Detection and Characterisation of Novel Vector-borne Viruses in Mozambique

Harindranath Cholleti
Faculty of Veterinary Medicine and Animal Science
Department of Biomedical Sciences and Veterinary Public Health
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2018

Acta Universitatis agriculturae Sueciae
2018:42

ISSN 1652-6880
ISBN (print version) 978-91-7760-224-8
ISBN (electronic version) 978-91-7760-225-5

Print: SLU Service/Repro, Uppsala 2018

Detection and Characterisation of Novel Vector-borne Viruses in Mozambique

Abstract

Arthropod vectors carry a wide variety of viruses that can cause vector-borne infectious diseases that affect the health of both humans and animals. The vector-borne viruses present in nature show high diversity and can infect a broad range of hosts. Moreover, some of these viruses may be transmitted to new hosts in the future and possibly become pathogenic and/or zoonotic.

This thesis describes the detection and characterisation of novel vector-borne viruses circulating in arthropod vectors from an understudied area, the Zambezi Valley of Mozambique, using a viral metagenomic approach. A combination of sequence-independent amplification, high-throughput sequencing and different bioinformatic tools were used to investigate the viromes of mosquitos and ticks in this area. The results of this study resulted in the identification of a broad range of viruses belonging to several viral families, including *Flaviviridae*, *Rhabdoviridae*, *Iflaviridae*, *Picornaviridae*, and *Dicistroviridae*, as well as a number of unclassified RNA viruses that were present in *Culex* and *Mansonia* spp. mosquitoes. The full-length genome of a novel flavivirus from *Mansonia* was characterised, and an analysis of conserved domains showed that it belongs to the insect-specific flaviviruses and is closely related to Nakiwogo virus. Further, a near full-length genome of a highly divergent picorna-like virus from *Culex* mosquitoes was characterised, and the phylogenetic analysis showed that the novel picorna-like virus clustered with members of iflaviruses, confirming that it belongs to the *Iflaviridae* viral family in the order *Picornavirales*. Analysis of the *Rhipicephalus* tick virome identified a number of reads classified as being in the *Orthomyxoviridae* family. The assembled contigs showed an amino acid identity (32-52%) towards known viruses in the genus *Quaranjavirus*.

Overall, viral metagenomics was successfully used to detect and characterise the viromes of important vectors (mosquitoes and ticks) from Mozambique. All the viral sequences identified from this study were divergent from previously known viruses and potentially represent novel viruses that are circulating in the Zambezi Valley of Mozambique. These results constitute the basis for further studies on viruses that are circulating in vectors and on the evolution, viral maintenance and transmission of pathogenic arboviruses.

Keywords: Vector-borne viruses, Viral metagenomics, Mosquitoes, Ticks, Insect-specific flaviviruses, Picorna-like virus, *Quaranjavirus*, RNA viruses, Mozambique

Author's address: Harindranath Cholleti, SLU, Department of Biomedical Sciences and Veterinary Public Health, P.O. Box 7028, 750 07 Uppsala, Sweden
Email: harindranath.cholleti@slu.se

Dedication

To my family and friends…

अस्माकं कार्याणि अस्मान्सावधीकरिष्यंति!
Only action will define us!

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I **Cholleti, H.***, Hayer, J., Abilio, A.P., Mulandane, F.C., Verner-Carlsson, J., Falk, K.I., Fafetine, J.M., Berg, M. & Blomstrom, A.L. (2016). Discovery of Novel Viruses in Mosquitoes from the Zambezi Valley of Mozambique. *PLoS ONE,* 11(9), p. e0162751.

II **Cholleti, H.***, Hayer, J., Fafetine, J., Berg, M. & Blomstrom, A.L. (2018). Genetic characterisation of a novel picorna-like virus in Culex spp. mosquitoes from Mozambique. *Virology Journal,* 15(1), p. 71.

III **Cholleti, H.***, Hayer, J., Mulandane, F.C., Falk, K., Fafetine, J., Berg, M. & Blomstrom, A.L. (2018). Viral metagenomics reveals the presence of highly divergent quaranjavirus in Rhipicephalus ticks from Mozambique. *Infection Ecology & Epidemiology,* 8(1), p. 1478585.

* Corresponding author

The contribution of HC to the papers included in this thesis was as follows:

I Designed experiments together with co-authors. Performed part of the lab work. Performed part of the analysis. Drafted the manuscript and handled correspondence with the journal.

II Designed the experiment. Performed the lab work. Performed most of the analysis, partly with the co-author. Drafted the manuscript and handled correspondence with the journal.

III Designed experiments together with co-authors. Performed the lab work. Performed most of the analysis, partly with the co-author. Drafted the manuscript and handled correspondence with the journal.

List of tables

List of figures

Abbreviations

BLAST	Basic local alignment search tool
BTV	Bluetongue virus
CCHFV	Crimean-Congo haemorrhagic fever virus
cDNA	Complimentary DNA
CHIKV	Chikungunya virus
cISF	Classical insect-specific flavivirus
CuCuV	Cuacua virus
CuPV-1	*Culex* picorna-like virus 1
DENV	Dengue virus
DNA	Deoxyribonucleic acid
DWV	Deformed wing virus
EIDs	Emerging infectious diseases
EILV	Eilat virus
ELISA	Enzyme-linked immune sorbent assay
Gb	Gigabase
HTS	High-throughput sequencing
ISF	Insect-specific flavivirus
ISV	Insect-specific virus
MoPRV	Mopeia rhabdovirus
NAKV	Nakiwogo virus
NCBI	National center for biotechnology information
NGS	Next-generation sequencing
ORF	Open reading frame
PCR	Polymerase chain reaction
PCV	Palm Creek virus
PRF	Programmed ribosomal frameshift
RACE	Rapid amplification of cDNA ends
RdRP	RNA-dependent RNA polymerase

rPCR	Random PCR
rRNA	Ribosomal RNA
RT-PCR	Reverse-transcriptase PCR
RVFV	Rift Valley fever virus
SISPA	Sequence-independent, single primer amplification
TBEV	Tick-borne encephalitis virus
UTR	Untranslated region
VDV-1	Varroa destructor virus 1
WNV	West Nile virus
YFV	Yellow fever virus
ZIKV	Zika virus

1 Introduction

Viruses are likely the most abundant and diverse biological entities on earth (Suttle, 2005) and are able to infect animals, plants and microorganisms. Viruses have both living characteristics, including the ability to reproduce, but only in living host cells, and non-living characteristics, as they are acellular and are dependent on the host cell machinery to synthesise and assemble the viral components (obligate intracellular parasites) (Flint *et al.*, 2015).

Viruses vary in size between 5 and 300 nm (Mimiviruses and Pandoraviruses have a diameter of 0.4 to 0.8 μm) and possess DNA or RNA, but not both, as their genetic material. Examples of ways in which viruses are transmitted include the air, water, food, blood and insect vectors, and viruses can cause diseases that range from the common cold and diarrhoea to more severe diseases such as encephalitis and haemorrhagic fever. Viruses systematically use a series of molecular mechanisms to infect and complete their life cycles, allowing them to cause diseases in a wide range of hosts, indicating that they have sophisticated machineries to avoid host defences. Therefore, it is essential to understand the diversity and nature of viruses to better understand how they cause illnesses.

The introduction of high-throughput sequencing has enabled us to discover novel viruses and to be able to predict future emerging and re-emerging infectious diseases, further re-shaping our knowledge on viral biodiversity (Shi *et al.*, 2016).

Analyses of previous human infectious disease outbreaks have has revealed that bacteria and viruses have been major disease causing agents and that most were zoonotic, including vector-borne diseases (Smith *et al.*, 2014). The emergence and re-emergence of vector-borne viruses has been widely reported across the globe, and these viruses are rapidly extending their boundaries and severely affecting public health. The emergence or re-emergence of these viruses may be due to several factors, such as increased urbanisation, global transportation, adaptation of vectors to new environments, etc. Therefore, it is

necessary to investigate disease causing agents in the environment and their vectors to combat the future emergence and re-emergence of infectious diseases and to improve the public health system. The current work focused on identifying and characterising novel vector-borne viruses in two important vectors, mosquitoes and ticks, from the central part of Mozambique using a metagenomics approach.

1.1 Vector-borne viruses

Vectors are the living organisms that can spread infectious agents between hosts (human to human or animal to human). Blood sucking insects are the most common vectors and transmit infectious agents by taking blood meals. Mosquitoes are considered one of the primary vectors for infectious agents, with others including ticks, sand-flies and flies. Vector-borne diseases have been estimated to constitute over 17% of all infectious diseases (WHO, 2014), which are generally caused by parasites, bacteria and viruses.

Vector-borne viruses are often referred to as arthropod-borne viruses and are termed arboviruses. The term arboviruses (**ar**thropod-**bo**rne **viruses**) was first introduced in 1942 to define members of a group of animal viruses that multiplied in an arthropod and were transmitted to a vertebrate host. Several of these viruses belong to different taxonomic groups. In 1963, the International Subcommittee on Viral Nomenclature endorsed the term "arbovirus" for arthropod-borne viruses. Arboviruses are defined as viruses that are maintained in nature through biological transmission between susceptible vertebrate hosts by haematophagous arthropods. Viral transmission in arthropods also occurs through vertical pathways by means of transovarial and transstadial transmission and possibly by venereal transmission. After an arthropod bites and a period of extrinsic incubation, viruses replicate and produce viremia in the new vertebrate host (Arthropod-borne and rodent-borne viral diseases. Report of a WHO Scientific Group, 1985).

Vector-borne viruses belong to a wide variety of viral families including *Flaviviridae*, *Phenuiviridae*, *Reoviridae*, *Togaviridae*, *Rhabdoviridae* and *Orthomyxoviridae*, *Asfaviridae* and *Poxviridae.* The majority of zoonotic arboviruses belong to the families *Flaviviridae* and *Togaviridae* (Dobler, 2010; Kurkela *et al.*, 2008), and other medically important arboviruses belong to the family *Phenuiviridae*, e.g., Crimean-Congo haemorrhagic fever virus (CCHFV) and Rift Valley fever virus (RVFV) (Pepin *et al.*, 2010; Hoogstraal, 1979). Colorado tick fever virus is another important arbovirus that belongs to the family *Reoviridae* and infects humans (Calisher *et al.*, 1985). Figure 1 illustrates

the classification of vector-borne viruses based on the viral genome characteristics.

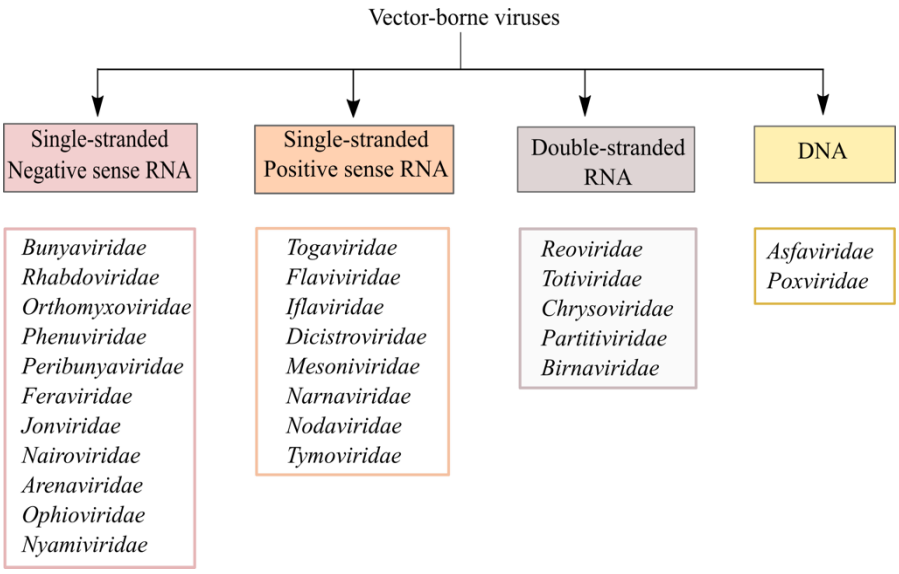

Figure 1. Classification of vector-borne viruses based on its genome characteristics. Vector-borne viruses are classified into 4 different groups, represented in different coloured boxes and boxes below represent viral families consist of vector-borne viruses.

