzania was used as an example to determine the value of whole genome resolution over sub-genomic (partial N-gene) data to uncover fine-scale population structure. Using sub-genomic data continental-scale patterns of population structure were identified, but whole genome resolution was required to distinguish between samples from a small spatio-temporal window and characterize endemic transmission dynamics within Tanzania with confidence. While RABV has defined spatial structure at large scales, increasingly frequent levels of admixture were observed at regional and sub-regional levels. Discrete phylogeographic analysis revealed long-distance dispersal within Tanzania, attributed to human-mediated movement. Much of the viral population structure within Tanzania appears to result from initial invasive waves that have persisted for long periods endemically, with structure eroding over time and due to human-mediated movement. Evidence of multiple persistent, co-circulating lineages at a very local scale in the Serengeti District, were also observed, which may reflect a more continuous dispersal dynamic in endemic landscapes where RABV has had a lengthy period of circulation and increased admixture due to human-mediated introductions. Samples in cluster Tz1 share identity with samples from Kenya indicating a shared evolutionary history that may relate to an initial outbreak across both countries with a southward dissemination of this lineage. Cluster Tz4 is related to samples from countries south of Tanzania, possibly South Africa or Mozambique, underlining the threat of re-invasion or introduction from external sources and the potential value of whole genome resolution to robustly and more accurately identify sources. To date, the vast majority of genetic data from Africa is partial genome, however our data suggest that whole genome characterization would be valuable and should be an aim for the future.

Sequences from the island of Pemba indicated multiple introductions from a diverse range of sources. Samples were scattered throughout the phylogeny and the most divergent lineage Tz5 consisted of two samples from Pemba. The distribution of these lineages may reflect earlier invasions of RABV into Pemba from elsewhere in Tanzania (Tz1 and 2) and the African continent (Tz5), resulting from Pemba's

location on an important trading route. However, since these lineages were not resampled and no cases have been detected from Pemba in over a year, RABV may have been only transiently circulating.

Using genetic data from a hierarchy of spatial scales and varying levels of genome coverage, the advantage of whole genome over sub-genomic data to describe the spatio-temporal dynamics of rabies virus in an endemic system was demonstrated.

Virus Evolution, 2015, 1(1): vev011

doi: 10.1093/ve/vev011
Research article

Elucidating the phylodynamics of endemic rabies virus in eastern Africa using whole-genome sequencing

Kirstyn Brunker,[1,2,3,*] Denise A Marston,[3] Daniel L Horton,[3,4] Sarah Cleaveland,[1,2] Anthony R Fooks,[3] Rudovick Kazwala,[5] Chanasa Ngeleja,[6] Tiziana Lembo,[1,2] Maganga Sambo,[7] Zacharia J Mtema,[7] Lwitiko Sikana,[7] Gavin Wilkie,[8] Roman Biek,[1,2,†] and Katie Hampson[1,2,†]

[1]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow G12 8QQ, UK, [2]The Boyd Orr Centre for Population and Ecosystem Health, University of Glasgow, Glasgow G12 8QQ, UK, [3]Animal and Plant Health Agency, Weybridge, Woodham Lane, KT15 3NB, UK, [4]School of Veterinary Medicine, University of Surrey, Guildford GU2 7XH, UK, [5]Department of Veterinary Medicine and Public Health, Sokoine University of Agriculture, Morogoro, United Republic of Tanzania, [6]Tanzania Veterinary Laboratory Agency, Dar es Salaam, United Republic of Tanzania, Temeke Veterinary, Mandela Road, P.O. BOX 9254, [7]Ifakara Health Institute, Ifakara, United Republic of Tanzania, P.O. Box 53 and [8]MRC Centre for Virus Research, University of Glasgow, Sir Michael Stoker Building, Garscube Campus, 464 Bearsden Road, Glasgow G61 1QH, UK

*Corresponding author: E-mail: k.brunker.1@research.gla.ac.uk
†These authors contributed equally to this work.

Abstract

Many of the pathogens perceived to pose the greatest risk to humans are viral zoonoses, responsible for a range of emerging and endemic infectious diseases. Phylogeography is a useful tool to understand the processes that give rise to spatial patterns and drive dynamics in virus populations. Increasingly, whole-genome information is being used to uncover these patterns, but the limits of phylogenetic resolution that can be achieved with this are unclear. Here, whole-genome variation was used to uncover fine-scale population structure in endemic canine rabies virus circulating in Tanzania. This is the first whole-genome population study of rabies virus and the first comprehensive phylogenetic analysis of rabies virus in East Africa, providing important insights into rabies transmission in an endemic system. In addition, sub-continental scale patterns of population structure were identified using partial gene data and used to determine population structure at larger spatial scales in Africa. While rabies virus has a defined spatial structure at large scales, increasingly frequent levels of admixture were observed at regional and local levels. Discrete phylogeographic analysis revealed long-distance dispersal within Tanzania, which could be attributed to human-mediated movement, and we found evidence of multiple persistent, co-circulating lineages at a very local scale in a single district, despite on-going mass dog vaccination campaigns. This may reflect the wider endemic circulation of these lineages over several decades alongside increased admixture due to human-mediated introductions. These data indicate that successful rabies control in Tanzania could be established at a national level, since most dispersal appears to be restricted within the confines of country borders but some coordination with neighbouring countries may be required to limit transboundary movements. Evidence of complex patterns of rabies circulation within Tanzania necessitates the use of whole-genome sequencing to delineate finer scale population structure that

can that can guide interventions, such as the spatial scale and design of dog vaccination campaigns and dog movement controls to achieve and maintain freedom from disease.

Key words: RNA virus; phylodynamics; zoonoses; endemic; rabies; translocation.

1 Introduction

The general trend of increasing incidence and expansion of emerging or re-emerging zoonotic diseases (e.g., Ebola, Chikungunya, and avian influenza) (Woolhouse 2002; Greger 2007; Jones et al. 2008) and persistence of established zoonoses, such as canine rabies, highlights the ongoing challenges faced as we attempt to characterize and control them. The processes that drive the spread and persistence of infectious diseases are reflected in a genetic signature in pathogen genomes (Biek and Real 2010). Understanding the processes that give rise to spatial population structure in pathogens can inform the management and control of infectious diseases. For example, analyses of evolutionary, epidemiological, and ecological data have recently demonstrated that global live swine trade strongly predicts the global dissemination of influenza A viruses in swine (Nelson et al. 2015), and air travel has been revealed as a major factor driving the intra-continental spread of Dengue virus (Nunes et al. 2014). Viral pathogens, particularly fast-evolving RNA viruses, are model systems to explore pathogen populations as they rapidly accumulate genetic diversity on a timescale similar to epidemiological processes (Drummond et al. 2003; Biek et al. 2015). Statistical phylogeographic approaches are available (Lemey et al. 2009; Bedford et al. 2014; Bielejec et al. 2014) to develop a quantitative understanding of the processes that give rise to spatial patterns in RNA viruses (Holmes and Grenfell 2009) on epidemiological time scales. Whole-genome sequencing (WGS) is increasingly being used as a means to extract these patterns, but it is unclear how much resolution can be gained and at what temporal and spatial scale. In this article, canine rabies virus (RABV) is used as a model to determine the spatiotemporal patterns of an endemic zoonotic virus using whole-genome data to distinguish structure at an increasingly fine scale.

Rabies is a globally distributed zoonotic disease caused by a single-stranded negative sense RNA virus from the *Lyssavirus* genus. Though capable of infecting any mammal, given virus variants are typically maintained in distinct species-specific cycles within the orders Carnivora and Chiroptera (Rupprecht, Hanlon, and Hemachudha 2002). The disease causes thousands of human deaths every year, predominantly in Asia and Africa where the virus circulates endemically in domestic dogs (*Canis lupus familiaris*) (Knobel et al. 2005; Shwiff, Hampson, and Anderson 2013). The majority of these deaths (~99%) are caused by bites from rabid dogs, instilling fear into the many communities that live under continuous threat from a disease that is almost invariably fatal but entirely preventable. Although the role of domestic dogs as key vectors of rabies is recognized, much less is known about the dog-associated RABV variant than wildlife variants such as raccoon or skunk RABVs circulating in North America (Brunker et al. 2012). Moreover, while epidemic expansions of wildlife RABV have been well documented and studied (e.g., Real et al. 2005; Biek et al. 2007; Kuzmina et al. 2013), we know little about the persistence and spread of rabies in endemic landscapes.

Characterizing the spatial scales of canine rabies dispersal is a critical step toward identifying the processes and factors driving its dynamics and the scale at which control strategies need to be implemented. On a global scale, canine RABV exhibits a strong phylogeographic structure with the distribution of seven distinct major clades reflecting the position of major barriers, such as oceans and mountain ranges, or historical mass human colonization/migration events (David et al. 2007; Bourhy et al. 2008). However, it is unclear whether this genetic structure will persist in endemic scenarios or at smaller scales, and how much it is influenced by human-mediated dispersal. Indeed, on a regional scale, this landscape structure becomes less distinct: some landscape features, for example, geopolitical boundaries can act as apparent barriers to movement, as seen in North Africa (Talbi et al. 2010), whilst contradictory patterns of synchronous cycles of RABV across multiple countries (Hampson et al. 2007) and repeated cross-border incursions (Hayman et al. 2011) have also been observed elsewhere in the continent.

While there is a growing understanding of the epidemiology of canine rabies in Africa (Hampson et al. 2007, 2009; Lembo et al. 2008), effective rabies control is still hindered by limited knowledge of some of the key drivers of viral transmission and spread. Mass dog vaccination is the mainstay of successful rabies control but requires sustained coverage of at least 70 per cent (WHO 2005; Townsend et al. 2013). In addition, spatial heterogeneity may affect how vaccine is most effectively distributed to interrupt key transmission corridors and target regions seeding RABV dispersal. An important aspect of this heterogeneity is the impact of human factors on RABV transmission, which has direct implications for control, including the design and scale of interventions necessary to interrupt transmission and maintain freedom from disease. For example, movement of people between urban and rural areas and the dog meat trade have been postulated as means of spreading RABV through human-mediated dog movements in rabies-endemic countries in Asia (Denduangboripant et al. 2005; Tao et al. 2009; Ahmed et al. 2015). Uncovering the viral population structure and dynamics of RABV in Tanzania may identify similar patterns attributed to human-mediated movements that can aid the identification of sources key to viral persistence.

There have been few in-depth spatial epidemiology studies of RABV in sub-Saharan Africa, probably owing to the lack of resources for effective surveillance including sample collection (Nel 2013). However, recent studies provide intriguing insights into the dynamics of rabies in certain parts of Africa (see Talbi et al. 2010; Mollentze et al. 2014), indicating a degree of spatial structure with evidence of long distance movements facilitated by humans. RABV spatiotemporal dynamics in East Africa in particular are poorly resolved and very few sequences are publically available. At present, little is known about the genetic diversity or structure of RABV in Tanzania other than coarse phylogenetic analyses of partial or full nucleoprotein gene (N gene) sequences, limited to well-studied regions (Kissi, Tordo, and Bourhy 1995; Lembo et al. 2007). Furthermore, whole-genome population studies of RABV have not yet been attempted despite their great potential to provide a better understanding of the processes determining rabies spread and persistence.

In this study, finely resolved space-time-genetic data, including whole-genome sequences, were used to determine the spatial and temporal dynamics of endemic RABV at both a regional and local scale in Tanzania. The utility of partial genome data was also demonstrated as a means to characterize large-scale phylogeographic patterns of RABV in Africa. Specifically, we aimed to characterize the dynamics of rabies virus in an endemic system including the role of human-mediated transport.

2 Materials and methods

2.1 Samples

For this study, fifty-nine new whole-genome sequences were obtained from animal hosts (primarily domestic dogs) from nine regions in Tanzania sampled between 2003 and 2012. A previously sequenced sample (RV2772; accession: KF155002) (Marston et al. 2013) from Southern Tanzania was included in the dataset and used as a reference sequence (sample details in Supplementary Table S1). The main study area encompassed the Serengeti District in northwest Tanzania where approximately half the samples ($n = 33$) were obtained from active surveillance, which enabled the collection of brain material from suspect rabid animals and GPS coordinates and dates recorded as described in Hampson et al. (2009). The remaining twenty-seven samples were obtained opportunistically from other regions in Tanzania as part of surveillance by the Tanzanian Ministry of Livestock and Fisheries Development and the Tanzanian Veterinary Laboratory Agency. All samples were sent to the Animal & Plant Health Agency (APHA) in Weybridge, UK, for processing.

In addition, fifty new partial N gene sequences (405 bp) from Tanzania were obtained via reverse transcription-polymerase chain reaction (RT-PCR) and Sanger sequencing using samples archived at APHA and submitted to GenBank with accession numbers KR534217–KR534266 (details in Supplementary Material and Table S2).

2.2 RNA extraction and WGS

Total RNA was extracted at APHA directly from brain tissue using TRIZOL, according to manufacturer's instructions (Invitrogen). Precipitated total RNA was re-suspended in molecular-grade water at a one in ten dilution and quantified using a NanoDrop spectrophotometer (Thermo Scientific). Samples were sequenced on a range of next generation sequencing platforms during NGS protocol optimization (see Supplementary Material for details). The majority of samples ($n = 48$) were sequenced by the following method: TRIzol-extracted viral RNA was depleted of host genomic DNA using the on-column DNase treatment in RNeasy plus mini kit (Qiagen) as per manufacturer's instructions with elution in 30 µl molecular grade water. This was followed by host ribosomal RNA depletion using Terminator 5'-phosphate-dependent exonuclease (Epicentre Biotechnologies), as detailed in (Marston et al. 2013). First- and second-strand cDNA was synthesized using a Roche cDNA synthesis system kit with random hexamers (Roche). Resultant cDNA was quantified using Picogreen dsDNA quantitation reagent (Invitrogen) and approximately 1 ng of each sample used in a 'tagmentation' reaction mix using a Nextera XT sample preparation kit (Illumina), according to the manufacturer's protocol (minus the bead normalization step). DNA libraries for each sample were quantified using a Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) or a Qubit assay kit (Life technologies), and average library size was measured with a high-sensitivity DNA Bioanalyzer chip on a model 2100 Bioanalyzer (Agilent). Sample libraries were transported to the MRC Centre for Virus Research at the University of Glasgow, UK, for the final steps of library preparation and sequencing. Individual libraries were pooled and normalized to equimolar concentrations at a suitable plexity (x24 for MiSeq runs). Libraries were sequenced as 150-bp paired-end reads on an Illumina MiSeq. Additional sequencing was conducted on a NextSeq 500 platform (Glasgow Polyomics at the University of Glasgow, Glasgow, UK) and reads merged with MiSeq reads to increase coverage for poorly sequenced samples (see Supplementary Material).

2.3 Bioinformatics and sequence analysis

Raw reads were assessed in FastQC (Andrews 2010), and Trimmomatic (Bolger, Lohse, and Usadel 2014) was used to trim 3'-ends, remove adapter contamination, and to filter based on quality with default parameters. Filtered reads were mapped to the previously sequenced genome of Tanzanian RABV sample RV2772 (accession: KF155002) with BWA mem version 0.7.10 (Li and Durbin 2009) and converted to bam file format using SAMtools v. 0.1.18 (Li et al. 2009).

A conservative single-nucleotide polymorphism (SNP) calling routine was implemented in GATK utilizing the UnifiedGenotyper tool to identify high confidence SNPs, which had passed GATK filters on strand bias ($FS > 60$, $SOR > 4$), mapping quality ($MQ < 40$, $MQRankSum < -12.5$), read position ($ReadPosRankSum < -8$), and depth of coverage ($DP < 5$). Indels were filtered if $FS > 200$ and $ReadPosRankSum < -20$ and further manually inspected for inclusion (e.g., dismissed if near a homopolymer run). Consensus sequences were built using a custom script in R, which called filtered SNPs with a 75 per cent consensus rule (positions with <75% consensus were given a IUPAC code for the corresponding ambiguous base call) and genome positions with a depth of coverage less than two were labelled 'N'. Potential SNP calls that failed only the depth filter, that is, had a depth < 5 but >1, were passed if the same polymorphism had been present as a high confidence SNP in at least two other samples. Otherwise, the position was given an IUPAC code representing the population-level calls and the potential SNP. In addition, a set of consensus sequences using a more relaxed approach to SNP calling was produced which involved strict calls of all SNPs with depth > 1 and gaps filled with the majority population consensus sequence. These relaxed consensus sequences were used to produce initial starting trees for BEAST analyses (see Section 2.6).

Sequencing resulted in 93–100 per cent coverage of the genome, with >99 per cent genome coverage achieved for 95 per cent of samples and a median depth of coverage of seventy-seven (range: 6–1,871, see Supplementary Table S1). Nextera XT is a transposase-based method of library preparation and sequence reads typically miss the ends of the genome; however, as the ends of lyssaviruses are highly conserved (Marston et al. 2007; Kuzmin et al. 2008), it is unlikely that any informative variation was missed. We, therefore, consider our analyses to be based on genome-wide variation and henceforth refer to our dataset as whole-genome sequences. Consensus sequences were aligned using MAFFT v7.149b (Katoh and Standley 2013) and submitted to GenBank (accession numbers: KR906734–KR906792).

2.4 Phylogenetic reconstruction

Initial datasets of (1) partial N gene 405 bp (1,317 sequences) and (2) full N gene 1,350 bp (674 sequences) sequences isolated in

Africa were constructed using sequences retrieved from GenBank and including new Tanzanian isolates sequenced for this study (59 new WGS samples and 50 new partial genome sequences). Following maximum likelihood (ML) phylogenetic reconstruction with the initial sequence datasets, subset trees were extracted for samples in the Africa 1b clade (430 samples in the partial N dataset and 100 in the full N dataset).

Alignments for whole genome, full N, and partial N gene were created in MAFFT (Katoh and Standley 2013) and estimated phylogenetic relationships using both ML and Bayesian methods. ML phylogenies were estimated in RAxML (Stamatakis et al. 2012) with a general time reversible (GTR) nucleotide substitution model and a gamma distribution model of among-site rate variation. A Chinese dog RABV sequence from GenBank (accession no: FJ712193) was used as an outgroup, and node support was evaluated with 1,000 bootstrap replicates. Bayesian phylogenetic reconstruction was conducted in BEAST v1.8.1 (Drummond et al. 2012) using a posterior distribution of trees (without a molecular clock model). Phylogenies were visualized and annotated in R using the packages adegenet (Jombart 2008) and APE (Paradis, Claude, and Strimmer 2004), and maps were made in R with Maptools (Lewin-koh et al. 2009) and sp packages (Pebesma and Bivand 2005; Bivand, Pebesma, and Gomez-Rubio 2008). The degree of spatial admixture at large phylogeographic scales, i.e. sub-continental and country level, was quantified by an association index (AI) using BaTS software with beast phylogenies (Parker et al. 2008).

2.5 Selecting an evolutionary model

An initial nucleotide substitution model was chosen based on the model selected by PartitionFinder (Lanfear et al. 2012). Our whole-genome alignments were partitioned into sixteen sets of nucleotides: one for each codon position (CP) in each gene (five genes) and one for concatenated non-coding regions. Model scheme selection was based on the best AIC score from a greedy search of substitution models, which favoured a GTR model with partitioning into three CPs (CP123).

In addition, model comparison based on marginal likelihood estimates using path sampling (PS) and stepping stone (SS) sampling implemented in BEAST (100 path steps and a chain length of 100,000 steps) were used to test varying levels of complexity in the substitution model. Non-coding sequence was concatenated and partitioned as a 'gene' with its own evolutionary model. Specifically, we tested the HKY model with: (1) a gene-specific nucleotide model with gene-specific rate variation; (2) a gene-linked CP partitioned model with among CP rate heterogeneity and homogeneous rates among genes; and (3) a gene-specific CP partitioned model with among CP and among gene rate heterogeneity. Model types 2 and 3 were also tested with CP112 and CP123 partitioning schemes. Following the results for the best HKY model, we did a final step comparing the most supported HKY model with the same structure but using a GTR model. Results (Supplementary Table S2) strongly favoured GTR and CP models (CP123 was best supported), but there was no support for partitions according to genes, which all had similar rates of substitution. This significantly reduced the complexity of the model and is an important finding in the context of analysing RABV whole-genome sequence.

2.6 Bayesian evolutionary analyses

Bayesian Markov-chain Monte Carlo analyses were performed using BEAST v1.8.1 (Drummond et al. 2012) and the BEAGLE library (Ayres et al. 2012). Based on model comparisons, the most supported evolutionary model was a GTR model with different substitution parameters for CPs one, two, and three ($GTR_{123} + CP_{123} + \Gamma_{123}$) and homogeneous rates among genes, with a GTR + G substitution model for non-coding sequence. A relaxed molecular clock with a lognormal distribution was used to model rate variation among branches with a Bayesian skyline model (Drummond et al. 2005) with ten groups as a flexible tree prior. A Bayesian skyline plot was used to estimate the viral effective population size through time (Drummond et al. 2005). To reconstruct the spatial dynamics of RABV spread in Tanzania, we implemented a discretized diffusion process among nine regional sampling locations, formalized as an asymmetric continuous time Markov chain (CTMC) model (Lemey et al. 2009). Three independent Markov-chain Monte Carlo chains with 50 million states and a sampling frequency of 50,000 were combined in LogCombiner after discarding at least 10 per cent burn. Posterior distributions were inspected in Tracer v1.6 (Rambaut et al. 2014) to ensure adequate mixing and convergence. Initial analyses revealed issues with tree likelihood convergence. Therefore, a CTMC model using relaxed consensus sequences (see above), which contained fewer ambiguities, was implemented first and the maximum clade credibility (MCC) tree used as a starting tree for the final CTMC models with conservative consensus sequences.

To estimate the most significant pathways of viral dispersal between regions, a stochastic search variable selection (BSSVS) procedure was implemented to identify the best supported diffusion rates through Bayes factor (BF) testing (Lemey et al. 2009; Bielejec et al. 2011). For the per lineage rate of migration (Kühnert, Wu, and Drummond 2011), a conditional reference prior (Ferreira and Suchard 2008) is commonly used but for our data resulted in convergence problems with some of the BSSVS parameters. Instead, we used an exponential prior with a mean of 0.01, which gave the most robust results out of a range of values tested (data not shown). The degree of spatial admixture was scored using a modified Association Index (Wang et al. 2001; Lemey et al. 2009) and quantified using the inferred number of lineage migration/movement events between locations estimated by Markov jump counts (Minin and Suchard 2008) along the branches of the posterior tree distribution. For BEAST models with Markov Jump counts implemented, a conditional reference prior on the per lineage rate of migration was chosen. A summarized history of Markov jump counts was used to identify movements between regions that occurred on very short branches and thus over unusually short time frames. Dogs rarely move 1 km from their homestead (Hampson et al. 2009; Woodroffe and Donnelly 2011), and Hampson et al. (2009) found a maximum distance of 20 km between linked RABV cases in the Serengeti District. Lineage migrations between regions, representing distances >100 km, were therefore considered unlikely to be attributable to dog movement alone if they occurred over a period of 2 years or less and were instead interpreted as being the result of human-mediated movement. In addition, the same form of discretized diffusion model was used to assess diffusion at a broader scale, that is, between the North and South of Tanzania, with Pemba Island classed as a third discrete state. A BF test in the program SPREAD (Bielejec et al. 2011) was used to identify well-supported migration pathways (BF > 3). Sampled trees were summarized as an MCC tree with median node heights using TreeAnnotator v.1.8.1, and Figtree v1.4.2 was used to visualize trees and the inferred ancestral locations for internal branches.

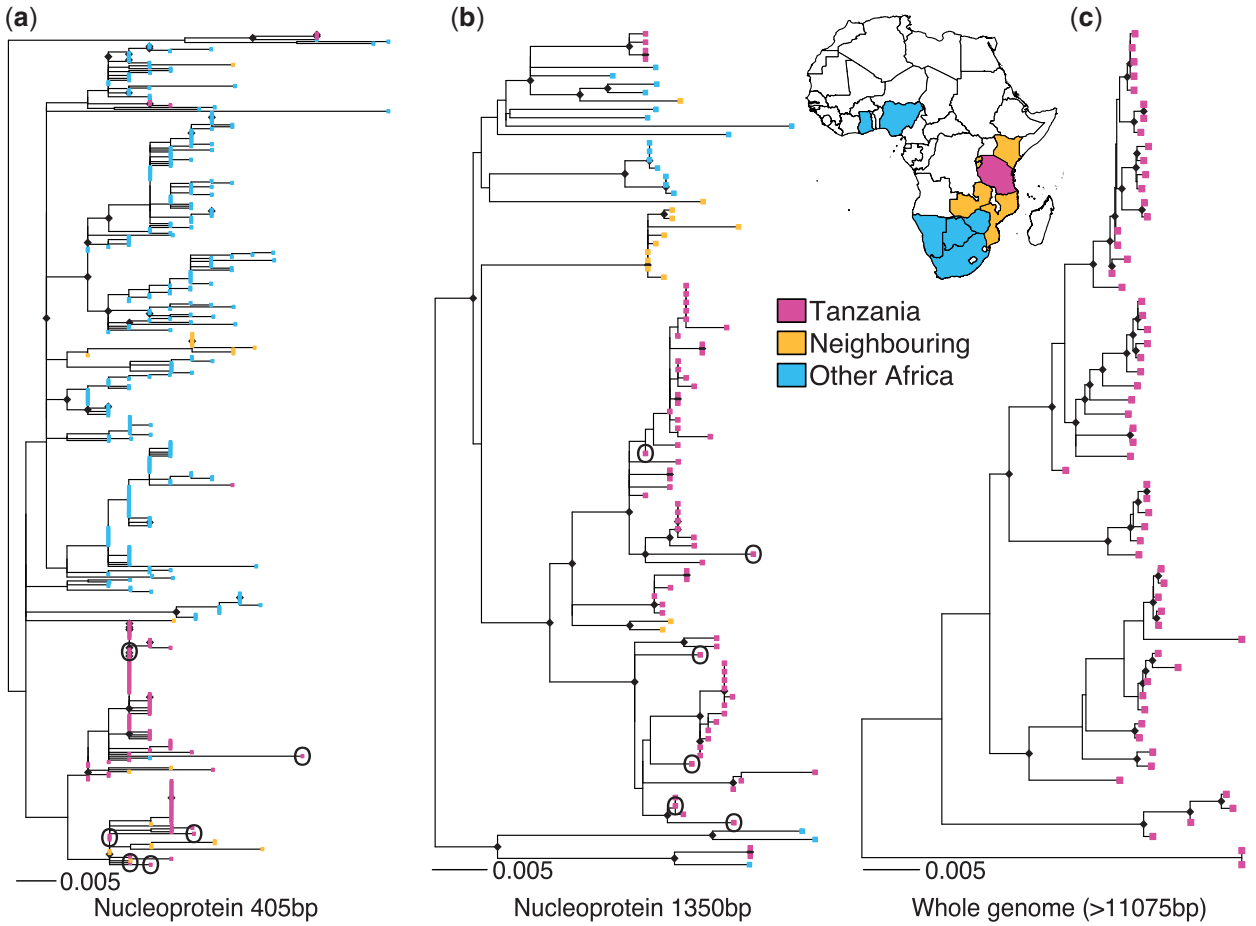


Figure 1. ML trees derived from datasets of rabies virus sequences from the Africa 1b clade for increasing levels of genome coverage: (a) 430 sequences from African countries highlighted on the map for a 405 bp fragment of the nucleoprotein gene, (b) 100 sequences of full 1,350 bp nucleoprotein gene from the same countries (except Botswana, Ghana, Kenya, and Zimbabwe); and (c) sixty full or near-full genome sequences (range: 11,076–11,923 bp) from Tanzania. Trees are scaled by number of substitutions per site and node symbols indicate nodes with bootstrap support ≥ 0.8. Historical samples from the Serengeti District (~20 years old) are circled in partial genome trees.

3 Results

3.1 Geographic resolution: partial versus full viral genomes

Consistent with previous large-scale phylogenetic studies, partial genome phylogenies indicated that RABV in sub-Saharan Africa falls into several regional groups with viruses from Eastern Africa generally being genetically distinct from those in western, central, and southern parts of the continent (Supplementary Figs S1 and S2). Of the four major lineages of RABV in Africa (David et al. 2007; Bourhy et al. 2008; Talbi et al. 2009), only the Cosmopolitan clade, and more specifically the Africa 1b lineage, was detected in Tanzania, as found previously (Lembo et al. 2007). Within the Africa 1b lineage, there was evidence of admixture between Tanzania and neighbouring countries and occasional long-range admixture at a continental scale (ML trees in Fig. 1 and Bayesian MCC trees in Supplementary Fig. S3). Sequences clustering most closely with Tanzanian sequences came from Kenya, which shares a border to the north. Although partial genome data were sufficient to identify such large-scale spatial patterns, these data did not provide adequate resolution to distinguish between samples at sub-national scales within Tanzania. The proportion of nodes with bootstrap support ≥80 per cent was only 0.11 (Bayesian posterior probability ≥ 90%: 0.18) for partial N and 0.27 (Bayesian: 0.41) for full N gene phylogenies. Furthermore, out of the sixty Tanzanian WGS, 60 per cent were identical at the partial N and 25 per cent at the full N gene level. In contrast, ML and Bayesian trees based on whole-genome sequences were fully resolved and well supported (proportion of nodes with ML bootstrap support ≥ 80%: 0.86; Bayesian posterior probability ≥90%: 0.89), even when samples had been taken in close spatial and temporal proximity. For example, RV2498 and RV2499 were sampled a day apart from the same village in Morogoro region: both samples only differed by a single SNP for the full N gene but were differentiated by twenty-five SNPs in the whole-genome alignment. This large divergence strongly suggests that the two samples are not from the same chain of transmission, which might have been the conclusion based on partial genome data. The median raw pairwise genetic distance between Tanzanian whole-genome sequences was 259 (range: 0–608) nucleotides and between the Serengeti District samples was 120 (range: 2–212) nucleotides, showing considerable diversity at even a small spatiotemporal scale. However, much of this high divergence is due to the

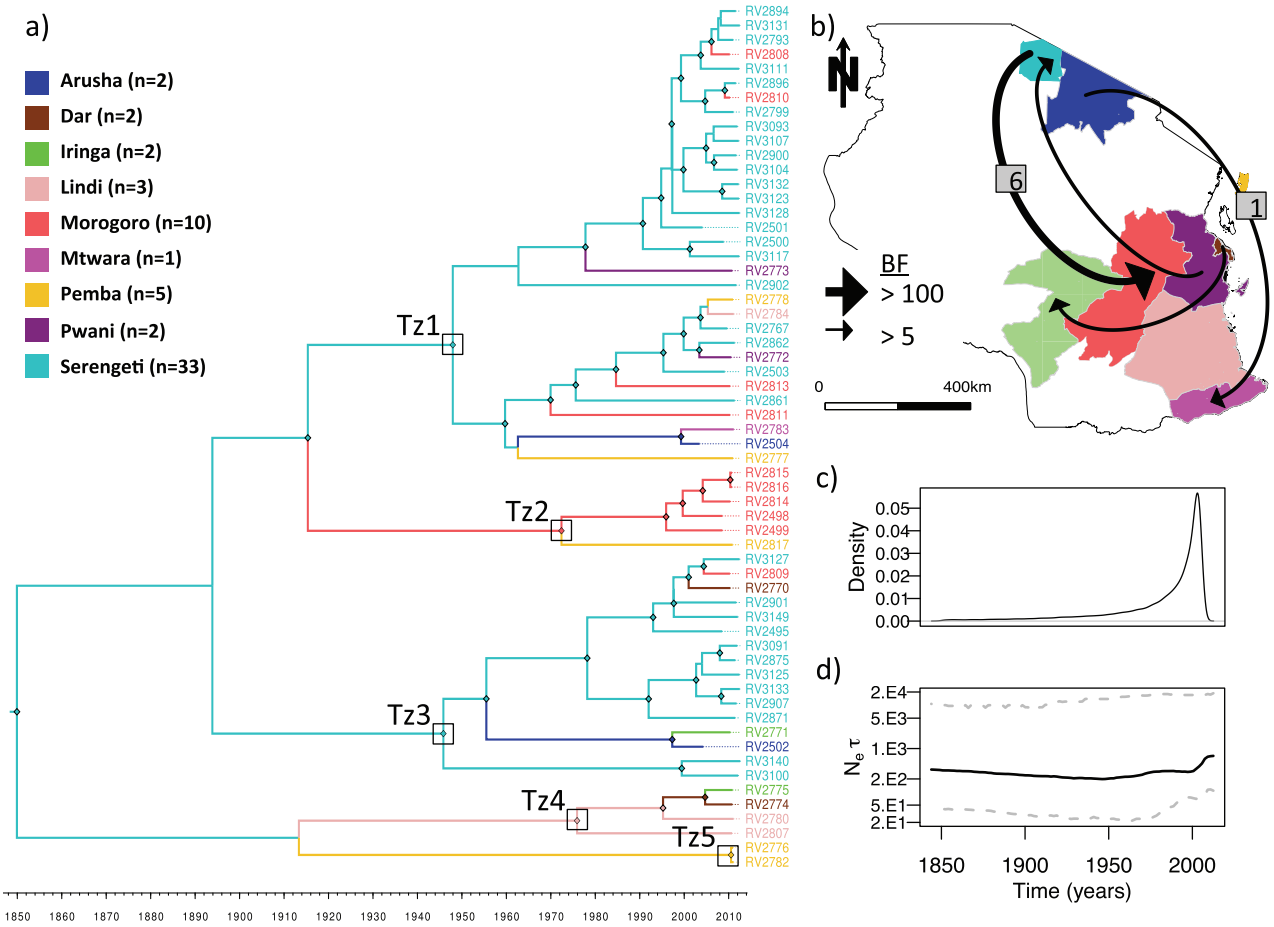


Figure 2. Regional phylogeography among sixty rabies virus whole-genome sequences sampled in Tanzania from 2003 to 2012: (a) an MCC tree with branches coloured according to the most probable posterior location of its descendent node inferred by discrete-state phylogeographic reconstruction in BEAST. Five major phylogenetic groups (Tz1-5) are annotated on the tree and node symbols indicate node posterior support ≥0.9. (b) The four most significant dispersal pathways indicated by BF results from a BSSVS procedure in BEAST with the median number of transitions estimated by Markov jump counts indicated in cases where posterior support for a transition was >0.7. (c) Markov jump densities for total number of transitions through time. (d) Bayesian Skyline plot showing $N_e\tau$, the product of the effective population size (N_e), and the generation time in years (τ) through time.

Table 1. Raw median genetic distance within each of the five main rabies virus lineages identified in Tanzania.

	Genetic distance	
RV lineage	No. of substitutions per site	No. of SNPs
Tz1	9.47E-03	110
Tz2	2.27E-03	27
Tz3	8.83E-03	95
Tz4	4.17E-03	49.5
Tz5	0	0

presence of multiple lineages, some of which are evident even based on partial genome data. Using WGS, we identified five distinct lineages with high posterior probability support (annotated in Fig. 2)—median pairwise genetic distance within each lineage is listed in Table 1. Bayesian phylogenetic reconstruction of WGS yielded a mean evolutionary rate of 1.44×10^{-4} substitutions/site/year (95% highest posterior density: 5.78×10^{-7} to 3.19×10^{-4}), similar to previous estimates for N gene and G gene evolution (David et al. 2007; Talbi et al. 2009; Ahmed et al. 2015).