1.2 Life cycle and emergence of vector-borne viruses

Vector-borne viruses are maintained in the environment by a complex life cycle that includes a primary invertebrate host and a vertebrate host. Transmission of viruses may be influenced by several factors, such as the host susceptibility for the virus, the affinity of the vectors for the host and the vector competence for a particular virus (Carver *et al.*, 2009; Lambrechts & Scott, 2009). The majority of arboviruses are maintained through an enzootic cycle (sylvatic cycle), where birds, rodents or non-human primates serve as reservoir hosts (Figure 2) and the virus transmission occurs by primary insect vectors. At the onset of viral infection, the virus replicates in the vertebrate host to higher titres and induces viremia. Upon feeding on this host, an uninfected vector will become infected and become able to transmit the virus to the next host. Viruses can also be transmitted between non-wild or domestic animals, such as pigs and equines, by vectors and lead to epidemics (epizootic rural cycle), which can be further

extended to humans (Weaver & Barrett, 2004). Spill over events of the enzootic cycle or the movement of humans into sylvatic habitats can trigger the emergence of epidemics in humans and domestic animals, probably due to the enzootic vectors involved in virus transmission. However, for some viruses in this cycle, humans and domestic animals are typically considered as dead-end hosts (e.g., West Nile virus (WNV)), as the amplification of the virus insufficient to allow for arthropod vectors to become infected (Gubler, 2001). Some other important arboviruses, such as dengue virus (DENV), chikungunya virus (CHIKV) and yellow fever viruses (YFV) alter their host range from non-human primates to humans, where it amplifies and becomes able to be transmitted to the next person by mosquitoes (urban epidemic cycle), leading to outbreaks without the use of an animal reservoir (Weaver & Reisen, 2010).

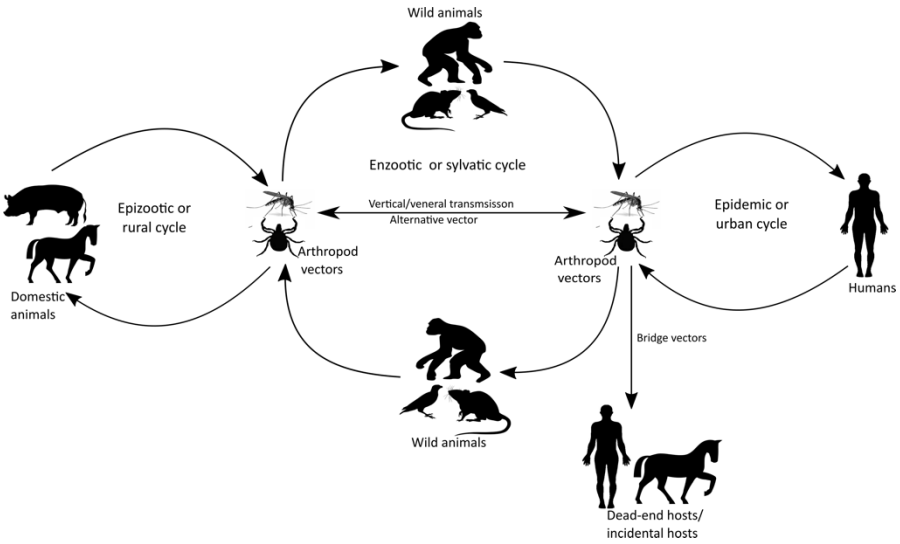

Figure 2. Transmission and maintenance cycles of vector-borne viruses

During the past two decades, the incidences of vector-borne viruses have been expanding geographically. It has been estimated that approximately 50% of the world's population is currently affected by at least one type of vector-borne pathogen. The diseases caused by these pathogens constitute 30% of all emerging infectious diseases (EIDs) (Jones *et al.*, 2008). A combination of socio-economical, environmental and ecological factors have contributed to the emergence of novel viruses, including expanding human population densities, deforestation, climate change, scattering of livestock, livestock-wildlife contacts

and viral adaptation to new hosts species (Pfeffer & Dobler, 2010; Morens *et al.*, 2004). Finally, globalisation, together with the complex web of factors mentioned above, facilitates the spread of viruses to new geographical locations, contributing to the emergence or re-emergence of vector-borne viruses (Simon *et al.*, 2008).

1.3 Genetic diversity of vector-borne viruses

Vector-borne viruses comprise a genetically diverse group of viruses that differ in the structure, composition and organisation of their genomes. This diversity is generally not only evident between viral families but also between individual viral species, which can have distinctive molecular mechanisms for replication, transmission, pathogenesis and evolution (Gray & Banerjee, 1999). The majority of vector-borne viruses contain RNA as their genetic material (positive or negative sense and single or double stranded) and DNA to a smaller extent (Figure 1). RNA viruses may be non-segmented (e.g., *Togaviridae*, *Flaviviridae*, and *Rhabdoviridae*) or segmented (e.g., *Phenuiviridae*, *Reoviridae*, and *Orthomyxoviridae*), with the number of genomic segments ranging from 3 to 12. Viral genomes may consist of single (e.g., *Flaviviridae*) or multiple open reading frames (e.g., *Rhabdoviridae*). In addition, other diverse characteristics have been reported in different vector-borne viruses, including varying lengths of the viral genome and UTRs, the presence of poly(A) tails, organisation of structural and non-structural proteins, the production of sub-genomic RNAs and the mechanism of protein synthesis.

Apart from ecological factors, certain genetic factors influence the diversity and emergence of vector-borne viruses, including: (i) the lack of proofreading activity and repair mechanisms of the RNA-dependent RNA polymerase (RdRP), resulting in the generation of random insertions, deletions and substitutions (point mutations) and new viral variants (Drake & Holland, 1999); (ii) the exchange of long stretches of genomic sequences between closely related viruses (genetic recombination), e.g., the Western equine encephalitis virus is a product of recombination between the Eastern equine encephalitis virus and a Sindbis-like virus (Weaver *et al.*, 1997; Hahn *et al.*, 1988), and *in vitro* studies have also shown the potential recombination within chikungunya virus species (He *et al.*, 2010); and (iii) the exchange of genome segments between segmented viruses during co-infections (genetic reassortment) that generates new genetic combinations, e.g., Thogoto virus (Davies *et al.*, 1987), Bluetongue virus (BTV) and Schmallenberg virus, the latter of which may be the result of a reassortment between Sathuperi and Shamonda viruses (Yanase *et al.*, 2012). Co-circulation

or simultaneous infections of different BTV serotypes can potentially generate novel reassortant viruses (Shafiq *et al.*, 2013; Maan *et al.*, 2012).

Arboviruses must be able to infect both invertebrate and vertebrate hosts to replicate and maintain their life cycle in nature such that these viruses often diversify and evolve to become novel variants. These variants may have the ability to alter the viral infection rate. For example, a single mutation in the envelop glycoprotein E1 enhanced CHIKV transmission by *Aedes albopictus* mosquitoes, i.e., it increased the competence of *Ae. Albopictus* (Tsetsarkin et al., 2007) and additional sequential mutations in CHIKV E2 increased the infection of *Ae. albopictus* (Tsetsarkin & Weaver, 2011). Viral emergence can also be significantly influenced by viral intra-host evolution. For example, viral sequences containing mutations may not be identified by the RNA interference defence system (RNAi, the primary antiviral defence mechanism in mosquitoes) (Brackney *et al.*, 2015; Brackney *et al.*, 2009). Because of the high genetic diversity of viruses, the application of improved molecular methods may be required to detect novel viruses as well as to characterise the viral populations, viral variants or quasispecies in different arthropod vectors.

1.4 Detection of vector-borne viruses

The vast diversity of vector-borne viruses present in nature makes their discovery and classification challenging and may require a combination of methods. In general, the primary focus in most studies has been the detection of pathogenic viruses that are medically important, such as DENV, WNV, tick-borne encephalitis virus (TBEV), and CHIKV and not on other insect-borne viruses, as many of them are asymptomatic in the vertebrate host. The choice of detection method is based on the known characteristics that are specific to each virus, such as incubation period, viremia pattern and antibody response. The identification of infection by antibody based serological methods is typically used at the onset of illness or weeks after the development of symptoms (Innis *et al.*, 1989; Gubler *et al.*, 1984). The classical methods of serology include haemagglutinin inhibition and complement fixation, and most frequently involves the use of enzyme-linked immunosorbent assays (ELISA) and immunofluorescence assays. Direct detection methods that are currently available include virus isolation, electron microscopy, molecular methods and viral antigen detection methods. Virus isolation has been a gold standard method despite the development of methods that are highly sensitive and rapid molecular techniques (Leland & Ginocchio, 2007). However, virus isolation and electron microscopy are laborious processes, requiring a long time for viral cultivation, which is sometimes not possible, and requires special laboratory facilities.

Molecular detection primarily includes nucleic acid based amplification methods, including polymerase chain reaction (PCR)-based methods, specifically, reverse transcriptase (RT)-PCR based assays, as most vector-borne viruses are RNA viruses (Tanaka, 1993). These methods offer a means of rapid viral detection during the viremic phase and they are highly sensitive (Lanciotti & Kerst, 2001; Lanciotti *et al.*, 2000). However, some viruses produce low and short-lived viremias, making it difficult to detect viruses such as WNV (Murray *et al.*, 2011). In addition to PCR, standard molecular methods, such as nucleic acid hybridisation methods and microarrays, have also been used as detection assays. All these methods are based on prior information of viral sequences and are commonly species specific. Thus, the detection of a virus is sometimes not possible if the virus has an unknown aetiology.

1.4.1 Viral metagenomics

Viral metagenomics is the study of viral genomes from primary samples, e.g., environmental samples, clinical material from humans and animals and insect tissue. This newly developed, culture- and sequence-independent method has been able to detect viruses of an unknown aetiology as well as allowing for the characterisation of the complete viral populations in a given sample (Delwart, 2007). The workflow of viral metagenomics includes the following steps: sample preparation, sequence-independent amplification, high-throughput sequencing, bioinformatics and follow-up studies, if necessary (Blomstrom, 2011).

Sample preparation and amplification

Sample preparation includes the combination of different methods that are used to enrich the virome in the sample, including filtration, ultracentrifugation, nuclease treatment and the removal of ribosomal RNA. This is an important step, as the ratio of viral nucleic acids will be much lower compared to the host genome (Hall *et al.*, 2014). Amplification of nucleic acids can be performed by different methods, including sequence-independent, single-primer amplification (SISPA), which is based on the ligation of adapters to nucleic acids (Reyes & Kim, 1991). SISPA has been combined with random PCR and nuclease treatment steps (Djikeng *et al.*, 2008; van der Hoek *et al.*, 2004; Allander *et al.*, 2001; Froussard, 1993) to amplify divergent viral sequences present in the sample. Random PCR (rPCR) (Froussard, 1992), linker-amplified shortgun library (LASL) (Breitbart *et al.*, 2002) and multiple displacement amplification (MDA), the latter of which uses the displacement DNA polymerase, e.g., the phi29 DNA polymerase (Dean *et al.*, 2002), are other amplification methods that

have been used. Although these methods have been successfully used to amplify the metagenomes, they have some limitations, such as an incomplete retrieval of viral genomes, an amplification bias towards the 3' end of the genomes and a biased distribution of sequencing depth (Myrmel *et al.*, 2017; Rosseel *et al.*, 2015; Karlsson *et al.*, 2013).

High-throughput sequencing (HTS)

A combination of Sanger sequencing and advanced fluorescent detection methods led to the development of next generation sequencing (NGS), often referred to as second-generation sequencing. The first high-throughput sequencing platform was introduced in 2005, which was 454 pyrosequencing by 454 Life Sciences (acquired by Roche in 2007 and later shut down in 2013). Several HTS platforms have been developed over the years that feature variable read lengths, type of sequencing, run times and throughput capacity (Goodwin *et al.*, 2016). The cost of sequencing for each reaction has been significantly reduced in recent years, and sequencing machines are able to generate massive sequence outputs, up to 1500 Gb. The Illumina method is based on a paired-end read chemistry and has numerous platforms (HiSeq, MiSeq, and NextSeq), each with different read lengths and run times while producing high throughput data. Ion-torrent (from Life technologies) runs as a single-read platforms and is the first semiconductor-based platform that can generate up to 1 Gb of data, with a longer read lengths of up to 400 bases. The newer versions, Ion proton and Ion S5, can generate up to 15 Gb of data with varying read lengths. The latest HTS platforms from Pacific Bio and Oxford Nanopore have been developed to generate longer sequences of up to 200 Kb. The choice of sequencing platform depends on the application, and each platform has it strengths and weaknesses that should be thoroughly studied at the beginning of the project. The benchtop instruments developed by Illumina and Ion Torrent have been largely used in various insect virome sequencing projects (Pettersson *et al.*, 2017; Cholleti *et al.*, 2016; Frey *et al.*, 2016; Xia *et al.*, 2015; Liu *et al.*, 2011). Table 1 summarises the HTS platforms available and their sequencing features.

Table 1. *Summary of high-throughput sequencing technologies available, the detection method, their output read length, throughput, and runtime. Information has been obtained from the sequencing company webpages. (PE: paired-end, SE: single-end, Kb, kilobase pairs, Mb: megabase pairs, Gb: gigabase pairs, Tb: terabase pairs, M: millions, B: billions, h: hours, NA: not available)*

Platform	Detection method	Read length	Throughput	Reads	Runtime
Short-read sequencing					
Illumina MiniSeq	Sequencing by Synthesis	150 (PE)	7.5 Gb	44-50 M	24 h
Illumina MiSeq	Sequencing by Synthesis	300 (PE)	13-15 Gb	44-50 M	21-56 h
Illumina NextSeq	Sequencing by Synthesis	150 (PE)	100-120 Gb	800 M	29 h
Illumina HiSeq	Sequencing by Synthesis	150 (PE)	650-750 Gb	2.5 B	1-3.5 d
Ion Torrent	Semiconductor Sequencing	400 (SE)	1-2 Gb	4-5.5 M	7.3 h
Ion Proton	Semiconductor Sequencing	200 (SE)	Up to 10 Gb	60-80 M	2-4 h
Ion S5	Semiconductor Sequencing	200 (SE)	10-15 Gb	60-80 M	2.5 h
Long-read sequencing					
Pacific Bio	Single–molecule real-time Sequencing	20 Kb	500Mb-1Gb	~55,000	4 h
Pacific Bio Sequel	Single–molecule real-time Sequencing	8-12 Kb	3.5-7Gb	~350,000	0.5-6 h
Oxford Nanopore MinION	Nanopore DNA sequencing	Up to 200Kb	40 Gb	>100,000	48 h
Oxford Nanopore PromethION (48 flow cells)	Nanopore DNA sequencing	Up to 200Kb	Up to 15 Tb	NA	64 h

Bioinformatics

Bioinformatics is the application of tools and computational analyses to understand and interpret biological data. Bioinformatics is an interdisciplinary field that has been widely applied in modern biology and medicine for data management (Bayat, 2002). The analysis of massive sequencing data generated from HTS platforms typically includes quality checking, assembly and taxonomic classification of reads or contigs produced by the assembly.

Quality checking involves trimming of sequences according to Phred quality scores, which are related to base calling error probabilities (Ewing & Green, 1998). Identifying and removing sequence duplicates that are produced by the HTS platform as a result of PCR amplification, PCR errors or sequencing errors is necessary to reduce the computational time, to accurately calculate an estimated species abundance and to improve the assembly. All these quality filtering conditions can be specified depending on the downstream analyses required (Bokulich *et al.*, 2013; Patel & Jain, 2012). Moreover, the sequences that are not a target of the output can be filtered out to eliminate misassemblies and to speed up the analysis. For example, the host sequences can be removed from a sample if the target sequences are viral-related reads (Bzhalava *et al.*, 2013). Possible contaminating sequences or sequences that are not relevant can also be removed by aligning against reference sequences, which can be performed using several alignment tools such as BWA, SOAP2 and Bowtie2 (Langmead & Salzberg, 2012; Li & Durbin, 2010; Li *et al.*, 2009).

The assembly of shorter sequences that have matching overlaps generates longer sequences called contiguous sequences (contigs), a method referred to as *de novo* assembly. These contigs can be further extended by merging shorter contigs. There are two primary types of *de novo* assembly programmes, Overlap/Layout/Consensus assemblers (e.g., MIRA, Celera, and VICUNA), which are widely used for longer reads (Yang *et al.*, 2012) and de Bruijn graph assemblers (e.g. Velvet, SOAP de novo, SPAdes) (Bankevich *et al.*, 2012; Luo *et al.*, 2012; Zerbino, 2010). However, the assembly process might generate 'chimeric' sequences involving the assembly of sequences from different organisms or species, which may be a problem in viral metagenomic studies as the biological sample may contains closely related viral sequences (Berthet *et al.*, 2016).

Taxonomic classification is the final step in the metagenomic analysis, where each sequence is assigned to a taxonomic group. The most commonly used similarity-based classification is Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990), where the sequences are compared to known genomes. Different versions of BLAST can be used, such as BLASTx and tBLASTx (McGinnis & Madden, 2004; Gish & States, 1993). Considering the time span

for sequence classification, several different tools have been developed that can reduce the time required from weeks to days, e.g., RAPsearch2, Diamond, and GHOSTX (Suzuki *et al.*, 2017a; Buchfink *et al.*, 2015; Zhao *et al.*, 2012).

Viral metagenomics provide basic information on which viruses are present in a sample. More extensive analysis or follow-up studies are necessary to understand the roles of the identified virus. These analyses may include obtaining full-length viral genomes by the primer walking approach, RACE analysis, virus isolation, viral characterisation and developing diagnostic assays, which all depends on the objective of the study (Rosseel *et al.*, 2014; Blomstrom, 2011).

1.5 Implications of invertebrate viromes in human and animal public health

Arthropods are invertebrates, including insects, spiders, crustaceans and centipedes, which often share common habitats with humans and animals. Haematophagous arthropods (require blood meals for survival) feed on animals or humans and may be able to transmit disease causing agents among them. These agents, especially vector-borne viruses, have been associated with significant morbidity and mortality rates in human and animal populations during past decades (Gubler, 2002). Vector-borne diseases have the ability to influence the ecosystem by causing a loss of biodiversity and can transmit severe diseases to new hosts, including humans and livestock. Such events with high infection rates can indirectly influence the socioeconomic status of countries.

In humans, several mosquito-borne epidemics have been reported across the globe, including the emergence of dengue fever by DENV serotypes 1-4 transmitted by *Aedes aegypti*, which has also been a vector for other disease-causing viruses including YFV, CHIKV, and Zika virus (ZIKV) (Gratz, 1999). YFV was controlled by mosquito abatement techniques in the Americas, but it remains a constant threat for re-emergence in new areas by *Aedes* mosquitoes. ZIKV transmission occurs primarily through bites of infected *Aedes* mosquitoes and has promoted major disease outbreaks in humans from different geographical areas (Ryan *et al.*, 2017; Musso *et al.*, 2014; Duffy *et al.*, 2009). WNV was introduced into the Americas in 1999 and was probably a derivative of an Israeli WNV strain, with this strain having rapidly spread across the United States (Lanciotti *et al.*, 1999). Similarly, CHIKV was introduced into Asia from Africa in the mid-2000s before subsequently spreading to the Caribbean region in 2013 (Burt *et al.*, 2017). RVFV, another medically important virus in livestock, has been identified in different mosquito species (*Aedes, Culex*, *Anopheles, etc.*) (Tantely *et al.*, 2015; Grobbelaar *et al.*, 2011) and can be

transmitted to humans from infected animals. Other than mosquitoes, several arthropods have been identified as vectors for different pathogenic viruses. For example, adult midges were shown to carry BTV (Brand & Keeling, 2017), sand-flies are able to transmit Toscana virus, and ticks can carry TBEV and CCHFV (Mansfield *et al.*, 2017; Charrel *et al.*, 2005), with many more vector-borne viruses expected to be identified that affect public health.

In addition to human and animal diseases caused by vector-borne viruses, these viruses can also affect invertebrate health, including that of honey bees, causing serious damage to food crops that results in huge economic losses for the agricultural industry (Brutscher *et al.*, 2016). Dicistroviruses (Acute bee paralysis virus, Israeli acute bee paralysis virus, Kashmir bee paralysis virus, etc.), iflaviruses (Deformed wing virus (DWV), Kakugo virus and Varroa destructor virus-1 (VDV-1), and Sacbrood virus (SBV), etc.), and other groups of viruses have been reported as pathogens that infect honey bee populations in different geographical locations (Chen & Siede, 2007).

With the use of metagenomics and transcriptomics, a broad range of unknown and highly divergent RNA viruses have been discovered from different invertebrate species. For example, a meta-transcriptomics analysis of 220 invertebrate species resulted in the discovery of 1445 RNA viruses, including probable new viral families (Shi *et al.*, 2016). In another study, 112 novel RNA viruses were reported from 70 arthropod species (Li *et al.*, 2015). These studies show that invertebrates harbour RNA viruses with greater genetic diversity than previously expected, and the identified viruses are likely to be ancestors of major viral groups, including those that infect vertebrates. Thus, analysing the biodiversity of invertebrate viromes may have important implications for our understanding of virus evolution, ecology and emergence.

Viruses that are restricted to insect hosts (insect-specific viruses, ISVs) have also been reported in different species of mosquitoes. Evolutionary relationships show that these viruses are closely related to arboviruses, and they may have been established as pathogenic arboviruses over a period of time (Bolling *et al.*, 2015). Different studies have reported that some of these ISVs are able to reduce the replication of certain arboviruses following pre-infection or coinfection (Kuwata *et al.*, 2015; Kenney *et al.*, 2014; Hobson-Peters *et al.*, 2013). Therefore, these viruses may be used as biological control agents for arboviruses. Exploiting the host range restriction at multiple levels in vertebrate cells (Junglen *et al.*, 2017; Nasar *et al.*, 2015b), ISVs may also be used in vaccines and in diagnostic platforms, e.g., a chimaera of Eilat virus (EILV)/CHIKV was used as an ELISA antigen (Erasmus *et al.*, 2015) and was also developed as a vaccine candidate for chikungunya fever (Erasmus *et al.*, 2017).

2 Aims of the thesis

The overall aim of this thesis was to investigate the composition of the viromes of important vectors and to characterise some of the novel vector-borne viruses from the Zambezi Valley of Mozambique by viral metagenomics.

The specific aims were to:

- Characterise the virome of *Culex* and *Mansonia* spp. mosquitoes from Mozambique (Paper I)
- Characterise a full-length novel insect-specific flavivirus in *Mansonia* mosquitoes (Paper I)
- Detect and characterise a novel picorna-like virus from *Culex* mosquitoes and study the presence of *Culex* picorna-like virus in individual *Culex* and *Mansonia* mosquito pools (Paper II)
- Characterise the virome of *Rhipicephalus* spp. ticks (Paper III)

3 Materials and Methods

The methods used to conduct studies I, II and III are summarised in this section. The materials and methods are comprehensively described in each paper (I-III).