3.2 Phylogeography of RABV in Tanzania

Within Tanzania, we found evidence of phylogeographic structure (AI = 0.70, $P < 0.001$), similar to other estimates of African intra-country AI values (Table 2) (Talbi et al. 2009). Compared with the strong spatial structure between countries and larger spatial aggregations (Table 3), this indicates more fluid and dynamic dispersal patterns within Tanzania, as has been found in other African countries (Talbi et al. 2010). Across the posterior distribution of trees, there were 24 (95% highest posterior density: 21–28) independent lineage movement events. Using a summarized history of Markov Jump counts across the phylogeny, we found that 43 per cent of these migrations occurred in the most recent ten years of the phylogeny (Fig. 2c). A BSSVS procedure in BEAST identified eighteen potential diffusion pathways to explain the observed phylogeographic patterns in the posterior distribution. However, only four received substantial support based on BF > 5 (Fig. 2). Support was particularly strong for dispersal from the Serengeti District to Morogoro

Table 2. Degree of within-country rabies virus spatial admixture in three African countries measured using an Association Index (AI), which ranges from 0 indicating strong population subdivision to 1 indicating complete panmixis. Results are shown for Tanzania (this study) and Algeria and Morocco (Talbi et al. 2010).

Country	AI [95% BCI]	P	No of sequences	No of locations	Median distance (km)	Min	Max
Algeria	0.67 [0.62–0.73]	<0.001	117	20	233.08	23.27	674.62
Morocco	0.55 [0.51–0.63]	<0.001	133	28	326.45	28.62	926.70
Tanzania	0.70 [0.60-0.79]	<0.001	60	9	439.76	39.05	1088.26

BCI, Bayesian confidence interval.

Table 3. Degree of spatial admixture between rabies virus samples from Africa according to an Association Index (AI). AI values can range from 0 indicating complete population subdivision to 1 indicating complete panmixis.

Data	Spatial clustering level	No. of groups	AI [95% BCI]	P
N405	Sub-continent	3	0.04 [0.02–0.05]	<0.01
	Country	14	0.06 [0.05–0.08]	<0.01
N1350	Sub-continent	3	0.06 [0.04–0.08]	<0.01
	Country	10	0.13 [0.09–0.13]	<0.01

BCI, Bayesian confidence interval. Datasets of partial (N405) and full (N1350) nucleoprotein sequences were tested at two levels of spatial aggregation: (1) sub-continent geographical partitions relative to Tanzania (three states: Tanzania, neighbouring country, and other African country) and (2) country of origin.

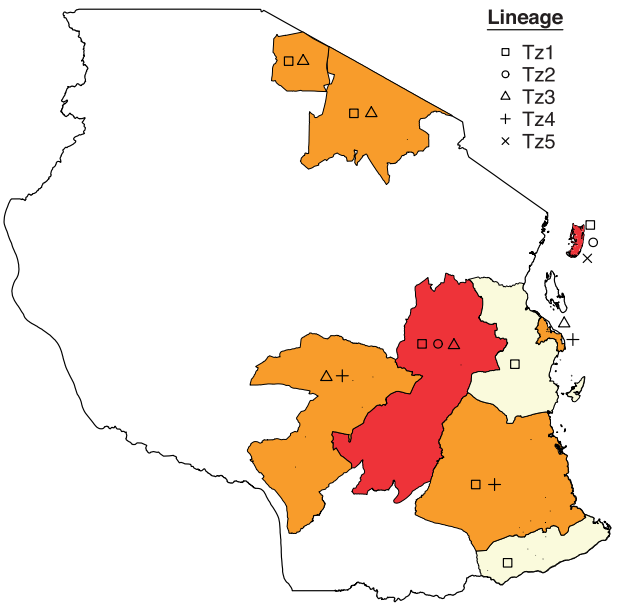


Figure 3. Spatial distribution of rabies virus lineages sampled from regions in Tanzania between 2003 and 2012 with a colour gradient (yellow to red) indicating the total number of lineages (low to high) sampled in each region.

(BF = 135.30), and Markov jump counts estimated a median of six (range: 3–10) migrations along this dispersal route, with at least one (range: 1–3) occurring on a branch representing a period of less than 2 years. Most lineages were sampled in more than one region in Tanzania, with some distributed across a larger geographic area than others (Fig. 3). The largest lineage, Tz1, contains not only a cluster of Serengeti samples but also encompassed samples from a larger geographic extent and was found in eight out of the nine districts sampled. The Bayesian skyline plot revealed that the effective population size has remained fairly constant over the past 150 years (Fig. 2d). Because of the much higher availability of samples from the Serengeti District, we also conducted a coarser phylogeographic analysis grouping sequences into 'North', 'South', or 'Pemba Island'. This identified a predominance of north to south dispersal (estimated thirteen independent movements compared with one movement south to north) and evidence of dispersal back and forth between the North of mainland Tanzania and Pemba Island (Supplementary Fig. S4).

4 Discussion

Using viral genetic data from a hierarchy of spatial scales and varying levels of genome coverage, we were able to demonstrate the advantages of whole-genome resolution to describe the spatio-temporal dynamics of endemically circulating canine rabies viruses.

We found a clear phylogeographic structure between countries in Africa, which could be identified with partial genome data. This suggests that the majority of dispersal occurs at a within-country scale, and control programs would be most appropriate at a national scale with strategies to deal with potential incursions from other countries. However, at regional and local scales within Tanzania, the discriminatory power of partial genome data became too limited to reveal fine-scale population structure that could aid the effectiveness of control interventions. In contrast, WGS provided the resolution to genetically distinguish between all samples and produced a well-supported phylogeny. Although sub-genomic information has utility at broad phylogeographic scales (e.g., Horton et al. 2015), this finding supports the application of WGS for studies aiming to discern population structure at a scale most relevant to control.

Although large-scale population structure according to sub-continental areas or country-specific lineages was apparent from sub-genomic data, we also found incidences of occasional large-scale admixture (Supplementary Figs S1 and S2). We continued to find evidence at increasingly fine scales: Tanzania and other countries within the Africa 1b clade showed signs of admixture particularly when countries shared a border and within country movements of RABV facilitated by humans were a feature of Tanzanian RABV. Even at a very local scale, within the Serengeti District, we found multiple co-circulating lineages. This recurrent theme may be a characteristic of endemic RABV in Africa reflecting decades of endemic circulation and human-mediated introductions. Similar patterns have been observed in canine rabies-endemic countries in South and Southeast Asia (Lumlertdacha et al. 2006; Matsumoto et al. 2013; Ahmed et al. 2015), where the presence of co-circulating lineages was attributed to a

combination of historical introductions from neighbouring countries and human-mediated movements.

Much of the viral population structure we found within Tanzania is consistent with initial invasive waves that have persisted endemically, with structure eroding over time aided by human-mediated movement. Historical accounts describe a rabies outbreak in southern regions of Tanzania in the mid-1950s, which spread throughout the country and was recorded in the Serengeti District in the late 1970s (Magembe 1985; Siongok and Karama 1985). This invasion perhaps shaped the initial RABV population structure in Tanzania. The timeline of regional scale migrations (Fig. 2c) indicates a strong rise in viral dispersal around this period, possibly reflecting increasing human connectivity. Samples in Tz1 (Fig. 2) share partial N gene identity with samples from Kenya (Fig. 1), indicating a shared evolutionary history that may relate to an initial outbreak across both countries. Lineages Tz1 and Tz3 appear to have a wide geographic distribution (Fig. 3), and we found evidence of a general north to south dispersal pattern in Tanzania (Supplementary Fig. S4). This highlights the potential for widespread dispersal should an incursion occur from Kenya and suggests a need for co-operative, cross-border rabies management once a national control program is established. Variation in the spread of different lineages across Tanzania may reflect the influence of heterogeneous landscape features or dog population structure in impeding or facilitating RABV dissemination. Identifying and quantifying such features is the logical next step to provide additional information that can inform control programmes, such as identifying and strengthening pre-existing barriers. Recent extensions of phylogeographic techniques have highlighted how this might be achieved using an integrated approach combining evolutionary and ecological analyses to quantify drivers of viral transmission (Lemey et al. 2014; Magee et al. 2014).

The ancestry of lineages Tz4 and 5 (denoting lineages found in the southern mainland and on the Island of Pemba, respectively) points toward an introduction from the north of Tanzania (posterior probability = 0.62), but the uncertainty of this ancestral location and the ancestral node itself (posterior = 0.45) suggest these may be distinct historical lineages with low-level persistence or undetected circulation elsewhere in Tanzania. Alternatively, these clusters may represent instances where lineages from external sources have more recently invaded and resulted in sustained transmission in Tanzania—partial genome phylogenies indicate Tz4 is related to clusters containing samples from countries south of Tanzania, South Africa and Mozambique. Furthermore, Tz5 shares a common ancestry with a Nigerian and several Central African Republic samples. This again underlines the threat of re-invasion or introduction from external sources and the potential value of whole-genome resolution to robustly and more accurately identify sources of new introductions. To date, the majority of RABV genetic data from Africa comes from partial genome analyses; however, our data suggest that whole-genome characterization would be valuable and should be an aim for the future.

Islands represent isolated landscapes with natural barriers to dispersal, but incidents of RABV outbreaks instigated by human-mediated introductions (e.g., to islands in Indonesia [Windiyaningsih et al. 2004; Susilawathi et al. 2012]) have often been recorded. Sequences from the island of Pemba were suggestive of multiple introductions from various sources with evidence of dispersal to and from the mainland (Supplementary Fig. S4). Samples from Pemba were scattered throughout the phylogeny, and the most divergent lineage Tz5 consisted of two samples from Pemba. The distribution of these lineages may reflect earlier invasions of RABV into Pemba from elsewhere in Tanzania (Tz1 and 2) and the African continent (Tz5), resulting from Pemba's location on an important trading route. However, since these lineages were not resampled and no cases have been detected from Pemba in over a year (Lushasi pers. comm.), RABV may have been only transiently circulating. Nevertheless, lessons from Indonesia, where lack of a swift and coordinated response to a rabies incursion led to an epidemic (Townsend et al. 2013), highlight the importance of active surveillance and rapid response to incursions.

Although we found evidence of RABV spatial structure within Tanzania, it was evident that dispersal was also frequent, with at least one long distance migration occurring within a small temporal window (<2 years). Lineage movements occurring on branches representing short evolutionary times, such as those identified between Serengeti and Morogoro (>750 km), indicate rates of dispersal much higher than those recorded for endemic wildlife rabies (Biek 2007) and imply human mediated movements as seen in parts of North Africa (Talbi et al. 2010). Movement of pastoralist and agro-pastoralist communities from the Lake zone in Northern Tanzania to southern regions, for example, Morogoro have been ongoing since the 1950s (Walsh 2008), with a major influx reported from 2003 to 2006 (PINGO's Forum 2013), attributed to climate change induced drought and forced evictions from newly protected areas (Kideghesho et al. 2013). During these migrations, pastoralists are accompanied by their dogs, which may facilitate the long distance movement of animals incubating the virus prior to transmission. On further inspection, we found that Morogoro samples indicated as instances of long-distance translocation came from rural areas where pastoralists are likely to migrate to, whereas other samples from urban Morogoro formed a distinct cluster (Fig. 2). We found three RABV lineages in Morogoro from only ten samples (Fig. 3). Quantifying networks of human-mediated movements (including livestock trade) would provide a valuable proxy for connectivity for many zoonotic diseases affecting domesticated animals.

Further to our findings of regional admixture in Tanzania, we also observed considerable diversity at a very local scale, that is, within the Serengeti District, with several lineages co-circulating. These lineages appear to have persisted for at least 20 years (older Serengeti samples obtained from GenBank also cluster within these lineages, indicated in Fig. 1) despite dog vaccination campaigns having been conducted in the district since 1996. While vaccination coverage has varied substantially across years and between villages (Viana et al. 2015) rabies incidence has at times significantly declined, for example, falling by 97 per cent in the late 1990s (Cleaveland et al. 2003). Yet our genetic data show that, despite substantial declines in incidence, these lineages must have persisted at very low levels within the district or subsequently reinvaded from neighbouring districts. Without sampling the surrounding districts, it is not possible to distinguish between these possibilities. The skyline plot indicated a stable effective RABV population size (N_e) through time (Fig. 3d). While this could be expected for an endemic pathogen, it is worth noting that there was no evidence of viral population size reductions in response to vaccination campaigns. Rabies control efforts across much of Tanzania have been very limited until recently and high turnover in the dog population (Hampson et al. 2009) likely contributes to the stable persistence of rabies in Tanzania. It has also been noted that geographic structure can obscure local fluctuations in subpopulations while

maintaining the appearance of a constant N_e in skyline plots (Carrington et al. 2005).

This study represents a snapshot of RABV dynamics in Tanzania, indicating that human movements have disseminated RABV out of locally endemic areas at scales relevant to control, that is, administrative units such as regions or districts. These frequent translocations have probably lead to the existence of multiple co-circulating lineages (Fig. 3), but relatively few introductions lead to sustained chains of transmission that are detectable as lineage movement events. However, in disease systems closely associated with human activities, the probability of successful translocation and establishment is likely to be much greater. This suggests that without some level of regional co-operation, Tanzania will be unable to eliminate rabies and maintain freedom from disease. Human movements are often in response to social drivers, which could be used as signals for increased vigilance and surveillance in at risk areas.

Our findings highlight the use of WGS to uncover fine scale transmission patterns that can directly inform control efforts. However, sub-genomic approaches can still have utility at a coarser scale and are more easily obtained, particularly when sample quality is an issue. In particular, they can be used to initially identify admixture between countries, which may indicate the necessity of coordinated regional control programs and surveillance. Co-circulation of multiple lineages and introductions facilitated by humans appear to be a feature of endemic rabies virus and complicate the design of a sustainable control strategy. However, using whole-genome data, we were able to identify sources of dispersal within Tanzania that could direct efforts toward surveillance and control. The finding that humans play an important role in the dynamics of RABV in Tanzania suggests that increasing awareness and dog vaccination in 'high-risk' communities such as pastoralists could help to reduce long-range dispersal. Moreover, the design of enhanced surveillance and containment strategies to mitigate human-mediated incursions and maintain disease freedom should be a priority once control programs are established and elimination is being targeted.

Supplementary data

Supplementary data is available at *VEVOLU Journal* online.

Acknowledgements

We acknowledge the Ministries of Livestock and Fisheries Development and of Health and Social Welfare, Tanzania National Parks, Tanzania Wildlife Research Institute, Ngorongoro Conservation Area Authority, Tanzania Commission for Science and Technology, and National Institute for Medical Research for permission and collaboration; and the Arusha, Mtwara and Mwanza Veterinary Investigation Centres, Pemba Veterinary Office and the Frankfurt Zoological Society for logistical and technical support. We are grateful to the veterinary and livestock field officers, healthworkers, and paravets throughout Tanzania and in particular to the following research assistants: Joel Changalucha, Zilpah Kaare, Gurdeep Kour, Ahmed Lugelo, Kennedy Lushasi, Mathias Magoto, Khadija Said, and Renatus Herman. We thank Pelagia Muchuruza, Matthew Maziku, Geoffrey Mchau, Eberhard Mbunda, Francois-Xavier Meslin, and Bernadette Abela-Ridder for their collaborations through the Bill & Melinda Gates/WHO rabies elimination demonstration project in Tanzania. This work was supported by the Wellcome Trust, the Medical Research Council, the UBS Optimus Foundation, Lincoln Park Zoo, and the Research and Policy for Infectious Disease Dynamics Program of the Science and Technology Directorate, Department of Homeland Security, Fogarty International Centre, National Institute of Health.

Conflict of interest: None declared.

References

Ahmed, K. et al. (2015) 'Molecular Epidemiology of Rabies Viruses Circulating in Two Rabies Endemic Provinces of Laos, 2011–2012: Regional Diversity in Southeast Asia', *PLoS Neglected Tropical Diseases*, 9: e0003645.

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/> accessed 4 Jun 2014.

Ayres, D. L. et al. (2012) 'BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics', *Systematic Biology*, 61: 170–3.

Bedford, T. et al. (2014) 'Integrating Influenza Antigenic Dynamics with Molecular Evolution', *eLife*, 2014: 1–26.

Biek, R. and Real, L. A. (2010) 'The Landscape Genetics of Infectious Disease Emergence and Spread', *Molecular Ecology*, 19: 3515–31.

—— et al. (2007) 'A High-Resolution Genetic Signature of Demographic and Spatial Expansion in Epizootic Rabies Virus', *Proceedings of the National Academy of Sciences of the United States of America*, 104: 7993–8.

—— et al. (2015) 'Measurably Evolving Pathogens in the Genomic Era', *Trends in Ecology & Evolution*, 30: 306–13.

Bielejec, F. et al. (2011) 'SPREAD: Spatial Phylogenetic Reconstruction of Evolutionary Dynamics', *Bioinformatics (Oxford, England)*, 27: 2910–2.

—— et al. (2014) 'Inferring Heterogeneous Evolutionary Processes through Time: From Sequence Substitution to Phylogeography', *Systematic Biology*, 63: 493–504.

Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. (2008) *Applied Spatial Data Analysis with R*. New York: Springer.

Bolger, A. M., Lohse, M., and Usadel, B. (2014) 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data', *Bioinformatics*, 30: 2114–20.

Bourhy, H. et al. (2008) 'The Origin and Phylogeography of Dog Rabies Virus', *The Journal of General Virology*, 89: 2673–81.

Brunker, K. et al. (2012) 'Integrating the Landscape Epidemiology and Genetics of RNA Viruses: Rabies in Domestic Dogs as a Model', *Parasitology*, 139: 1899–913.

Carrington, C. et al. (2005) 'Invasion and Maintenance of Dengue Virus Type 2 and Type 4 in the Americas', *Journal of Virology*, 79: 14680–7.

Cleaveland, S. et al. (2003) 'A Dog Rabies Vaccination Campaign in Rural Africa: Impact on the Incidence of Dog Rabies and Human Dog-Bite Injuries', *Vaccine*, 21: 1965–73.

David, D. et al. (2007) 'Identification of Novel Canine Rabies Virus Clades in the Middle East and North Africa', *The Journal of General Virology*, 88: 967–80.

Denduangboripant, J. et al. (2005) 'Transmission Dynamics of Rabies Virus in Thailand: Implications for Disease Control', *BMC Infectious Diseases*, 5: 52.

Drummond, A. J. et al. (2003) 'Measurably Evolving Populations', *Trends in Ecology & Evolution*, 18: 481–8.

—— et al. (2005) 'Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences', *Molecular Biology and Evolution*, 22: 1185–92.

—— et al. (2012) 'Bayesian Phylogenetics with BEAUti and the BEAST 1.7', *Molecular Biology and Evolution*, 29: 1969–73.

Ferreira, M. A. R., and Suchard, M. A. (2008) 'Bayesian Analysis of Elapsed Times in Continuous-Time Markov Chains', *Canadian Journal of Statistics*, 36: 355–68.

Greger, M. (2007) 'The Human/Animal Interface: Emergence and Resurgence of Zoonotic Infectious Diseases', *Critical Reviews in Microbiology*, 33: 243–99.

Hampson, K. et al. (2007) 'Synchronous Cycles of Domestic Dog Rabies in Sub-Saharan Africa and the Impact of Control Efforts', *Proceedings of the National Academy of Sciences of the United States of America*, 104: 7717–22.

—— et al. (2009) 'Transmission Dynamics and Prospects for the Elimination of Canine Rabies', *PLoS Biology*, 7: e53.

Hayman, D. T. S. et al. (2011) 'Evolutionary History of Rabies in Ghana', *PLoS Neglected Tropical Diseases*, 5: e1001.

Holmes, E. C., and Grenfell, B. T. (2009) 'Discovering the Phylodynamics of RNA Viruses', *PLoS Computational Biology*, 5: e1000505.

Horton, D. L. et al. (2015) 'Complex Epidemiology of a Zoonotic Disease in a Culturally Diverse Region: Phylogeography of Rabies Virus in the Middle East', *PLOS Neglected Tropical Diseases*, 9: e0003569.

Jombart, T. (2008) 'adegenet: A R Package for the Multivariate Analysis of Genetic Markers', *Bioinformatics (Oxford, England)*, 24: 1403–5.

Jones, K. E., et al. (2008) 'Global Trends in Emerging Infectious Diseases', *Nature*, 451: 990–3.

Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.

Kideghesho, J. et al. (2013) 'Emerging Issues and Challenges in Conservation of Biodiversity in the Rangelands of Tanzania', *Nature Conservation*, 6: 1–29.

Kissi, B., Tordo, N., and Bourhy, H. (1995) 'Genetic Polymorphism in the Rabies Virus Nucleoprotein Gene', *Virology*, 209: 526–37.

Knobel, D. L. et al. (2005) 'Re-Evaluating the Burden of Rabies in Africa and Asia', *Bulletin of the World Health Organization*, 83: 360–8.

Kühnert, D., Wu, C.-H., and Drummond, A. J. (2011) 'Phylogenetic and Epidemic Modeling of Rapidly Evolving Infectious Diseases', *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 11: 1825–41.

Kuzmin, I. V. et al. (2008) 'Complete Genomes of Aravan, Khujand, Irkut and West Caucasian Bat Viruses, with Special Attention to the Polymerase Gene and Non-Coding Regions', *Virus Research*, 136: 81–90.

Kuzmina, N. A. et al. (2013) 'The Phylogeography and Spatiotemporal Spread of South-Central Skunk Rabies Virus', *PloS One*, 8: e82348.

Lanfear, R. et al. (2012) 'Partitionfinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses', *Molecular Biology and Evolution*, 29: 1695–701.

Lembo, T. et al. (2007) 'Molecular Epidemiology Identifies Only a Single Rabies Virus Variant Circulating in Complex Carnivore Communities of the Serengeti', *Proceedings. Biological Sciences/ The Royal Society*, 274: 2123–30.

—— et al. (2008) 'Exploring Reservoir Dynamics: A Case Study of Rabies in the Serengeti Ecosystem', *Journal of Applied Ecology*, 45: 1246–57.

Lemey, P. et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLoS Computational Biology*, 5: e1000520.

—— et al. (2014) 'Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2', *PLoS Pathogens*, 10: e1003932.

Lewin-koh, A. N. J. et al. (2009) *Maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.8-23. <http://cran.r-project.org/web/packages/maptools/index.html> accessed 1 Sept 2014.

Li, H., and Durbin, R. (2009) 'Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform', *Bioinformatics (Oxford, England)*, 25: 1754–60.

—— et al. (2009) 'The Sequence Alignment/Map Format and SAMtools', *Bioinformatics (Oxford, England)*, 25: 2078–9.

Lumlertdacha, B. et al. (2006) 'Complex Genetic Structure of the Rabies Virus in Bangkok and Its Surrounding Provinces, Thailand: Implications for Canine Rabies Control', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100: 276–81.

Magee, D. et al. (2014) 'Combining Phylogeography and Spatial Epidemiology to Uncover Predictors of H5N1 Influenza A Virus Diffusion', *Archives of Virology*, 160: 215–24.

Magembe, S. (1985) 'Epidemiology of Rabies in United Republic of Tanzania', in Kuwer, E. et al. (eds) *Rabies in the Tropics*, pp. 392–8. New York: Springer Berlin Heidelberg.

Marston, D. A. et al. (2007) 'Comparative Analysis of the Full Genome Sequence of European Bat Lyssavirus Type 1 and Type 2 with Other Lyssaviruses and Evidence for a Conserved Transcription Termination and Polyadenylation Motif in the G-L 3' Non-Translated Region', *The Journal of General Virology*, 88: 1302–14.

—— et al. (2013) 'Next Generation Sequencing of Viral RNA Genomes', *BMC Genomics*, 14: 444.

Matsumoto, T. et al. (2013) 'Molecular Epidemiology of Human Rabies Viruses in Sri Lanka', *Infection, Genetics and Evolution*, 18: 160–7.

Minin, V. N., and Suchard, M. A. (2008) 'Counting Labeled Transitions in Continuous-Time Markov Models of Evolution', *Journal of Mathematical Biology*, 56: 391–412.

Mollentze, N. et al. (2014) 'A Bayesian Approach for Inferring the Dynamics of Partially Observed Endemic Infectious Diseases from Space-Time-Genetic Data', *Proceedings. Biological Sciences/ The Royal Society*, 281: 20133251.

Nel, L. H. (2013) 'Discrepancies in Data Reporting for Rabies, Africa', *Emerging Infectious Diseases*, 19: 529–33.

Nelson, M. I. et al. (2015) 'Global Migration of Influenza A Viruses in Swine', *Nature Communications*, 6: 6696.

Nunes, M. R. T. et al. (2014) 'Air Travel is Associated with Intracontinental Spread of Dengue Virus Serotypes 1-3 in Brazil', *PLoS Neglected Tropical Diseases*, 8: e2769.

Paradis, E., Claude, J., and Strimmer, K. (2004) 'APE: Analyses of Phylogenetics and Evolution in R Language', *Bioinformatics*, 20: 289–90.

Parker, J., Rambaut, A., and Pybus, O. G. (2008) 'Correlating Viral Phenotypes with Phylogeny: Accounting for Phylogenetic Uncertainty', *Infection, Genetics and Evolution*, 8: 239–46.

Pebesma, E. J., and Bivand, R. S. (2005) 'Classes and Methods for Spatial Data in R', *R News*, 5: 9–13.

PINGO's Forum (2013) *Pastoralists Indigenous Non Governmental Organizations' Forum Annual Report 2012-2013* <http://www.pingosforum.or.tz> accessed 23 Nov 2014.

Rambaut, A. et al. (2014) 'Tracer v1.6.' Computer program and documentation distributed by the author <http://beast.bio.ed.ac.uk/Tracer> accessed 4 Feb 2015.

Real, L. A. et al. (2005) 'Unifying the Spatial Population Dynamics and Molecular Evolution of Epidemic Rabies Virus', *Proceedings of the National Academy of Sciences of the United States of America*, 102: 12107–11.

Rupprecht, C. E., Hanlon, C. A., and Hemachudha, T. (2002) 'Rabies Re-examined', *The Lancet Infectious Diseases*, 2: 327–43.

Shwiff, S., Hampson, K., and Anderson, A. (2013) 'Potential Economic Benefits of Eliminating Canine Rabies', *Antiviral Research*, 98: 352–6.

Siongok, T., and Karama, M. (1985) 'Epidemiology of Human Rabies in Kenya', in E., Kuwer et al. (eds) *Rabies in the Tropics*, pp. 445–50. New York: Springer Berlin Heidelberg.

Stamatakis, A. et al. (2012) 'RAxML-Light: A Tool for Computing Terabyte Phylogenies', *Bioinformatics (Oxford, England)*, 28: 2064–6.

Susilawathi, N. M. et al. (2012) 'Epidemiological and Clinical Features of Human Rabies Cases in Bali 2008-2010', *BMC Infectious Diseases*, 12: 81.

Talbi, C. et al. (2009) 'Evolutionary History and Dynamics of Dog Rabies Virus in Western and Central Africa', *The Journal of General Virology*, 90: 783–91.

—— et al. (2010) 'Phylodynamics and Human-Mediated Dispersal of a Zoonotic Virus', *PLoS Pathogens*, 6: e1001166.

Tao, X. Y. et al. (2009) 'Molecular Epidemiology of Rabies in Southern People's Republic of China', *Emerging Infectious Diseases*, 15: 1192–8.

Townsend, S. E. et al. (2013) 'Designing Programs for Eliminating Canine Rabies from Islands: Bali, Indonesia as a Case Study', *PLoS Neglected Tropical Diseases*, 7: e2372.

Viana, M. et al. (2015) 'Dynamics of a Morbillivirus at the Domestic–Wildlife Interface: Canine Distemper Virus in Domestic Dogs and Lions', *Proceedings of the National Academy of Sciences of the United States of America*, 112: 1464–9.

Walsh, M. (2008) *Pastoralism and Policy Processes in Tanzania: Mbarali Case Study*, Report to the Tanzania Natural Resource Forum, Arusha.

Wang, T. H. et al. (2001) 'Identification of Shared Populations of Human Immunodeficiency Virus Type 1 Infecting Microglia and Tissue Macrophages Outside the Central Nervous System', *Journal of Virology*, 75: 11686–99.

WHO. (2005) 'WHO Expert Consultation on Rabies', *World Health Organization technical report series*, 931: 1–121.

Windiyaningsih, C. et al. (2004) 'The Rabies Epidemic on Flores Island, Indonesia (1998-2003)', *Journal of the Medical Association of Thailand = Chotmaihet thangphaet*, 87: 1389–93.

Woodroffe, R., and Donnelly, C. A. (2011) 'Risk of Contact between Endangered African Wild Dogs *Lycaon pictus* and Domestic Dogs: Opportunities for Pathogen Transmission', *Journal of Applied Ecology*, 48: 1345–54.

Woolhouse, M. E. J. (2002) 'Population Biology of Emerging and Re-emerging Pathogens', *Trends in Microbiology*, 10(10 Suppl): S3–7.

6.3: Article 31 in review with Virus Evolution

Analysis of a rabies virus host shift event reveals an increase in sub-viral populations within the new host reservoir

Summary: This article represents a comprehensive WGS analysis of a cohort of RABVs from Turkish red foxes. It is discussed twice within this thesis because two distinct analyses are undertaken. The first discussed here, is the utilisation of Bayesian inferences to investigate the source and timing of the CST event. Transmission of the fox RABV back into dogs is also investigated. The second, discussed in Chapter 7.2, is the utilisation of the deep sequence data to investigate the influence of sub-consensus populations on virus adaptation.

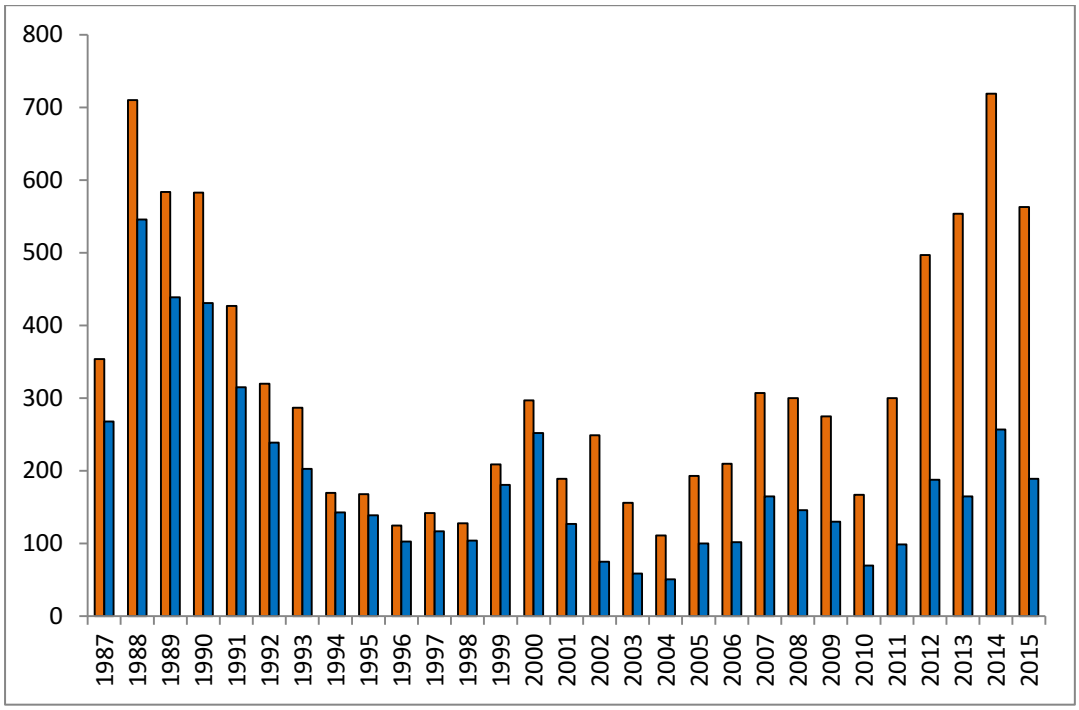
The fox samples were sequenced without amplification, using our developed methods (see Chapters 2 and 3) resulting in an unbiased dataset without the introduction of PCR errors. This methodology also provided sequences from the host, used to confirm the host species of each sample, which was essential to confirm that canine RABV was the origin of the fox epizootic. Complete genome sequences from 31 RABV positive brain samples, were obtained; 21 of which were RABV cases from the fox epizootic covering the time period from its emergence to the present day. Prior to 1999, fox rabies was rarely detected in Turkey, however, after 1999, the cases of fox rabies disproportionately increased and although control measures have reduced fox rabies, it continues to the present day (Figure 12). Analysis of the RABV dataset indicated that the virus population in Turkey reduced from the 1990s. This decrease corroborated the epidemiological data and coincided with the concerted rabies control efforts in dogs at the time.

By utilising WGS and Bayesian inferences, using host as a discrete trait, we demonstrated that the most likely source of the fox enzootic was from dogs, and furthermore the host shift occurred in 1997 (+/- 1 year), which corroborates the epidemiological information. Previous conclusions that the fox epizootic in Turkey was a result of a single host shift are challenged here. There are two highly related fox clades, rooted by RV2975D (Izmir 1999). Both clades are genetically distinct from contemporaneous fox viruses isolated from other regions (Erzurum). Closest viruses to this ancestor were detected in dogs in Izmir/Manisa, and epidemiological information indicated this was likely to be the location where the host shift occurred. The 2 clades circulated in the same area and time, including contemporary RABV samples.

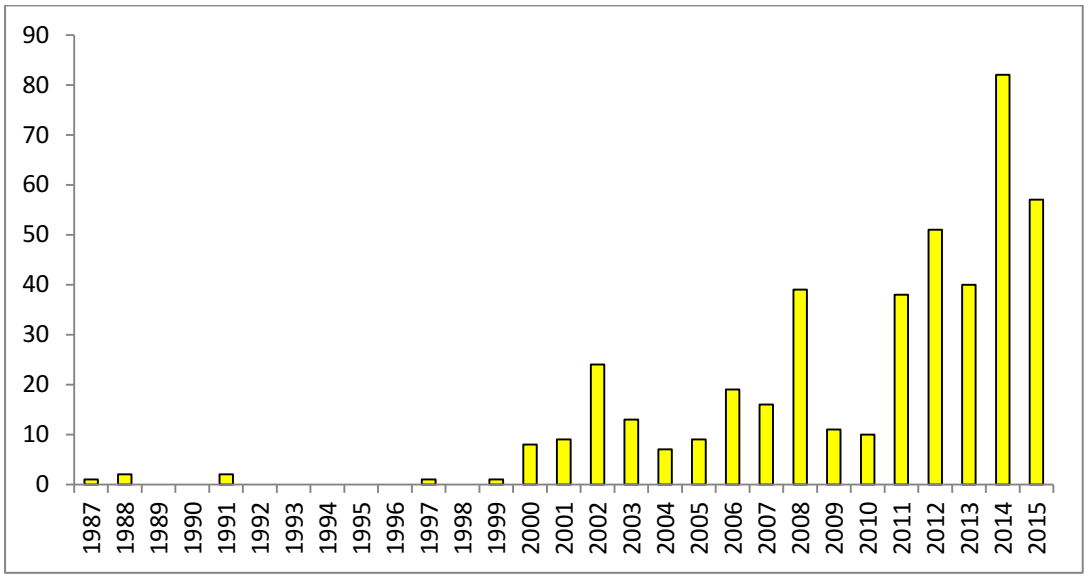
Figure 12: Incidence of animal rabies cases in Turkey between 1987 and 2015.