3.1 Sample collection and morphological classification

Mosquitoes (540) and ticks (51) were collected from a private land area in Cuacua village in the Zambezia province of Mozambique during Oct-Nov 2014 with the permission of the land-owner and local farmers. The mosquitoes were collected using CDC light traps and were classified up to the genus level according to morphological identification keys (Hopkins, 1952; Edwards, 1941), after which they were stored in RNAlater (Invitrogen) until further use. Ticks were collected from the surface skin of small ruminants (sheep and goats) and transported to the lab in RNAlater where they were classified according to Walker *et al*, 2003 (Walker *et al.*, 2003). All the samples were shipped to The Public Health Agency of Sweden, Solna, on dry ice and were stored at -80°C until further processing.

3.2 Sample processing

Prior to the nucleic acid extraction the mosquitoes and ticks were pooled, based on genera, into 23 and 25 pools, respectively, containing up to 20 mosquitoes or 3 ticks. Each pool was mechanically homogenised in Trizol LS reagent (Invitrogen) using a TissueLyser II (Qiagen), and RNA was then extracted from the homogenate according to the manufacturer's protocol. RNA (5 µl) from each pool was further pooled into a single pool (one mosquito pool and one tick pool) and concentrated before a DNase treatment. The eluted RNA from the DNase treatment was treated to remove ribosomal RNA using a RiboZero kit (Ribo-

Zero Gold (Human/Mouse/Rat), Illumina) and RNA was further concentrated. The RNA quality and concentration were measured with an Agilent 2100 Bioanalyser (Agilent technologies) using an Agilent RNA 6000 Nano kit according to manufacturer's protocol.

3.3 cDNA synthesis and pre-amplification

The whole metagenome in the sample was amplified by SISPA in combination with rPCR (Allander *et al.*, 2001). The detailed protocol is described in Paper I, briefly, the first strand cDNA was synthesised using SuperScript III (Invitrogen) and the FR26RV-N primer. Second strand synthesis was performed with a Klenow fragment (3' to 5' exo-) (New England BioLabs Inc, NEB). Random amplification of the ds-cDNA was performed in triplicate with AmpliTaq Gold DNA polymerase (Applied Biosystems) and the FR20RV primer targeting the tagged cDNA sequence. The tagged sequence was cleaved off with the restriction enzyme *Eco*RV (NEB), and the final product was subsequently purified using a GeneJet PCR purification kit (Thermo Fisher Scientific).

3.4 Library preparation and high-throughput sequencing

Sequencing libraries were prepared, and sequenced at the National Genomics Infrastructure (NGI), SciLifeLab, Uppsala, Sweden. The metagenome was sequenced on an Ion Torrent PGM platform using an Ion 318^{TM} Chip v2 with 400-bp read length chemistry. The raw sequencing data has been deposited at the Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra), and the project numbers are provided in papers I-III.

3.5 Bioinformatics analysis of sequencing data

The raw sequencing data was analysed using a combination of bioinformatics tools. Sequences were filtered to remove those with quality scores lower than Q<20, and exact duplicates and the low-quality ends of the sequences were trimmed using PRINSEQ. The mosquito host sequences were removed by mapping the sequences to the *Anopheles*, *Aedes* and *Culex* genomes using the mapping tool Bowtie2. The unmapped sequences were classified by BLAST querying the NCBI nucleotide (nt) and protein (nr) databases. The E-value cut-off used for similarity searches was 0.001. Sequences that were classified as viruses were assembled into longer sequences (contigs) by different *de novo* assemblers, including CodonCode Aligner 6.02 (CodonCode Corporation),

SeqMan 11.2.1 (DNASTAR) and SPAdes. The viral contigs were aligned to the viral reference genomes using CodonCode Aligner 6.02 and Bowtie2.

3.6 Sequence confirmation and recovery of genomic ends by RACE analysis

The selected viral sequences were confirmed by PCR and Sanger sequencing. Primers were designed manually from contigs and individual HTS reads to fill the gaps and obtain longer sequences. The Sanger sequences were confirmed with BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and the ORFs of viral genomes were predicted using ORF finder at NCBI (https://www.ncbi.nlm.nih.gov/orffinder/). The RACE method was used to recover the 5' and 3' sequencing ends of the viral genomes. The detailed protocol used is described in Paper III. Sanger sequencing was carried out at Macrogen Europe (Macrogen Inc.). The verified sequences were deposited at NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/), and the GenBank accession numbers are provided in papers I-III.

In study II, RT-PCR assays were used to test individual mosquito pools for the presence of specific viruses. First strand cDNA was synthesised using SuperScript III (Invitrogen) and random hexamers, with virus specific primers and AmpliTaq Gold DNA polymerase used in the PCR reactions.

3.7 Prediction of conserved protein motifs and phylogenetic analysis

The complete ORFs of closely related viruses were aligned with ClustalW using BioEdit software 7.0.4 and conserved domains were predicted. To determine the sequence similarities and to predict the cleavage sites and evolutionary relationships of the identified viruses, viral nucleotide and protein sequences were downloaded from GenBank and UniprotKB, representing viruses belonging to different viral families or viral orders. The flaviviral sequences used in this study are listed in the Table 2. Sequence alignments were performed using the multiple sequence alignment tool ClustalW, and phylogenetic trees were constructed by the neighbor-joining (NJ) or maximum likelihood (ML) methods with 500 or 1000 bootstraps using MEGA.

Table 2. Insect-specific flaviviruses used in this study for sequence identity matrix and protein cleavage site predictions.

Virus name	Abbreviation	GenBank Accession number
Cuacua virus	CuCuV	KX245154.1
Nakiwogo virus	NAKV	GQ165809.2
Palm Creek virus	PCV	KC505248.1
Nienokoue virus	NIEV	KX879625.1
Culex flavivirus	CxFV	JQ308188.1
Culex theleri flavivirus	CTFV	HE574573.1
Quang binh virus	QBV	NC_012671.1
Cell fusing agent virus	CFAV	NC_001564.2
Hanko virus	HANKV	NC_030401.1
Kamiti river virus	KRV	NC_005064.1
Aedes flavivirus	AEFV	AB488408.1

4 Results and Discussion

4.1 Viral metagenomics of *Culex* and *Mansonia* mosquitoes from Mozambique

In study I, *Culex* and *Mansonia* spp. mosquitoes, collected from the Zambezi Valley of Mozambique, were analysed to characterise the virome using a viral metagenomic approach. The collected mosquitoes were morphologically classified in the field as *Culex* and *Mansonia* species. The Ion Torrent HTS-generated data were analysed with different bioinformatics tools. After the quality check and host mapping, the unmapped reads were taxonomically assigned using a local NCBI BLAST search. The majority of the reads were observed to be from the host genomes and bacteria. The dominant bacterial genera in the dataset include *Wolbachia* and *Salmonella*. The number of reads belonging to viruses from the taxa *Culex* and *Mansonia* were 3269 and 983, constituting 0.19 and 0.03% of the total reads in these datasets, respectively. The viral sequences primarily belonged to the families *Flaviviridae*, *Rhabdoviridae*, *Iflaviridae*, *Dicistroviridae*, and other unclassified RNA viruses (Figure 2). These findings are in agreement with previous studies that used metagenomic approaches, where a broad range of viruses belonging to different families were identified, including *Flaviviridae*, *Totiviridae*, *Densoviridae*, *Rhabdoviridae*, *Reoviridae*, *Anelloviridae*, *Bunyaviridae* and many unclassified RNA viruses (Fauver et al., 2016; Shi et al., 2015; Coffey et al., 2014). Collectively, these results suggest that the mosquitoes are reservoirs for a variety of RNA viruses. These types of studies facilitate the identification of unknown viruses and increase our understanding of viral biodiversity in arthropod vectors.

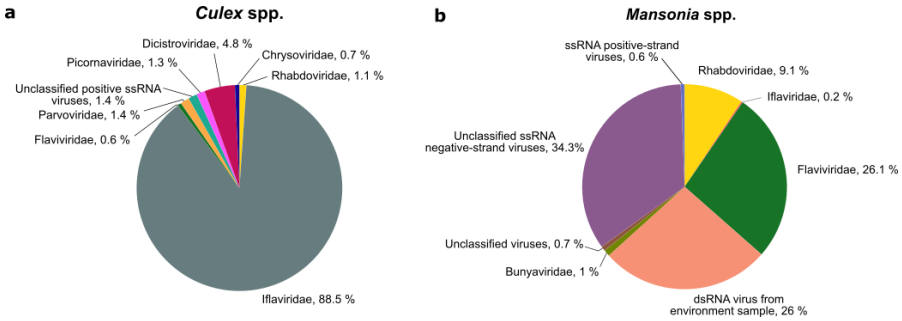

Figure 3. Taxonomic classification of good-quality Ion Torrent reads from (a) *Culex* and (b) *Mansonia* mosquito species, from Paper I (Cholleti *et al.*, 2016).

Mosquito-borne flaviviruses are important pathogens of humans and animals, although some flaviviruses (called insect-specific flaviviruses (ISFs)) are specific to the mosquito vector. Due to the potential pathogenicity of flaviviruses, the sequences related to the *Flaviviridae* family were further analysed. Both mosquito species analysed had sequences related to the *Flaviviridae* family, although most were identified as *Mansonia* species (257 reads), while only 19 reads were identified in the *Culex* data set. At the amino acid level, the flaviviral reads were closely related to Palm Creek virus (PCV) and Nakiwogo virus (NAKV). PCV and NAKV are two ISFs that have previously been identified in *Coquillettidia* and *Mansonia* mosquito species, respectively (Hobson-Peters *et al.*, 2013; Cook *et al.*, 2009). The other flaviviral reads identified were related to Nienokoue virus (NIEV) and *Culex* flaviviruses, which were both reported as ISFs. Advancements in genomics and sequencing methods during the past decade have enabled us to characterise a wide range of insect-specific viruses (ISVs) in different mosquito species and from different geographical areas. The first ISV reported was Cell fusing agent virus (CFAV) in 1975, which was isolated from *Aedes aegypti* cells (Stollar & Thomas, 1975). Many years after this first discovery, Kamiti river virus (KRV) was characterised from field-caught mosquitoes in 2003 (Crabtree *et al.*, 2003). Most of the ISVs characterised so far phylogenetically belong to the *Flaviviridae* family, although ISVs have also been identified in other viral families, such as *Reoviridae*, *Phenuiviridae*, *Rhabdoviridae*, and *Togaviridae* (Hobson-Peters *et al.*, 2016; Auguste *et al.*, 2015; Bolling *et al.*, 2015; Nasar *et al.*, 2012; Quan *et al.*, 2010). The majority of the ISVs in mosquitoes were detected as part of arbovirus surveillance projects in which mosquitoes are the most important target (Vasilakis & Tesh, 2015). To understand the genetic diversity and molecular evolution of the identified ISF from *Mansonia* mosquitoes, in this study, the virus was further characterised by complete genome sequencing.

4.1.1 Full-length genome characterisation of a novel ISF from *Mansonia* mosquitoes

The assembly of the viral reads related to the *Flaviviridae* family generated contigs that were related to NAKV (10 contigs), PCV (11 contigs) and NIEV (2 contigs), with an amino acid identity of between 45-87%. The contigs and individual reads that aligned to NAKV covered up to 70% of the genome. The full-length genome was obtained by filling the gaps between contigs through conventional PCR reactions and subsequent sequencing. The genomic ends were sequenced via 5' and 3' RACE. A BLAST search of the full-length genome showed 75 and 86% of identity at the nt and aa levels to NAKV, respectively. The identified genome was named Cuacua virus (CuCuV) (based on the village name where the mosquitoes were collected) and represents the first insect-specific flavivirus to be characterised in Mozambique.

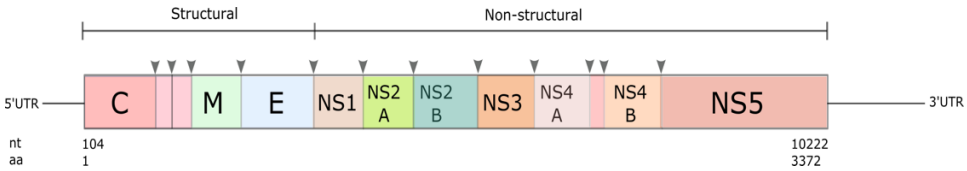

Figure 4. Schematic representation of the full-length CuCuV genome and its ORF. The structural and non-structural proteins are represented on the top. Predicted cleavage sites for each protein are indicated by filled arrows. The nucleotide and amino acid positions are shown for the ORF in the bottom of the figure.

Genome organisation of CuCuV

The complete CuCuV genome, including the 5' and 3' UTRs, was shown to be 10,725 nucleotides (nt) in length and the predicted single polyprotein was 3372 amino acids (Figure 4). The lengths of the 5' and 3' UTRs were 103 and 503 nt, respectively, and the predicted ORF was 10,119 nt in length. The 5' UTR length of CuCuV is consistent with those of most of the previously characterised ISFs and other flaviviruses (91 to 113 nt), although the 3' UTR of ISFs are highly variable, ranging from 350 to 1200 nt (Blitvich & Firth, 2015; Markoff, 2003). The predicted polyprotein consists of three structural proteins (C, prM, and E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). The specific nucleotide positions for each gene have been predicted by multiple sequence analysis (Table 3).

Table 3. *The predicted genomic positions of each gene and UTRs, and its protein size*

Region	Gene	Position in genome (nt range)	Protein size (aa)
5'UTR		1-103	
Structural	C	104-451	116
	pr	452-763	104
	M	764-955	64
	E	956-2236	427
Non-structural	NS1	2237-3433	399
	NS2A	3434-4063	210
	NS2B	4064-4504	147
	NS3	4505-6265	587
	NS4A	6266-6724	153
	NS4B	6725-7552	276
	NS5	7553-10222	890
3'UTR		10223-10725	

The BLASTx search of the complete ORF showed the closest protein identity to NAKV (85.9%), the pairwise identity to other classical ISFs (cISFs) both at aa and nt level are presented in Table 4.

Table 4. *Pairwise identity percentages between the polyprotein sequence of CuCuV and other classical ISFs. Top right and bottom left values represent the identity percentages at both amino acid level and nucleotide level, respectively.*

	CuCuV	**NAKV**	**PCV**	**NIEV**	**CxFV**	**CTFV**	**QBV**	**CFAV**	**HANKV**	**KRV**	**AEFV**
CuCuV	*	85.9	67	56.1	53.3	52.5	53.3	43.1	41.9	40.7	40.5
NAKV	75.3	*	67.3	56.3	53.1	52.5	53.5	43.5	42.2	41.1	40.7
PCV	65.8	65.3	*	57.4	54.8	54.5	55.6	44.4	42.4	40.7	41
NIEV	60	59.7	60.7	*	54.8	53.9	54.7	44.3	42.4	41.3	40.7
CxFV	58.2	58.5	59.6	59.7	*	70.9	69.8	47.2	42	41.1	40.5
CTFV	58.1	58.5	59.2	58.7	67.6	*	73.4	48.8	41.4	40.8	39.8
QBV	58.4	58.6	59.5	59.6	66.9	68.5	*	47.4	41.4	39.9	40.8
CFAV	54.8	54.8	55	55.4	57.2	58	57	*	44	66.7	61.1
HANKV	54.4	54.4	54.5	54.8	54.5	54.7	54.4	54.4	*	43	42.3
KRV	53.5	53.7	54	53.9	53.9	54.1	54.2	65.4	54.3	*	69
AEFV	53.2	53.6	53.4	53.3	53.8	54	53.4	62.1	54.3	66	*

Table 5. *Predicted cleavage sites of CuCuV and other classical insect-specific flaviviruses (Bolling et al., 2015)*

	CuCuV	**NAKV**	**PCV**	**NIEV**	**CTFV**	**QBV**	**CxFV**	**KRV**	**AEFV**	**CFAV**	**HANKV**
Virion C/Anch	LEKRR↓GAGVS	LEKRR↓GVWSP	LEKKR↓DGRAA	LEQRR↓GAQRG	LEVRR↓SANNP	LENRR↓SANPL	LEAKR↓SAKNA	LEKQR↓SGPNL	LEAQR↓SHSPV	LESRR↓TTGDP	LEKER↓SHPRK
C/prM	VGIFS↓INVID	VGIFS↓LNVVD	FGVMG↓VVVID	MVTFA↓AVVDV	VLCGC↓VIIDM	TLCGT↓MVIDM	MMVLG↓AVVID	GLCYG↓EMLRY	GLALS↓ETLRY	VLCGC↓VVIDM	IVVTG↓LSIEL
pr/M	TRRQR↓DVKEE	NRKQR↓SVKDE	TRAKR↓VAPDG	RPVRR↓DVTPA	KRVKR↓APETP	KRVKR↓ATEQP	KRERR↓VASTN	VRRRR↓APQPQ	PRKRR↓SSQP	KREKR↓SREP	ERETR↓QKVDD
prM/E	TVVKA↓EFMEP	TTVRG↓EFMEP	TTVRG↓EYMEP	TTVSG↓EYLEP	TTVKG↓EFVEP	STVKG↓EFVEP	TTVKG↓EFVEP	NVVKA↓SSIEP	NVVRA↓TSIEP	TTVKG↓EFVEP	NVVKG↓EFVEP
E/NS1	YFVKG↓DFGCG	YTVRA↓DFGCG	YFVRA↓DFGCG	YYVRA↓DVGCG	YFARA↓DVGCG	YYTRA↓DVGCG	VYTKA↓DVGCG	RSVSA↓DVGCG	RRVAG↓DIGCG	YYVRA↓DLGCG	VYVKA↓DVGCG
NS1/NS2A	SPGEA↓AKIQT	PPSGA↓EKLQQ	PMGET↓AKIQN	GGAEA↓TQSFF	PGTGA↓FPDFQ	PGAEA↓LLQDF	PPVEG↓SYPDF	GKAHA↓CSDFR	GKADA↓TADFH	GKANA↓QSDFR	YRVPS↓TNAED
NS2A/NS2B	ASGLR↓RPRPH	ASGLR↓KPRPH	GDGLR↓APRPH	KSGLR↓SITSW	KSGLR↓ASKSS	KSGLR↓ASKRS	RSGLR↓ASRRS	KNGYR↓DYGAS	KSSYR↓TSGRS	RNGYR↓DSGAN	RSGYR↓ALCSS
NS2B/NS3	EFSQR↓STSEL	EFAQR↓SSSEL	AMSQR↓ANSEL	STAQR↓SDLLL	STAYR↓AGVND	TSNRR↓SGVND	VSVFR↓SNEVN	SEQNR↓SDDLL	NEHCR↓SDDLL	TANNR↓SDDLL	TNAFR↓SDELI
NS3/NS4A	FLKQR↓SAIPI	FLKQR↓SVLPF	FLKQR↓SLYFD	FLKQR↓SLFID	FLKQR↓SGANF	FLKQR↓SVLNF	YLKQR↓SNFNF	YLNCR↓SSQTF	YINTR↓SSASL	YMNCR↓GGPSL	YMGTR↓SFLSV
NS4A/2K	GSSQR↓GLLDS	GGSQR↓GILDS	GGSQR↓GVLDS	ANSQR↓GFAEN	NNVHR↓AYTGD	TNVHR↓AYTGD	NNVHR↓AYTTD	AIGNR↓SYMDS	AAGNR↓SYLDS	SIGNR↓SYMDS	SAGQR↓SYVDI
2K/NS4B	IGIAA↓WELQL	IGIAA↓WELQL	IGVTA↓WELEL	SAVVA↓WELNL	MGVVA↓WELNL	MGIVA↓WELEL	MGVVA↓WEMDL	CGVLA↓WEMRL	CSVLA↓WEMRL	CGVLA↓WEMRM	IGVIC↓WELRL
NS4B/NS5	RLSVR↓SLVKS	RLSVR↓SLVKS	RLGVR↓SLVKS	LDMRR↓SLMKT	RGGLR↓SLVKT	RLATR↓SLVKT	RMALR↓SLVKT	FNQFR↓ALEKS	FSKFR↓ALEKS	FNQFR↓ALEKS	NITTR↓SLEKS

Predicted polyprotein cleavage sites in CuCuV

The polyprotein amino acid sequence of CuCuV was aligned with other cISFs to predict the cleavage sites (Table 5). The predicted cleavage sites in CuCuV were similar to the other cISFs. Cleavage sites of cISFs have been shown to follow rules established for dual-host flaviviruses, with some exceptions (Blitvich & Firth, 2015). However, the predicted cleavage sites in CuCuV need to be confirmed.

Ribosomal frameshifting in CuCuV

Several RNA viruses use programmed -1 ribosomal frameshifting (-1 PRF) to increase the protein products translated from limited mRNA transcripts. The -1 PRF conserved sequence motif in cISFs has been predicted to be between the NS1 and NS2A regions (Firth et al., 2010). A slippery heptanucleotide sequence motif where the -1 nt shift takes place has been predicted in the CuCuV genome, and this motif is between 3445-3453 nt (Figure 4), close to the NS1 and NS2A junction. An overlapping ORF (termed *fifo*, Fairly Interesting Flavivirus ORF), generated as a product of frameshifting, was present in the CuCuV genome and consisted of 280 codons, identical in size to that of the predicted *fifo* ORF in NAKV (Blitvich & Firth, 2015). The length of the frame shifting ORF in flaviviruses (termed as *foo*, Flavivirus Overlapping ORF) is relatively short (45 codons) (Firth & Atkins, 2009) compared to ISFs (approximately 300 codons). The ribosomal frameshift sites were previously identified in the JEV serogroup of flaviviruses, and it was experimentally shown that the ribosomal frame shift leads to the production of the NS1 protein, which plays a role in neuroinvasiveness (Melian *et al.*, 2010). However, the role of the *fifo* ORF in insect-specific flaviviruses has yet to be described.

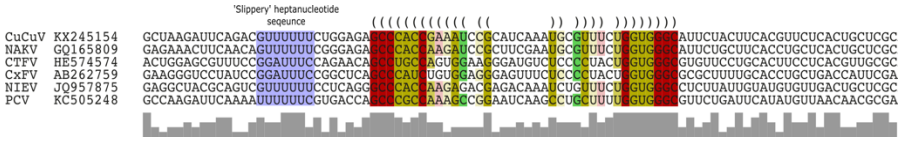

Figure 5. Predicted ribosomal -1 frameshift sequences and stimulatory elements in CuCuV. The nt sequence in the blue-coloured box shows the conserved slippery heptanucleotide in cISFs. The CuCuV RNA stem loop structure was compared with other cISFs: predicted base-pairing nucleotides are shown with red-, yellow-, green- and pink-coloured boxes based on identity, predicted base-pairing is shown with "()", and the conserved positions are indicated by the grey coloured histogram in the bottom.

Predicted fusion peptide motif in CuCuV

Membrane fusion between flaviviruses and the host cell is thought to be mediated by an amino acid motif consisting of 14 aa in the E protein. This fusion motif was identified in the E protein of CuCuV, and this motif seems to be highly conserved among the cISFs (Figure 6).

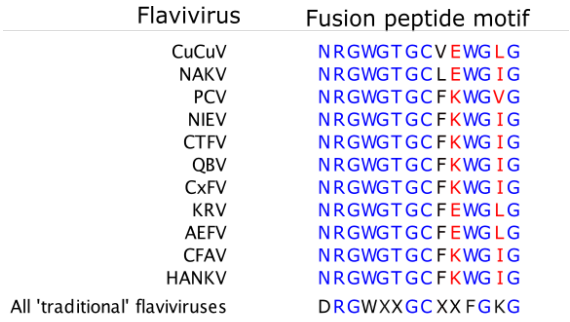

Figure 6. Conserved fusion peptide motifs in CuCuV and other cISFs

Phylogenetic relationship of CuCuV with other flaviviruses

The evolutionary relationships of different ISFs have been analysed with viruses in the genus '*Flavivirus*', and these analyses have shown that cISFs form two distinct clades. The first clade contains viruses associated with *Aedes* spp., while the second clade contains viruses associated with *Culex* spp., in addition to NAKV and PCV, which were isolated from *Mansonia* and *Coquillettidia* (Blitvich & Firth, 2015; Cook *et al.*, 2012). Phylogenetic analysis of the CuCuV NS5 aa sequence showed that it clustered with ISFs and is closely related to NAKV, which is in the cISF group, suggesting that CuCuV is a novel member in the cISF group.