The total number of (A) animal cases (orange) and dog cases (blue), (B) fox cases (yellow) are shown. (Source: Rabies bulletin Europe).

A



B



Evidence presented here suggested that transmission between dog and fox populations occurs readily. The inclusion of dog virus sequences from similar regions to the fox samples provided important data regarding the possible frequency of transmissions between fox and dog populations. Combined with evidence for lack of adaptation at the consensus level between the Cubuk dog/fox pair, a clear threat of this RABV lineage re-emerging in dogs is apparent if vaccination of dogs is not sustained and effective at local levels.

The high sequence identity between fox and dog RABVs demonstrated transmission is not dependent on host specific residues in the RABV genome. More importantly, comprehensive analysis of the fox sequences over 15 years after a host shift from dog to fox, has demonstrated a lack of evidence for host adaptation at the consensus level. It was generally believed that non-fox RABV lineages need higher doses than fox RABVs to successfully infect a new host, indicating a species specificity of the respective reservoir host strain encrypted in the prevailing genetic setting. In this study, however, the results imply that consensus level 'fixed' adaptations in the virus genome are not necessary for successful onward maintenance in new host species, i.e. foxes.

Data such as complete genome length, and SNP analysis were essential to obtain the fine-resolution required when comparing highly related viruses.

Analysis of a rabies virus host shift event reveals a transitory increase in sub-viral populations within the new host reservoir.

Denise A. Marston[1,2*], Daniel L. Horton[3], Javier Nunez[4], Richard J. Ellis[4], Richard J. Orton[5,6], Nicholas Johnson[1,7], Ashley C. Banyard[1], Lorraine M. McElhinney[1,8], Conrad M. Freuling[9], Müge Firat[10], Nil Unal[10], Thomas Müller[9], Xavier de Lamballerie[2], Anthony R. Fooks[1,8]

[1]Wildlife Zoonoses & Vector-Borne Diseases Research Group, Animal and Plant Health Agency, New Haw, Addlestone, Surrey, United Kingdom

[2] UMR "Émergence des Pathologies Virales" (EPV: Aix-Marseille Univ – IRD 190 – Inserm 1207 – EHESP – IHU Méditerranée Infection), Marseille, France

[3]School of Veterinary Medicine, University of Surrey, Guildford, United Kingdom

[4]Surveillance and Laboratory Services Department, Animal and Plant Health Agency, New Haw, Addlestone, Surrey, United Kingdom

[5]Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom.

[6]MRC-University of Glasgow, Centre for Virus Research, University of Glasgow, Glasgow, United Kingdom

[7]Faculty of Health and Medical Science, University of Surrey, Guildford, United Kingdom.

[8]Institute of Infection and Global Health, University of Liverpool, United Kingdom

[9]Friedrich-Loeffler-Institute, (FLI), Institute of Molecular Virology and Cell Biology, Greifswald-Insel Riems, Germany

[10]Etlik Veterinary Control Central Research Institute, Etlik-Ankara, Turkey

*Corresponding Author: Email: Denise.Marston@apha.gsi.gov.uk

Total word count: 7,816

Abstract

Host shift events play an important role in epizootics as adaptation to new hosts can profoundly affect the spread of the disease and the measures needed to control it. During the late 1990s, an epizootic in Turkey resulted in a sustained maintenance of rabies virus (RABV) within the fox population. Utilisation of Bayesian inferences to investigate whole genome sequences from a cohort of fox and dog brain tissues from Turkey demonstrated that the epizootic occurred in 1997 (+/- 1 year). Furthermore, these data indicate that the epizootic was most likely due to a host shift from locally infected domestic dogs, rather than an incursion of a novel fox or dog RABV. No evidence was detected for virus adaptation to foxes at consensus sequence level; therefore, the deep sequence data was analysed to investigate the influence of sub-consensus populations on host shift events. Viral heterogeneity was measured in all RABV samples; viruses in the early phase after the host shift had increased heterogeneity, in relation to those in the later stage, possibly indicating a role in establishing transmission within a new host. The dynamics of majority and minority variants are consistent with genetic drift, rather than positive selection. The transient expansion of sub-consensus viral populations in the new host species likely represents the virus adapting to a new environment, perhaps due to increased replication within the CNS resulting in a larger population of viruses, or reflecting the lack of host constraints present in the new host reservoir.

Introduction

The lyssavirus genus is a group of negative-strand RNA viruses characterised by their ability to cause fatal encephalitis. The type species for lyssaviruses, *Rabies lyssavirus* is transmitted by a wide range of mammalian hosts within the *Carnivora* and *Chiroptera* orders and has a global distribution. Unlike all other lyssaviruses, rabies viruses (RABVs) have established independent transmission cycles in a broad range of carnivore and meso-carnivore host reservoirs, where particular RABV lineages circulate within host conspecifics. Phylogenetic analyses of RABV sequences have demonstrated

particular RABV lineages associate with individual host species (Nel, et al. 1993; Kuzmin, Hughes, et al. 2008; Hanke, et al. 2016). Host shifts (where the virus is maintained in a new host) infrequently occur, however a study in the US determined that cross-species transmissions (CSTs, which do not result in sustained onward transmission within the new host) involving raccoon variant RABV increased over a 20 year period, in both wildlife and domestic animals posing a significant risk to public health (Wallace, et al. 2014). Identifying successful host shifts using molecular inferences is a useful tool to understand RABV evolution (Daoust, et al. 1996; Leslie, et al. 2006; Kuzmin, et al. 2012; Borucki, et al. 2013). Importantly for lyssaviruses, the molecular mechanisms influencing virus spread between different species and adaptation to a new potential reservoir host are poorly understood, but it is clear that host, viral and ecological factors all play important roles (Streicker, et al. 2010; Mollentze, et al. 2014).

RNA virus populations exhibit a heterogeneous population of viruses within single individuals, often referred to as 'quasispecies' but perhaps more accurately termed 'viral heterogeneity' (Holmes and Moya 2002). This is widely accepted to be due to the absence of proof reading, repair and post-replicative error correction observed for RNA viruses (Domingo and Holland 1997). Viral heterogeneity, represented by nucleotide sequence heterogeneity of the viral RNA, was originally proposed for RABV as a mechanism for viral evolution and adaptation to a new host (Benmansour, et al. 1992). The heterogeneous population of a RABV street strain (European fox isolate) has been investigated previously (Kissi, et al. 1999) using cloned PCR products covering 19% of the genome. This study identified sub-consensus sequences and concluded no host specific consensus level changes were observed, although the amount of heterogeneity observed varied depending on the host used to passage the virus. Furthermore, the fixed RABV strain CVS was passaged experimentally in BHK cells, resulting in selection of a dominant variant that differs genetically and phenotypically from the consensus sequence present in mice or neuroblastoma cell passaged CVS (Morimoto, et al. 1998).

Since the late 1990's until the development of Next Generation Sequencing (NGS) techniques, very little progress in the area of viral heterogeneity was made.

NGS techniques generate unprecedented cost-effective large-scale data, enabling both the generation of complete genomes (Marston, et al. 2013; McGinnis, et al. 2016; Montmayeur and Ng 2017; Parker and Chen 2017) and the analysis of viral sub-consensus populations (Eriksson, et al. 2008; Wright, et al. 2011; Poh, et al. 2013; Raghwani, et al. 2016). The inter-host dynamics of multiple different viral infections have been examined. These range from persistent infections, representing narrow transmission bottlenecks, where few virus particles are transmitted (Fischer, et al. 2010), to acute infections of equine influenza virus and norovirus, which have broad transmission bottlenecks (Murcia, et al. 2010; Bull, et al. 2012). Detailed analysis of Foot and mouth disease virus (FMDV) traced minor variants within and between animal hosts using NGS data (Morelli, et al. 2013). In the lyssavirus field, a number of studies have used NGS to obtain consensus whole genome sequences for phylogenetic analysis (Hoper, et al. 2015; Hanke, et al. 2016; Troupin, et al. 2016). The application of NGS to study lyssavirus host shift events is becoming more commonplace, however, utilising the deep sequence data to investigate viral heterogeneity is an area still largely unexplored (Borucki, et al. 2013; Nadin-Davis, et al. 2017).

Despite concerted control efforts, rabies remains a significant public health problem in many regions of Turkey (Johnson, et al. 2010). The first phylogenetic study to focus on carnivore rabies in Turkey, compared a small conserved region of the RABV nucleoprotein gene (327 base pairs) of eighteen samples from across the country and identified three clades of RABV within Turkey that clustered geographically (Johnson, et al. 2003). More recently, a comprehensive study of RABV in the Middle East, confirmed the clades and highlighted the importance of trans boundary movements in the epidemiology of rabies in the region (Horton, et al. 2015).

The original phylogenetic study demonstrated that fox rabies was emerging in western Turkey and was related to rabies circulating in dogs from the same part of the

country (Johnson, et al. 2003). It is widely accepted that a similar host shift caused the fox enzootic in Europe in the second half of the 20th century (McElhinney, et al. 2011). Determining whether the virus represented a host shift from a dog RABV into a sylvatic species, multiple CSTs without a host shift, or had been circulating undetected in foxes was not possible previously using partial gene data and lack of genetic barcoding to confirm the host species. Understanding the source of infection had direct implications on the control strategy implemented, as wildlife reservoirs required a completely different approach to that used to control RABV in dog populations (Fusaro, et al. 2013). Epidemiological records showed that rabies spread eastwards and southwards within the fox population from a presumed epicenter near the western city of Izmir prompting the introduction of oral rabies vaccination (ORV) campaigns targeting wildlife species, covering the provinces of Izmir and Manisa in 2008 (Un, et al. 2012). As a direct result, epidemiological records suggested fox rabies was eliminated within the vaccination zone by 2010, however, re-emergence of the disease in the original vaccination zone was observed by 2012 (Un, et al. 2012).

In the present study, a cohort of RABV samples from this Turkish fox host shift dataset from western and central Turkey, including all available fox and dog samples between 1999 and 2015 (n=21) were analysed. The principal objectives were to utilise whole genome sequencing (WGS) to (i) confirm the most likely source of the fox enzootic, (ii) refine the date of the host shift, (iii) infer the spatio-temporal phylogenetic relationships based on concatenated coding sequence data and (iv) investigate adaptation in the new fox host resulting in successful onward transmission in the Turkish fox population. Together, these data will provide information useful for rabies control policies and informing our understanding of virus emergence and maintenance in new host reservoirs.

Results

RABV positive brain samples (n=30), confirmed by fluorescent antibody test (FAT), were obtained from rabies suspect animals in Turkey between 1999 and 2015 and one sample from Russia which is included as an outgroup for some analyses (Fig 1 and Table 1). During this period, Turkey reported 434 fox rabies cases to Rabies Bulletin Europe (RBE) (RBE 2017). Twenty viruses represented the host shift dataset which had spread as far East as Ankara and surrounding areas by 2015 including 3 dog samples (RV3163D, Sub 5791D, RV3161D) collected from canine rabies free towns around Ankara, suspected to be infected with the fox RABV lineage. For further comprehensive analyses, the 20 viruses were divided into two different temporal phases: The early phase, 1999-2007 (n=7) including RV1124F the first confirmed fox RABV in the area representing the early phase Clade 2 viruses and RV1126F representing the early phase Clade 1 viruses (Fig 2). The late phase comprised of isolates from 2008-2015 (fox (n=10) and dog (n=3)) (Table 1). RV2975D represented the dog strain circulating in Izmir at the time of the documented host shift and is referred to as 'pre-host shift' in the analyses. Additionally, 10 isolates, from geographic locations outside the host shift region were included (non-host shift dataset) to provide some context to the wider rabies virus epidemiology in Turkey. Reads from the host's brain samples were mapped against the Cytochrome C Oxidase 1 (COI) gene to confirm the host species. The host species of all 31 samples were confirmed using this method (Table 1).

Whole genome sequencing from clinical brain samples proves adequate reads to obtain consensus sequences

All 31 brain samples were sequenced directly from extracted RNA that were subsequently depleted of gDNA and rRNA. The total number of RABV reads, average coverage and percentage to total reads varied between samples (Table 1, Table S1). The lowest number of specific RABV viral reads obtained for a sample was 8,408 reads representing 0.19% of total reads, which was sufficient to obtain a full genome consensus sequence (Table 1). The coverage of reads across the genome was not

uniform, but was consistent between different samples, suggesting that conformation of the native RNA molecules might affect the evenness of the coverage, or other physical constraints affect the efficiency of sequencing across the genome (Table S1, Fig S1).

The source of the fox host shift was locally circulating dog RABV resulting in a successful maintenance in the fox population

The RABVs circulating in Turkey all cluster within previously described Middle East clades (Text S1, Fig S2) (Horton, et al. 2015). The RABVs circulating in Turkey cluster according to geographic location, and the host shift RABV sequences cluster with RV2975D in two distinct clades from a common ancestor (Fig 2, inset). Assessment of models implemented in BEAST indicated very similar AICM (Akaike information criteria model) values and mean substitution rates at 3.10×10^{-4} subs/site/year (95% highest posterior density (HPD) 2.68-3.56 $\times 10^{-4}$ subs/site/year) (Table 2). All models produced Maximum Clade Credibility (MCC trees) with similar topology and time to most recent comment ancestor (TMRCA) at key nodes.

All models support the occurrence of a common ancestor virus of the fox clades that first emerged in the fox population (probability 98%) between 1996 and 1998 (Fig 2). These data strongly support circulation of rabies in dogs, followed by a spill over into foxes at that time. The emergent fox clade is genetically distinct from contemporaneous fox viruses isolated from other regions in Turkey (Erzurum). The closest viruses to this ancestor were detected in Izmir/Manisa, corroborating epidemiological information, indicating that this location is likely to be the region where the host shift occurred. Furthermore, the MCC analysis confirms the spread of rabies from Izmir and surrounding areas westwards to the Ankara region.

There was overall drop in the RABV population size starting in the 1990s (Fig 3a), which coincides with the reduced number of lineages observed in Turkey and reflects the concerted rabies control efforts in dogs at the time. In contrast, the population size of the host shift sequences alone (1999-2015) increased, reflecting the expansion of the virus

in a new host reservoir (Fig 3b). The increase in population size stabilises by 2001 and continues as a stable population to the present day.

Analysis of non-coding regions does not reveal host specific viral adaptation

An alignment of all 31 complete genome sequences identified 2 intergenic regions with single nucleotide (nt) indels in homopolymer regions, most likely due to polymerase error. These indels result in differing genomic lengths between 11,923 and 11,925 depending on presence or absence of one or both of these indels (Table S2). The first indel (position 3192) a homopolymeric region of either 5 or 6 adenosine residues (As) in the M-G intergenic region (Table S2 and S3). This indel differentiates the two fox clades: clade 1 viruses (and RV2975D) are 11,923 nts, clade 2 are 11,924-5 nts long (Fig 2). The sequences from outside the host- shift dataset (n=9) have 5As.

The second indel (position 5348) a homopolymeric region of 7 or 8 As in the poly A signal of the glycoprotein gene in the G-L intergenic region, (Table S2 and S4). All but 2 sequences (sub5792D and RV3160D, both Clade 2) have 7 As in the consensus. The additional 'A' at position 5348 in sub5792F and RV3160F result in a total genome length of 11,925 nts. The majority of clade 2 samples have sub-populations of both 7 and 8 As indicating that this region is not fixed, resulting in different length consensus sequences for highly related viruses (Table S2).

Analysis of coding regions demonstrates no positive selection at consensus level

Concatenated genes (total length 10,815 nts) from the 20 host shift RABV samples were aligned with the pre-host shift RV2975D sequence to investigate SNPs. Eight positions, 2 located in the nucleoprotein, 1 in the phosphoprotein and 5 in the RdRp were identified (Table 3). 2/8 were non-synonymous, located in the RdRp ($977^{Val/Ile}$ and $1871^{Leu/Arg}$). However, these residues are unlikely to be species specific, as the majority of the dog sequences from outside the host-shift region had the fox SNP, rather than RV2975D SNP.

Due to the lack of host specific residues in the dataset, SNPs observed within the host shift dataset were analysed. Across the concatenated genes, 26 SNPs were identified, which represent the clade 1 and clade 2 SNPs and subsequent substitutions accumulating over time (Fig 2 and Table S5). The majority of the substitutions are synonymous; of which one (382^{Leu} RdRp) had 3 different nucleotides in the third position all encoding Leucine (CT**G,** CT**T** and CT**A**). All 4 non-synonymous substitutions observed with the host shift dataset were located where RV1124F clusters from the remaining clade 2 viruses. There is a 5 year time period between RV1142F and the next sampled clade 2 virus (RV2976F -2004) during which time the 4 non-synonymous substitutions occurred (Fig 2, Table S5). It is striking that all 4 non-synonymous changes occurred during the same time period, however, its importance is unclear. Taken together, the majority of substitutions observed within the consensus sequences over time are not due to positive selection after host switching, rather the result of genetic drift over time.

Selection pressure was evaluated by computing the non-synonymous to synonymous ratio (dN/dS) for the consensus sequence of each gene of the samples in this study. The dN/dS ratios of the 5 genes were generated from the dataset, using a reference sequence evolutionarily unrelated to the host shift viruses (RV303RD). Across the entire data set the dN/dS ratios are substantially < 1 with N, M and L genes obtaining ratios below 0.02. P-gene had the highest observed ratio for all viruses (Fig S3A). The average dN/dS ratio across the 5 genes for all viruses analysed was between 0.02 and 0.025 regardless of the sample date (Fig S3B). Such low dN/dS ratios across the genes across the dataset suggest genetic drift is driving the consensus level variation observed, similar results were obtained when using RV1142D (Istanbul, 2001) as the reference sequence (data not shown). However, we note that dN/dS is a relatively conservative measure that may be unable to detect directional selection indicative of a host switch event.

Although the consensus sequences for all 5 genes, either separately or concatenated strongly suggested purifying selection as the main effect on the viruses, this did not take into account selection at specific residues, known to be important in the glycoprotein, which is the main target of the host antibody responses (Evans, et al. 2012). Analysis of individual codons across the glycoprotein did not detect evidence for a change in selection pressure on the glycoprotein associated with the host shift event (data not shown).

Sub-viral population heterogeneity changes during the early phase of the fox epizootic

To further investigate the RABV host shift within the fox population, the deep sequence data was utilised, representing the sub-consensus viral population present in the brain samples. For each sample dataset, the viral reads were mapped against the corresponding consensus sequence. The mapped viral reads were analysed for alternative calls in relation to the reference sequence using a Heterogeneity Index (H Index). The H Index was calculated on the viral reads as the number of alternative calls per million base calls sequenced. The resulting calculation controls for the variation of the depth of coverage within and between virus populations, enabling a direct comparison between individual viruses.

The H Index for all sequences was calculated (Table S6). No correlation was found between the H Index and the total number of viral reads obtained, for samples with total viral reads >25,000 (Table 1 and Table S6). Samples with reads <25,000 had sufficient reads to provide complete genome consensus sequence data, but had a strong correlation between H Index and viral reads, possibly due to low frequency variants being missed, particularly using strict quality criteria. Therefore, heterogeneity and dN/dS analysis of the sub-consensus viral population was undertaken on samples with >25,000 viral reads. Repeat H Index analysis on the same read dataset provided reproducible results (RV2975D: 713 (710); RV2983F: 427 (426)) as did analysis from an independent

RNA extraction from same sample (RV1124F: 713 and 710). Shannon entropy (SE) has been previously used as a measure of diversity with FMDV intra-host data (Morelli, et al. 2013). Independent analysis of the Turkey RABV read dataset to determine the entropy scores resulted in similar trends between both methods. Individual virus heterogeneity was measured using both methods with good correlation, albeit that SE scores had less range between the highest and lowest (SE range: 252 - 957; H Index range: 384 – 1841) (Table S6 and Fig S4). The only exception was RV2982F where the SE value was lower than expected compared to the H Index (Entropy 529; H Index 1048) (Fig S4). The average H-Index scores were compared between the whole host shift dataset, the early stage and late stages of the host shift dataset and the non-host shift viruses (Fig 4). The early and late virus phases were defined using the H Index scores and fits with the start of the vaccination campaign in 2008. The highest heterogeneity was present in the sub-consensus population of the early stage viruses (Average H-Index 1523; StDev 300.4), compared to the viruses within the late stage (861.5; 188.4 p=0.0014). The average heterogeneity score of the host shift dataset was still significantly higher than the non-host shift viruses (p=0.0112) (Fig 4). To further investigate the differences in the heterogeneity present, the H Index of all viruses was plotted over time. These data indicated that the early phase viruses exhibited higher heterogeneity than those in the later phase (p=0.0007) (Fig S5). Furthermore, the non-host shift viruses and pre-host shift RV2975D heterogeneity scores were all significantly lower than the early phase viruses (p=0.0002). The entropy data analysis from the same dataset was in concordance with the H-Index results (data not shown). Due to the strong phylogenetic association between the early and late phase viruses, phylogenetic dependency was investigated using a phylogenetic generalised least squares regression using the Brownian model for correlation structure. The temporal phase and H Index scores were not considered significantly correlated (p=0.2708). Together, these results indicate an expansion of the sub-consensus virus population after the bottleneck of a host switch event, which continued until the virus was established in the fox population. The

heterogeneity subsequently stabilised approximately 10 years after the initial host shift at a level observed in other RABVs sampled across Turkey.

dN/dS analysis indicates increased synonymous changes in the sub-viral population in the early phase host shift viruses

The dynamics of the heterogeneity observed was evaluated by computing the non-synonymous to synonymous ratio (dN/dS) at each position for each gene of the samples in this study. Analysis of the dN/dS ratios observed within the NGS read data including all substitutions was undertaken (Fig S6). In comparison to non-host-shift viruses, which largely have a dN/dS =1 (Fig S6B), a reduction in dN/dS (particularly in the N gene) is observed in the host shift viruses in the early phase, which is not apparent in the later phase (Fig S6A). The matrix (M)-protein and N-protein of RV2982F have significant dN/dS values in comparison to the other viruses in the dataset. Investigation into these regions revealed a SNP with higher than average ratio (C^{30}/T^{136}: Val/Val) and 4 SNPs with 3 bases called, suggesting increased heterogeneity in the N-protein, but not the M-protein (data not shown). The increase in dN/dS ratio in the M-protein of RV2982F indicates a large increase in non-synonymous changes in the sub-consensus viral population, which is present in > 0.5% of the reads (data not shown). Analysis using average dN/dS ratios across the entire concatenated coding region, reveals that the dN/dS ratio was lower in the early phase (3 of the 4 viruses have dN/dS ratios lower than non-host shift viruses) and increased over time to approximately 1, suggesting that the early phase viruses sub-consensus population had more synonymous changes than non-synonymous changes in the years after the host shift in comparison to the non-host shift and late phase RABVs (Fig S6C).

Discussion

RABV host shift events have rarely been documented, and until recently, investigations have been mostly epidemiological in nature, with less emphasis on

understanding the viral molecular adaptations required to establish an infection within a new host reservoir. This study has utilised WGS to investigate the host shift that occurred in Turkey at the end of the 20th Century, estimating the date of the host shift into foxes to within one year, and investigating viral adaptation at the sub-consensus population level. The preparation of the viral RNA for WGS, and the methods employed to obtain deep sequence data will inevitably affect the downstream analysis. Here, the samples were sequenced without amplification, resulting in an unbiased dataset without the introduction of PCR errors. A limitation to this approach is the reduction in depth of coverage across the genome due to the reduced number of viral reads in a clinical sample in comparison to cultured or PCR amplified sample. Despite the variability in number of reads between each sample, consensus sequences from all samples were obtained (Fig S1, Table 1, Table S1). This methodology also provided host reads, used to confirm the host species of each sample (Table 1). The host information was used as a discrete trait in the Bayesian analysis to investigate the origin of the host shift, therefore confirmation of the host species utilising host COI sequence, was an added benefit of using clinical samples. DNA barcoding is a robust, widely applied technique to confirm species, and analysis of sequences obtained from amplicons and non-amplified NGS data can equally be used (Hebert, et al. 2003; Zhou, et al. 2013).

Complete genome sequences from 31 RABV positive brain samples, were obtained; 20 of which were RABV cases from the host shift dataset covering the time period from its emergence in foxes to the present day. Prior to 1999, fox rabies was rarely detected in Turkey, however, after 1999, the cases of fox rabies disproportionately increased and although control measures have reduced fox rabies, it continues to the present day (Johnson, et al. 2010; RBE 2017). The samples collected, from the first fox case diagnosed in 1999 (RV1124F), to those from the central and Ankara region in 2015 represent the geographical range affected in Turkey and provide valuable data regarding the adaptation of RABV in a new host species. However, as with all non-experimental datasets, only a proportion of infected foxes have been sampled. Analysis of the RABV

dataset indicated that the virus population in Turkey reduced from the 1990s (Fig 3). This decrease corroborated the epidemiological data and coincided with the concerted rabies control efforts in dogs at the time. This overall reduction was also parsimonious with the suggestion from the Bayesian analysis (Fig 2) that several clades detected in the early 2000's in Bursa, Istanbul and Ardahan were no longer detected.

By utilising WGS and Bayesian inferences, using host as a discrete trait, we have demonstrated that as the most probable ancestor species of the host shift clade was dog, the most likely source of the fox enzootic was from dogs. Furthermore the TMRCA, for the host shift was 1997 (+/- 1 year), corroborating the epidemiological information (Fig 2). The substitution rate 3.10 (2.68 - 3.56) $x10^{-4}$ subs/site/year was observed regardless of the model used (Table 2). This rate is higher than other rates calculated from similar (concatenated genome or complete genome) datasets, including a comprehensive canine RABV dataset 2.44 (2.10-2.80) $x10^{-4}$ subs/site/year (Troupin, et al. 2016), and a fox RABV dataset from Greenland 2.5 (1.9-3.1) $x10^{-4}$ subs/site/year (Hanke, et al. 2016). Higher substitution rates and larger variation in substitution rates were observed in viruses from mongoose and ferret badgers up to 7.82 (3.13 – 13.17) $x10^{-4}$ subs/site/year. The higher substitution rate observed in this dataset could be a consequence of the smaller dataset analysed, however, it may have also been influenced by the host shift and the population increase associated with it. There are two highly related fox clades, rooted by RV2975D (Izmir 1999). Both clades are genetically distinct from contemporaneous fox viruses isolated from other regions (Erzurum). Closest viruses to this ancestor were detected in dogs in Izmir/Manisa, and epidemiological information indicated this was likely to be the location where the host shift occurred. The two clades circulated in the same area and time, including contemporary RABV samples (Fig 2). Interestingly, viruses most closely related to the early host shift viruses continued to circulate during the vaccination campaign (2008 and 2010) (Text S2). These observations highlight the importance of maintaining rabies-free areas using continued vaccination campaigns. Data such as complete genome length, and SNP analysis

provide the fine-resolution required to differentiate such highly related viruses confirming the continued presence of early phase variants during and after the vaccination campaigns.

The inclusion of dog RABV sequences from similar regions to the fox samples provided important data regarding the possible frequency of transmissions between fox and dog populations (Text S3). Of the three RABV dog sequences from the same region as the fox outbreak, all three were highly related to the fox sequences from the same region and distinct from dog RABVs in other regions of Turkey (Fig 1 and Fig 2). As seen in other European countries, these dog cases are likely to be individual spill over cases in unvaccinated dogs from the ongoing fox enzootic, rather than a host shift to the dog population. In addition, despite consistent surveillance, the number of confirmed dog rabies cases in these regions is low (5 in Eregi, 10 in Aksaray in 2014). The re-emergence of this RABV lineage in dogs is possible if vaccination of dogs is not sustained and effective at local levels.

The high sequence identity between fox and dog (Sub5790F and Sub5791D) demonstrated transmission is not dependent on host specific residues in the RABV genome. More importantly, comprehensive analysis of the fox host shift sequences over 15 years, has demonstrated a lack of evidence for positive selection at the consensus level for sustained onward transmission. A number of fixed substitutions were identified in the sequence dataset, although none were unique to specific hosts, therefore unlikely to be responsible for host species switching. A comprehensive analysis using WGS of RABVs from carnivores also concluded that positive selection was not clearly associated with host switching (Troupin, et al. 2016). Furthermore, similar observations have been documented previously for host switching events between bats and recipient species (Kuzmin, et al. 2012) and skunks and foxes (Borucki, et al. 2013).

It was generally believed that non-fox RABV lineages need higher doses than fox RABVs to successfully infect foxes, indicating a species specificity of the respective reservoir host strain encrypted in the prevailing genetic setting (Blancou and Aubert

1997). In this study, however, the results imply that consensus level 'fixed' adaptations in the virus genome are not necessary for successful onward maintenance in new host species, i.e. foxes. Perhaps the generally high susceptibility of foxes to RABV infection as experimentally demonstrated (Blancou and Aubert 1997) are one reason for successful host switching events. This is further supported by an extensive analysis of RABV isolates from the Middle East identifying regular host switching and trans-boundary movements of RABV (Horton, et al. 2015). This implies that any rabies control effort in this region should consider wildlife reservoirs as a potential source of host switching and vice versa, i.e that if the reservoir in dogs is not eliminated by sufficient population immunity, there is the risk of establishment of a subsequent wildlife RABV reservoir.

Viral heterogeneity has been long proposed as a mechanism for viruses to adapt to new environments. Therefore, deep sequence data from the WGS analysis was used to investigate the heterogeneity in the sub-consensus viral population, and to investigate whether sub-consensus level changes were driving adaptation of RABV in a new host. Utilising NGS data from clinical samples is advantageous for a number of reasons including lack of introduced error due to amplification, a methodology which is applicable to all viruses regardless of homology (because no primers are required), and utilising host data to confirm the host species. However, the depth of coverage across the genome is not uniform (Fig S1), and is lower than the coverage obtained by amplification (Borucki, et al. 2013; Stapleford, et al. 2016). The H Index was developed to analyse this dataset because it did not take depth of coverage into consideration, which is more appropriate for NGS data obtained from clinical samples. Investigating low level variants would not be appropriate with this methodology, however, these data indicate investigating population heterogeneity is reliable with samples containing >25,000 viral reads. Utilising a heterogeneity index (H Index), RABV heterogeneity was compared across all the Turkey RABVs. Shannon entropy scores were also generated and indicated that H Index and entropy analyses were comparable (Table S6). Analysis of the

H Index dataset demonstrated the combined H Index values from outside the study area are significantly lower than the host shift RABVs. Furthermore, higher H Index scores were observed in the early phase viruses compared to the late phase viruses. The dN/dS ratios of both the consensus and the sub-consensus viral population were examined to investigate the increased heterogeneity observed in the early phase of the host shift. A difference between the consensus sequence dN/dS values and the sub-consensus viral ratios was observed. The average ratios across the 5 genes using consensus sequences analysed against RV303RD (Russian outgroup) do not exceed 0.025 (Fig S3B), whereas comparison of the sub-consensus sequences indicates higher ratios for all viruses, with an average close to 1. The higher ratio reflects the heterogeneity present in the sub-consensus population, which contains non-infectious genomes and genomes with indels and deleterious residues, in addition to those which may contain advantageous mutations. Across the 5 genes, the dN/dS ratio of the sub-consensus reads indicated more synonymous than non-synonymous substitutions in the early phase viruses, than the later and non-host shift viruses (Fig S6C), corresponding to the observed increase in heterogeneity in these viruses. The only exception to this was observed with RV2982F 2012 (Ankara) where non-synonymous substitutions were greater than synonymous, in the M-protein for reasons which are currently unclear.

Due to the complexities surrounding surveillance of wildlife, including the geography and timescales involved, the sample numbers in this dataset are less than optimum. Although there are only four viruses representing the early phase host shift viruses, statistically the difference between this group and the other groups is significant. Furthermore, the number of samples within the late phase and non-host shift groups is larger, with statistically significant differences in the H Index values. Despite the smaller than optimum sample sizes, this dataset is unique, representing one of the few recorded, sampled RABV host shifts. With the exception of a comprehensive analysis investigating bat host shifts, which found evidence for positive selection in a number of genes (Streicker, et al. 2010); analysis of host shifts involving non-volant mammals, have failed

to provide definitive evidence for positive selection at the consensus level related to the host shift event (Kuzmin, et al. 2012; Borucki, et al. 2013; Troupin, et al. 2016). Therefore investigation of the sub-consensus data, and the spectrum of viral heterogeneity contained within, is a logical progression. The mechanisms involved in modulating the diversity observed at the sub-consensus level are currently undefined; however there are a number of plausible options. Firstly, the increase in heterogeneity observed in the early phase viruses could be a result of an increased incubation time within the new host, resulting in the generation of a diverse virus population. Although it is generally understood that replication rate is low during this time, the diversity within the virus population would be predicted to be greater given an extended incubation period. A previously study comparing European fox RABVs, isolated 10 years apart, showed the more recently isolated RABV had a decreased time in the onset of clinical disease when infected into naïve foxes suggesting adaptation of the virus over time in that population (Aubert, et al. 1991). We suggest that increased incubation (either at the site of inoculation, within the CNS or both) early on in a host shift could result in increased viral heterogeneity, which may reduce as the virus transmits within the new host, adapting over time. Although lyssaviruses are acute viral pathogens, due to a high mortality rate they are unlike other acute viruses such as Influenza A, which have been described to have a 'smash and grab' strategy for transmission (Lythgoe, et al. 2017). Therefore, reducing viral heterogeneity in established host reservoirs such as bats and dogs might be a valid strategy to optimise transmission within the host population. After a CST event, it is understood that the majority of RABVs do not continue within the new host population (Mollentze, et al. 2014), which in part is likely due to the host succumbing to the disease before onward transmission, in addition to other ecological and host factors. When onward transmission does occur, reducing the virulence, by reducing the viral heterogeneity could aid the virus to establish within the new host population. The mechanism utilised by RABV to achieve this is unclear, perhaps involving the fidelity of the RdRp.

The observation that the variation in the viral sub-population, rather than at a consensus level, is associated with success of a RABV host shift, is worthy of further investigation, particularly where NGS data already exists (Borucki, et al. 2013). In addition to viral genetic diversity, there are many additional factors which influence the success of a host shift, including ecological and host factors. With the increased availability of NGS platforms and decrease in cost, WGS will be pivotal in investigating and unravelling the genetic mechanisms in which RABVs evolve and emerge in new host reservoirs.

Material and Methods

Rabies Virus samples

The Turkey dataset brain samples (*n*=30) used in this study were obtained from infected dogs (*Canis lupus familiaris*) and red foxes (*Vulpes vulpes*) submitted to the Etlik Central Veterinary Control and Research Institute (Ankara, Turkey), RV303RD was held within APHA archives and originated from Russia. The majority of samples (*n*=21) were collected within the host shift areas, another 9 samples were located in other geographic regions of Turkey (Fig 1) and one (RV303RD) from Russia.'Total RNA from infected brain was extracted using TRIzol (Invitrogen, UK), following manufacturer's instructions, resuspending the RNA and adjusting to 1µg/µl in molecular grade water. All samples were confirmed RABV positive using either a hemi-nested RT-PCR (Heaton, et al. 1997) or a real-time TaqMan differential RT-PCR (Wakeley, et al. 2005). The Ct values of the samples tested by real-time TaqMan RT-PCR range from 19-26, consistent with similarly high viral load.