Insect-specific viruses are thought to be maintained in nature primarily by vertical transmission and to a lesser extent by horizontal transmission (by contact). For example, it has been demonstrated that CxFV can transmit from mother to offspring by vertical/transovarial and venereal transmission (from infected males to naïve females) in *Culex* mosquitoes (Bolling *et al.*, 2012; Bolling *et al.*, 2011; Saiyasombat *et al.*, 2011). Vertical transmission was also observed in laboratory-bred *Aedes* mosquitoes that were orally infected with KRV (Lutomiah *et al.*, 2007). Many insect-associated viruses are maintained in nature through vertical transmission, suggesting that these viruses have been associated with insects for a long period of time and that they probably co-evolved and diversified with their insect hosts (Li *et al.*, 2015; Marklewitz *et al.*,

2015). Some of the viral sequences appear to be integrated into the insect host genome (Suzuki *et al.*, 2017b; Fort *et al.*, 2012; Crochu *et al.*, 2004). It has also been hypothesised that some viruses possibly evolved from being ISVs to being dual-host viruses that are able to infect vertebrates, suggesting that the current ISVs may have the potential to evolve as future emerging pathogens.

ISVs have been observed to lack the ability to replicate in mammalian cell lines. Several studies have demonstrated host restriction of ISVs at different points of its life cycle in vertebrate cells. For example, NIEV has been inhibited at the entry-, replication- and assembly levels in BHK-21, Vero, and HEK-293 cells (Junglen *et al.*, 2017). Host-restriction of EILV at both the entry- and replication- levels has been demonstrated in different mammalian cell lines (Nasar *et al.*, 2015b). Studies have also shown that ISVs can modulate the replication of dual-host pathogenic viruses. For example, primary infection of PCV suppressed the replication of secondarily infected viruses, including WNV and Murray Valley Encephalitis virus, in C6/36 cells (Hobson-Peters *et al.*, 2013). It has also been demonstrated that CxFV can reduce WNV replication during the early stages of infection (Bolling *et al.*, 2012), and prior EILV infection of *Ae. agypti* mosquitoes delayed dissemination of CHIKV (Nasar *et al.*, 2015a). These studies suggest that mosquito-specific viruses can interfere with pathogenic arboviruses and influence the vector competence, suggesting that these viruses can potentially be used as biological control agents. The characteristics of ISV host-restriction at multiple levels in mammalian cells can be taken advantage of in vaccine and diagnostic applications. Erasmus *et al.* (2015) showed that a chimeric virus of EILV and CHIKV, which was rescued from C7/10 mosquito cells, could be used as an ELISA antigen to detect CHIKV antibodies from infected human serum samples, eliminating the need for high-level biosafety facilities and inactivation of the antigen (Erasmus *et al.*, 2015). Furthermore, the EILV/CHIKV chimaera virus produced a long lasting and complete protection against CHIKV fever in mouse models and in non-human primates, suggesting that ISVs may be used in vaccine development (Erasmus *et al.*, 2017). Based on these studies, the pathogenicity and host-restriction of CuCuV can be evaluated in vertebrate cells, and further validate its effect on pathogenic arboviruses.

4.1.2 Characterisation of Rhabdoviruses

HTS reads related to the *Rhabdoviridae* family were identified in both the *Culex* (36 reads) and *Mansonia* (89 reads) mosquitoes. The BLAST search of the assembled contigs showed the closest similarity (40-56% at amino acid level) towards the N, G, and L proteins of *Culex tritaeniorhynchus* rhabdovirus

(CTRV), which is an unclassified rhabdovirus isolated from Japanese *Culex tritaeniorhynchus* mosquitoes. The L protein sequence of a novel rhabdovirus identified in *Mansonia* mosquitoes (tentatively named as Mopeia rhabdovirus (MoPRV)) phylogenetically clustered with an unassigned group in the *Rhabdoviridae* family and is closely related to CTRV and Beaumont virus. Rhabdoviruses have a broad range of hosts, including vertebrates, invertebrates and plants. Arthropods such as, mosquitoes, ticks, sand-flies, and midges have all been reported as vectors for rhabdoviruses (Longdon *et al.*, 2015). A number of rhabdoviruses have been discovered in mosquitoes, suggesting that they are common in mosquitoes/arthropods (Shahhosseini *et al.*, 2017; Sun *et al.*, 2017; Reuter *et al.*, 2016), and some of these viruses are insect-specific rhabdoviruses (Kuwata *et al.*, 2011). Rhabdovirus-like sequences have been identified in different arthropod genomes, indicating that these viruses may have been associated with arthropod species for a long period of time (Aiewsakun & Katzourakis, 2015).

4.2 Characterisation of a novel picorna-like virus from *Culex* mosquitoes from Mozambique

In study I, a large number of sequences from the *Culex* mosquitoes were classified as members of the *Iflaviridae* and *Dicistroviridae* families, which belong to the order *Picornavirales*. Viruses from this order have not been well characterised in mosquitoes. Thus, in study II, a near full-length novel picorna-like virus was characterised from the *Culex* mosquito pool, and individual mosquito pools were tested for its presence.

4.2.1 Sequencing the near full-length genome of CuPV-1

The assembled viral contigs from the *Iflaviridae*, *Dicistroviridae* and *Picornaviridae* reads (from Study I) were primarily related to Hubie picorna-like virus 35, with an identity of 37-49% at the amino acid level. The sequence information from these contigs was used to sequence the near-full length genome of a picorna-like virus. Rapid amplification of cDNA ends (RACE) and Sanger sequencing were used to obtain the complete coding region and a partial 5' UTR. The novel virus sequence has been tentatively named *Culex* picorna-like virus 1 (CuPV-1). A single ORF was identified in the 5' to 3' direction, and no longer ORFs were identified in the reverse direction, suggesting that the genome is a single stranded positive sense RNA. The BLASTx search of the ORF sequence showed an amino acid identity of 38% to Hubei picorna-like virus 35, which was previously identified in dragonflies from China (Li *et al.*, 2015). These results

indicate that CuPV-1 is highly divergent from other picorna-like viruses and represents the first discovery of a novel picorna-like virus from Mozambican mosquitoes.

4.2.2 Genome organisation and conserved domains in CuPV-1

The genome organisation of viruses in the *Picornavirales* order is highly variable. For example, iflaviruses and picornaviruses encode a single large polyprotein, while dicistroviruses encodes multiple proteins (Le Gall *et al.*, 2008). The coding region of structural proteins in iflaviruses precedes the coding region for the non-structural proteins, whereas in dicistroviruses, the 5' ORF encodes the non-structural proteins and structural proteins encoded by the second ORF (van Oers, 2010). A multiple sequence alignment of the complete polyprotein of CuPV-1 and other viruses from the *Iflaviridae* family identified both structural and non-structural proteins. The conserved motifs or their equivalents for 1C(VP3), 1D(VP1), helicase, protease and eight RdRP domains were identified. The presence of three replication domain block helicase-protease-RdRP is a hallmark of all viruses in the order *Picornavirales* (Koonin & Dolja, 1993). These domains are thought to be involved in several cellular activities, such as NTP binding (helicase), polyprotein cleavage (protease) and viral genome replication (RdRP) (Le Gall *et al.*, 2008). The presence of these motifs in CuPV-1 confirms that it belongs to the order *Picornavirales*.

The genome organisation of CuPV-1 was similar to iflaviruses by having structural proteins at the 5' side and non-structural proteins at the 3' side of the genome. The structural proteins of iflaviruses may be preceded by a leader peptide (L), as observed in VDV-1 (Ongus *et al.*, 2004), However, no conserved domains for L protein were identified in CuPV-1. The equivalent conserved domains for two capsid proteins (1C(VP3) and 1D(VP1)) were predicted in CuPV-1, although other domains were not predicted, which may be due to a high genomic diversity. Capsid protein domains were previously identified in many insect iflaviruses, including SBV, VDV-1, and DWV, including for mosquito iflavirus (*Armigeres* iflavirus, ArIFV). The equivalent of all three conserved domains in non-structural proteins were predicted in CuPV-1 at the distal part of the polyprotein, confirming that CuPV-1 is a novel virus in the *Iflaviridae* family. Several iflaviruses that infect different insect species have previously been characterised. Most of these viruses do not show any apparent sublethal infection and do not cause severe effects to their hosts, resulting in the number of characterised iflaviruses being proportionately low. However, some of these viruses are lethal to hosts, including SBV and DWV (van Oers, 2010). The number of iflaviruses detected in different insects is increasing, and it has been

predicted that there are many more to be discovered. Mosquitoes have also been shown to harbour iflaviruses, including a novel iflavirus that was recently isolated from *Armigeres* spp. mosquitoes in the Philippines (Kobayashi *et al.*, 2017). The pathogenicity of CuPV-1 has yet to be evaluated by isolating it from C6/36 cells and infecting different mammalian cell lines.

4.2.3 Phylogenetic analysis of CuPV-1

In the order *Picornavirales*, the highest sequence similarity can be observed in non-structural proteins. Therefore, these protein sequences, most often RdRP regions, are used in phylogenetic analyses. A maximum likelihood analysis of the RdRP region from different viruses in the order *Picornavirales* and of CuPV-1 was performed, which placed CuPV-1 together with viruses in the *Iflaviridae* family. Hubei picorna-like virus 34 and 35, and SBV closely grouped with CuPV-1, further suggesting that CuPV-1 is a novel member of the *Iflaviridae* family. It appears that the evolution of iflaviruses is different from their insect hosts. For example, Infectious flacherie virus (IFV), *Perina nuda* virus and Ectropus obliqua picorna-like virus have all been characterised in Lepidopteran species, but they do not share a clade in phylogenetic analyses. Similarly, *Varroa* destructor virus-1, Kakugo virus, and DWV form separate groups from the Slow bee paralysis virus and SBV, although they all infect honey bees (van Oers, 2010). The evolutionary patterns may be resolved by understanding the host range and characterising many more genomes of iflaviruses.

4.2.4 Detection of CuPV-1 in *Culex* and *Mansonia* mosquitoes

Seven mosquito pools out of 23 were tested positive by RT-PCR for CuPV-1, including pools of *Culex* and *Mansonia* species, suggesting that both the species can be infected by CuPV-1. The Sanger sequencing results of the positive PCR products were 97% identical, indicating the presence of different variants (quasispecies) of this virus. Variation in the iflavirus genomes among strains has been previously reported for several viruses. For example, 1% of the variation in RdRP nucleotide sequences between different IFV strains from different geographical locations was observed (van Oers, 2010). Polymorphisms were reported in DWV isolates from individual brood cells of Chinese hives (Forsgren *et al.*, 2009), which supports quasispecies theory of iflaviruses.

In summary, a novel highly divergent picorna-like virus was characterised from *Culex* spp. mosquitoes from Mozambique, and it was further determined that this virus belongs to the *Iflaviridae* family through genome organisation and phylogenetic analyses.