Preparation of total RNA for WGS

Total RNA was depleted of host genomic DNA (gDNA) and ribosomal RNA (rRNA) following methods described previously (Marston, et al. 2013; Marston, et al. 2015). Briefly, gDNA was depleted using the on-column DNase digestion protocol in RNeasy plus mini kit (Qiagen) following manufacturer's instructions, eluting in 30µl molecular grade water. Subsequently, rRNA was depleted, using Terminator 5'-phosphate-dependent exonuclease (Epicentre Biotechnologies). Briefly, 30µl of gDNA depleted RNA was mixed with 3µl of Buffer A, 0.5µl of RNAsin Ribonuclease inhibitor (20-40 U/µl) and incubated at 30°C for 60 minutes. The depleted RNA was purified to remove the enzyme using the RNeasy plus mini kit as above, without the DNase digestion, eluting in 30µl of molecular grade water. Double stranded cDNA (ds-cDNA) was synthesised using random hexamers and a cDNA synthesis kit (Roche) following manufacturer's instructions. The resulting ds-cDNA was purified using AMPure XP magnetic beads (Beckman Coulter), quantified using Quantifluor (Promega) and approximately 1ng of each sample was used in a 'tagmentation' reaction mix using a Nextera XT DNA sample preparation kit (Illumina) following manufacturer's instructions – without the bead normalisation step. DNA libraries were quantified using Quantifluor (Promega). Individual libraries were pooled (approximately 8-12 samples per MiSeq run) and normalised to equimolar concentrations. Libraries were sequenced as 2 x 150bp paired-end reads on an Illumina MiSeq platform.

Mapping to reference sequence to obtain complete genome sequences

Short reads from the first sample to be sequenced (RV1124F) were mapped to the most genetically related RABV available (RV427 - Estonian Raccoon dog RABV - KF154997). The resulting RV1124F full genome sequence was subsequently used to remap the RV1124F raw data, and to map all the Turkey dataset. Reads were mapped using the Burrow-Wheeler Aligner (BWA version 0.7.5a-r405) (Li and Durbin 2010) and

were visualised in Tablet (Milne, et al. 2013). A modified SAMtools/vcfutils (Li, et al. 2009) script was used to generate an intermediate consensus sequence in which any indels and SNPs relative to the original reference sequence were appropriately called. The intermediate consensus sequence was used as the reference for four subsequent iterations of mapping and consensus calling. Sequencing resulted in 98-100% coverage for all samples, with an average depth of coverage of 82,588 (range 1,523 – 450,449) (Table S1). Due to the transposase-based method of the Nextera XT kit, and the complementary nature of the lyssavirus genomic ends, the genome ends were not obtained for all samples. However, as the lyssavirus genomic ends are highly conserved (Marston, et al. 2007; Kuzmin, Wu, et al. 2008), and the Turkish RABV sequence dataset revealed a large degree of homology at the genomic termini, genomic termini with no mapped reads were added using a representative sequence, resulting in a dataset of complete genome sequences. Consensus sequences were submitted to Genbank (accession numbers KY860583-KY860613) and raw data deposited in European Nucleotide Archive (Accession number: PRJEB22173).

Multiple alignments

Complete genome sequences and concatenated sequences for each RABV were aligned using Megalign Pro (DNAStar) and saved as multiple sequence formats (msf). The msf files were inspected for accuracy, including correct open reading frames (ORFs) and presence of indels using GeneDoc to confirm indels and identify positions of variation.

COI analysis

Sequence reads for each sample were mapped to Cytochrome C Oxidase 1 reference sequences obtained from Barcode of Life Data Systems (http://www.boldsystems.org/; *Canis lupus familiaris*: GTENK021-11 and *Vulpes vulpes*:

GBMA1928-09 with BWA (Li and Durbin 2010). Consensus sequences were called and taxonomic assignment confirmed by using the Identification feature of the BOLD website (http://www.boldsystems.org/index.php/IDS_OpenIdEngine).

Bayesian reconstructions for host and temporal analysis

Analysis was undertaken to infer evolutionary relationships and the probable host and date of ancestral viruses. Each concatenated sequence was given a discrete trait corresponding to the species in which the virus sequences were detected (dog or fox). Phylogenetic inference was implemented using Bayesian Markov Chain Monte Carlo simulation, without partitioning, in the BEAST package v1.8.1 [43]. A TN93 nucleotide substitution model with rate variation among sites was chosen as the best nucleotide substitution model available using Bayesian Information Criterion in MEGA 6.0. Markov Chain Monte Carlo (MCMC) simulations were undertaken with strict and relaxed molecular clock models, and with constant or skyline population priors for 10,000,000 iterations, sampling every 1,000 states to give effective sample sizes of over 200. Molecular clock and population coalescent models were compared using a modified AIC in Tracer for computational efficiency, as described previously (Baele, et al. 2012). Maximum clade credibility trees were then annotated using TreeAnnotator (v1.8.1) after 10% of trees were discarded and were then visualised using Fig Tree (v1.4.0) (Fig. 3). Branches are colored by most probable host at the ancestral node, and node labels are posterior probabilities.

In addition to the concatenated coding sequence analysis, partial N-gene sequences (400bp) from a wide range of RABV cases from the Middle East (Horton, et al. 2015), were analysed alongside the same region from the Turkey sequence dataset to provide context to the samples (Fig S2, Text S1). The same conditions were used as described for coding region sequences above.

Bayesian skyline plot analyses were conducted as above on two datasets: whole Turkey dataset (1980-2015) and whole host shift dataset (1999-2015), using the Coalescent Bayesian Skyline as the tree prior (ESS>200) (Drummond, et al. 2005)

Heterogeneity Index analysis

For each NGS sample, short reads were filtered to discard duplicates in order to reduce the chances of having overrepresented reads due to the library preparation PCR. Additionally, in order to reduce sequencing errors, the sort reads were end trimmed when the quality score of the base call dropped below 20 (range 0 to 40) (Bolger, et al. 2014). Reads shorter than 36 bases were also discarded. The resulting set of reads was mapped onto the homologous viral consensus genome using SMALT (SANGER Institute). Only reads with a top mapping quality of 60 (range 0 to 60) were considered from this point.

A Heterogeneity Index (H Index) was calculated on the remaining viral reads as the number of alternative calls per million of base calls sequenced. This is the ratio of the sum of the number of alternative calls for every read divided by the sum of the number of calls of each read. This value is then multiplied by 1 $x10^6$. In order to minimise false alternative calls due to sequencing errors, the base calls with a sequencing score lower than a threshold equal to 30 (0-40) are ignored during the calculation. Additionally, alternative base calls at a given genomic position are ignored if the proportion of reads agreeing with that particular call is lower than 5%.

The thresholds used during data processing were set to minimise incorporating sequencing errors without being too restrictive. The purpose of this H Index is not the calculation of the absolute heterogeneity within the sample, rather the H Index is comparable from sample to sample and alterations on the thresholds do not significantly affect the relativeness of the H Index between samples (particularly because all the samples have been sequenced using the same equipment). The two-tail student t-tests

were performed based the assumptions that H indices are normally distributed with equal variances.

Entropy and dN/dS analysis

The heterogeneity in each virus sample was characterised by computing the Shannon entropy at each site and then averaging over every site in the genome as described previously (Morelli, et al. 2013). The resulting value was multiplied by 100,000 to obtain equivalent values to H index. Similarly to H Index, in order to minimise false alternative calls due to sequencing errors, the base calls with a sequencing score lower than a threshold equal to 30 (0-40) are ignored during the calculation.

For each sample's consensus sequence, the synonymous to non-synonymous ratio (dN/dS) for each codon in the 5 RABV ORFs was determined by counting the number of synonymous and non-synonymous mutations with respect to the consensus sequence of sample RV303RD which is an evolutionary distinct lineage of RABV to the host shift samples, using the program SNAP (https://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html) (Korber 2000). The dN/dS ratio of the sub-consensus sequences was obtained as described previously (Morelli, et al. 2013). Briefly, in order to estimate the dN/dS ratio from NGS data, for each codon the observed number of synonymous mutations are counted (only considering reads that cover the codon entirely), divided by the expected number of mutations for the codon, and subsequently divided by the read coverage to give a pS value for each codon; a similar process is used for non-synonymous mutations to give a pN value for each codon. Individual pN and pS values for each codon are then summed and averaged to give an overall pN and pS for the ORF. dN/dS was determined from pN and pS as described previously (Nei and Gojobori 1986).

Acknowledgements

The authors wish to thank Emma Wise (Public Health England), Isobel Jarvis, Dr. Guanghui Wu and Adam Ashton (APHA) for technical support and Kirstyn Brunker (University of Glasgow) for constructive and informative discussions.

Funding

This work was supported by an initiative of an OIE funded laboratory twinning project on rabies between the Friedrich-Loeffler-Institut and the Etlik Veterinary Control and Research Institute Ankara [File Ref: GKB/KH/2009/22] and was financially supported by the UK Department for Environment, Food and Rural Affairs (Defra),Scottish Government and Welsh Government by grants [SE0427,SE0431]; the European Union H2020-funded Research Infrastructure Grant 'European Virus Archive Global (EVAg)' [H2020 - grant agreement n°653316] and by an intramural collaborative research grant from the Friedrich-Loeffler-Institut.

References

Aubert MF, Blancou J, Barrat J, Artois M, Barrat MJ. 1991. [Transmissibility and pathogenicity in the red fox of two rabies viruses isolated at a 10 year interval]. Ann Rech Vet 22:77-93.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol Biol Evol 29:2157-2167.

Benmansour A, Brahimi M, Tuffereau C, Coulon P, Lafay F, Flamand A. 1992. Rapid sequence evolution of street rabies glycoprotein is related to the highly heterogeneous nature of the viral population. Virology 187:33-45.

Blancou J, Aubert MF. 1997. [Transmission of rabies virus: importance of the species barrier]. Bull Acad Natl Med 181:301-311; discussion 311-302.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Borucki MK, Chen-Harris H, Lao V, Vanier G, Wadford DA, Messenger S, Allen JE. 2013. Ultra-deep sequencing of intra-host rabies virus populations during cross-species transmission. PLoS Negl Trop Dis 7:e2555.

Bull RA, Eden JS, Luciani F, McElroy K, Rawlinson WD, White PA. 2012. Contribution of intra- and interhost dynamics to norovirus evolution. J Virol 86:3219-3229.

Daoust PY, Wandeler AI, Casey GA. 1996. Cluster of rabies cases of probable bat origin among red foxes in Prince Edward Island, Canada. J Wildl Dis 32:403-406.

Domingo E, Holland JJ. 1997. RNA virus mutations and fitness for survival. Annu Rev Microbiol 51:151-178.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22:1185-1192.

Eriksson N, Pachter L, Mitsuya Y, Rhee SY, Wang C, Gharizadeh B, Ronaghi M, Shafer RW, Beerenwinkel N. 2008. Viral population estimation using pyrosequencing. PLoS Comput Biol 4:e1000074.

Evans JS, Horton DL, Easton AJ, Fooks AR, Banyard AC. 2012. Rabies virus vaccines: is there a need for a pan-lyssavirus vaccine? Vaccine 30:7447-7454.

Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T, Han CS, Gleasner CD, Green L, Lo CC, et al. 2010. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. PLoS One 5:e12303.

Fusaro A, Monne I, Salomoni A, Angot A, Trolese M, Ferre N, Mutinelli F, Holmes EC, Capua I, Lemey P, et al. 2013. The introduction of fox rabies into Italy (2008-2011) was due to two viral genetic groups with distinct phylogeographic patterns. Infect Genet Evol 17:202-209.

Hanke D, Freuling CM, Fischer S, Hueffer K, Hundertmark K, Nadin-Davis S, Marston D, Fooks AR, Botner A, Mettenleiter TC, et al. 2016. Spatio-temporal Analysis of the Genetic Diversity of Arctic Rabies Viruses and Their Reservoir Hosts in Greenland. PLoS Negl Trop Dis 10:e0004779.

Heaton PR, Johnstone P, McElhinney LM, Cowley R, O'Sullivan E, Whitby JE. 1997. Heminested PCR assay for detection of six genotypes of rabies and rabies-related viruses. J Clin Microbiol 35:2762-2766.

Hebert PD, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. Proc Biol Sci 270:313-321.

Holmes EC, Moya A. 2002. Is the quasispecies concept relevant to RNA viruses? J Virol 76:460-465.

Hoper D, Freuling CM, Muller T, Hanke D, von Messling V, Duchow K, Beer M, Mettenleiter TC. 2015. High definition viral vaccine strain identity and stability testing using full-genome population data--The next generation of vaccine quality control. Vaccine 33:5829-5837.

Horton DL, McElhinney LM, Freuling CM, Marston DA, Banyard AC, Goharrriz H, Wise E, Breed AC, Saturday G, Kolodziejek J, et al. 2015. Complex epidemiology of a zoonotic disease in a culturally diverse region: phylogeography of rabies virus in the Middle East. PLoS Negl Trop Dis 9:e0003569.

Johnson N, Black C, Smith J, Un H, McElhinney LM, Aylan O, Fooks AR. 2003. Rabies emergence among foxes in Turkey. J Wildl Dis 39:262-270.

Johnson N, Un H, Fooks AR, Freuling C, Muller T, Aylan O, Vos A. 2010. Rabies epidemiology and control in Turkey: past and present. Epidemiol Infect 138:305-312.

Kissi B, Badrane H, Audry L, Lavenu A, Tordo N, Brahimi M, Bourhy H. (15510 co-authors). 1999. Dynamics of rabies virus quasispecies during serial passages in heterologous hosts. J Gen Virol 80 (Pt 8):2041-2050.

Korber B. 2000. HIV signature and sequence variation analysis. In: Rodrigo AG, Learn GH, editors. Computational Analysis of HIV Molecular Sequences Netherlands: Kluwer Academic Publishers. p. 55-72.

Kuzmin IV, Hughes GJ, Botvinkin AD, Gribencha SG, Rupprecht CE. 2008. Arctic and Arctic-like rabies viruses: distribution, phylogeny and evolutionary history. Epidemiol Infect 136:509-519.

Kuzmin IV, Shi M, Orciari LA, Yager PA, Velasco-Villa A, Kuzmina NA, Streicker DG, Bergman DL, Rupprecht CE. 2012. Molecular Inferences Suggest Multiple Host Shifts of Rabies Viruses from Bats to Mesocarnivores in Arizona during 2001-2009. PLoS Pathog 8:e1002786.

Kuzmin IV, Wu X, Tordo N, Rupprecht CE. 2008. Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions. Virus Res 136:81-90.

Leslie MJ, Messenger S, Rohde RE, Smith J, Cheshier R, Hanlon C, Rupprecht CE. 2006. Bat-associated rabies virus in Skunks. Emerg Infect Dis 12:1274-1277.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589-595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Lythgoe KA, Gardner A, Pybus OG, Grove J. 2017. Short-Sighted Virus Evolution and a Germline Hypothesis for Chronic Viral Infections. Trends Microbiol 25:336-348.

Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR. 2013. Next generation sequencing of viral RNA genomes. BMC Genomics 14:444.

Marston DA, McElhinney LM, Johnson N, Muller T, Conzelmann KK, Tordo N, Fooks AR. (30 co-authors). 2007. Comparative analysis of the full genome sequence of European bat lyssavirus type 1 and type 2 with other lyssaviruses and evidence for a conserved transcription termination and polyadenylation motif in the G-L 3' non-translated region. J Gen Virol 88:1302-1314.

Marston DA, Wise EL, Ellis RJ, McElhinney LM, Banyard AC, Johnson N, Deressa A, Regassa F, de Lamballerie X, Fooks AR, et al. 2015. Complete genomic sequence of rabies virus from an ethiopian wolf. Genome Announc 3.

McElhinney LM, Marston DA, Freuling CM, Cragg W, Stankov S, Lalosevic D, Lalosevic V, Muller T, Fooks AR. 2011. Molecular diversity and evolutionary history of rabies virus strains circulating in the Balkans. J Gen Virol 92:2171-2180.

McGinnis J, Laplante J, Shudt M, George KS. 2016. Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. J Clin Virol 79:44-50.

Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D. 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform 14:193-202.

Mollentze N, Biek R, Streicker DG. 2014. The role of viral evolution in rabies host shifts and emergence. Curr Opin Virol 8:68-72.

Montmayeur AM, Ng TF. 2017. High-Throughput Next-Generation Sequencing of Polioviruses. 55:606-615.

Morelli MJ, Wright CF, Knowles NJ, Juleff N, Paton DJ, King DP, Haydon DT. 2013. Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts. Vet Res 44:12.

Morimoto K, Hooper DC, Carbaugh H, Fu ZF, Koprowski H, Dietzschold B. 1998. Rabies virus quasispecies: implications for pathogenesis. Proc Natl Acad Sci U S A 95:3152-3156.

Murcia PR, Baillie GJ, Daly J, Elton D, Jervis C, Mumford JA, Newton R, Parrish CR, Hoelzer K, Dougan G, et al. 2010. Intra- and interhost evolutionary dynamics of equine influenza virus. J Virol 84:6943-6954.
Nadin-Davis SA, Colville A, Trewby H, Biek R, Real L. 2017. Application of high-throughput sequencing to whole rabies viral genome characterisation and its use for phylogenetic re-evaluation of a raccoon strain incursion into the province of Ontario. Virus Res.
Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-426.
Nel LH, Thomson GR, Von Teichman BF. 1993. Molecular epidemiology of rabies virus in South Africa. Onderstepoort J Vet Res 60:301-306.
Parker J, Chen J. 2017. Application of next generation sequencing for the detection of human viral pathogens in clinical specimens. J Clin Virol 86:20-26.
Poh WT, Xia E, Chin-Inmanu K, Wong LP, Cheng AY, Malasit P, Suriyaphol P, Teo YY, Ong RT. 2013. Viral quasispecies inference from 454 pyrosequencing. BMC Bioinformatics 14:355.
Raghwani J, Rose R, Sheridan I. 2016. Exceptional Heterogeneity in Viral Evolutionary Dynamics Characterises Chronic Hepatitis C Virus Infection. 12:e1005894.
RBE reported rabies cases in Europe database [Internet]. 2017. cited 14/01/2017]. Available from http://www.who-rabies-bulletin.org/site-page/queries.
Stapleford KA, Moratorio G, Henningsson R, Chen R, Matheus S, Enfissi A, Weissglas-Volkov D, Isakov O, Blanc H, Mounce BC, et al. 2016. Whole-Genome Sequencing Analysis from the Chikungunya Virus Caribbean Outbreak Reveals Novel Evolutionary Genomic Elements. PLoS Negl Trop Dis 10:e0004402.
Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. 2010. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science 329:676-679.
Troupin C, Dacheux L, Tanguy M, Sabeta C, Blanc H, Bouchier C, Vignuzzi M, Duchene S, Holmes EC, Bourhy H. 2016. Large-Scale Phylogenomic Analysis Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts. PLOS Pathogens 12:e1006041.
Un H, Eskiizmirliler S, Unal N, Freuling CM, Johnson N, Fooks AR, Muller T, Vos A, Aylan O. 2012. Oral vaccination of foxes against rabies in Turkey between 2008 and 2010. Berl Munch Tierarztl Wochenschr 125:203-208.
Wakeley PR, Johnson N, McElhinney LM, Marston D, Sawyer J, Fooks AR. 2005. Development of a real-time, TaqMan reverse transcription-PCR assay for detection and differentiation of lyssavirus genotypes 1, 5, and 6. J Clin Microbiol 43:2786-2792.
Wallace RM, Gilbert A, Slate D, Chipman R, Singh A, Cassie W, Blanton JD. 2014. Right place, wrong species: a 20-year review of rabies virus cross species transmission among terrestrial mammals in the United States. PLoS One 9:e107539.
Wright CF, Morelli MJ, Thebaud G, Knowles NJ, Herzyk P, Paton DJ, Haydon DT, King DP. 2011. Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. J Virol 85:2266-2275.

Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q. 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. Gigascience 2:4.

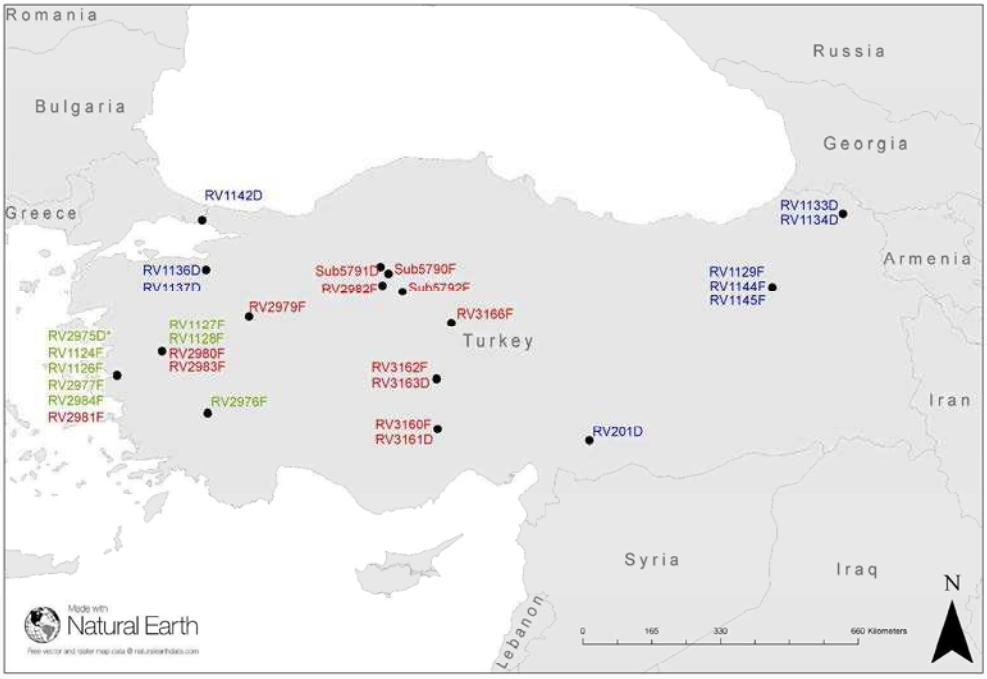
Figure Legends

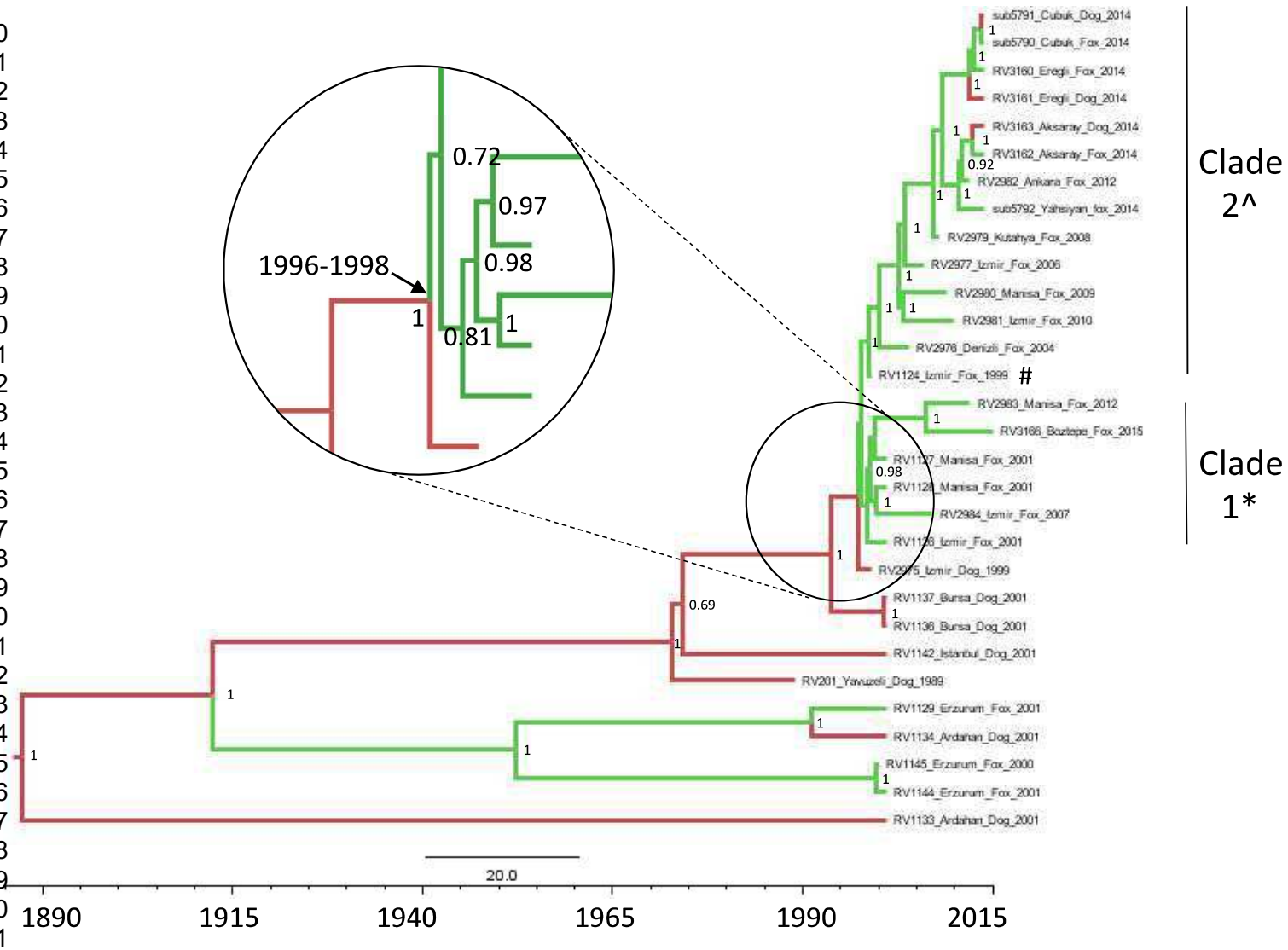
Fig 1: Map of Turkey and surrounding countries with location of the viruses from the Turkey dataset. pre-host-shift canine RABV (*); early phase host shift RABVs (red), late phase host shift (green); and non-host shift RABVs (blue). RV303RG not included as outside of map area. See Table 1 for sample details.

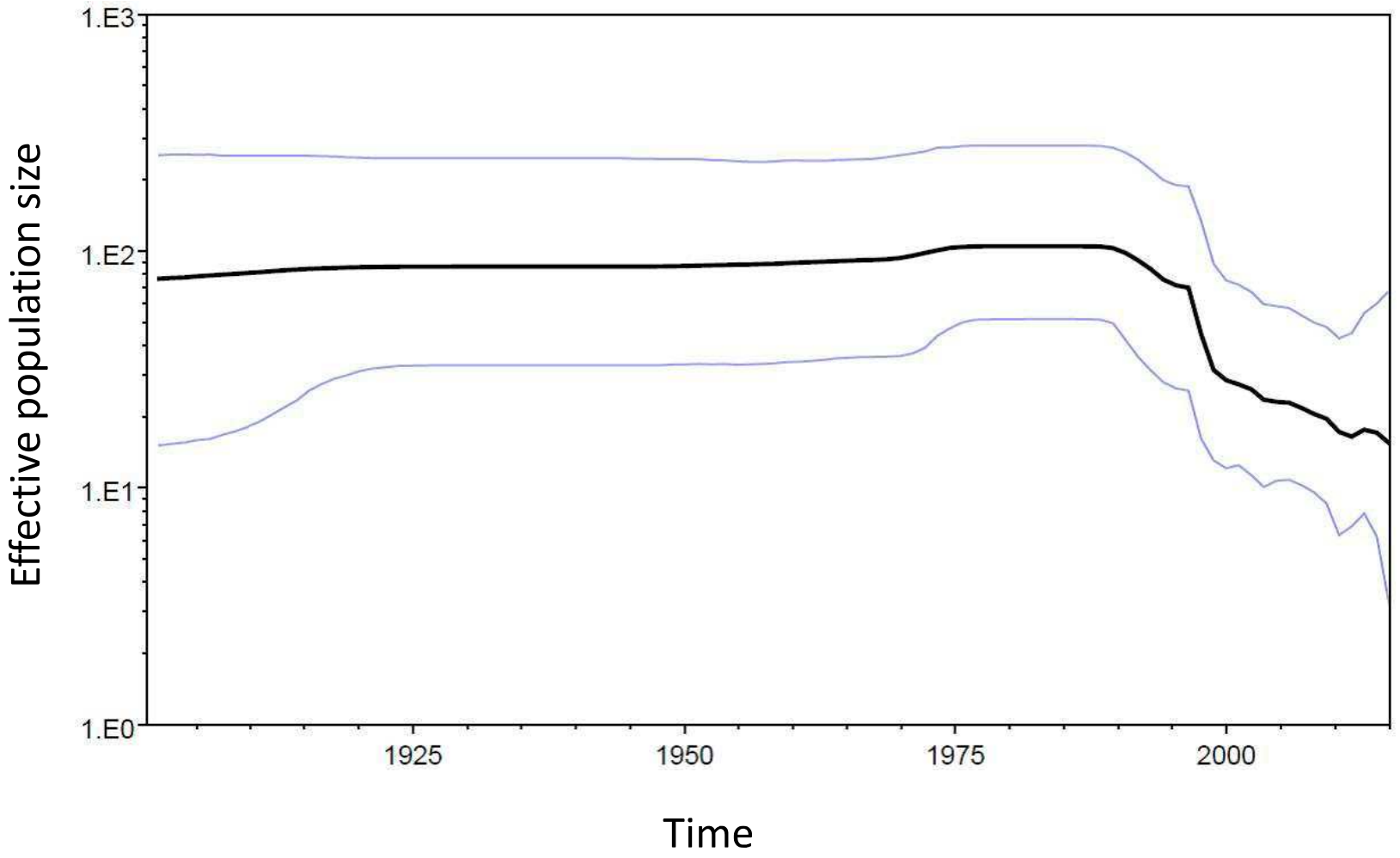
Fig 2: Bayesian Maximum clade credibility phylogeny of Turkey concatenated coding sequences. Branch colored according to reservoir host species: dog (red) or fox (green). Posterior probabilities are indicated at each node. The node for the MRCA of the host shift event is enlarged and the dates indicated (Highest Posterior Density - HPD 95%). The fox host shift viruses are divided into 2 distinct clades indicated, Clade 1 genome length is 11,923* and clade 2 genome length is 11,924 or 11,925^ (Table S2). SNPs between clade 1 and 2 are described in Table S5. # indicates the position where the non-synonymous changes are observed.

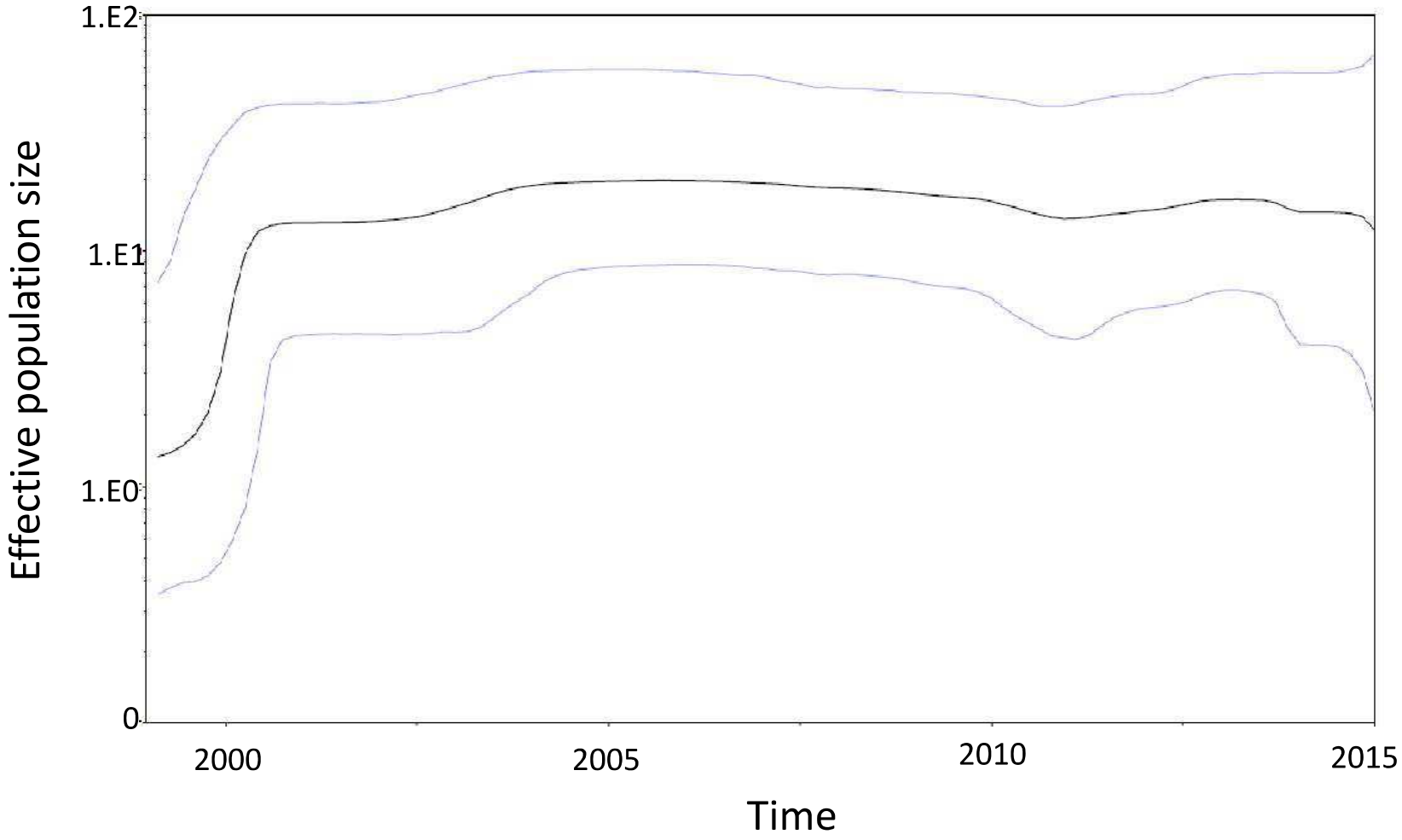
Fig 3: Bayesian skyline plots showing the effective population size over time of Turkey RABV population with HPD 95% limits for (**A**) whole Turkey dataset (1980-2015), (**B**) Fox host shift dataset (1999-2015).

Fig 4: Graphical representation of sub consensus heterogeneity using H Index for all viruses with viral reads >25,000. All Turkey viruses grouped as follows: early phase (n= 4), late phase (n=9; fox 7, dog 2), all host shift dataset (n=13) and non-host shift (n=6, fox 1, dog 5). Sample standard deviation is indicated for each group, unpaired t-test p values provided.


Romania
Bulgaria
Greece
Russia
Georgia
Armenia
Turkey
Iran
Syria
Iraq
Lebanon
RV1142D
RV1133D
RV1134D
RV1136D
RV1137D
Sub5791D
Sub5790F
RV2982F
Sub5792F
RV1129F
RV1144F
RV1145F
RV2979F
RV3166F
RV1127F
RV1128F
RV2980F
RV2983F
RV2975D*
RV1124F
RV1126F
RV2977F
RV2984F
RV2981F
RV3162F
RV3163D
RV2976F
RV3160F
RV3161D
RV201D
Made with
Natural Earth
0
165
330
660 Kilometers
N


sub5791_Cubuk_Dog_2014
sub5790_Cubuk_Fox_2014
RV3160_Eregli_Fox_2014
RV3161_Eregli_Dog_2014
RV3163_Aksaray_Dog_2014
RV3162_Aksaray_Fox_2014
RV2982_Ankara_Fox_2012
sub5792_Yahsiyan_fox_2014
RV2979_Kutahya_Fox_2008
RV2977_Izmir_Fox_2006
RV2980_Manisa_Fox_2009
RV2981_Izmir_Fox_2010
RV2976_Denizli_Fox_2004
RV1124_Izmir_Fox_1999 #
RV2983_Manisa_Fox_2012
RV3166_Boztepe_Fox_2015
RV1127_Manisa_Fox_2001
RV1128_Manisa_Fox_2001
RV2984_Izmir_Fox_2007
RV1126_Izmir_Fox_2001
RV2975_Izmir_Dog_1999
RV1137_Bursa_Dog_2001
RV1136_Bursa_Dog_2001
RV1142_Istanbul_Dog_2001
RV201_Yavuzeli_Dog_1989
RV1129_Erzurum_Fox_2001
RV1134_Ardahan_Dog_2001
RV1145_Erzurum_Fox_2000
RV1144_Erzurum_Fox_2001
RV1133_Ardahan_Dog_2001
0.72
0.97
0.98
1996-1998
1
0.81
1
0.69
0.92
0.98
Clade 2^
Clade 1*
20.0
1890
1915
1940
1965
1990
2015


1.E3
1.E2
1.E1
1.E0
Effective population size
1925
1950
1975
2000
Time


1.E2
1.E1
1.E0
0
Effective population size
2000
2005
2010
2015
Time

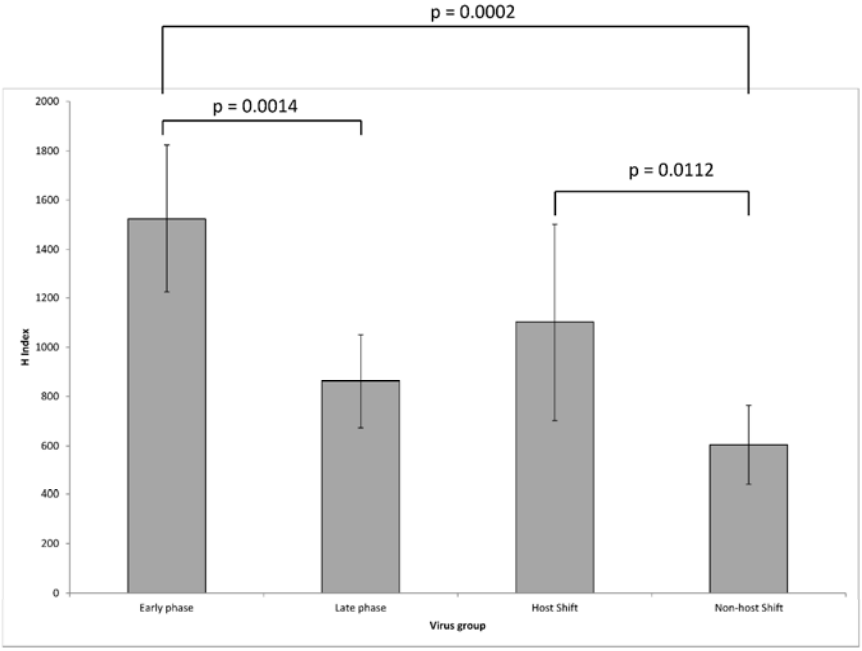

p = 0.0002
p = 0.0014
p = 0.0112
H Index
2000
1800
1600
1400
1200
1000
800
600
400
200
0
Early phase
Late phase
Host Shift
Non-host Shift
Virus group

Table 1: Turkish RABV isolates analysed in this study.

ID	Area	Host	Year	viral reads	% viral reads	Genbank Accession
RV2975D*	Izmir	Dog	1999	48,801	0.532	KY860583
RV1124F	Izmir	Fox	1999	17,010	0.24	KY860584
RV1126F	Izmir	Fox	2001	27,740	0.26	KY860585
RV1127F	Manisa	Fox	2001	125,950	5.395	KY860586
RV1128F	Manisa	Fox	2001	15,435	0.18	KY860587
RV2976F	Denizli	Fox	2004	30,993	1	KY860588
RV2977F	Izmir	Fox	2006	20,791	0.31	KY860589
RV2984F	Izmir	Fox	2007	45,162	4.771	KY860590
RV2979F	Kutahya	Fox	2008	31,332	0.257	KY860591
RV2980F	Manisa	Fox	2009	10,845	0.22	KY860592
RV2981F	Izmir	Fox	2010	113,385	5.526	KY860593
RV2982F	Ankara	Fox	2012	54,074	0.83	KY860594
RV2983F	Manisa	Fox	2012	11,565	0.438	KY860595
RV3162F	Aksaray	Fox	2014	77,904	3.5	KY860596
RV3163D	Aksaray	Dog	2014	30,321	1.45	KY860597
Sub5790F	Ankara-Cubuk	Fox	05/05/14	79,651	3.17	KY860598
Sub 5791D	Ankara-Cubuk	Dog	16/05/14	318,799	17.65	KY860599
Sub 5792F	Kirkkale-Yahsiyan	Fox	2014	43,663	2.01	KY860600
RV3160F	Konya-Eregli	Fox	27/05/14	95,237	3.65	KY860601
RV3161D	Konya-Eregli	Dog	26/05/14	21,225	1	KY860602
RV3166F	Kirsehir-Boztepe	Fox	21/05/15	11,945	0.26	KY860603
RV1145F	Erzurum	Fox	2000	13,552	0.35	KY860604
RV1144F	Erzurum	Fox	2001	33,268	0.79	KY860605
RV1129F	Erzurum	Fox	2001	8,408	0.19	KY860606
RV1136D	Bursa	Dog	2001	27,835	1.2	KY860607

RV1137D	Bursa	Dog	2001	43,891	1.36	KY860608
RV1142D	Istanbul	Dog	2001	140,830	3.3	KY860609
RV1133D	Ardahan	Dog	2001	450,449	8.7	KY860610
RV1134D	Ardahan	Dog	2001	169,262	3.15	KY860611
RV201D	Yavuzeli	Dog	1989	333,831	9.4	KY860612
RV303RD	Russia (East)	Raccoon Dog	1980	107,066	3.52	KY860613

Samples are listed in date / location order and divided as follows: RV2975* pre-host-shift canine RABV; host-shift dataset, 20 isolates (17 fox and 3 dog) subdivided by dashed line to indicate early (green) and late (red) phases of the host shift; and non-host shift (blue) from other regions of Turkey with Russia as outgroup.

Table 2: Assessment of model fit using a modified AIC for RABV full genome concatenated coding sequences.

Molecular clock	Population prior	AICM	Mean substitution rate (95% HPD)
Strict	Bayesian skyline	50443	3.10 (2.68-3.56) $x10^{-4}$
Strict	Constant	50444	3.11 (2.69-3.58) $x10^{-4}$
Uncorrelated lognormal	Constant	50450	3.21 (2.69-3.79) $x10^{-4}$
Uncorrelated lognormal	Bayesian skyline	50449	3.14 (2.68-3.67) $x10^{-4}$

An MCMC chain length of 10 million iterations using 31 representative sequences in BEAST. AICM compared using Tracer (v1.6). Lower AICM values indicate better fit

Table 3: SNPs between pre-host shift canine RABV RV2975D Izmir 1999 and all host shift RABV sequences.

Nuc position	AA Position within gene	Gene	Codon RV2975D	Codon Host shift	AA RV2975D	AA Host shift
75	25	N	CA**G**	CA**A**	Gln	Gln
792	264	N	AA**T**	AA**C**	Asn	Asn
2182	277	P	**T**TA	**C**TA	Leu	Leu
5998	523	L	**C**TG	**T**TG	Leu	Leu
7360	977	L	GTT	ATT	Val	Ile
8604	1391	L	TT**G**	TT**A**	Leu	Leu
10043	1871	L	C**T**A	C**G**A	Leu	Arg
10050	1873	L	TC**T**	TC**C**	Ser	Ser

Nucleotide (Nuc) positions relate to concatenated sequences. Amino acid (AA) positions relate to specific gene. Nucleotides that differ are in bold and non-synonymous changes are underlined. N – nucleoprotein gene, P – phosphoprotein gene, and L – Large protein gene (RNA dependent RNA polymerase).

Fig S1: Read depth of coverage across genome for a selcetion of eight host shift viruses representing the range of total viral reads obtained in the total dataset.

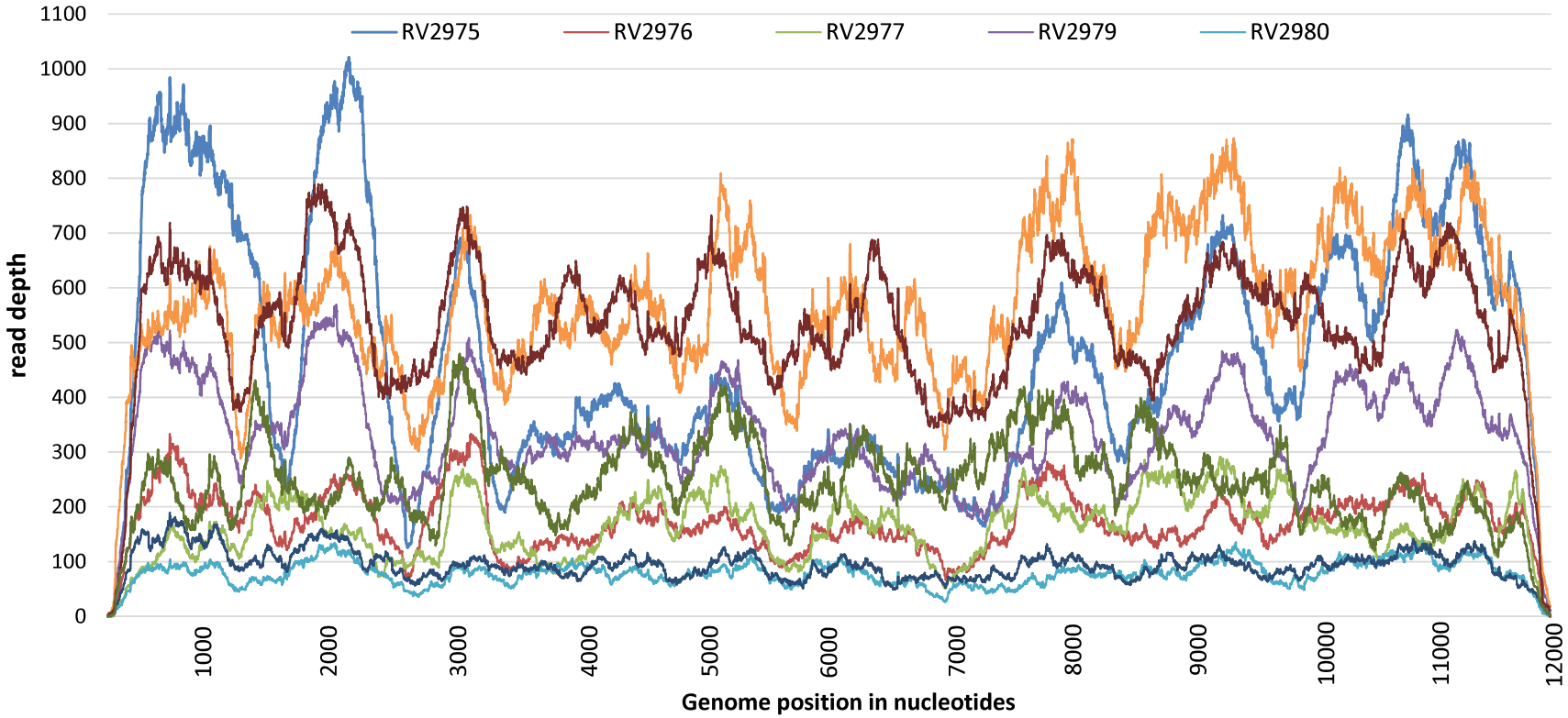


Fig S2: Bayesian Maximum clade credibility phylogeny of 400bp nucleoprotein gene from RABVs from Turkey and neighbouring countries in the Middle East.

c r b ccor g o pr v ou u (Hor o *et al* 2015), u b v ru s c u o co x clu v ru rom urop A r ca. Po r or prob b s r c ch o . urkey qu ces r pr om y w cl C r rg box.

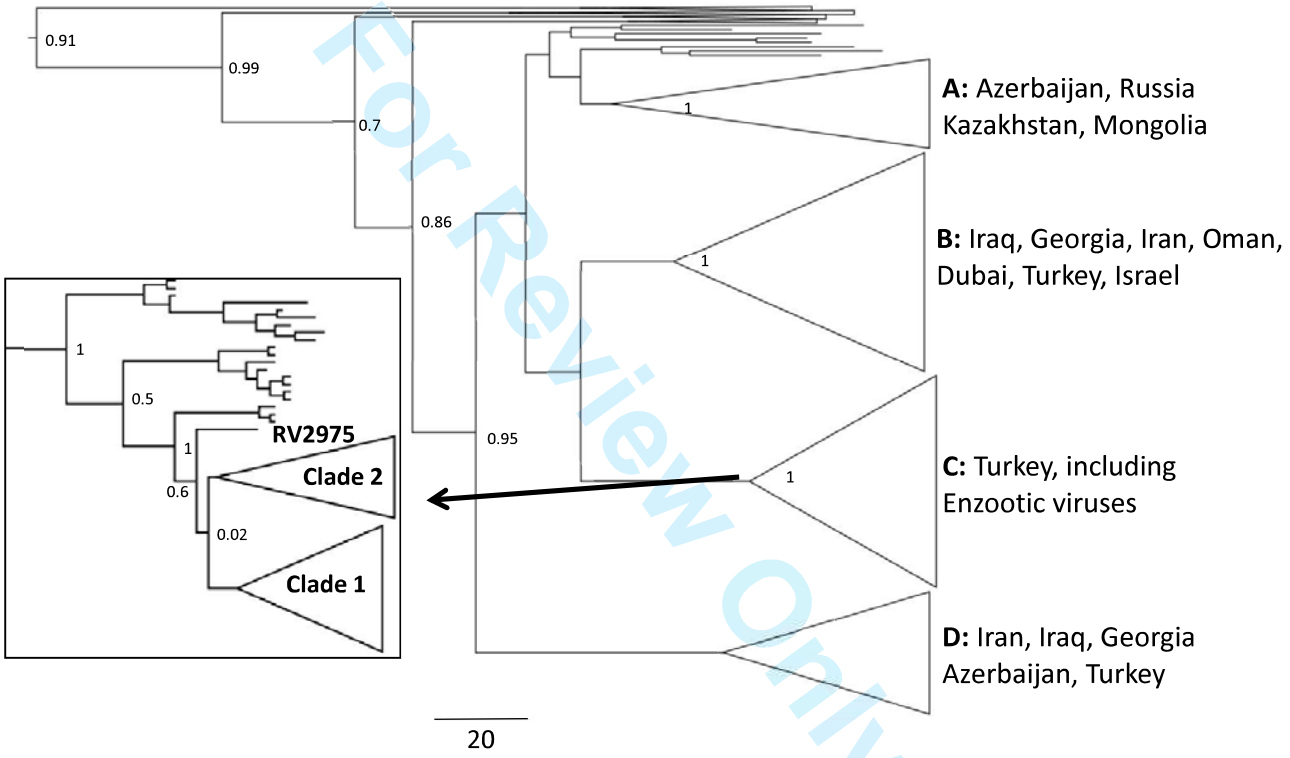
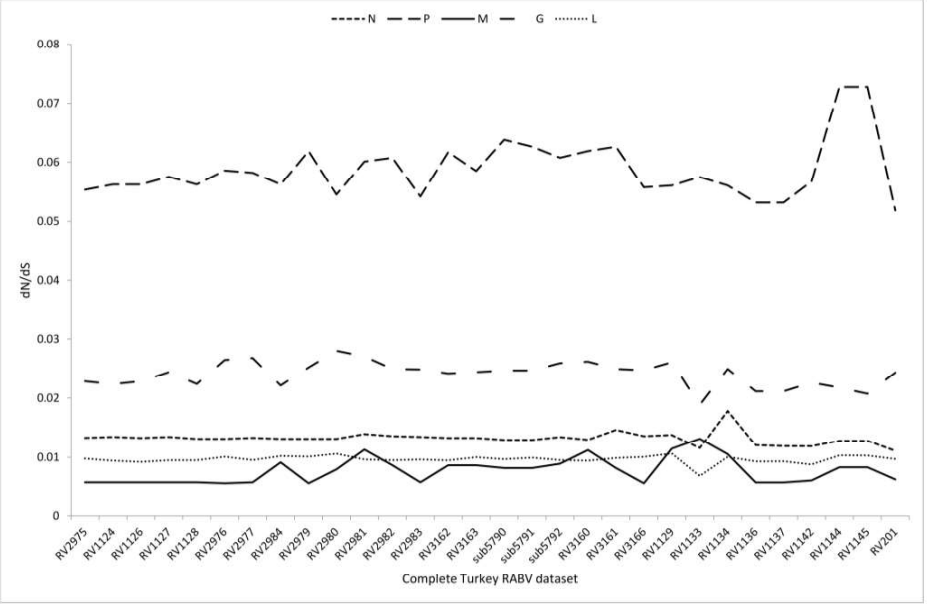
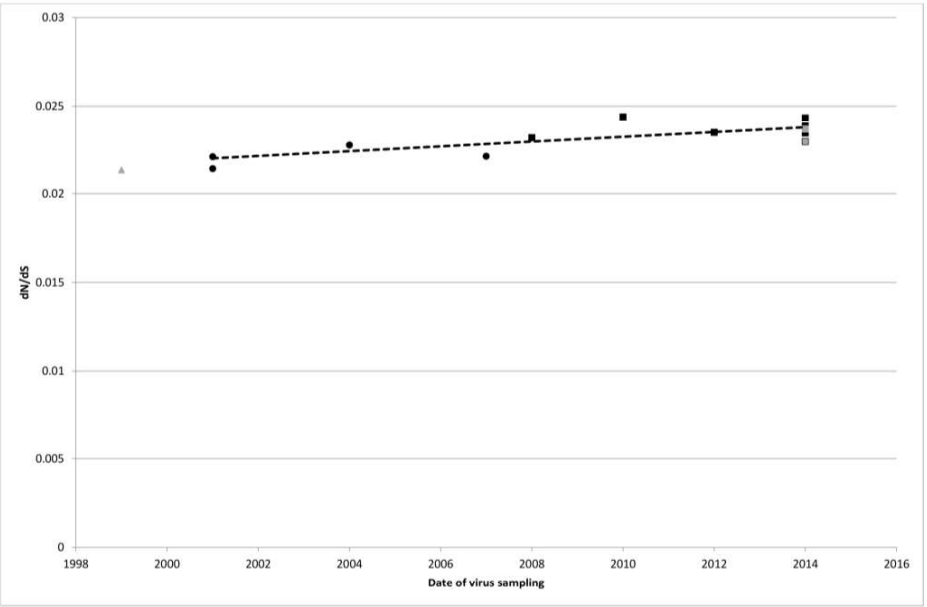


Fig S3: dN/dS ratios of consensus sequences using RV303 as reference, across all Turkey viruses ordered by collection date. (**A**) ! v u g , (**B**) v r g ! ! ! r os cro 5 g . V ru v o o ow g groups; pr - o (r g), r y p (c rcl), p (qu r). Fox mp s (b ck), og mp (gr y).

A.



&.



F g ! 4: H rog y ! x (so) ! o ropy ( ) or virus qu ce .

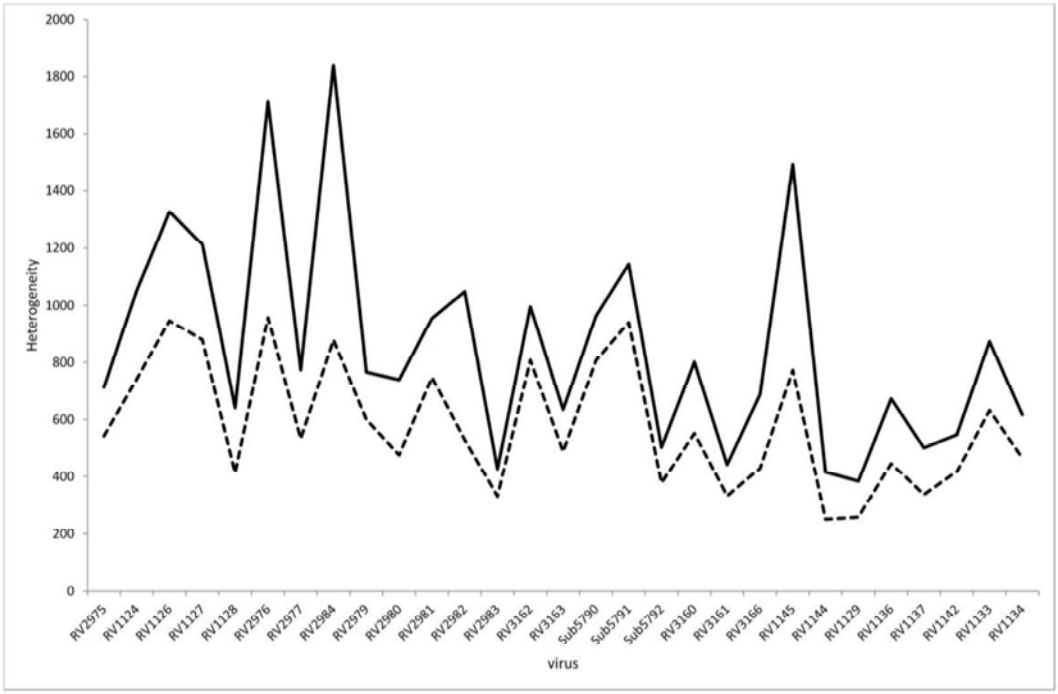


Fig S5: Graphical representation of sub consensus heterogeneity using H Index for all viruses with viral reads >25,000. A urk y v ru s group o ow : pr - o (r g), r y p (c rcl), p (qu r), o - o rom urk y (mo). r s or r y p v ru s c by . Fox mp s (b ck), og mp (gr y). p r - p 0.001 , p 0.05 .

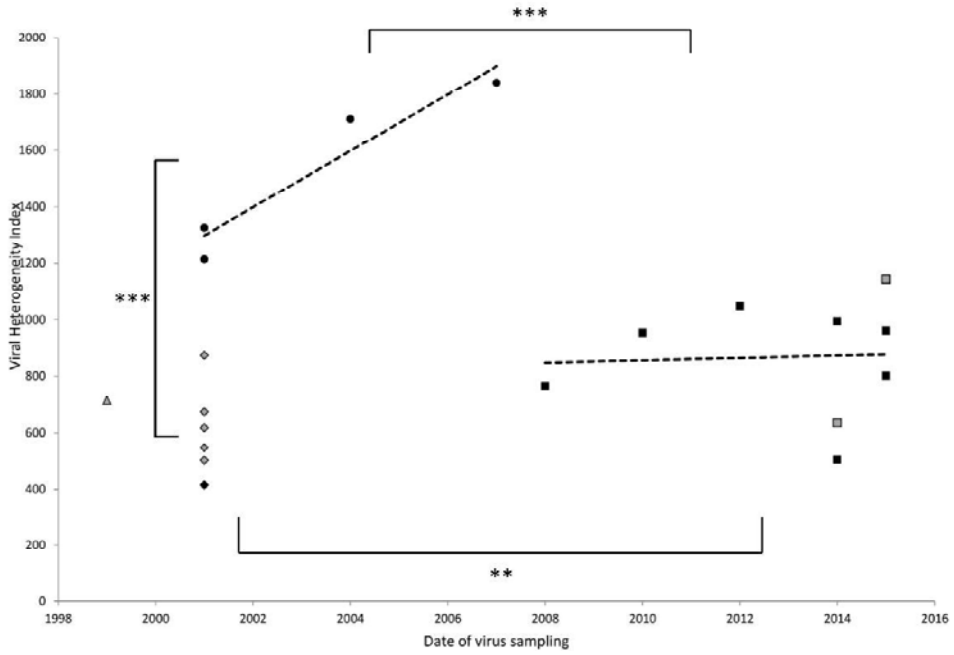
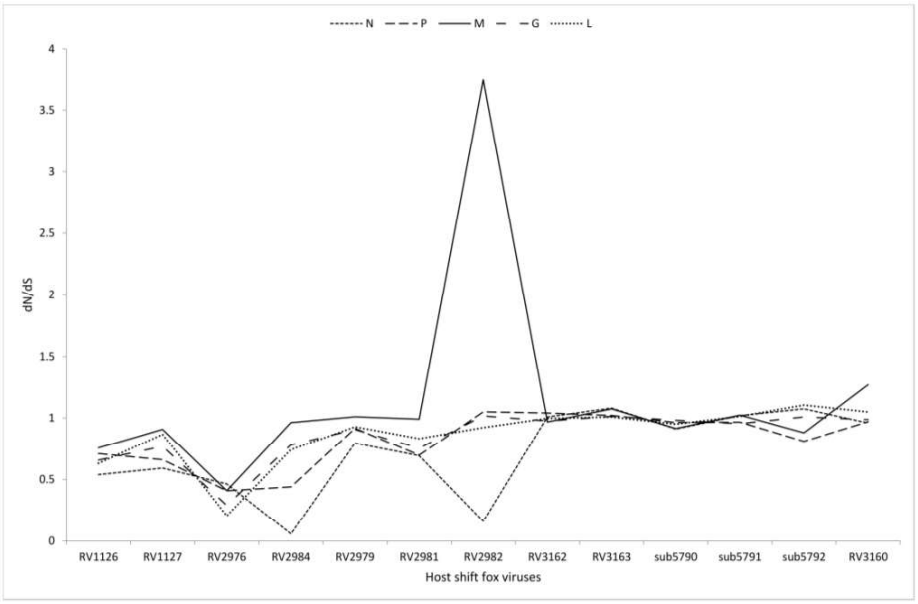
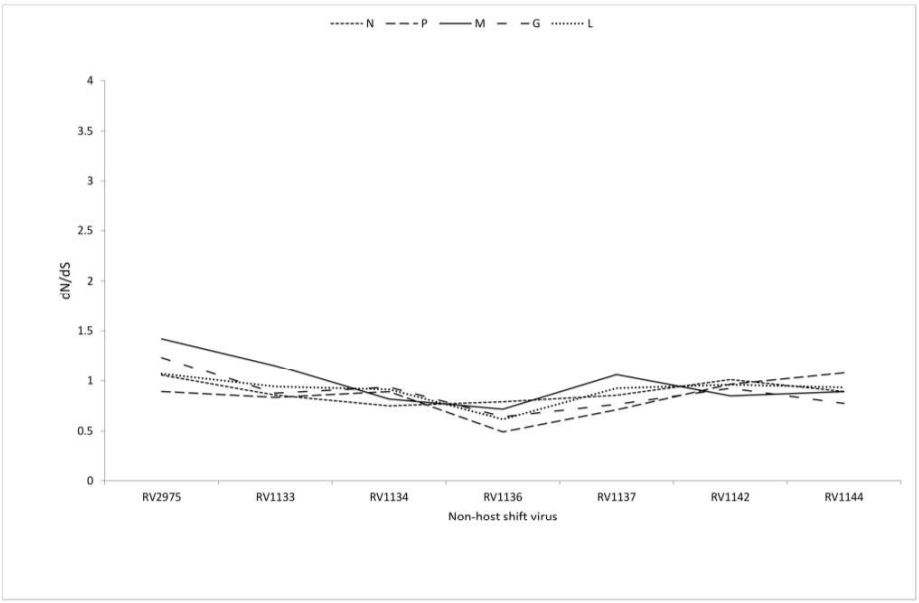


Fig S6: dN/dS ratios of sub-viral population across all viruses ordered by collection date, with viral reads >25,000. (**A**) Individual genes from host shift viruses, (**B**) individual genes from non-host shift viruses and pre-host shift virus RV2975D Izmir. (**C**) Average dN/dS ratios of host shift viruses, plotted against time, across all genes, pre- host shift virus RV2975D Izmir (triangle), early phase viruses (circle), late phase viruses (square). Fox samples (black), dog samples (grey).

A



B



C

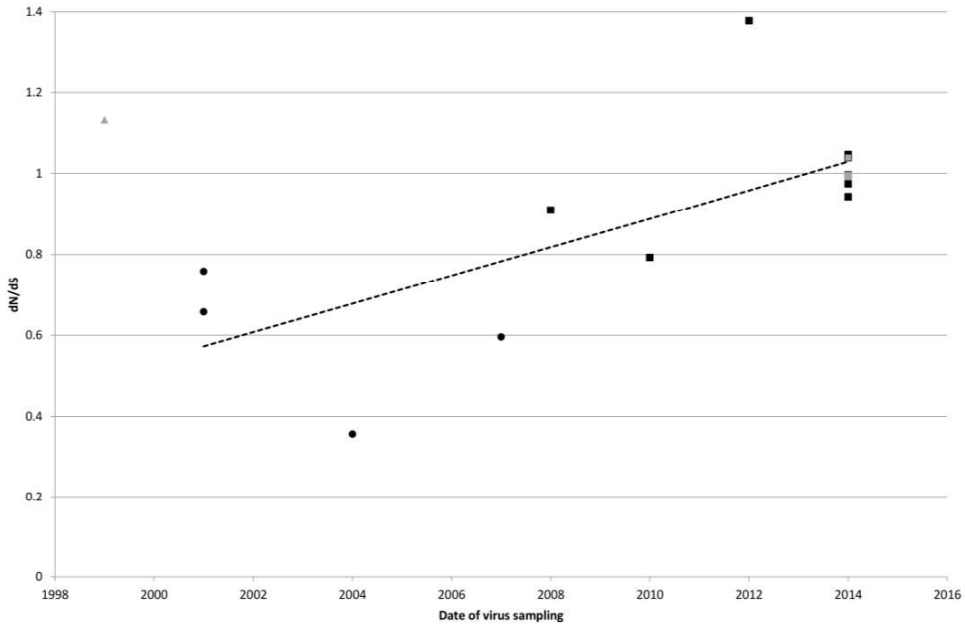

1.4
1.2
1
0.8
0.6
0.4
0.2
0
dN/dS
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Date of virus sampling

Table S1: All virus samples with NGS read data

ID	Area	Year	Viral Reads	% viral reads	Average depth coverage	% genome covered
RV2975D*	Izmir	1999	48,801	0.532	863	99.9
RV1124F	Izmir	1999	17,010	0.24	300	100
RV1126F	Izmir	2001	27,740	0.26	261	100
RV1127F	Manisa	2001	125,950	5.395	1225	100
RV1128F	Manisa	2001	15,435	0.18	161	100
RV2976F	Denizli	2004	30,993	1	341	99.9
RV2977F	Izmir	2006	20,791	0.31	225	100
RV2984F	Izmir	2007	45,162	4.771	361	99.8
RV2979F	Kutahya	2008	31,332	0.257	571	100
RV2980F	Manisa	2009	10,845	0.22	118	99.9
RV2981F	Izmir	2010	113,385	5.526	1070	100
RV2982F	Ankara	2012	54,074	0.83	578	100
RV2983F	Manisa	2012	11,565	0.438	176	99.9
RV3162F	Aksaray	2014	77,904	3.5	777	100
RV3163D	Aksaray	2014	30,321	1.45	338	100
Sub5790F	Cubuk	2015	79,651	3.17	760	100
Sub5791D	Cubuk	2015	318,799	17.65	2,795	100
Sub5792F	Yahsiyan	2014	43,663	2.01	484	100
RV3160F	Eregli	2015	95,237	3.65	1030	100
RV3161D	Eregli	2015	21,225	1	242	100
RV3166F	Boztepe	2015	11,945	0.26	55	100
RV1145F	Erzurum	2000	13,552	0.35	143	100
RV1144F	Erzurum	2001	33,268	0.79	376	99.9
RV1129F	Erzurum	2001	8,408	0.19	97	99.9
RV1136D	Bursa	2001	27,835	1.2	308	100
RV1137D	Bursa	2001	43,891	1.36	490	100
RV1142D	Istanbul	2001	140,830	3.3	1438	100
RV1133D	Ardahan	2001	450,449	8.7	3797	100
RV1134D	Ardahan	2001	169,262	3.15	1698	100

Samples are listed in date / location order and divided as follows: RV2975* pre-host-shift canine RABV; host-shift dataset, 20 isolates (17 fox and 3 dog) subdivided by dashed line to indicate early (green) and late (red) phases of the host shift; and non-host shift (blue) from other regions of Turkey

Table S2: Coding and non-coding lengths for the Turkey RABV genome sequence dataset.

Region	Length in nucleotides (nts)	
	Clade 1 (and other Turkey genomes)	Clade 2
3' UTR	70	70
N protein	1353	1353
N-P	90	90
P protein	894	894
P-M	88	88
M protein	609	609
M-G	**211**	**212**
G protein	1575	1575
G-L	**518**	**518^**
L protein	6384	6384
5' UTR	131	131
Total	11923 (10,815)	11924-5 (10,815)

The total genome length is indicated, concatenated coding length in brackets. The regions with differences between the dataset are in bold. All clade 2 viruses have an extra base (A) at position 3193 in the M-G intergenic region (See Table 3 for deep sequence analysis of this position). ^RV3160F and Sub5792D have extra base (A) at position 5348 (See Table S5 for deep sequence analysis of this position).

Table S3: Deep sequence data for position 3192, a homopolymeric region of either 5 or 6 As in the M-G intergenic region.

%A&V !%	% p cov r g	4 A	5 A's C 1	6 A's C 2
%V2975%	445	5	**440**	0
%V1124F>	216	0	23	**193**
%V1126F	255	0	**255**	0
%V2977F>	180	0	0	**180**
%V1127F	1,089	0	**1,089**	0
%V1128F	157	0	**157**	0
%V2984F	613	0	**613**	0
%V2976F>	214	0	24	**190**
%V2979F>	539	0	37	**502**
%V2980F>	110	0	0	**110**
%V2981F>	758	0	53	**705**
%V2982F>	546	0	47	**499**
%V2983F	916	0	**916**	0
%V3162F>	662	0	7	**655**
%V3163%>	283	0	4	**279**
ub5790F>	580	0	4	**576**
ub5791%>	2009	0	18	**1991**
ub5792F>	368	0	2	**366**
%V3160F>	722	0	3	**719**
%V3161%>	164	0	0	**164**
%V3166F	127	2	**125**	0
%V1145F	133	0	**133**	0
%V1144F	356	2	**354**	0

%V1129F	89	0	**89**	0
%V1136%	289	1	**288**	0
%V1137%	356	0	**356**	0
%V1142%	1,078	3	**1,075**	0
%V1133%	2,833	8	**2,825**	0
%V1134%	1,451	4	**1,447**	0
%V201%%	1904	6	**1898**	0

For ch qu c r p po o 3192 s ow o g umb r o r
ve 5As 6As. Co our co b 1, c 1( ) cl 2(>) bo o s
m jor y (co su) popu o .

Table S4: Deep sequence data for position 5348, a homopolymeric region of either 7 or 8 As in the G-L intergenic region.