4.3 Characterisation of the virome of *Rhipicephalus* species ticks from Mozambique

Ticks are another important arthropod vector for many pathogenic viruses. Considering their haematophagous nature, ticks have established a close relationship with animals and humans and represent an important risk factor for tick-borne viral disease transmission. Therefore, it is important to investigate the viromes of ticks, especially from under-sampled areas, to understand the viral diversity associated with ticks and to predict future emerging pathogens. The viromes of ticks in the Zambezi Valley of Mozambique has not been studied before. Therefore, in study III, a *Rhipicephalus* tick virome was characterised via viral metagenomics approaches. The tick-sample nucleic acids were pre-amplified by SISPA and rPCR, and HTS was performed with an Ion Torrent PGM. The bioinformatics analysis of the sequences showed that a large number of the sequences (58%) could not be classified by BLAST, while the remaining sequences classified into eukaryota (38.23%), bacteria (3.61%) and viruses (0.09%). The viral reads showed similarity towards 7 different viral families, in which 98% of the reads were classified into *Orthomyxoviridae* and were most closely related to different viruses in the *Quaranjavirus* genus. Metagenomic approaches have been previously used to analyse the viral diversity of ticks and to characterise novel viruses, for example, two novel phenuiviruses were detected in *Dermacentor* spp. ticks from California (Bouquet *et al.*, 2017) and another phenuivirus (Hearthland virus, a potential causative agent of febrile illness) was characterised in *Amblyomma* spp. ticks from Missouri (McMullan *et al.*, 2012). *Rhipicephalus* spp. ticks from China were shown to harbour a broad range of viruses belong to 24 different viral families (Xia *et al.*, 2015). The results of these studies suggest that ticks can be reservoirs for many different viruses.

Assembly of the quaranjaviral reads from the current study generated 6 contigs, and BLASTx searches of these sequences showed an amino acid identity of 32-55% towards segments of different quaranjaviruses (Wellfleet bay virus (WFBV) and Tjuloc virus (TLV)), indicating that these are novel quaranjaviral sequences. No sequences related to the matrix protein were identified in the data set, which may be due to high diversity in the sequences. The genus *Quaranjavirus* consists of two recognised viruses, Quaranfil virus (QRFV) and Johnston Atoll virus (JAV), as well as a tentative member, Lake Chad Virus (LKCV). QRFV and JAV were previously isolated from different tick species (*Argas arboreus* and *Ornithodoros capensis*) in different

geographical regions (Egypt and Johnston Atoll, central Pacific, respectively) (Clifford *et al.*, 1968; Taylor *et al.*, 1966). QRFV was also isolated from febrile children in Egypt (Presti *et al.*, 2009) and was shown to be lethal and cause respiratory disease when inoculated in mice (Baskerville & Lloyd, 1976; Taylor *et al.*, 1966). Furthermore, JAV and LKCV inoculation was pathogenic to newborn mice (Clifford *et al.*, 1968). A novel orthomyxovirus, Cygnet river virus, was isolated from a fatal disease outbreak in captive Australian ducks and showed similarity to other quaranjaviral segments (Kessell *et al.*, 2012). A recently isolated novel quaranjavirus, WFBV, was isolated from dead common eiders in the United States and is suspected to be transmitted by ticks (Allison *et al.*, 2015). Using HTS, Li *et al*. (2015) have identified viral sequences related to quaranjaviruses in a number of different arthropod vectors in China, including spiders, horseflies, true flies and mosquitoes (Li *et al.*, 2015). These studies indicate that quaranjaviruses have broad host ranges and are widely distributed. The individual ticks, from this study, can be further analysed for the prevalence of novel quaranjavirus. Recent studies showed that a novel nairovirus and phlebovirus were detected in more than 54% of ticks that were tested. The observed prevalence is more than those corresponding to tick-borne viruses known to be pathogenic to humans, and it is suspected that these viruses are either commensals or endosymbionts of ticks (Bouquet *et al.*, 2017). Altogether, previous studies have demonstrated that viruses from the *Quaranjavirus* genus can be transmitted by different arthropod vectors and can cause illness in humans and mortality in avian species. Since some of these viruses are pathogenic and have been reported from different countries, human and animal populations may be at risk of exposure to these viruses.

Parvovirus-related sequences have been previously shown to be largely integrated in different invertebrate genomes, including *Rhipicephalus* ticks (Francois *et al.*, 2016). Contigs from *Parvoviridae* reads were also assembled in this study, which showed similarity to different densoviruses, with an amino acid identity of 26-62%. It is possible that these sequences come from the host genome and appeared in the sequencing data due to incomplete nuclease treatment of the sample or from the viruses present in the sample.

In summary, viral metagenomics was used to explore the viral genomic diversity in ticks from Mozambique, and highly divergent viral sequences similar to quaranjaviruses were discovered in this study. Further investigations, including full genome sequencing, cell culture, and serological surveillance studies are warranted to characterise, isolate and determine the pathogenicity of the novel quaranjavirus.

5 Concluding remarks

- Viral metagenomics, using the SISPA method and Ion Torrent sequencing to amplify the nucleic acids and for high-throughput sequencing, has been successfully applied to detect and characterise novel vector-borne viruses in Mozambique.

- *Culex* and *Mansonia* spp. mosquitoes have been shown to harbour a diverse range of RNA viruses, including viruses belonging to the families *Flaviviridae*, *Rhabdoviridae*, *Iflaviridae*, *Dicistroviridae*, *Picornaviridae*, and unclassified RNA viruses, which are all shown to be divergent from previously reported viruses.

- The full-length genome of a novel insect-specific flavivirus (CuCuV) was characterised from *Mansonia* spp. mosquitoes. Potential cleavage sites, -1 PRF sequence and a conserved fusion peptide have been determined. Phylogenetic analysis showed that CuCuV clustered with other ISFs and is closely related to NAKV from Uganda.

- A highly divergent rhabdovirus was characterised in *Mansonia* mosquitoes (MoPRV). The MoPRV sequences showed the highest identity (40-60%) towards the N, G and L proteins of *Culex tritaeniorhynchus* rhabdovirus. Phylogenetic analysis of the L protein of MoPRV placed it in an unassigned group of viruses but within the *Dimarhabdovirus* supergroup.

- The near full-length genome of a highly divergent picorna-like virus (CuPV-1) belonging to the order *Picornavirales* was characterised from *Culex* mosquitoes. The virus sequence showed an amino acid identity of 38% to known viruses, and the conserved protein motifs of helicase-

protease-RNA-dependent RNA polymerase was identified in the polyprotein. CuPV-1 phylogenetically clustered with iflaviruses.

- Screening PCR studies for CuPV-1 indicated that the virus can infect both *Culex* and *Mansonia* mosquitoes.

- Analysis of the adult *Rhipicephalus* tick virome showed the presence of a large number of sequences related to the *Orthomyxoviridae* family, as well as a small number of *Parvoviridae* sequences. Six novel, near-full-length segments related to quaranjaviruses were characterised and were observed to be highly divergent (32-52% aa identity).

6 Future perspectives

- Sampling different mosquito species and screening for insect-specific viruses and picorna-like viruses may be needed to better understand the diversity, host range, transmission patterns and their evolutionary relationships to human or animal pathogenic viruses.

- CuCuV can be recovered from C6/36 cells by reverse genetics, and the host range can be determined by infecting different vertebrate cell lines. Host restriction barriers at different levels in vertebrates can be further investigated with chimeric flaviviruses to validate the possibility of their use in vaccine and diagnostic applications.

- The influence of CuCuV on pathogenic flaviviruses can be experimentally studied both *in vitro* and *in vivo*, and these studies may direct research towards using ISFs as biological control agents.

- Full genome characterisation of the identified quaranjaviral segments and its effect on mammalian/avian cells should be investigated, as previous studies have shown that quaranjaviruses have the capacity to infect avian species and lead to mortality.

- Surveillance of vector-borne viruses in Mozambique should be intensified to predict future emerging pathogens.

- Virus-enrichment methods can be further improved, which will enable the detection of viruses with low titres.

- Bioinformatics tools, such as assemblers and taxonomic classifiers, can be further improved to achieve full genome sequences and to

detect highly divergent viruses. Besides, curation and annotation of complete genomes in the genomic databases will also facilitate accurate detection of novel viruses.

Popular science summary

Viruses are the most abundant organisms on earth. These tiny organisms can cause serious health complications in animal and human populations. Viruses spread from one host to another in different ways, including through the air, water, or by specific vectors. Arthropods, such as mosquitoes and ticks, are the important vectors for different viruses (vector-borne viruses), some of which cause a significant number of deaths every year and affect public health worldwide. Dengue virus, yellow fever virus, chikungunya virus, Japanese encephalitis virus, tick borne encephalitis virus and Zika virus are a few examples of dangerous vector-borne viruses. Several environmental factors such as urbanisation, the movement and growth of animal and human populations, global climate changes, and alterations to ecosystems drive the spread and emergence of new viruses.

Vector-borne viruses belong to different viral families and can infect a broad range of arthropods. The genetic material of vector-borne viruses is primarily RNA and is present in form of different genomic organisations. The rate of genetic modifications in RNA viruses is typically very high. These changes in the genome, apart from the environmental issues, can also influence the spread of viruses to new habitats and hosts and lead to the emergence of novel viruses, which may become a serious threat to public health. Therefore, it is important to track the viruses circulating in arthropod vectors to predict new emerging pathogens and to understand their diversity, host range and evolutionary history.

The detection of vector-borne viruses is a challenging task considering their vast diversity in nature. The choice of detection method is important, as some of the methods can only detect viruses that have been previously well described. Viral metagenomics, a genomic sequencing based method, is a useful tool to identify all the viruses present in a sample, including novel viruses. The workflow of this method includes a combination of laboratory and computational approaches. This thesis describes the detection and characterisation of novel viruses from two arthropod vectors (mosquitoes and

ticks) from the Zambezi valley of Mozambique by viral metagenomics, which is a first study of its kind in Mozambique.

In study I, viral sequences from *Culex* and *Mansonia* species mosquitoes were characterised via viral metagenomics. The sequences that were categorised as viruses belonged to different viral taxa, such as *Flaviviridae*, *Rhabdoviridae*, *Iflaviridae*, *Dicistroviridae, etc*. Furthermore, a complete genome of a novel flavivirus (tentatively named Cuacua virus) was obtained from *Mansonia* mosquitoes. This virus was shown to be highly similarity with a previously identified flavivirus, Nakiwogo virus, which was identified in a Ugandan mosquito population. Evolutionary relationships of this virus with different flaviviruses showed that it belongs to an insect-specific group of flaviviruses that are restricted to mosquitoes. In study II, the near-complete genome of a novel and highly divergent picorna-like virus was characterised from *Culex* mosquitoes. This virus has been tentatively named as *Culex* picorna-like virus 1. The bioinformatic analysis of the protein sequence of this virus showed the presence of helicase-protease-RNA dependent RNA polymerase domains, which are well-preserved domains and are a hallmark of viruses in the order *Picornavirales*. The evolutionary relationships of *Culex* picorna-like virus 1 with other *Picornavirales* viruses in showed that it belongs to *Iflaviridae* family. In study III, highly divergent viral genomic segments belonging to the family Orthomyxoviridae were identified in *Rhipicephalus* species ticks. These segments showed the highest similarity to Wellfleet Bay virus, which was previously identified in dead seabirds from the United States. The conserved polymerase protein sequence of this virus was evolutionarily placed with viruses belonging to the genus *Quaranjavirus* in the family *Orthomyxoviridae*.

Collectively, the results show that mosquitoes and ticks can be infected by a wide range of viruses and the viruses characterised in this study were previously unidentified. Additional genomic studies of different arthropods will provide evidence for the diversity, evolution and transmission of vector-borne viruses.

Populärvetenskaplig sammanfattning

Virus är de vanligaste organismerna på jorden. Dessa små organismer kan orsaka allvarliga sjukdomar hos både djur och människor. Virus sprider sig från en värd till en annan på olika sätt, inklusive via luft, vatten eller med hjälp av specifika vektorer. Arthropoder, som myggor och fästingar, är bland de viktigaste vektorerna för olika virus (vektorburna virus). Några orsakar ett stort antal dödsfall varje år och påverkar därmed folkhälsan över hela världen. Denguevirus, gula febernvirus, chikungunyavirus, japansk encefalitvirus, fästingburen encefalitvirus och Zika-virus är några exempel på allvarliga vektorburna virus. Flera miljöfaktorer som urbanisering, transport av djur, intensifierad djurproduktion, ökad human population, globala klimatförändringar och förändringar i ekosystem driver spridningen och utvecklingen av nya virus.