%A&V !%	% p cov r g	7 A's	8 A's	9 A
%V2984F	833	**814**	19	0
%V1127F	1,0156	**995**	21	0
%V2981F>	1,399	**1,372**	27	0
%V2979F>	493	**347**	146	0
%V2982F>	592	**563**	29	0
ub5791%>	2,359	**2,312**	47	0
ub5792F>	434	6	**373**	55
%V3162F>	887	**876**	11	0
%V3160F>	942	8	**838**	104
%V1142%	1,417	**1,402**	15	0
%V1133%	3,812	**3,772**	40	0
%V1134%	1,831	**1,811**	20	0

For ch qu c w r rog y w ob rv ub-co su r p po o 5348 s ow o g umb r o r v 7-9As. Co our co b 1, cl 1( ) c 2(>) bo o s m jor y (co u) popu o .

Table S5: SNPs distinguishing clade 1 and 2 within host shift virus dataset across the concatenated genes.

Gene	N			P		G							L												
Position (nt)	129	297	639	1503	2179	3027	3057	3084	3204	3976	4173	4327	5163	5577	5598	6312	6729	7050	8166	8223	8430	8457	8984	10111	10335	10524
Position (aa)	43	99	213	50	276	57	67	76	116	378	439	491	244	382	389	627	766	873	1245	1264	1333	1342	1518	1894	1968	2031
Clade 1 nt	CTA	AAA	TTT	CCT	CTA	CTT	GTT	AAC	GCG	GTG	CTT	CGC	CTT	CTG	GTT	AGA	ATC	GTT	TCT	ATC	CAG	AAA	ACG	AGC	ATT	AAC
Clade 1 aa	Leu	Lys	Phe	Pro	Leu	Leu	Val	Asn	Ala	Val	Leu	Arg	Leu	Leu	Val	Arg	Ile	Val	Ser	Ile	Gln	Lys	Thr	Ser	Ile	Asn
Clade 2 nt	CTG	AAG	TTC	CCC	TTA	CTC	GTC	AAT	GCA	ATG	CTC	AGC	CTC	CTT	GTC	AGG	ATA	GTC	TCC	ATT	CAA	AAG	ATG	GGC	ATC	AAT
Clade 2 aa	*	*	*	*	*	*	*	*	*	Met	*	Ser	*	*	*	*	*	*	*	*	*	*	Met	Gly	*	*

Positions (nt) relate to concatenated sequences.
Positions (aa) relate to the individual protein residues.
Synonymous changes are indicated by *, non-synonymous changes are underlined.

Table S6: All virus samples with H Index and Entropy scores to compare the relative heterogeneity between viruses over time.

ID	Area	Year	Viral Reads	Heterogeneity Index	Shannon Entropy
RV2975D	Izmir	1999	48,801	713 (710)	542
RV1124F	Izmir	1999	17,010	1052 (1044)	742
RV1126F	Izmir	2001	27,740	1327	946
RV1127F	Manisa	2001	125,950	1214	879
RV1128F	Manisa	2001	15,435	641	414
RV2976F	Denizli	2004	30,993	1712	957
RV2977F	Izmir	2006	20,791	774	533
RV2984F	Izmir	2007	45,162	1841	879
RV2979F	Kutahya	2008	31,332	766	600
RV2980F	Manisa	2009	10,845	740	476
RV2981F	Izmir	2010	113,385	955	748
RV2982F	Ankara	2012	54,074	1048	529
RV2983F	Manisa	2012	11,565	427 (426)	332
RV3162F	Aksaray	2014	77,904	995	808
RV3163D	Aksaray	2014	30,321	635	488
Sub5790F	Cubuk	2015	79,651	963	807
Sub5791D	Cubuk	2015	318,799	1144	940
Sub5792F	Yahsiyan	2014	43,663	503	378
RV3160F	Eregli	2015	95,237	801	552
RV3161D	Eregli	2015	21,225	442	333
RV3166F	Boztepe	2015	11,945	688	429
RV1145F	Erzurum	2000	13,552	1493	773
RV1144F	Erzurum	2001	33,268	415	252
RV1129F	Erzurum	2001	8,408	384	259
RV1136D	Bursa	2001	27,835	673	446
RV1137D	Bursa	2001	43,891	501	338
RV1142D	Istanbul	2001	140,830	547	417
RV1133D	Ardahan	2001	450,449	873	632
RV1134D	Ardahan	2001	169,262	617	465

! umb r br cke r r o m ys o co occ o (xc p %V1124F w r m v ru w qu c o p ru b or y s). r y p ox mp (gr), p ox og mp (r) o - o v ru (b u). r v ru s 25,000 v r r w r o clu or rog y y .

Text S1: Partial N-gene Bayesian reconstructions provide geographical context

Initially, the RABV sequences were analysed with published sequences from neighboring countries to establish a wider geographical context. To enable the most inclusive analysis, a 400 base pair region of the nucleoprotein (N) gene was selected. Using Bayesian reconstructions the Turkish RABV sequences all cluster within the Middle East clades previously described, indicating the diversity of the RABV circulating in the country (Fig S2) (Horton, et al. 2015). However, the fox host-shift dataset all cluster within Middle East clade C (Fig S2 inset), separated into the two clades, with low posterior probabilities separating the clades (Fig S2 insert) and the sequences within (data not shown). Repeating the analysis using concatenated gene sequences (N, P, M, G and L) confirmed the topology with a much higher resolution and significant posterior probabilities (Fig 2). Clade 1 comprises of fox RABV sequences from the Izmir and Manisa region from 2001 until 2012, with a single sequence from the Ankara region (RV3166F, Boztepe, 2015), whereas Clade 2 contains sequences from the Izmir region but also from Ankara regions representing the subsequent spread of the disease (Fig 2).

References

Horton DL, McElhinney LM, Freuling CM, Marston DA, Banyard AC, Goharrriz H, Wise E, Breed AC, Saturday G, Kolodziejek J, et al. 2015. Complex epidemiology of a zoonotic disease in a culturally diverse region: phylogeography of rabies virus in the Middle East. PLoS Negl Trop Dis 9:e0003569.

Text S2: WGS provides evidence of ongoing infection within vaccination areas

The sequences cluster according to both geographic location and sample collection date (Fig 2 and S2). However, RV2980F (Manisa 2009) and RV2981F (Izmir 2010) were collected within the vaccination area and have a common ancestor with the early phase viruses that had circulated in the same area 10 years previously, rather than contemporary viruses circulating in the Ankara region. In addition, RV2983F (Manisa 2012), was collected within the vaccination area, after the vaccination ceased and has a recent common ancestor with the early phase Clade 1 viruses that circulated in Manisa 11 years previously, yet is more closely related to RV3166F (Boztepe 2015). Taken together these data indicate that the fox RABV was not eliminated within the vaccination zone by the third and final vaccination campaign in 2010, and that the virus continued to spread eastward to the Ankara region.

Text S3: WGS provides evidence that the fox RABV circulating in the Ankara region has transmitted back into dogs

The emergent fox clades spread, with repeated spill-over events back into dogs, evident in the three pairs of dog and fox rabies sequences that are highly similar, and closest neighbours on the phylogenetic tree: Cubuk sub5790F/sub5791D, Eregli RV3160F/RV3161D and Aksaray RV3162F/RV3163D (Fig 2). The Cubuk pair were collected 11 days apart (Table 1) and are highly related, with only 2 single nucleotide polymorphisms (SNPs) that vary over the entire genome: A/T^{1843} (phosphoprotein gene) and G/A^{9314} (RNA dependent RNA polymerase - RdRp gene), the latter resulting in R/K conservative amino acid substitution. The Eregli pair was sampled 1 day apart, although with 17 SNPs across the genome these viruses are less likely to represent a direct transmission event, albeit that RV3161D was infected with a fox RABV that is highly related to RV3160F, but not represented in this dataset. Similarly, the final pair, Aksaray, do not have detailed sample date data, but with 18 SNPs across the genome, are also highly related, yet unlikely to represent a direct transmission.

6.4: Article 32 in review with Molecular Ecology

Landscape attributes governing local transmission of an endemic zoonotic virus: rabies in domestic dogs

Summary: Based on a unique dataset of genetic, epidemiological and landscape data, including village-level vaccination coverage and dog density, the landscape attributes influencing the spread and persistence of RABV were investigated. Specifically, the mechanisms underlying transmission and the magnitude of their effect on phylogeographic signatures were assessed. Dog presence showed the strongest and most consistent level of support in each measure of the diffusion process, indicating that the spatial structure of the dog population is an important determinant of RABV transmission. Elevation also ranked highly, which can be regarded partially as an indicator of host structure given that human settlements (and therefore dogs) are less likely at higher elevations. There was also considerable support for the impact of physical attributes affecting host movement with evidence for rivers as barriers and roads as facilitators of RABV movement. Total dog density appeared to have limited effect on measures of RABV movement but susceptible dog density delivered a marginally better overall score. There was some evidence to support the impact of control, with average vaccination coverage and the susceptible dog density both performing better than the null model. However, the consistency of vaccination campaigns over a 10-year period had no apparent effect on RABV movement, making no improvement on the null model or generating any significant results (BF>3) in the model. The apparent failure of vaccination to reduce RABV movement is surprising, not least because its success in other areas has been demonstrated. However, vaccination is localised in Tanzania, to protect the Serengeti national park, therefore the results here indicate that regional vaccination efforts are less effective at controlling the spread of RABV over national scale vaccination campaigns.

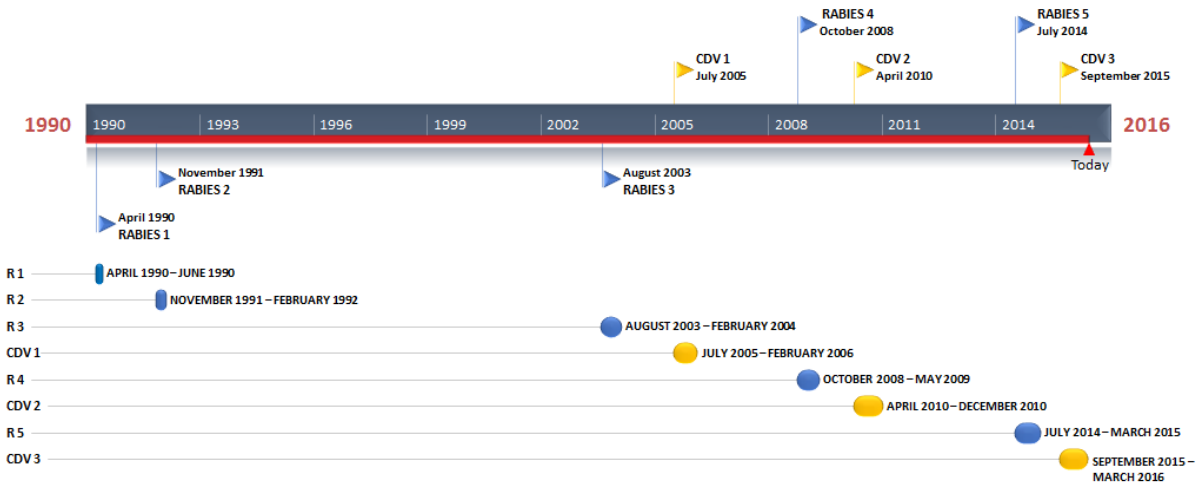
6.5: **Tracking RABV transmission in Ethiopian Wolf packs reveals multiple independent RABV introductions in the same outbreak**

6.5.1: Summary

The Ethiopian wolf (*Canis simensis*) is recognized as the rarest canid species in the world and as the most threatened carnivore in Africa. Fewer than 500 adult and sub-adult wolves remain in six Afroalpine habitat ranges. The largest Ethiopian wolf population is in the Bale Mountains National Park (BMNP) in southeastern Ethiopia, where wolf populations reach densities of up to 1.6 adults and sub-adults/km$_2$. On average, family packs contain 6 adult and sub-adults (range 2–20) and protect a home range of between 2 and 12km^2 (Marino et al., 2012). Ethiopian wolves are highly social and during periods of high wolf densities; increased packs numbers, and intense social behaviours increase the risks for disease transmission. In addition, Ethiopia has the highest incidence of rabies in Africa, people living in close proximity to the wolf habitat keep dogs to protect their livestock, which act as reservoirs for canine viruses such as rabies virus (RABV) and canine distemper virus (CDV). Over the past 3 decades there have been 5 rabies outbreaks and 3 CDV outbreaks in these endangered populations. Each outbreak occurred when wolf populations were at a high density and each one severely reduced the population size (Figure 13).

Figure 13: Timeline of disease outbreaks in Bale Ethiopian wolves.

Based on the APHA sample log and literature search – adapted from Miss Jemma Watson.

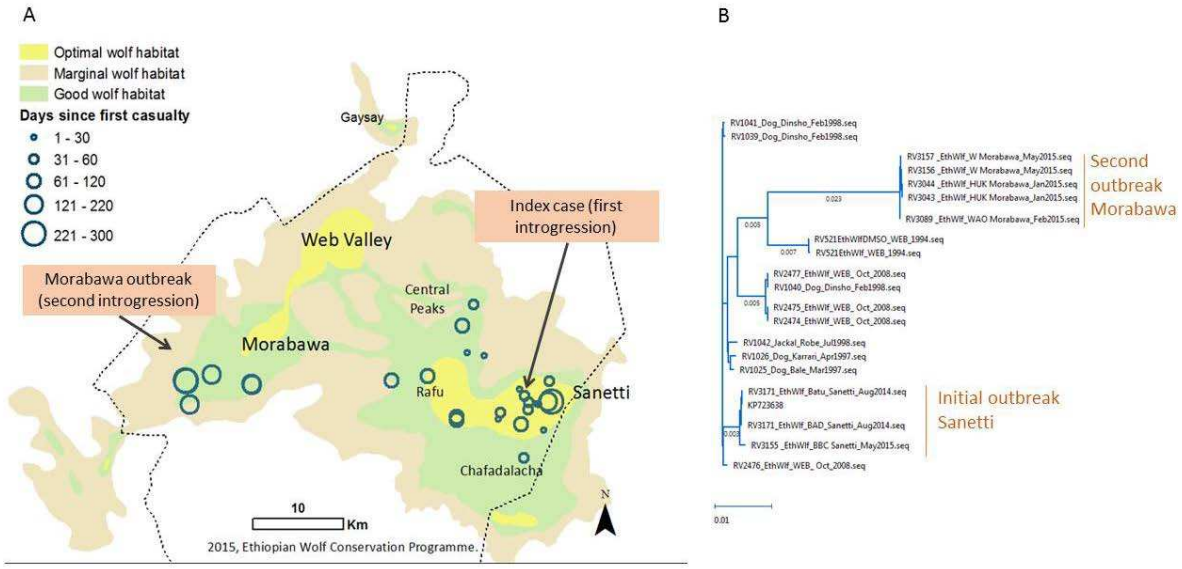


In order to understand the transmission and origin of the most recent outbreak, Ethiopian wolf rabies positive samples were fully genome sequenced along with wolf samples from previous outbreaks and canine samples from the same country. Unfortunately, representative dog rabies samples from Bale and surrounding regions are not available for sequencing, limiting our interpretation on where the RABVs originated.

The wolf population is closely monitored by a dedicated team of ecologists, conservationists and both regional and national government officials. Consequently each pack composition is recorded and therefore any disease outbreaks are quickly identified. During disease outbreaks post-mortems are completed on all carcasses and samples sent for diagnosis to APHA (CDV and rabies). The detailed epidemiological data is used to inform decision on vaccination and other control measures. The 2014-15 epizootic began in Sanetti and involved multiple packs lasting 13 months. The Morawaba epizootic affected multiple packs but only lasted 4 months. Epidemiologically, the last cases in Senetti and the first case in Morabawa overlap, suggesting that the virus had spread from Senetti to the Morabawa pack. However, the phylogenetic analysis utilising complete genomes indicated that two separate incursions occurred into the wolf population by two genetically distinct RABVs (Figure 14). Furthermore, analysis with previous wolf epizootics revealed that the Morabawa RABV was more closely related genetically to the virus involved in the 1992 wolf outbreak, 20 years earlier, than to the virus that infected the Sanetti packs, indicating that the RABV source of the Sanetti outbreak might be from a different region. Analysis of the 2008 outbreak viruses held in APHA also revealed two separate introductions of RABV into the wolf population, one representing a local RABV strain similar to the Morabawa and 1992 outbreaks and another more similar to the Sanetti RABV. The complete genome sequence obtained in article 6 was used as the reference sequence to map each genome against.

Figure 14: Map of Bale Mountains with 2014-15 epizootic cases.

(A) Bale Mountains map detailing the wolf habitat and geographic locations. RABV index case of the most recent epizootic (2014-15) in each area indicated by arrow and all subsequent cases are indicated by a circle, the size proportional to the number of days since first case in that area. (B) phylogenetic analysis of the complete genome sequences of the positive RABV cases from the most recent outbreak alongside historic outbreak representatives and other available positive RABV animals. Adapted from figure courtesy of Claudio Sillero.

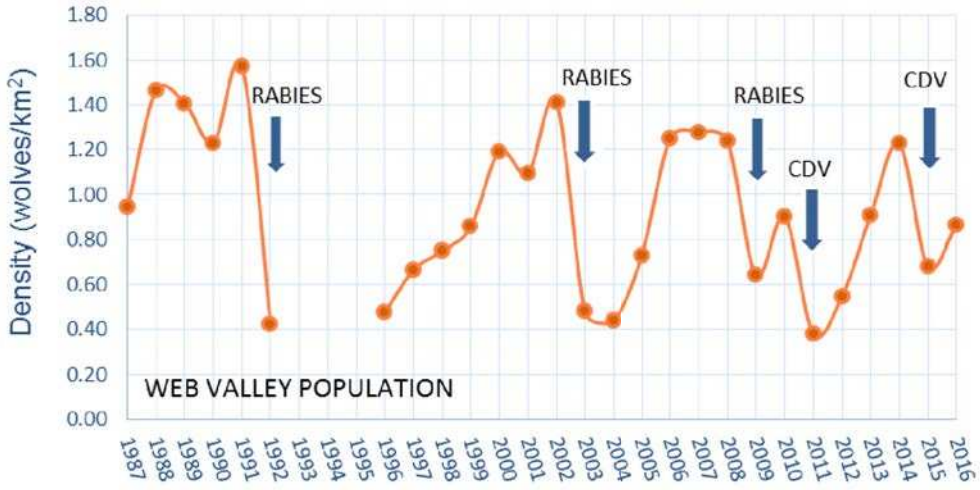


The low wolf population size found in the Bale Mountains is likely to be too small to maintain endemic infection. Wolf population density graphs produced by the EWCP as seen in Figure 15, show growth peaks followed by rapid declines associated with disease outbreaks. The change in population densities due to disease strongly suggests an intermittent source transmission and has been described for CDV previously (Bedin et al., 2015). Disease occurrence is therefore most likely to be associated with increased human settlement around park boundaries. Since 1994 the human population in two areas in southern Bale has increased almost threefold from 80,000 to >200,000 (Wakjira et al., 2015). Between 1990 and 2016 the population in Ethiopia has grown from 48 million to 101 million, 81% of which is rural (UN, 2015). If human settlement continues unabated, wolf territory boundaries will contract, resulting in higher contact rates with domestic dogs.

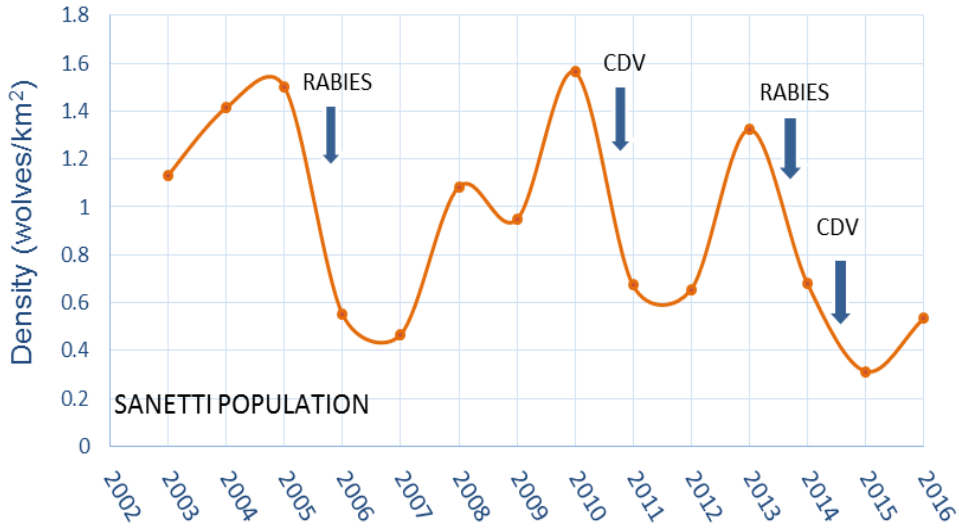
Figure 15: Timeline showing density of wolves/km^2.

Sharp declines appear to occur just after peaks in population density and in association with confirmed disease outbreaks suggesting point source infection rather than disease maintenance within the population. (A) Web valley and (B) Sanetti Ethiopian wolf populations. Source: EWCP Annual Report 2016 and adapted from figure courtesy of Claudio Sillero.

A)



B)



Vaccination of the local dog population around the Bale Mountains occurs annually and approximately 3-4,000 dogs are vaccinated each year. However, with the regular introduction of dogs from other regions and regular movement of people and dogs, the risk of RABV from the dog population is still present. The vaccination of local dogs has prevented RABV maintenance in the local dog population. RABV suspect dogs samples from the Bale Mountains are not diagnosed RABV positive, often they are CDV positive. It is also possible that other wildlife such as Jackals could be a reservoir for RABV. However, Jackals are not usually found in the Ethiopian wolf habitat. Vaccination of the wolves occurs in response to each epizootic, although this is costly and time consuming. Trails using oral vaccination are a promising and economic alternative (Gordon et al., 2015).

6.5.2: Preliminary Conclusions

Ethiopian wolves are highly endangered and viral diseases such as rabies and canine distemper are threatening their existence. Due to the small population sizes and also the vaccination campaigns in response to rabies epizootics, the wolves do not maintain rabies within the population; rather each epizootic is a result of an introduction of rabies and subsequent spread. Epizootics are observed within the wolf population when the density is above 1.2 wolves/km, indicating that this is a critical density to enable spread of the virus between animals. Full genome sequencing of the RABVs from the 2014-5 epizootic was undertaken, along with sequencing of other historic epizootic samples and archive dog and jackal RABVs from the Bale Mountain area. The 9 sequences from the 2014-15 outbreaks grouped into 2 distinct clades according to geographic location, strongly suggesting two separate introductions of RABV, the first in Spring 2014 in Senetti and the second at the beginning of 2015 in Morabawa. Furthermore the 2008 epizootic samples clustered into 2 separate clades, indicating multiple incursions. Within the two 2014-15 epizootic groups, the sequences were highly related with single nucleotide polymorphisms (SNPs) differentiating them. Less than 10 SNPs were identified between the 5 Morabawa sequences and less than 20 between the Sanetti sequences, the majority are non-synonymous or non-coding, strongly suggesting genetic drift rather than positive selection is occurring during transmission within the wolf population. Revealing that canine RABVs can infect and maintain within the Ethiopian wolf populations without requiring genetic adaptation is an important finding. Moreover, understanding the number of incursions and providing context to the reservoir of these incursions is vital to ensure the correct control strategies are implemented.

6.6: Conclusions

The papers described in this chapter represent comprehensive analyses of RABV population dynamics. WGS was used for the analyses whether possible, utilising partial sequence analysis only for contextualising the samples according to the global RABV situation. Understanding the transmission of RABV within dog, fox and other wildlife populations is paramount to ensure effective control measures are implemented.

From a practical standpoint, in a UK rabies outbreak scenario APHA are in a position to obtain complete genome sequences within days of submission, and with fine resolution transmission analyses we will be able to: a) confirm the source of the infection (for example from endemic EBLV-2 or from imported canine strain from a certain country), b) reconstruct transmission pathways to identify possible missing cases, human mediated translocations, or the presence of a different rabies source and c) if any CSTs occur, we can analyse identify any adaptations which may indicate a more permanent host shift. Combined, these data will be critical to inform policy, to ensure the correct preventative actions are taken to eradicate the disease.

Perhaps the biggest caveat to the success of these and future analyses, is understanding the sampling bias, particularly whether there are unsampled variants circulating in the same regions. Failure to include a clade or lineage from an unsampled area will result in ancestral reconstructions that may not capture the dispersal pattern well (Dellicour et al., 2017). Due to the diligence and long association our collaborators in Glasgow University have with the Tanzanian authorities and local people, the RABV samples collected within Ngorongoro and surrounding Serengeti regions are extremely comprehensive, with detailed records including any biting incidences. Furthermore, similar detailed information is gathered pertaining to the Ethiopian wolves, as highly endangered animals, conservationists monitor the wolves constantly, resulting in over 90% of all wolf carcasses being retrieved after death, providing complete data regarding the number affected by each RABV outbreak. Having access to such unique RABV datasets provides an unrivalled opportunity to study the transmission dynamics of RABV.

Chapter 7: Utilising NGS data to investigate viral heterogeneity

Chapter 7: Utilising NGS data to investigate viral heterogeneity

7.1: Introduction

Viral heterogeneity, representing nucleotide sequence heterogeneity of the viral RNA, arises from frequent nucleotide misincorporations during replication, which is a direct result of the lack of proof-reading by the viral RNA-dependent RNA polymerase (RdRP). Combined with a rapid replication rate, RNA viruses exist as heterogeneous populations call quasispecies, or mutant swarms. Some areas of virology have made huge progress in this fascinating area, particularly chronic viruses such as HIV and hepatitis, where drug resistance mutations can be identified. However for acute virus infections such as rabies, viral heterogeneity has hardly been addressed. Before the advent of NGS, cloned PCR products were sequenced to identify sub-consensus variants. Limited studies were conducted on lyssaviruses, perhaps the most comprehensive was an investigation into the glycoprotein after passaging an European fox RABV in different hosts (Kissi et al., 1999). No host specific consensus level changes were observed, although the amount of sub consensus heterogeneity observed varied depending on the host used to passage the virus. Furthermore, the fixed RABV strain CVS was passaged experimentally in BHK cells, resulting in selection of a dominant variant that differs genetically and phenotypically from the consensus sequence present in mice or neuroblastoma cell passaged CVS (Morimoto et al., 1998). The main limitations with these studies were the time consuming nature of the laboratory work, and the inherent issues with the introduction of PCR errors, resulting in limitations to the validity of the data. Since the late 1990's until the development of NGS techniques, little progress in the area of lyssavirus heterogeneity was made. Indeed, even now, only a limited number of articles have been published looking at sub-viral populations in lyssavirus infections. A search in PubMed with the terms 'quasispecies AND rabies' retrieves 4 records (Benmansour et al., 1992; Borucki et al., 2013; Kissi et al., 1999; Morimoto et al., 1998). An interesting article detailing the presence of DI genomes in CVS using de novo assembly is also applicable in the context of sub-genomic variations, but as a fixed laboratory strain the question of how applicable DI particles are in natural infections remains (Hoper et al., 2012). A single manuscript, published on the subject of viral heterogeneity where NGS is utilised, is the investigation following a CST from striped skunks to grey foxes in California (Borucki et al., 2013). This comprehensive paper obtained complete genome sequences from 40 of the infected grey foxes, investigating the transmission and adaptation of the virus in the new host. However, PCR amplicons were sequenced using NGS technologies, which introduced additional error into the dataset. In addition to this drawback, the deep sequence data

was limited to analysis of the consensus level changes observed, to identify the minority variants in previous samples, rather than quantify the amount of heterogeneity within the sample.

With such a limited publication history beforehand, this chapter includes articles either in the review process or in preparation which attempt to tackle some important questions:

1) Is there evidence for viral population heterogeneity in lyssaviruses

2) Does the level of viral population heterogeneity in lyssaviruses vary between virus species and can this be linked to a restricted host reservoir

3) Do lyssaviruses use viral heterogeneity to adapt to a new host species?

These questions have been tackled in the following articles

7.2: Article 31 in review with Virus Evolution

Analysis of a rabies virus host shift event reveals an increase in sub-viral populations within the new host reservoir.

(see section 6.3, page 84 for printed copy)

Summary: This article represents a comprehensive WGS analysis of a cohort of RABVs from Turkish red foxes. It is discussed twice within this thesis because two distinct analyses are undertaken. The first discussed previously in Chapter 6.3, is the utilisation of Bayesian inferences to investigate the source and timing of the host shift event. Transmission of the fox RABV back into dogs is also investigated. The second, discussed here, is the utilisation of the deep sequence data to investigate the influence of sub-consensus populations on virus adaptation.

Viral heterogeneity has been long proposed as a mechanism for viruses to adapt to new environments; therefore we used the deep sequence data obtained from the WGS analysis to investigate the heterogeneity in the sub-consensus viral population, in particular to investigate whether sub-consensus level changes were driving adaptation of RABV in a new host. As WGS was undertaken directly, without using PCR amplicons, the deep sequence data was analysed directly without having to normalise for PCR errors.

Utilising a heterogeneity index (H Index), fox virus heterogeneity was compared across all the Turkey viruses, in particular investigating the fox enzootic viruses in relation to the other Turkey viruses. Shannon entropy scores were also generated and indicated that H Index and entropy analyses were comparable. Analysis of the H Index dataset demonstrated the combined H Index values from outside the study area are significantly lower than the early phase host shift fox values. Furthermore, higher H Index scores were observed in the early phase viruses compared to the late phase. To investigate the significance of the increase heterogeneity observed in the early phase of the host shift, the dN/dS ratios of both the consensus and the sub-consensus viral population were examined. The sub-consensus sequences had higher dN/dS values in comparison to the consensus sequences. The higher ratio reflects the heterogeneity present in the sub-consensus population, which contains non-infectious genomes and genomes with indels and deleterious residues unable to establish a new infection, in addition to those which may contain advantageous mutations. Across the 5 genes, the dN/dS ratio of the sub-consensus reads indicated more synonymous than non-synonymous

substitutions in the early phase viruses, than the late phase and non-host shift viruses, corresponding to the observed increase in heterogeneity in these viruses.

Despite the smaller than optimum sample sizes, this dataset is unique, representing one of the few recorded, sampled RABV host shifts. Of the few published studies on RABV host shifts, none have provided definitive evidence for positive selection at the consensus level related to the host shift event. Therefore investigation of the sub-consensus data, and the spectrum of viral heterogeneity contained within, is a logical progression. The mechanisms involved in modulating the diversity observed at the sub-consensus level are currently undefined; however there are a number of plausible options. Firstly, the increase in heterogeneity observed in the early phase viruses could be a result of an increased incubation time within the new host, resulting in increased viral replication. It is possible that increased incubation (either at the site of inoculation, within the CNS or both) early on in a host shift could result in increased viral heterogeneity, which may reduce as the virus transmits within the new host, adapting over time. Alternatively, short sighted virulence, first described for bacterial meningitis, poliomyelitis and HIV/AIDS, considers that mutation and positive selection within a host increases virulence, but provides no advantage to the transmission of the pathogen. Therefore, reducing viral heterogeneity in established host reservoirs such as bats and dogs might also be a valid strategy to increase transmission within the host population. It is understood that the majority of RABV CSTs do not continue within the new host population, which in part is likely due to the host succumbing to the disease before onward transmission. However, if onward transmission does occur, minimising the effect of 'short-sighted virulence' will be important for the virus to establish within the new host population. These data presented here would fit this hypothesis, as the reduction of viral heterogeneity over time becomes evident in the new fox population.

The observation that the variation in the viral sub-population, rather than at a consensus level, is associated with success of a RABV host shift, is worthy of further investigation, particularly where NGS data already exists (Borucki et al., 2013). With the increased availability of NGS platforms and decrease in cost, WGS will be pivotal in investigating and unravelling the mechanisms in which RABVs evolve and emerge in new host reservoirs.

See Section 6.3, Page 84 for the printed article.

7.3: Ultra deep WGS of Lagos bat virus reveals cell and host adaptation

7.3.1: Summary

A Lagos bat virus (LBV) from the brain of an infected *Eidolon helvum* (straw-coloured fruit bat) was passaged 4 times in BHK cells and then used to inoculate captive straw-coloured fruit bats within specially designed bat cages in Ghana. This study is part of an international collaboration investigating the *in vivo* characteristics of LBV. Due to the amount of inoculum required for the in vivo experiments the LBV virus required propagating within an *in vitro* system. Conscious of the issues surrounding adaptation of viruses during propagation in cells the minimum number of passages were undertaken to amplify the virus. Four passages resulted in a high titre virus, subsequently used to infect *E. helvum* peripherally. One clinical bat (Bat 19) brain was harvested once clinical signs were observed and RNA extracted and sequenced using methodologies described in Chapter 2. Comparison of the passaged material (P2 and P4) alongside the original bat genome and Bat19 genome revealed two consensus level changes, which resulted in amino acid substitutions; and a third position where substantial variation is observed within the deep sequence data, although the consensus did not alter (Figure 16). This third position is within the genomic promoter, and may be important for regulation of transcription and replication. Furthermore, the percentage of reads that have either the original or the altered SNP were calculated for all three positons.

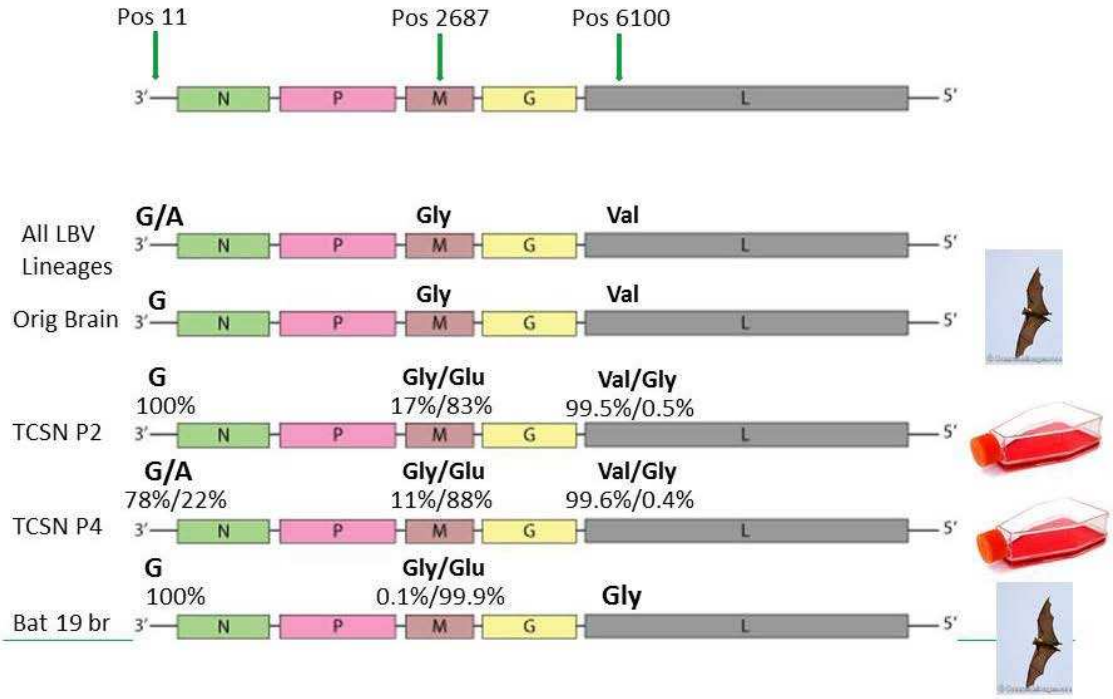
The first site where heterogeneity is observed, at position 11 of the genome, is G in the consensus sequence for all samples analysed. However, in the sub consensus population in P4 22% of reads have A at this position. Whether further rounds of passaging in cells would have fixed this change is unknown but 100% of reads in the Bat 19 were G, indicating that G is important *in vivo*.

The second SNP at position 2687 (Matrix protein) Gly/Glu^{57} is a Glycine (Gly) in all known LBV sequences, yet by P2 the consensus sequence was Glutamic acid (Glu). The proportion of reads with Glu^{57} remained at a fairly constant level during cell culturing (83% in P2 and 88% in P4) and appeared to fix in Bat 19 (99.9%) suggesting that either this change is due to natural drift, resulting in neither a positive or negative effect on the virus in both environments, or that *in vitro* Glu^{57} was advantageous, then tolerated within the natural host resulting in fixing Glu^{57}.

The final position at 6100 in the RNA dependent RNA polymerase (RdRP) Val/Gly^{205} is interesting because a low level variant Glycine (Gly) present at 0.5% in both passages in cells becomes dominant when the virus is inoculated back into the host reservoir. As the original bat brain sample

was sequenced by Roche 454 pyrosequencing and no further material is available for resequencing on Illumina we are unable to know whether Gly[205] was also present at a low level in the original virus.

Figure 16: Schematic of the consensus level changes observed during the passaging with percentage of reads representing each SNP.



7.3.2: Preliminary Conclusions

Despite 4 passages of LBV in cells the consensus sequence remains remarkably similar to the original LBV genome sequence. Two non-synonymous SNPs in the consensus sequences have been observed. The first, M Gly/Glu[57] is not situated in any known domains and is likely due to natural drift as the SNP due to passaging in cells remained when introduced into the natural host. The second SNP, *RdRP* Val/Gly[205], is also outside of any known functional domains, and fixed in Bat 19 whilst only observed within the cell cultured samples at a low (but detectable) level. The sub-consensus SNP observed within the cultured samples, at 22% in P4, might be a cell adaptation and therefore would have fixed to 100% if the passaging *in vitro* continued, but was not detected in Bat 19, suggesting that this change was not tolerated within the natural host.

To continue this work, the H Index and dN/dS ratios from the in vitro and in vivo viruses will be obtained and compared in relation to each other to see if any adaptation at the sub-consensus

level is occurring. Samples from a second in vivo experiment (intracranial inoculation) will also be analysed.

7.4: **Investigating heterogeneity across lyssavirus species**

7.4.1: Summary

One of the original hypotheses of this thesis was that there could be a link between the number of CST and host shifts documented for each lyssavirus species and the amount of heterogeneity observed in the sub-consensus population within the host reservoirs for those species (Question 2 in section 7.1). The observation that RABV has multiple host reservoirs both volant and non-volant species, whereas other lyssaviruses are highly host specific, has been explored in detail in Article 1. In Article 1 we hinted that one explanation could be lyssaviruses that are host restricted (such as EBLV-2) could have a reduced heterogeneity within the host due to the long association with that host. The long associations with the same host species results in the sub-viral population stabilizing as no selective pressures are being exerted on the population. This is in direct contrast to RABV, which employs a strategy of 'hit and run' whereby propagation is not reliant on host association, rather by ensuring successful transmission to a new host, be it the same species or different. The heterogeneity observed in the sub-viral population in RABVs, would be higher as selective pressures are present in the new hosts. Every transmission from one animal to another, for every lyssavirus is a bottleneck with little heterogeneity in the inoculum, therefore the heterogeneity observed in the brain of each host is a result of replication from an almost clonal inoculum. The increased diversity observed after a host shift might indeed by a direct result of increased replication time required by the virus which had be optimized in the previous host reservoir.

Samples from 6 lyssavirus species (some with more than one representative) were deep sequenced and H Index values calculated as described in Article 31 (Table 3). The H Index values range from 672 (LBV) and 1340 (EBLV-1).

Table 3: Representative H Index values for lyssavirus species

ID	Species	host	sample	multiple host reservoirs	H Index	Comments
CVS	RABV	NA	MB	Fixed strain	1059	
SHBV	RABV	bat	TCSN	Yes	909	
Turk Dog	RABV	Dog	brain	Yes	720.86	Average (n=6)
Turk Fox	RABV	Fox	brain	Yes	1099.18	Average (n=7)

host shift						
Turk Fox non-host shift	RABV	Fox	brain	Yes	760.67	Average (n=9)
LBV Bat 19	LBV	Bat	brain	No	672	
RV266	EBLV-1	Bat	brain	No	1340	
EBLV-2	EBLV-2	bat	brain	No	929.8	Average (n=5)
GBLV	GBLV	bat	brain	unknown	1203.8	Average (n=4)
LLEBV	LLEBV	bat	brain	unknown	679	

7.4.2: Preliminary Conclusions

The hypothesis that viral heterogeneity is linked to host specificity was interrogated in the analysis of the Turkey host shift dataset (Article 31). These data suggest that heterogeneity is not high in the host reservoirs per se, rather it increases after a CST event in the new host and for a number of years afterwards, without clear selection at the consensus sequence level. Although it is unclear as to how and why this occurs, one can consider that this is more of a reaction to a new environment rather than a strategy employed by the virus itself. With the data presented here, there is no obvious correlation between lyssavirus species that have successfully switched hosts and H Index values, corroborating the observation of the Turkey host shift analyses (Article 31), that increased heterogeneity is a result of a host switch, rather than a feature present in the host reservoir population. However, a major limitation of the analyses is that the majority of viruses that are believed to be host restricted only have a small number of isolations; therefore the host reservoir numbers are not known. For example GBLV has only recently been identified within a fruit bat species (*Pteropus*) in Sri Lanka, all four samples had high H Index scores (similar to the early phase host shift viruses from Turkey). EBLV-2 represents perhaps the best example of a lyssavirus that is believed to be host restricted, having only been isolated from *Myotis* species. Of the five EBLV-2 samples analyses, three had lower H Index values (456, 521 and 846), the other two EBLV-2 viruses had higher H Index values (1501, 1325), with no obvious explanation for these differences.

That the range of H index values in the lyssavirus sequences analysed span the range observed in the Turkey fox host shift viruses, despite the host shift potential raises a number of questions. Firstly does each lyssavirus species have its own H Index range so that they cannot be compared between

lyssavirus species, secondly, as these samples are all clinical samples is the H Index reflecting the stage of infection in the brain, e.g an artefact of viral load, or thirdly is there another unrecognized factor that determines the H Index values. Further work is necessary to determine which of these options, if any, are correct.

7.5: Conclusions

At the beginning of this chapter, three questions were posed which are discussed in more detail below:

Is there evidence for viral population heterogeneity in lyssaviruses?

In this chapter I have presented evidence of the presence of viral heterogeneity within lyssavirus samples. Both in vitro grown lyssaviruses (LBV culturing section 7.3) and clinical samples have been analysed and found to contain heterogeneity at the sub-consensus level (Section 7.2 and 7.4).

Does the level of viral population heterogeneity in lyssaviruses vary between virus species and can this be linked to a restricted host reservoir?

From the preliminary data analysed here, there is not a clear difference between the level of heterogeneity observed between lyssavirus species. That is to say, host restricted lyssavirus (such as EBLV-2) do not have an obviously different H Index in comparison to RABV. However, the evidence presented here does suggest that the H Index increases in clinical samples following a successful host switch. Whether the increase in heterogeneity is a beneficial strategy or a consequence of an undefined biological function is currently unclear.

Do lyssaviruses use viral heterogeneity to adapt to a new host species?

Evidence from the in vitro propagation of LBV indicates that the viral sub-population over the passages altered, presumably to adapt to the culture environment. Furthermore, it has become apparent that in nature host shift samples, positive selection of particular amino acid changes in the new host is not occurring. Therefore, either RABV is truly a virus which can effectively be transmitted amongst all mammalian host reservoirs with a truly universal receptor / host immune evasion functions, or the virus is using a different strategy to adapt to the new host. Altering the sub-consensus population is one such alternative strategy.

Chapter 8: Conclusions

Chapter 8: Conclusions

This thesis has focussed on increasing the understanding of lyssavirus evolution, particularly evolution within host reservoirs and host switching at the molecular level. At the beginning of this PhD, the NGS technologies which were developed and refined for human and bacterial genomes were still in their infancy for viral genomics. One might be forgiven for thinking that obtaining a complete genome for a virus which is 12,000 bases long might be easier than for a bacterium; however there are a number of virus specific critical issues which required specific thought. For example, the mapping programs such as BWA discarded reads with SNPs in comparison to the consensus sequence, as these were considered poor data in the field of bacteriology and human genetics. Whereas in Virology the presence of sub-consensus viral populations with heterogeneity in comparison to the consensus sequence is not only accepted, but vital to be included to analyse those viral populations. Other issues such as complementary genomic ends, low levels of virus in clinical samples and segmented genomes are all specific to the field of Virology and are still being optimised. Before attempting investigation of viral heterogeneity, developing sample preparation methodologies, followed by bioinformatics to analyse the 'big data' obtained from the sequencing machines were essential.

The methodologies developed within this thesis are hugely beneficial because they can be applied to any RNA or indeed DNA virus (if DNAse I treatment is removed). Within APHA, full genome sequences of over 11 virus genera have been successfully sequenced, including important pathogens such as avian influenza, coronaviruses, rift valley fever virus and indeed over 70 lyssaviruses. The analysis of the genome sequences, including phylogeography, virus transmission and deep sequence virus heterogeneity analysis all are made possible due to the refinement of these methodologies.

Three novel lyssaviruses have been identified during this PhD and complete genome sequence obtained using true de novo assembly. The first lyssavirus, IKOV has been fully characterised and accepted as a new species *Ikoma lyssavirus*. LLEBV, more recently identified by collaborators in Spain and GBLV, identified in a fruit bat colony in Sri Lanka have been approved by the ICTV committee and are awaiting ratification in early 2018. Complete genome sequence was pivotal to the proposals to the ICTV and would have taken many months, if not years to have obtained using traditional 'primer walking' methodologies. The importance of these lyssavirus discoveries and genome characterisation are perhaps best illustrated in the analysis of complete genomes from all lyssavirus species isolated from bats. Article 23 'The Global Phylogeography of Lyssaviruses - Challenging the 'Out of Africa' Hypothesis', utilises the whole genome sequences obtained to investigate the origins of bat

lyssaviruses. These analyses were undertaken using Bayesian Inferences testing the hypothesis using a probabilistic phylogeographical approach. The data presented suggested that the MRCA of the contemporary lyssaviruses originated in the Palearctic region. This is a controversial discovery as Africa contains the most diverse range of lyssavirus species, including representative from each phylogroup, therefore a more intuitive hypothesis would be that lyssaviruses evolved 'out of Africa' (Rupprecht et al., 2017). However, more lyssavirus species have been isolated from the Palearctic region than any other region, including representatives of phylogroup I and III. The data suggested that there have been 3 independent transmission events to the Afrotropical region representing the three major lyssavirus phylogroups. Although these conclusions are controversial, it should be noted that the Palearctic region includes North Africa. The close evolutionary relationship lyssaviruses have with bats is important and could hold the key to understanding the origins and dispersal of lyssaviruses. A recent paper investigating the evolution of hantaviruses by using phylogenetic analyses from both the viruses and their reservoir hosts provided 'fixed points' to anchor the viral phylogenies resulting in a more realistic time scale of evolution (Castel et al., 2017). Utilising bat phylogenetic data to interpret lyssavirus phylogenies could aid calibration of the lyssavirus ancestry.

The future of lyssavirus phylogeography will undoubtedly be shaped by the increasing availability of genome sequences. The articles included in Chapter 5 are good examples of utilising complete genome sequences to interrogate the evolution and dispersal of viruses. EBLV-2 analysis of complete genome sequences across Europe is a case in point where highly related viruses (identical when compared using partial N gene data) were distinguished and using date and location as priors were utilised in a large European dataset to investigate the time of the most recent common ancestor.

Reconstructing endemic transmission of RABV within dog, fox and other wildlife populations is useful to ensure effective control measures are implemented. In the UK FMDV outbreak were traced at a farm level and missing premises were identified using full genome sequences (Cottam et al., 2008). If a rabies outbreak occurred in the UK, it is now possible to obtain complete genome sequences within days of submission, and with fine resolution transmission analyses we will be able to: a) confirm the source of the infection (for example from endemic EBLV-2 or from imported canine strain from a certain country), b) reconstruct transmission pathways to identify possible missing cases, human mediated translocations, or the presence of a different rabies source and c) if a host switch occurs, investigate any adaptations which may indicate a more permanent host shift. Combined, these data will be critical to inform policy, to ensure the correct preventative actions are taken to eradicate the disease.

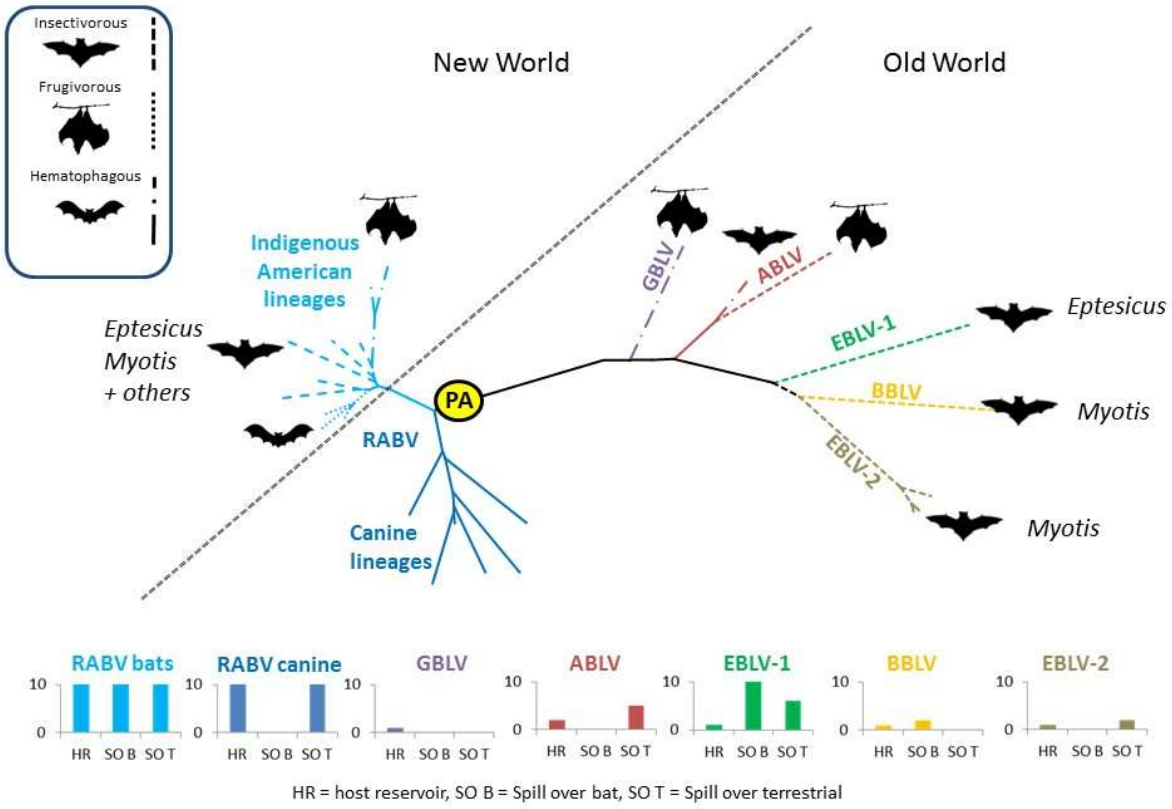
As a proof of principle, the Ethiopian wolf epizootic analysis provides an excellent dataset to hone our analysis. The dedication to these beautiful endangered creatures is impressive. Every wolf is recorded and monitored. Samples are taken from every carcase and therefore any disease outbreak is recorded from the first to last case. Genomic analysis is still underway, but initial findings indicate a lack of positive selection during the transmission of RABV within the wolves.

Having access to such unique RABV datasets and the methodologies to obtain complete genome sequences provides an unrivalled opportunity to study the transmission dynamics of RABV. Understanding bat to bat transmission for other lyssaviruses will be the next stage in understanding lyssavirus transmission; particularly as maintenance in bats populations could be different from terrestrial mammals. The EBLV-2 Stokesay dataset is perhaps the most comprehensive intrahost dataset available to date (n=4). Preliminary analysis reveals a low evolutionary rate, indicative of a long establish host/virus relationship.

The main focus of this thesis is to investigate the role of viral heterogeneity in the adaptation of lyssaviruses to new hosts. The association of all lyssavirus species, with the exception of RABV, to a single or highly restricted host species is compelling. For each lyssavirus species an element of geographical restriction, alongside an apparent host restriction, is observed (Banyard et al., 2011). Comprehensive surveillance schemes exist for bat lyssaviruses across Europe, yet the detection of EBLV-1, EBLV-2 and BBLV in any bat species other than the natural host reservoir is rare (BBLV and EBLV-1) or absent (EBLV-2) (Figure 17). Incidences of bat-associated rabies (RABV) in the Americas are more evident than the rest of the world. Moreover, hundreds of RABV spill over events are recorded, including multiple non-volant (Wallace et al., 2014) and volant species (Velasco-Villa et al., 2017).

Figure 17: Host restriction and reported spill over events for lyssaviruses (from article 1).

Unrooted, phylogenetic tree of six lyssavirus species with bat reservoir host indicated. Graphs indicate approximate relative numbers, where a value of 10 = 10 or more, of recorded host reservoirs (HR), spill overs in bat (SOB) and spill overs in terrestrial mammals (SOT) for each lyssavirus (and for RABV the bat and canine lineages). The RAC/SK lineage of RABV is not included for clarity). PA, suggested pre-adaptation stage.



The same bat species in the Americas and Old World (for example *Myotis* and *Eptesicus* species) are host reservoirs for different lyssaviruses. In the New World multiple *Myotis* and *Eptesicus* species are reservoirs for RABVs circulating in a large range of bat species, whereas in the Old World, specific *Eptesicus* species are reservoir hosts for EBLV-1 and specific *Myotis* species are reservoir hosts

for BBLV and EBLV-2, despite multiple bat species co-roosting and feeding. These observations suggest that the virus, rather than the host is driving this co-evolution.

It has become apparent that RABV host shifts are not accompanied by obvious adaptive mutations in the virus genomes, with no definitive positive selection identified correlating with host shifts (Borucki et al., 2013; Hanke et al., 2016; Kuzmin et al., 2012; Troupin et al., 2016). That 'pre-adaptation' has resulted in successful host shifts has been previously suggested (Kuzmin et al., 2012; Mollentze et al., 2014a). I have proposed that pre-adaptation is not something that occurs before each host shift; rather that RABV's pre-adaptation has been acquired at the virus species level, which has increased the success of onward transmission after a CST. Apart from RABV, ABLV is the only other lyssavirus species known to circulate in more than one host reservoir (*Pteropus* spp - fruit bats and *Saccolaimus* spp - insectivorous bats) (Weir et al., 2014). The closely related GBLV has recently been identified circulating within *Pteropus* spp in Sri Lanka (Gunawardena et al., 2016), although other host reservoirs are currently unknown. It is tempting to speculate that if lyssavirus surveillance in bat species within the Indian sub-continent were implemented, GBLV would be detected in other bat species, including insectivorous bats (Mani et al., 2017; Pal et al., 1980).

What the pre-adaptation might be is an interesting question. Perhaps a changed function in the viral machinery, for example, in the phosphoprotein or glycoprotein, or replication efficiency? Perhaps it is due to recombination? The controversial topic of recombination in lyssaviruses was proposed, to explain the RABV host shift from bats to raccoons in the Americas, although the requirement for a nonvolant species (skunk) RABV glycoprotein does not provide explanation for how the skunk variant came into existence (Ding et al., 2017). Without experimental evidence for recombination in lyssaviruses, it is difficult to understand how a virus with such an acute disease progression can co-infect the same animal. Furthermore, how can recombination occur when the RNP protects the genome for the majority of the life cycle? My hypothesis is that, an increase in the virus sub-population heterogeneity (quasispecies) occurs after the bottleneck which occurs during a CST event, as observed in arboviruses during mosquito transmission (Stapleford et al., 2014).

To investigate this hypothesis a unique cohort of clinical brain samples from foxes believed to be infected with a canine RABV strain (therefore representing a host shift event) in Turkey were analysed to investigate the viral heterogeneity present. In particular, to investigate whether sub-consensus level changes were driving adaptation of RABV in a new host. We identified a significant difference in the amount of heterogeneity present in the viruses immediately after the host shift event when compared to the viruses in the late phase, and non-host shift viruses. The higher heterogeneity

was accompanied by higher synonymous substitutions in the early phase viruses, than the late phase and non-host shift viruses.

The mechanisms modulating the diversity observed at the sub-consensus level are currently undefined; and are discussed in section 7.2. An increased incubation time within the new host, resulting in increased viral replication, therefore a more expansive sub-consensus population is a logical proposal and fits with historic in vivo data where the incubation time in a new host is longer than the host reservoir (Aubert et al., 1991; Blancou and Aubert, 1997). Alternatively, one can view the viruses adapted to a particular host reservoir as having reduced viral heterogeneity as a result of reducing virulence and optimising transmission within the host population. A recent paper discusses strategies to minimise the effect of 'short-sighted virulence' for certain chronic disease viruses by having low within host evolution to ensure onward transmission (Lythgoe et al., 2017). For an acute disease viruses with high fatality rates such as RABV it would be important to minimise virulence, and therefore within host heterogeneity to ensure the infect animal survives long enough to transmit to another animal. The majority of RABV CSTs do not continue within the new host population, which in part is likely due to the host succumbing to the disease before onward transmission. However, if onward transmission does occur, minimising the effect of 'short-sighted virulence' will be important for the virus to establish within the new host population.

Although we have discovered a difference between the heterogeneity observed in clinical brain samples after a host shift event for RABV, comparison of heterogeneity between lyssavirus species does not reveal obvious differences. For instance, host restricted lyssavirus such as EBLV-2, do not have different H Index values in comparison non-host restricted lyssaviruses such as RABV. Perhaps this is not surprising, the amount of heterogeneity in the brain sample will be relative for each lyssavirus species, therefore comparing across species is less likely to be informative than between samples involved in a host shift and those in the host reservoir.

In these exciting times of deep sequencing data, characterising novel viruses, tracking transmission of viruses in endemic and epizootic landscapes and investigating molecular mechanisms involved in host switching are all possible. Viruses will continue to adapt, evolve and emerge. We will continue to respond by developing control strategies and implementing global eradication campaigns. Understanding the molecular adaptations viruses employ to cross host species barriers will be key to predicting and preventing future viral emergence.

References

References

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

Amarasinghe, G.K., and Y. Bao. 2017. Taxonomy of the order Mononegavirales: update 2017.

Arechiga Ceballos, N., S. Vazquez Moron, J.M. Berciano, O. Nicolas, C. Aznar Lopez, J. Juste, C. Rodriguez Nevado, A. Aguilar Setien, and J.E. Echevarria. 2013. Novel lyssavirus in bat, Spain. *Emerg. Infect. Dis.* 19:793-795.

Arredondo, J., V.T. Nguyen, A.I. Chernyavsky, D. Bercovich, A. Orr-Urtreger, W. Kummer, K. Lips, D.E. Vetter, and S.A. Grando. 2002. Central role of alpha7 nicotinic receptor in differentiation of the stratified squamous epithelium. *The Journal of cell biology*. 159:325-336.

Aubert, M.F., J. Blancou, J. Barrat, M. Artois, and M.J. Barrat. 1991. [Transmissibility and pathogenicity in the red fox of two rabies viruses isolated at a 10 year interval]. *Annales de recherches veterinaires. Annals of veterinary research*. 22:77-93.

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

Banerjee, A.K. 1987. Transcription and replication of rhabdoviruses. *Microbiol. Rev.* 51:66-87.

Banyard, A.C., D.T. Hayman, N. Johnson, L. McElhinney, and A.R. Fooks. 2011. Bats and Lyssaviruses. *In* Research Advances in Rabies. Vol. 79. A.C. Jackson, editor. Elsevier/Academic Press, London. 239-289.

Bedin, E., J. Marino, and C. Sillero-Zub. 2015. EWCP Annual report 2015.

Belikov, S.I., G.N. Leonova, I.G. Kondratov, E.V. Romanova, and E.V. Pavlenko. 2009. Isolation and genetic characterisation of a new lytssavirus strain in the Primorskiy kray. *East Siberian J. Infect. Pathol*. 16:68-69.

Bellinger, D.A., J. Chang, T.O. Bunn, J.R. Pick, M. Murphy, and R. Rahija. 1983. Rabies induced in a cat by high-egg-passage Flury strain vaccine. *Journal of the American Veterinary Medical Association*. 183:997-998, 965.

Benmansour, A., M. Brahimi, C. Tuffereau, P. Coulon, F. Lafay, and A. Flamand. 1992. Rapid sequence evolution of street rabies glycoprotein is related to the highly heterogeneous nature of the viral population. *Virology*. 187:33-45.

Benmansour, A., H. Leblois, P. Coulon, C. Tuffereau, Y. Gaudin, A. Flamand, and F. Lafay. 1991. Antigenicity of rabies virus glycoprotein. *J. Virol.* 65:4198-4203.

Blancou, J., and M.F. Aubert. 1997. [Transmission of rabies virus: importance of the species barrier]. *Bulletin de l'Academie nationale de medecine*. 181:301-311; discussion 311-302.

Blanton, J.D., D. Palmer, J. Dyer, and C.E. Rupprecht. 2011. Rabies surveillance in the United States during 2010. *Journal of the American Veterinary Medical Association*. 239:773-783.

Borucki, M.K., H. Chen-Harris, V. Lao, G. Vanier, D.A. Wadford, S. Messenger, and J.E. Allen. 2013. Ultra-deep sequencing of intra-host rabies virus populations during cross-species transmission. *PLoS Negl Trop Dis*. 7:e2555.

Botchkareva, N.V., V.A. Botchkarev, L.H. Chen, G. Lindner, and R. Paus. 1999. A role for p75 neurotrophin receptor in the control of hair follicle morphogenesis. *Developmental biology*. 216:135-153.

Both, L. 2013. Production and molecular characterization of monoclonal antibodies against rabies virus. Vol. Doctor of Philosophy (PhD). St Georges University of London, London. 173.

Both, L., C. van Dolleweerd, E. Wright, A.C. Banyard, B. Bulmer-Thomas, S. D., F. Altmann, A.R. Fooks, and M.J.K. C. 2013. Production, characterization, and antigen specificity of recombinant 62-

71-3, a candidate monoclonal antibody for rabies prophylaxis in humans *FASEB Journal*. 27:2055-2065.

Botvinkin, A.D., E.M. Poleschuk, I.V. Kuzmin, T.I. Borisova, S.V. Gazaryan, P. Yager, and C.E. Rupprecht. 2003. Novel lyssaviruses isolated from bats in Russia. *Emerging infectious diseases*. 9:1623-1625.

Brookes, S.M., D.M. Healy, and A.R. Fooks. 2006. Ability of rabies vaccine strains to elicit cross-neutralising antibodies. *Dev Biol (Basel)*. 125:185-193.

Brookes, S.M., G. Parsons, N. Johnson, L.M. McElhinney, and A.R. Fooks. 2005. Rabies human diploid cell vaccine elicits cross-neutralising and cross-protecting immune responses against European and Australian bat lyssaviruses. *Vaccine*. 23:4101-4109.

Brzozka, K., S. Finke, and K.K. Conzelmann. 2006. Inhibition of interferon signaling by rabies virus phosphoprotein P: activation-dependent binding of STAT1 and STAT2. *J. Virol.* 80:2675-2683.

Bull, R.A., J.S. Eden, F. Luciani, K. McElroy, W.D. Rawlinson, and P.A. White. 2012. Contribution of intra- and interhost dynamics to norovirus evolution. *J Virol*. 86:3219-3229.

Carpenter, J.A., L.P. Keegan, L. Wilfert, M.A. O'Connell, and F.M. Jiggins. 2009. Evidence for ADAR-induced hypermutation of the Drosophila sigma virus (Rhabdoviridae). *BMC genetics*. 10:75.

Castel, G., N. Tordo, and A. Plyusnin. 2017. Estimation of main diversification time-points of hantaviruses using phylogenetic analyses of complete genomes. *Virus Res*. 233:60-69.

Chen, L., B. Liu, Z. Wu, Q. Jin, and J. Yang. 2017. DRodVir: A resource for exploring the virome diversity in rodents. *Journal of genetics and genomics = Yi chuan xue bao*. 44:259-264.

Ciota, A.T., D.J. Ehrbar, G.A. Van Slyke, G.G. Willsey, and L.D. Kramer. 2012. Cooperative interactions in the West Nile virus mutant swarm. *BMC evolutionary biology*. 12:58.

Cottam, E.M., J. Wadsworth, A.E. Shaw, R.J. Rowlands, L. Goatley, S. Maan, N.S. Maan, P.P. Mertens, K. Ebert, Y. Li, E.D. Ryan, N. Juleff, N.P. Ferris, J.W. Wilesmith, D.T. Haydon, D.P. King, D.J. Paton, and N.J. Knowles. 2008. Transmission pathways of foot-and-mouth disease virus in the United Kingdom in 2007. *PLoS Pathog*. 4:e1000050.

Cummings, M.P., S.A. Handley, D.S. Myers, D.L. Reed, A. Rokas, and K. Winka. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Systematic biology*. 52:477-487.

Dacheux, L., F. Larrous, A. Mailles, D. Boisseleau, O. Delmas, C. Biron, C. Bouchier, I. Capek, M. Muller, F. Ilari, T. Lefranc, F. Raffi, M. Goudal, and H. Bourhy. 2009. European bat Lyssavirus transmission among cats, Europe. *Emerg. Infect. Dis.* 15:280-284.

Dechant, G., and Y.A. Barde. 2002. The neurotrophin receptor p75(NTR): novel functions and implications for diseases of the nervous system. *Nature Neuroscience*. 5:1131-1136.

Dellicour, S., R. Rose, N.R. Faria, L.F.P. Vieira, H. Bourhy, M. Gilbert, P. Lemey, and O.G. Pybus. 2017. Using viral gene sequences to compare and explain the heterogeneous spatial dynamics of virus epidemics. *Molecular biology and evolution*.

Dietzschold, B., J. Li, M. Faber, and M. Schnell. 2008. Concepts in the pathogenesis of rabies. *Future Virol*. 3:481-490.

Dietzschold, B., T.J. Wiktor, R. Macfarlan, and A. Varrichio. 1982. Antigenic structure of rabies virus glycoprotein: ordering and immunological characterization of the large CNBr cleavage fragments. *Journal of Virology*. 44:595-602.

Ding, N.Z., D.S. Xu, Y.Y. Sun, H.B. He, and C.Q. He. 2017. A permanent host shift of rabies virus from Chiroptera to Carnivora associated with recombination. *Scientific reports*. 7:289.

Drake, J.W. 1993. Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci U S A*. 90:4171-4175.

Easton, A.J., and C.R. Pringle. 2012. Mononegavirales. *In* Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. A.M. King, E. Lefkowitz, M.J. Adams, and E.B. Carstens, editors. Elsevier Academic Press, San Diego. 653-657.

Eriksson, N., L. Pachter, Y. Mitsuya, S.Y. Rhee, C. Wang, B. Gharizadeh, M. Ronaghi, R.W. Shafer, and N. Beerenwinkel. 2008. Viral population estimation using pyrosequencing. *PLoS computational biology*. 4:e1000074.

Evans, J.S., D.L. Horton, A.J. Easton, A. Fooks, and A.C. Banyard. 2012. Rabies virus vaccines: Is there a need for a pan-lyssavirus vaccine? *Vaccine*. 30:7447-7454.

Faber, M., M.L. Faber, A. Papaneri, M. Bette, E. Weihe, B. Dietzschold, and M.J. Schnell. 2005. A single amino acid change in rabies virus glycoprotein increases virus spread and enhances virus pathogenicity. *J. Virol.* 79:14141-14148.

Familusi, J.B., and D.L. Moore. 1972. Isolation of a rabies related virus from the cerebrospinal fluid of a child with 'aseptic meningitis'. *African Journal of Medical Sciences*. 3:93-96.

Familusi, J.B., B.O. Osunkoya, D.L. Moore, G.E. Kemp, and A. Fabiyi. 1972. A fatal human infection with Mokola virus. *American Journal of Tropical Medicine and Hygiene*. 21:959-963.

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

Ferguson, E.A., K. Hampson, S. Cleaveland, R. Consunji, R. Deray, J. Friar, D.T. Haydon, J. Jimenez, M. Pancipane, and S.E. Townsend. 2015. Heterogeneity in the spread and control of infectious disease: consequences for the elimination of canine rabies. *Scientific reports*. 5:18232.

Finke, S., and K.K. Conzelmann. 2003. Dissociation of rabies virus matrix protein functions in regulation of viral RNA synthesis and virus assembly. *J. Virol.* 77:12074-12082.

Finke, S., and K.K. Conzelmann. 2005. Replication strategies of rabies virus. *Virus Research*. 111:120-131.

Finke, S., R. Mueller-Waldeck, and K.K. Conzelmann. 2003. Rabies virus matrix protein regulates the balance of virus transcription and replication. *J. Gen. Virol.* 84:1613-1621.

Fischer, W., V.V. Ganusov, E.E. Giorgi, P.T. Hraber, B.F. Keele, T. Leitner, C.S. Han, C.D. Gleasner, L. Green, C.C. Lo, A. Nag, T.C. Wallstrom, S. Wang, A.J. McMichael, B.F. Haynes, B.H. Hahn, A.S. Perelson, P. Borrow, G.M. Shaw, T. Bhattacharya, and B.T. Korber. 2010. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PloS one*. 5:e12303.

Fooks, A.R., A.C. Banyard, D.L. Horton, N. Johnson, L.M. McElhinney, and A.C. Jackson. 2014. Current status of rabies and prospects for elimination. *Lancet*. 384:1389-1399.

Fooks, A.R., L.M. McElhinney, D.J. Pounder, C.J. Finnegan, K. Mansfield, N. Johnson, S.M. Brookes, G. Parsons, K. White, P.G. McIntyre, and D. Nathwani. 2003. Case report: isolation of a European bat lyssavirus type 2a from a fatal human case of rabies encephalitis. *J. Med. Virol.* 71:281-289.

Freuling, C.M., B. Abendroth, M. Beer, M. Fischer, D. Hanke, B. Hoffmann, D. Höper, F. Just, T.C. Mettenleiter, J. Schatz, and T. Müller. 2013. Molecular diagnostics for the detection of Bokeloh bat lyssavirus in a bat from Bavaria, Germany. *Virus Research*. 177:201-204.

Freuling, C.M., M. Beer, F.J. Conraths, S. Finke, B. Hoffmann, B. Keller, J. Kliemt, T.C. Mettenleiter, E. Muhlbach, J.P. Teifke, P. Wohlsein, and T. Muller. 2011. Novel lyssavirus in Natterer's bat, Germany. *Emerg. Infect. Dis.* 17:1519-1522.

Freuling, C.M., D. Klöss, R. Schröder, A. Kliemt, and T. Müller. 2012. The WHO Rabies Bulletin Europe: a key source of information on rabies and a pivotal tool for surveillance and epidemiology. *Rev. sci. tech. Off. int. Epiz.* 31:799-807.

Frey, B.J., and D. Dueck. 2007. Clustering by passing messages between data points. *Science*. 315:972-976.

Gastka, M., J. Horvath, and T.L. Lentz. 1996. Rabies virus binding to the nicotinic acetylcholine receptor alpha subunit demonstrated by virus overlay protein binding assay. *Journal of General Virology*. 77 (Pt 10):2437-2440.

Gaudin, Y., R.W. Ruigrok, M. Knossow, and A. Flamand. 1993. Low-pH conformational changes of rabies virus glycoprotein and their role in membrane fusion. *Journal of Virology*. 67:1365-1372.

Gaudin, Y., R.W. Ruigrok, C. Tuffereau, M. Knossow, and A. Flamand. 1992. Rabies virus glycoprotein is a trimer. *Virology*. 187:627-632.

Gordon, C.H., A.C. Banyard, A. Hussein, M.K. Laurenson, J.R. Malcolm, J. Marino, F. Regassa, A.M. Stewart, A.R. Fooks, and C. Sillero-Zubiri. 2015. Canine distemper in endangered Ethiopian wolves. *Emerging infectious diseases*. 21:824-832.

Gosztonyi, G. 1994. Reproduction of lyssaviruses: ultrastructural composition of lyssavirus and functional aspects of pathogenesis. *Current Topics in Microbiology and Immunology*. 187:43-68.

Graham, S.C., R. Assenberg, O. Delmas, A. Verma, A. Gholami, C. Talbi, R.J. Owens, D.I. Stuart, J.M. Grimes, and H. Bourhy. 2008. Rhabdovirus matrix protein structures reveal a novel mode of self-association. *PLoS Pathog*. 4:e1000251.

Gunawardena, P.S., D.A. Marston, R.J. Ellis, E.L. Wise, A.C. Karawita, A.C. Breed, L.M. McElhinney, N. Johnson, A.C. Banyard, and A.R. Fooks. 2016. Lyssavirus in Indian Flying Foxes, Sri Lanka. *Emerging infectious diseases*. 22:1456-1459.

Gupta, A.K., D. Blondel, S. Choudhary, and A.K. Banerjee. 2000. The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C. *J. Virol.* 74:91-98.

Hampson, K., L. Coudeville, T. Lembo, M. Sambo, A. Kieffer, M. Attlan, J. Barrat, J.D. Blanton, D.J. Briggs, S. Cleaveland, P. Costa, C.M. Freuling, E. Hiby, L. Knopf, F. Leanes, F.X. Meslin, A. Metlin, M.E. Miranda, T. Muller, L.H. Nel, S. Recuenco, C.E. Rupprecht, C. Schumacher, L. Taylor, M.A. Vigilato, J. Zinsstag, and J. Dushoff. 2015. Estimating the global burden of endemic canine rabies. *PLoS Negl Trop Dis*. 9:e0003709.

Hanke, D., C.M. Freuling, S. Fischer, K. Hueffer, K. Hundertmark, S. Nadin-Davis, D. Marston, A.R. Fooks, A. Botner, T.C. Mettenleiter, M. Beer, T.B. Rasmussen, T.F. Muller, and D. Hoper. 2016. Spatio-temporal Analysis of the Genetic Diversity of Arctic Rabies Viruses and Their Reservoir Hosts in Greenland. *PLoS Negl Trop Dis*. 10:e0004779.

Hanlon, C.A., I.V. Kuzmin, J.D. Blanton, W.C. Weldon, J.S. Manangan, and C.E. Rupprecht. 2005. Efficacy of rabies biologics against new lyssaviruses from Eurasia. *Virus Research*. 111:44-54.

Hanna, J.N., I.K. Carney, G.A. Smith, A.E. Tannenberg, J.E. Deverill, J.A. Botha, I.L. Serafin, B.J. Harrower, P.F. Fitzpatrick, and J.W. Searle. 2000. Australian bat lyssavirus infection: a second human case, with a long incubation period. *Medical Journal of Australia*. 172:597-599.

Hayman, D.T., A.R. Fooks, D. Horton, R. Suu-Ire, A.C. Breed, A.A. Cunningham, and J.L. Wood. 2008. Antibodies against Lagos bat virus in megachiroptera from West Africa. *Emerging infectious diseases*. 14:926-928.

Hodges, E.N., B.S. Heinrich, and J.H. Connor. 2012. A vesiculovirus showing a steepened transcription gradient and dominant trans-repression of virus transcription. *Journal of Virology*. 86:8884-8889.

Holmes, E.C., and A. Moya. 2002. Is the quasispecies concept relevant to RNA viruses? *J Virol*. 76:460-465.

Hoper, D., S. Finke, C.M. Freuling, B. Hoffmann, and M. Beer. 2012. What can we learn about lyssavirus genomes using 454 sequencing? *Berliner und Munchener tierarztliche Wochenschrift*. 125:242-248.

Hoper, D., C.M. Freuling, T. Muller, D. Hanke, V. von Messling, K. Duchow, M. Beer, and T.C. Mettenleiter. 2015. High definition viral vaccine strain identity and stability testing using full-genome population data--The next generation of vaccine quality control. *Vaccine*. 33:5829-5837.

Horton, D.L., A.C. Banyard, D.A. Marston, E. Wise, D. Selden, A. Nunez, D. Hicks, T. Lembo, S. Cleaveland, A.J. Peel, I.V. Kuzmin, C.E. Rupprecht, and A.R. Fooks. 2014. Antigenic and genetic characterization of a divergent African virus, Ikoma lyssavirus. *J. Gen. Virol.* 95:1025-1032.

Horton, D.L., L.M. McElhinney, D.A. Marston, J.L. Wood, C.A. Russell, N. Lewis, I.V. Kuzmin, R.A. Fouchier, A.D. Osterhaus, A.R. Fooks, and D.J. Smith. 2010. Quantifying antigenic relationships among the Lyssaviruses. *J. Virol.* 84:11841-11848.

ICTV. 2012. Official Taxonomy: Updates since the 8th Report, 2009. Vol. 2012.

Jackson, A.C., H. Ye, C.C. Phelan, C. Ridaura-Sanz, Q. Zheng, Z. Li, X. Wan, and E. Lopez-Corella. 1999. Extraneural organ involvement in human rabies. *Laboratory Investigation*. 79:945-951.

Jacob, Y., H. Badrane, P.E. Ceccaldi, and N. Tordo. 2000. Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein. *Journal of Virology*. 74:10217-10222.

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

Kgaladi, J., N. Wright, J. Coertse, W. Markotter, D. Marston, A.R. Fooks, C.M. Freuling, T.F. Muller, C.T. Sabeta, and L.H. Nel. 2013. Diversity and epidemiology of Mokola virus. *PLoS Negl Trop Dis*. 7:e2511.

Kiddle, S.J., O.P. Windram, S. McHattie, A. Mead, J. Beynon, V. Buchanan-Wollaston, K.J. Denby, and S. Mukherjee. 2010. Temporal clustering by affinity propagation reveals transcriptional modules in Arabidopsis thaliana. *Bioinformatics*. 26:355-362.

Kissi, B., H. Badrane, L. Audry, A. Lavenu, N. Tordo, M. Brahimi, and H. Bourhy. 1999. Dynamics of rabies virus quasispecies during serial passages in heterologous hosts. *J Gen Virol*. 80 (Pt 8):2041-2050.

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

Komarova, A.V., E. Real, A.M. Borman, M. Brocard, P. England, N. Tordo, J.W. Hershey, K.M. Kean, and Y. Jacob. 2007. Rabies virus matrix protein interplay with eIF3, new insights into rabies virus pathogenesis. *Nucleic acids research*. 35:1522-1532.

Koser, M.L., J.P. McGettigan, G.S. Tan, M.E. Smith, H. Koprowski, B. Dietzschold, and M.J. Schnell. 2004. Rabies virus nucleoprotein as a carrier for foreign antigens. *Proc. Natl. Acad. Sci. U. S. A.* 101:9405-9410.

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

Kuzmin, I.V., M. Shi, L.A. Orciari, P.A. Yager, A. Velasco-Villa, N.A. Kuzmina, D.G. Streicker, D.L. Bergman, and C.E. Rupprecht. 2012. Molecular Inferences Suggest Multiple Host Shifts of Rabies Viruses from Bats to Mesocarnivores in Arizona during 2001-2009. *PLoS Pathog*. 8:e1002786.

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

Lafon, M. 2005. Rabies virus receptors. *Journal of Neurovirology*. 11:82-87.

Lahaye, X., A. Vidy, C. Pomier, L. Obiang, F. Harper, Y. Gaudin, and D. Blondel. 2009. Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: Evidence that NBs are sites of viral transcription and replication. *Journal of Virology*. 83:7948-7958.

Lewis, P., Y. Fu, and T.L. Lentz. 2000. Rabies virus entry at the neuromuscular junction in nerve-muscle cocultures. *Muscle and Nerve*. 23:720-730.

Lionel Harischandra, P.A., A. Gunesekera, N. Janakan, G. Gongal, and B. Abela-Ridder. 2016. Sri Lanka takes action towards a target of zero rabies death by 2020. *WHO South-East Asia journal of public health*. 5:113-116.

Liu, W., Y. Liu, J. Liu, J. Zhai, and Y. Xie. 2011. Evidence for inter- and intra-clade recombinations in rabies virus. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 11:1906-1912.

Liu, Y., Q. Chen, F. Zhang, S. Zhang, N. Li, H. Lian, Y. Wang, J. Zhang, and R. Hu. 2013a. Evaluation of rabies biologics against Irkut virus isolated in China. *Journal of clinical microbiology*. 51:3499-3504.

Liu, Y., S. Zhang, J. Zhao, F. Zhang, and R. Hu. 2013b. Isolation of Irkut virus from a Murina leucogaster bat in China. *PLoS Negl Trop Dis*. 7:e2097.

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

Lumio, J., M. Hillbom, R. Roine, L. Ketonen, M. Haltia, M. Valle, E. Neuvonen, and J. Lahdevirta. 1986. Human rabies of bat origin in Europe. *Lancet*. 1:378.

Lythgoe, K.A., A. Gardner, O.G. Pybus, and J. Grove. 2017. Short-Sighted Virus Evolution and a Germline Hypothesis for Chronic Viral Infections. *Trends in microbiology*. 25:336-348.

Macfarlan, R.I., B. Dietzschold, and H. Koprowski. 1986. Stimulation of cytotoxic T-lymphocyte responses by rabies virus glycoprotein and identification of an immunodominant domain. *Molecular Immunology*. 23:733-741.

Mani, R.S., D.P. Dovih, M.A. Ashwini, B. Chattopadhyay, P.K. Harsha, K.M. Garg, S. Sudarshan, R. Puttaswamaiah, U. Ramakrishnan, and S.N. Madhusudana. 2017. Serological Evidence of Lyssavirus Infection among Bats in Nagaland, a North-Eastern State in India. *Epidemiology and infection*. 145:1635-1641.

Mannen, K., K. Hiramatsu, K. Mifune, and S. Sakamoto. 1991. Conserved nucleotide sequence of rabies virus cDNA encoding the nucleoprotein. *Virus genes*. 5:69-73.

Marino, J., C. Sillero-Zubiri, P.J. Johnson, and D.W. Macdonald. 2012. Ecological bases of philopatry and cooperation in Ethiopian wolves. *Behavioral Ecology and Sociobiology*. 66:1005-1015.

Markotter, W., I. Kuzmin, C.E. Rupprecht, and L.H. Nel. 2008. Phylogeny of Lagos bat virus: challenges for lyssavirus taxonomy. *Virus Research*. 135:10-21.

Marston, D.A., R.J. Ellis, D. Horton, I. Kuzmin, E.L. Wise, L. McElhinney, A.C. Banyard, C. Ngeleja, J. Keyyu, S. Cleveland, T. Lembo, C. Rupprecht, and A. Fooks. 2012a. Complete genomic sequence of Ikoma Lyssavirus. *Journal of Virology*. 86:10242-10243

Marston, D.A., D.L. Horton, C. Ngeleja, K. Hampson, L.M. McElhinney, A.C. Banyard, D. Haydon, S. Cleaveland, C.E. Rupprecht, M. Bigambo, A.R. Fooks, and T. Lembo. 2012b. Ikoma lyssavirus, highly divergent novel lyssavirus in an African civet. *Emerging infectious diseases*. 18:664-667.

Marston, D.A., L.M. McElhinney, R.J. Ellis, D.L. Horton, E.L. Wise, S.L. Leech, D. David, X. de Lamballerie, and A.R. Fooks. 2013. Next generation sequencing of viral RNA genomes. *BMC genomics*. 14:444.

Marston, D.A., L.M. McElhinney, N. Johnson, T. Muller, K.K. Conzelmann, N. Tordo, and A.R. Fooks. 2007. Comparative analysis of the full genome sequence of European bat lyssavirus type 1 and type 2 with other lyssaviruses and evidence for a conserved transcription termination and polyadenylation motif in the G-L 3' non-translated region. *J. Gen. Virol.* 88:1302-1314.

McCall, B.J., J.H. Epstein, A.S. Neill, K. Heel, H. Field, J. Barrett, G.A. Smith, L.A. Selvey, B. Rodwell, and R. Lunt. 2000. Potential exposure to Australian bat lyssavirus, Queensland, 1996-1999. *Emerging infectious diseases*. 6:259-264.

McGinnis, J., J. Laplante, M. Shudt, and K.S. George. 2016. Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. *J Clin Virol*. 79:44-50.

Mebatsion, T., M. König, and K.-K. Conzelmann. 1996a. Budding of rabies virus particles in the absence of the spike glycoprotein. *Cell*. 84:941-951.

Mebatsion, T., M.J. Schnell, J.H. Cox, S. Finke, and K.-K. Conzelmann. 1996b. Highly stable expression of a foreign gene from rabies virus vectors. *Proceedings of the National Acadamy of Sciences USA*. 93:7310-7314.

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

Mollentze, N., R. Biek, and D.G. Streicker. 2014a. The role of viral evolution in rabies host shifts and emergence. *Current opinion in virology*. 8:68-72.

Mollentze, N., L.H. Nel, S. Townsend, K. le Roux, K. Hampson, D.T. Haydon, and S. Soubeyrand. 2014b. A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. *Proceedings. Biological sciences / The Royal Society*. 281:20133251.

Montmayeur, A.M., and T.F. Ng. 2017. High-Throughput Next-Generation Sequencing of Polioviruses. 55:606-615.

Morelli, M.J., G. Thebaud, J. Chadoeuf, D.P. King, D.T. Haydon, and S. Soubeyrand. 2012. A Bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data. *PLoS computational biology*. 8:e1002768.

Morelli, M.J., C.F. Wright, N.J. Knowles, N. Juleff, D.J. Paton, D.P. King, and D.T. Haydon. 2013. Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts. *Veterinary research*. 44:12.

Morimoto, K., H.D. Foley, J.P. McGettigan, M.J. Schnell, and B. Dietzschold. 2000. Reinvestigation of the role of the rabies virus glycoprotein in viral pathogenesis using a reverse genetics approach. *J. Neurovirol.* 6:373-381.

Morimoto, K., D.C. Hooper, H. Carbaugh, Z.F. Fu, H. Koprowski, and B. Dietzschold. 1998. Rabies virus quasispecies: implications for pathogenesis. *Proc Natl Acad Sci U S A*. 95:3152-3156.

Morimoto, K., D.C. Hooper, S. Spitsin, H. Koprowski, and B. Dietzschold. 1999. Pathogenicity of different rabies virus variants inversely correlates with apoptosis and rabies virus glycoprotein expression in infected primary neuron cultures. *J. Virol.* 73:510-518.

Moustafa, A., C. Xie, E. Kirkness, and W. Biggs. 2017. The blood DNA virome in 8,000 humans. 13:e1006292.

Mrak, R.E., and L. Young. 1993. Rabies encephalitis in a patient with no history of exposure. *Human pathology*. 24:109-110.

Muller-Rover, S., E.J. Peters, V.A. Botchkarev, A. Panteleyev, and R. Paus. 1998. Distinct patterns of NCAM expression are associated with defined stages of murine hair follicle morphogenesis and regression. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*. 46:1401-1410.

Muller, T., J. Cox, W. Peter, R. Schafer, N. Johnson, L.M. McElhinney, J.L. Geue, K. Tjornehoj, and A.R. Fooks. 2004. Spill-over of European bat lyssavirus type 1 into a stone marten (Martes foina) in Germany. *J Vet Med B Infect Dis Vet Public Health*. 51:49-54.

Muller, T., B. Dietzschold, H. Ertl, A.R. Fooks, C. Freuling, C. Fehlner-Gardiner, J. Kliemt, F.X. Meslin, R. Franka, C.E. Rupprecht, N. Tordo, A.I. Wanderler, and M.P. Kieny. 2009. Development of a mouse monoclonal antibody cocktail for post-exposure rabies prophylaxis in humans. *PLoS Negl Trop Dis*. 3:e542.

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

Nadin-Davis, S. 2007. Molecular Epidemiology. *In* Rabies. A.C. Jackson and W. Wunner, editors. Elsevier/Academic Press, London. 341-381.

Nadin-Davis, S.A., A. Colville, H. Trewby, R. Biek, and L. Real. 2017. Application of high-throughput sequencing to whole rabies viral genome characterisation and its use for phylogenetic re-evaluation of a raccoon strain incursion into the province of Ontario. *Virus Res*.

Nelson, J.R. 2014. Random-primed, Phi29 DNA polymerase-based whole genome amplification. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*. 105:Unit 15.13.

Neville, J. 2004. Rabies in the ancient world. *In* Historiacal perspective of Rabies in Europe and the Mediterranean Basin. A.A. King, A.R. Fooks, M. Aubert, and A.I. Wanderler, editors. OIE Publications. 1-12.

Ni, Y., Y. Tominaga, Y. Honda, K. Morimoto, S. Sakamoto, and A. Kawai. 1995. Mapping and characterization of a sequential epitope on the rabies virus glycoprotein which is recognized by a neutralizing monoclonal antibody, RG719. *Microbiology and immunology*. 39:693-702.

OIE. 2013. Vaccine Bank for Rabies.

Pal, S.R., B. Arora, P.N. Chhuttani, S. Broor, S. Choudhury, R.M. Joshi, and S.D. Ray. 1980. Rabies virus infection of a flying fox bat, Pteropus policephalus in Chandigarh, Northern India. *Tropical and geographical medicine*. 32:265-267.

Parker, J., and J. Chen. 2017. Application of next generation sequencing for the detection of human viral pathogens in clinical specimens. *J Clin Virol*. 86:20-26.

Paweska, J.T., L.H. Blumberg, C. Liebenberg, R.H. Hewlett, A.A. Grobbelaar, P.A. Leman, J.E. Croft, L.H. Nel, L. Nutt, and R. Swanepoel. 2006. Fatal human infection with rabies-related Duvenhage virus, South Africa. *Emerging infectious diseases*. 12:1965-1967.

Peh, W.C., and K.H. Ng. 2010. Writing a book review. *Singapore medical journal*. 51:685-688.

Picard-Meyer, E., A. Servat, E. Robardet, M. Moinet, C. Borel, and F. Cliquet. 2013. Isolation of Bokeloh bat lyssavirus in Myotis nattereri in France. *Archives of virology*.

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

Poh, W.T., E. Xia, K. Chin-Inmanu, L.P. Wong, A.Y. Cheng, P. Malasit, P. Suriyaphol, Y.Y. Teo, and R.T. Ong. 2013. Viral quasispecies inference from 454 pyrosequencing. *BMC bioinformatics*. 14:355.

Posada-Cespedes, S., D. Seifert, and N. Beerenwinkel. 2016. Recent advances in inferring viral diversity from high-throughput sequencing data. *Virus Res*.

Pringle, C.R., and A.J. Easton. 1997. Monopartite negative strand RNA genomes. *Seminars in Virology*. 8:49-57.

Raghwani, J., R. Rose, and I. Sheridan. 2016. Exceptional Heterogeneity in Viral Evolutionary Dynamics Characterises Chronic Hepatitis C Virus Infection. 12:e1005894.

Raux, H., A. Flamand, and D. Blondel. 2000. Interaction of the rabies virus P protein with the LC8 dynein light chain. *Journal of Virology*. 74:10212-10216.

Rupprecht, C., I. Kuzmin, and F. Meslin. 2017. Lyssaviruses and rabies: current conundrums, concerns, contradictions and controversies. *F1000Research*. 6:184.

Rupprecht, C.E., J. Barrett, D. Briggs, F. Cliquet, A.R. Fooks, B. Lumlertdacha, F.X. Meslin, T. Muler, L.H. Nel, C. Schneider, N. Tordo, and A.I. Wandeler. 2008. Can rabies be eradicated? *Dev Biol (Basel)*. 131:95-121.

Rupprecht, C.E., C.A. Hanlon, and T. Hemachudha. 2002. Rabies re-examined. *The Lancet. Infectious diseases*. 2:327-343.

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

Sacramento, D., H. Bourhy, and N. Tordo. 1991. PCR technique as an alternative method for diagnosis and molecular epidemiology of rabies virus. *Molecular and Cellular Probes*. 5:229-240.

Sanjuan, R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. 365:1975-1982.

Sanjuan, R. 2012. From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. *PLoS Pathog*. 8:e1002685.

Scarnecchia, D.L. 2004. Writing Book Reviews for the Journal Of Range Management and Rangelands. *Journal of Range Management*. 57:418-421.

Schnell, M.J., J.P. McGettigan, C. Wirblich, and A. Papaneri. 2010. The cell biology of rabies virus: using stealth to reach the brain. *Nat Rev Microbiol*. 8:51-61.

Seifert, D., and N. Beerenwinkel. 2016. Estimating Fitness of Viral Quasispecies from Next-Generation Sequencing Data. *Curr Top Microbiol Immunol*. 392:181-200.

Shakin-Eshleman, S.H., A.T. Remaley, J.R. Eshleman, W.H. Wunner, and S.L. Spitalnik. 1992. N-linked glycosylation of rabies virus glycoprotein. Individual sequons differ in their glycosylation efficiencies and influence on cell surface expression. *Journal of Biological Chemistry*. 267:10690-10698.

Shi, M., X.D. Lin, J.H. Tian, L.J. Chen, X. Chen, C.X. Li, X.C. Qin, J. Li, J.P. Cao, J.S. Eden, J. Buchmann, W. Wang, J. Xu, E.C. Holmes, and Y.Z. Zhang. 2016. Redefining the invertebrate RNA virosphere. *Nature*.

Shwiff, S., K. Hampson, and A. Anderson. 2013. Potential economic benefits of eliminating canine rabies. *Antiviral research*. 98:352-356.

Spann, K.M., P.L. Collins, and M.N. Teng. 2003. Genetic Recombination during Coinfection of Two Mutants of Human Respiratory Syncytial Virus. *Journal of Virology*. 77:11201-11211.

Stapleford, K.A., L.L. Coffey, S. Lay, A.V. Borderia, V. Duong, O. Isakov, K. Rozen-Gagnon, C. Arias-Goeta, H. Blanc, S. Beaucourt, T. Haliloglu, C. Schmitt, I. Bonne, N. Ben-Tal, N. Shomron, A.B. Failloux, P. Buchy, and M. Vignuzzi. 2014. Emergence and transmission of arbovirus evolutionary intermediates with epidemic potential. *Cell host & microbe*. 15:706-716.

Symons, R. 2013. Australian bat lyssavirus update communiqué. *State of Queensland, Department of Agriculture, Fisheries and Forestry*.

Talbi, C., P. Lemey, M.A. Suchard, E. Abdelatif, M. Elharrak, J. Nourlil, A. Faouzi, J.E. Echevarria, S. Vazquez Moron, A. Rambaut, N. Campiz, A.J. Tatem, E.C. Holmes, and H. Bourhy. 2010. Phylodynamics and human-mediated dispersal of a zoonotic virus. *PLoS Pathog*. 6:e1001166.

Tan, G.S., M.A. Preuss, J.C. Williams, and M.J. Schnell. 2007. The dynein light chain 8 binding motif of rabies virus phosphoprotein promotes efficient viral transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 104:7229-7234.

Taylor, L.H., K. Hampson, A. Fahrion, B. Abela-Ridder, and L.H. Nel. 2017. Difficulties in estimating the human burden of canine rabies. *Acta tropica*. 165:133-140.

Thoulouze, M.I., M. Lafage, M. Schachner, U. Hartmann, H. Cremer, and M. Lafon. 1998. The neural cell adhesion molecule is a receptor for rabies virus. *Journal of Virology*. 72:7181-7190.

Tjornehoj, K., A.R. Fooks, J.S. Agerholm, and L. Ronsholt. 2006. Natural and experimental infection of sheep with European bat lyssavirus type-1 of Danish bat origin. *J. Comp. Pathol.* 134:190-201.

Tordo, N. 1996. Characteristics and molecular biology of the rabies virus. *In* Laboratory techniques in rabies. F.X. Meslin, M.M. Kaplan, and H. Koprowski, editors. WHO, Finland. 28-45.

Tordo, N., and A. Kouknetzoff. 1993. The rabies virus genome: an overview. *Onderstepoort Journal of Veterinary Research*. 60:263-269.

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

Troupin, C., L. Dacheux, M. Tanguy, C. Sabeta, H. Blanc, C. Bouchier, M. Vignuzzi, S. Duchene, E.C. Holmes, and H. Bourhy. 2016. Large-Scale Phylogenomic Analysis Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts. *PLOS Pathogens*. 12:e1006041.

Tsiang, H. 1993. Pathophysiology of rabies virus infection of the nervous system. *Advances in Virus Research*. 42:375-412.

Tsiang, H., P.E. Ceccaldi, and E. Lycke. 1991. Rabies virus infection and transport in human sensory dorsal root ganglia neurons. *Journal of General Virology*. 72 (Pt 5):1191-1194.

Turner, P.E., and L. Chao. 1998. Sex and the evolution of intrahost competition in RNA virus phi6. *Genetics*. 150:523-532.

Ugolini, G. 2008. Use of rabies virus as a transneuronal tracer of neuronal connections: implications for the understanding of rabies pathogenesis. *Dev Biol (Basel)*. 131:493-506.

UN. 2015. World Population Prospects (Worldometers). United Nations Department of Economic and Scoical Affairs.

Vazquez Moron, S., J. Juste, C. Ibanez, J.M. Berciano, and J.E. Echevarria. 2011. Phylogeny of European bat Lyssavirus 1 in Eptesicus isabellinus bats, Spain. *Emerging infectious diseases*. 17:520-523.

Velasco-Villa, A., M.R. Mauldin, M. Shi, L.E. Escobar, N.F. Gallardo-Romero, I. Damon, V.A. Olson, D.G. Streicker, and G. Emerson. 2017. The history of rabies in the Western Hemisphere. *Antiviral research*.

Wakjira, D.T., F. D'Udine, and A. Crawford. 2015. Migration and conservation in the Bale Mountains ecosystem. International Institute for Sustainable Development.

Wallace, R.M., A. Gilbert, D. Slate, R. Chipman, A. Singh, W. Cassie, and J.D. Blanton. 2014. Right place, wrong species: a 20-year review of rabies virus cross species transmission among terrestrial mammals in the United States. *PloS one*. 9:e107539.

Warrell, M.J., and D.A. Warrell. 2004. Rabies and other lyssavirus diseases. *Lancet*. 363:959-969.

Weir, D.L., E.J. Annand, P.A. Reid, and C.C. Broder. 2014. Recent observations on Australian bat lyssavirus tropism and viral entry. *Viruses*. 6:909-926.

Weir, D.L., I.L. Smith, K.N. Bossart, L.F. Wang, and C.C. Broder. 2013. Host cell tropism mediated by Australian bat lyssavirus envelope glycoproteins. *Virology*. 444:21-30.

Whelan, S.P., J.N. Barr, and G.W. Wertz. 2004. Transcription and replication of nonsegmented negative-strand RNA viruses. *Current Topics in Microbiology and Immunology*. 283:61-119.

Wiktor, T. 1985. Historical aspects of rabies treatment. *In* World's debt to Pasteur. H. Koprowski and S.A. Plotkin, editors. Alan R. Liss, New York. 141-151.

Wright, C.F., M.J. Morelli, G. Thebaud, N.J. Knowles, P. Herzyk, D.J. Paton, D.T. Haydon, and D.P. King. 2011. Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. *J Virol*. 85:2266-2275.

Wunner, W.H. 1991. The chemical composition and molecular structure of rabies viruses. *In* The natural history of rabies. G.M. Baer, editor. CRC Press, Boca Raton. 31-67.

Wunner, W.H., and K.-K. Conzelmann. 2013. Rabies Virus. *In* Rabies: Scientific basis of the disease and its management. A.C. Jackson, editor. Elsevier Academic Press, Oxford. 17-49.

Wunner, W.H., K.J. Reagan, and H. Koprowski. 1984. Characterization of saturable binding sites for rabies virus. *Journal of Virology*. 50:691-697.

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

Appendices

Appendix 1: Table of Oral presentations given during PhD

Date	Conference	City	Title
Nov 2012	CoVetLab	Addlestone	Obtaining full length viral genomes from clinical samples and cultured material using pyrosequencing
Mar 2013	PhD Student Day	Addlestone	Viral genetic heterogeneity is utilised by lyssaviruses to adapt to specific host reservoirs and cross species barriers
Mar 2013	Society for General Microbiology	Manchester	Next-generation sequencing of RNA viral genomes from clinical and cultured samples: Application in discovery of novel a lyssavirus in Africa.
Nov 2014	APHA Conference	Addlestone	Virology - PHE and our Interactions.
Nov 2014	APHA Science talk		Next Generation Sequencing – challenges and rewards in Virology.
Dec 2014	Tri-Partite meeting	Madrid	Full genome sequencing using NGS - Applications and plans.
Feb 2016	Hantavirus workshop	Liverpool	Next Generation Sequencing – Applications to Hantaviruses
Jun 2016	19th Conference in Genomics and Proteomics	London	Investigation of lyssavirus CST events using complete genome sequences
Jun 2016	Birds and bats workshop	Guildford	Introduction to NGS and its applications in Virology
Oct 2016	EVAg NGS workshop	London	NGS at APHA and its applications in Virology
Jan 17	Collaborative meeting	Glasgow	NGS at APHA and its applications in Virology
May 2017	APHA Science talk	Addlestone	Tracking rabies viruses using whole genome sequences.
May 2017	Tri-Partite meeting	Germany	Current and future plans for NGS technologies Molecular diagnostic improvements Submission of GBLV and LLEBV to ICTV

Appendix 2: Table of Poster presentations given during PhD

Date	Conference	City	Title
Jul 2014	PhD Student Day	Addlestone	Next Generation Sequencing of RNA viral genomes.
Sept 2014	VEME workshop	Rome	Next Generation Sequencing of RNA viral genomes.
Jun 2015	23rd annual PhD student conference	Marseille	Investigation of a Rabies Cross-Species Transmission and Maintenance in Foxes
Jun 2015	Negative Strand Virus conference	Siena	Investigation Of A Rabies Epizootic Following Cross-Species Transmission Of Virus From Dogs To Foxes
Jun 2017	MedVetNet One health conference	Guildford	Investigation Of A Rabies Epizootic Following Cross-Species Transmission Of Virus From Dogs To Foxes

Appendix 3: Table of Lectures given during PhD

Date	Course	University	Title
Dec 2015	MSc Veterinary Microbiology	University of Surrey	Transmission and control of rabies.
Feb 2016	MSc in Virology	Barts and the London School of Medicine	Rhabdoviruses, with species emphasis on rabies.

Appendix 4: Contributions to manuscripts within this thesis

Library preparation and Sequencing Machine work was undertaken by staff at APHA Central Sequencing Unit (CSU).

Article	Experimental plan	Sample prep	PCR design	PCR optimisation	Analysis of raw reads	Post mapping analysis	Phylogenetic Analysis	Preparation of whole article	Preparation of sections for article	Review of article	1st author	corresponding author
1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	N/A	N/A	Yes	No
2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No	No	Yes	No	No
3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	N/A	N/A	Yes	Yes
4	Yes	Yes	Yes	Yes	Yes	Yes	N/A	Yes	N/A	N/A	Yes	No
5	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No	Yes	Yes	Yes	No
6	Yes	Partial	N/A	N/A	No	Yes	N/A	Yes	N/A	N/A	Yes	Yes
7	Yes	No	N/A	N/A	Partial	Yes	N/A	Yes	Yes	Yes	No	No
8	Yes	Yes	N/A	N/A	No	Yes	N/A	Yes	N/A	N/A	Yes	Yes
9	Yes	No	N/A	N/A	No	No	N/A	No	No	Yes	No	No
10	No	No	N/A	N/A	No	Yes	N/A	No	Yes	No	No	No
11	Yes	Yes	N/A	N/A	No	Yes	N/A	Yes	N/A	N/A	Yes	Yes
12	N/A	Yes	Yes	Yes	Yes	Yes	N/A	Yes	N/A	N/A	Yes	No
13	N/A	No	N/A	N/A	No	No	N/A	No	No	Yes	No	No
14	No	Yes	No	Yes	Yes	Yes	N/A	No	Yes	Yes	No	No
15	Yes	Yes	Yes	Yes	Yes	N/A	Partial	Yes	Yes	Yes	Yes	No
16	Yes	Yes	Yes	Yes	Yes	Yes	N/A	Yes	N/A	N/A	Yes	No
17	Partial	No	N/A	N/A	N/A	N/A	N/A	No	Yes	Yes	No	No
18	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	N/A	Yes	Yes	Yes
19	Yes	Yes	Yes	Yes	No	Yes	N/A	Yes	N/A	N/A	Yes (joint)	No
20	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	N/A	Yes	Yes	Yes
21	Yes	Yes	Yes	Yes	No	Yes	N/A	Yes	N/A	N/A	Yes	No

Article	Experimental plan	Sample prep	PCR design	PCR optimisation	Analysis of raw reads	Post mapping analysis	Phylogenetic Analysis	Preparation of whole article	Preparation of sections for article	Review of article	1st author	corresponding author
22	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	N/A	Yes	Yes	Yes
23	No	No	No	No	partially	Yes	N/A	No	Yes	Yes	No	No
24	No	Partial	Yes	Yes	No	Yes	N/A	No	Yes	Yes	No	No
25	Partial	Yes	Yes	Yes	Yes	Yes	Partial	No	Yes	Yes	No	No
26	No	N/A	N/A	N/A	N/A	N/A	No	No	Yes	Yes	No	No
27	No	No	N/A	N/A	No	No	No	No	Yes	Yes	No	No
28	Yes	Yes	N/A	N/A	No	Yes	Partial	No	Yes	Yes	No	No
29	No	Yes for APHA	N/A	N/A	Yes for APHA	Yes for APHA	No	No	No	Yes	No	No
30	No	No	N/A	N/A	No	No	No	No	No	Yes	No	No
31	Partial	Yes	N/A	N/A	No	Yes	Yes	Yes	N/A	Yes	Yes	Yes
32	No	No	N/A	N/A	No	No	No	No	No	Yes	No	No