Vektorburna virus hör till olika virala familjer och kan infektera ett brett spektrum av leddjur. Arvsmassan av vektorburna virus är primärt RNA och ser olika ut med tanke på dess genetiska organisation. Graden av genetiska modifieringar, till exempel mutationer, i RNA-virus är typiskt mycket hög. Dessa förändringar i genomet, förutom miljöproblemen, kan också påverka spridningen av virus till nya miljöer och värdar och på så sätt leda till utveckling av nya virus som kan bli ett allvarligt hot mot folkhälsan. Därför är det viktigt att spåra virus som cirkulerar i artropodvektorer för att kunna förutsäga hur nya patogener uppstår, att förstå deras mångfald, värdspektrum och evolutionära historia.

Detektionen av vektorburna virus är en utmanande uppgift med tanke på deras stora mångfald i naturen. Valet av detekteringsmetod är viktig, eftersom vissa metoder endast kan upptäcka virus som tidigare redan beskrivits. Virus metagenomik, en genomisk sekvensbaserad metod, är ett användbart verktyg för att identifiera alla virus som finns i ett prov, inklusive nya virus. Arbetsflödet för denna metod innefattar en kombination av laboratorie- och dataanalyser. Denna avhandling beskriver upptäckten och karakteriseringen av helt nya virus från två artropodvektorer, myggor och fästingar, från Zambezi-dalen i Moçambique genom virus metagenomik. Detta är den första studien av sitt slag i Moçambique.

I studie I studerades den virala populationen från *Culex* och *Mansonia*-myggarter. Virus sekvenser kategoriserades från respektive art och tillhörde olika virala familjer, såsom *Flaviviridae*, *Rhabdoviridae*, *Iflaviridae*, *Dicistroviridae*, etc. Vidare erhölls ett komplett genom av ett helt nytt flavivirus (kallat Cuacua-virus) från *Mansonia*-myggor. Detta virus visade sig vara mest likt ett tidigare identifierat flavivirus, Nakiwogo-virus, vilket identifierades i en myggpopulation i Uganda. Släktskapsstudier av detta virus visade att det tillhör en insektsspecifik grupp av flavivirus som är bara kan infektera myggor. I studie II bestämdes det fullständiga genomet av ett nytt och mycket unikt picornaliknande virus från *Culex*-myggor. Detta virus har tentativt kallats ”*Culex* picornaliknande virus 1”. Den bioinformatiska analysen av proteinsekvensen för detta virus visade närvaron av helikas-proteas-RNA-beroende RNA-polymerasdomäner, som är välbevarade domäner och är ett typiskt kännetecken för virus i *Picornavirales*. Släktskapet mellan *Culex* picorna-liknande virus 1 och andra virus från *Picornavirales* visade att det tillhör *Iflaviridae*-familjen. I studie III identifierades ett antal unika virala gensegment hos fästingar av *Rhipicephalus*-arten som hör till familjen *Orthomyxoviridae*-virus. Dessa segment visade det största släktskapet med Wellfleet Bay virus, som tidigare identifierades hos döda sjöfåglar från USA. Den konserverade polymerasproteinsekvensen för detta virus var evolutionärt sett mest likt andra virus som tillhör släktet *Quaranjavirus* i familjen *Orthomyxoviridae*.

Sammantaget visar resultaten att myggor och fästingar kan infekteras av ett brett spektrum av olika virus och de virus som upptäcktes och studerades i denna studie var tidigare helt okända. Ytterligare studier av olika arthropoder kommer att ge oss ännu mer kunskap om diversiteten, utvecklingen och överföringen av vektorburna virus.

Acknowledgements

This study was financially supported by the Swedish Research council (VR) (project no. SWE-2012-138) and Swedish Research council for Environment, Agricultural Sciences and Spatial planning, FORMAS (project no. 2012-586), awarded to Assoc. Prof. Anne-Lie Blomström and Prof. Mikael Berg, respectively. The authors would like to thank National Genomics Infrastructure (NGI)/ Uppsala Genomic Center and UPPMAX for the support in high-throughput sequencing and computational resources. The authors would also thank SLU Global Bioinformatics Center, Swedish University of Agricultural Sciences (SLU), Uppsala for providing computational time for the data analysis. The funders had no role in study design, data collection, and analysis, decision to publish, or reprinting of the manuscripts.

I would like to express my sincere gratitude to:

My main supervisor- **Prof. Mikael Berg**, for giving me the opportunity to pursue my doctoral studies under your supervision. It has been a great experience and really enjoyed working with you. Your suggestions and scientific freedom made me to think, learn and express myself in the lab, which provided successful results. The timely scientific meetings and feedback on projects and manuscripts made this work successful. I really thank for all the support and encouragement for making this thesis possible.

My co-supervisor- **Assoc. Prof. Anne-Lie Blomström**, for the opportunity to work with your projects. Your guidance has always been valuable for this project. Your supervision made me to think independently and grow myself professionally and personally. Thank you for all your patience for correcting manuscripts and feedback which has really improved the quality of the manuscripts. Also, thank you for introducing me to viral metagenomics. This work would not have been possible if it wasn't you! I am very fortunate and glad having you as my co-supervisor.

My co-supervisor- **Dr. Juliette Hayer**, I could not have imagined having a better mentor for bioinformatics. Your guidance and support are highly appreciated during my studies. You are a wonderful teacher and I have learnt so much from you. Our long discussions about data analysis were really fruitful and it was fun! Hope we continue this association in future. I also had a great time in Montpellier, especially with grape-stomping ☺ , thank you and **Jack** for all the arrangements.

My co-supervisor- **Dr. Jose Fafetine**, I am very grateful to have you as my supervisor, I would like to thank for all your help in Maputo, driving me to the work place every day, and providing me all the resources for sample collection and the lab work. Without your help I would not have been survived in Mozambique. Hope you can make it for my thesis defense.

My co-supervisor- **Prof. Erik-Bongcam Rudlof**, thank you for all your suggestions during metaviromics meetings which have been always valuable.

Prof. Sándor Belák, it has been my pleasure to know you ever since I started working at SVA and you have always been encouraging. Thanks for the valuable feedback from my half time defense.

Dr. Maja Malmberg, you are an excellent colleague, I have really enjoyed your discussions on different topics, sharing science to life in general and thank you for all your concern and help during my studies.

Assoc. Prof. Fredrik Granberg, thank you for your suggestions during my initial days of sequence analysis, and introducing me to CodonCode Aligner. It was really nice working with you.

Emeli Torsson, you are a superb colleague, friend and a nice office-mate. Thank you for all your discussions and sharing ideas, and for organizing lunch seminars. I should really thank you for the new office, which has brought the light in our lives ☺. I wish you good luck with your thesis.

Dr. Giorgi Metreveli- for all your help and advices, you have been a great colleague and friend.

Pontus Öhlund, thank you for lunch time discussions where we talk many things, include sharing a Noble prize in the future, hope we reach there! I really enjoyed playing ping-pong and hope we continue this. Good luck for your PhD! **Hedviq Stenberg, Hanna Lundén** and **Ylva L**- thank you for your company during lunch and fikas and was really nice working with you, wish you all good luck with your doctoral studies.

Former colleagues at Virology section- **Oskar KL**, I have really enjoyed working with you, had great discussions about science and life. Your valuable advices made my job easy on several occasions. You are a great asset to any institute! **Belisario**, thank you for your guidance, company and time in Mozambique. **Marlene, Xingyu**, it was really nice to working with you.

Fernando Chanisso and **Ana Paula,** for all the help during field work in Mozambique, I have really enjoyed your company. **Vladimero**, thank you for classifying tick samples. **Kerstin Falk** and **Jenny-Verner Carlsson**, for the help in sample shipping and preparation at Public Health Agency of Sweden.

Hadrien G, thank you for your company during lunch and fikas, suggestions in making graphs and of course taking care of planetsmasher! **Magnus Å**, thank you for your lunch time discussions and hope we continue serving Friday fika. **Stefan Örn**, for cheering every time we meet and your kind support as a Director of Postgraduate Education. **Jonas Wensman**, for sharing your scientific knowledge and encouragement!

This study was conducted at the Department of Biomedical Sciences and Veterinary Public Health (BVF), SLU, Uppsala. My sincere thanks to former and current Head of BVF, **Prof. Leif Norrgren** and **Prof. Ivår Vågsholm**, and **Prof. Caroline Fossum**, **Assoc. Prof. Sofia Boqvist** for their encouragement and support throughout my studies. I would also like to thank all the **BVF colleagues** and **fellow PhD students** for their continuous support during my studies.

Colleagues from previous work place SVA- **Claudia B**- You provided me with everything, I am always grateful to you for your help and support. **Alia**, I really thank for your care and concern during THOSE difficult days, you made me feel home. **Misha, Karin U, Fredrik W**, **Jay and Hongyan** thank you all for your company and support during my stay at SVA. **Lihong** and **Siamak**, I am always grateful for your encouragement and company.

Friends in India, have always been supportive and taking care of many things even without my presence. **YOU** are always there for me!

Asst. Prof. Nagarjun Vijay (IISER-Bhopal, India), really thankful for your patience and help to get started with my bioinformatics and solving silly technical issues.

Indian friends in Uppsala, Stockholm and Gothenburg, too many to name and don't want to miss anyone. You have been so kind, caring, helping, encouraging and supportive during my stay in Sweden. You made me feel home, without you all I would not have enjoyed my life in Sweden! I would also like

to thank all the Cricket and badminton friends in Uppsala for their time and made my stay joyful.

I thank my parents **Rama Chary** and **Vijaya sree**, sister and brother in-law-**Mrunalini** and **Santhosh Kumar**, and brother-**Mahendranath** for their endless support and love through this journey, I am truly blessed! To my cousins who scattered across the world and uncles for their concern and encouragement. I owe a debt of gratitude to **Srinivas Maroju**- without your initial support I would not have been landed in Sweden.

I owe my deepest gratitude for my wife **Sharanya**- for your eternal support, love and understanding of my goals, you made my life simple. Last but not the least, to my newly arrived daughter- **Vasikha**, you are an absolute treasure of my life!!.

References

Aiewsakun, P. & Katzourakis, A. (2015). Endogenous viruses: Connecting recent and ancient viral evolution. *Virology,* 479-480, pp. 26-37.

Allander, T., Emerson, S.U., Engle, R.E., Purcell, R.H. & Bukh, J. (2001). A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci U S A,* 98(20), pp. 11609-14.

Allison, A.B., Ballard, J.R., Tesh, R.B., Brown, J.D., Ruder, M.G., Keel, M.K., Munk, B.A., Mickley, R.M., Gibbs, S.E., Travassos da Rosa, A.P., Ellis, J.C., Ip, H.S., Shearn-Bochsler, V.I., Rogers, M.B., Ghedin, E., Holmes, E.C., Parrish, C.R. & Dwyer, C. (2015). Cyclic avian mass mortality in the northeastern United States is associated with a novel orthomyxovirus. *J Virol,* 89(2), pp. 1389-403.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol,* 215(3), pp. 403-10.

Arthropod-borne and rodent-borne viral diseases. Report of a WHO Scientific Group (1985). *World Health Organ Tech Rep Ser,* 719, pp. 1-116.

Auguste, A.J., Kaelber, J.T., Fokam, E.B., Guzman, H., Carrington, C.V., Erasmus, J.H., Kamgang, B., Popov, V.L., Jakana, J., Liu, X., Wood, T.G., Widen, S.G., Vasilakis, N., Tesh, R.B., Chiu, W. & Weaver, S.C. (2015). A newly isolated reovirus has the simplest genomic and structural organization of any reovirus. *J Virol,* 89(1), pp. 676-87.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. & Pevzner, P.A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol,* 19(5), pp. 455-77.

Baskerville, A. & Lloyd, G. (1976). The pathogenesis and pathology of experimental Quaranfil virus infection. *Br J Exp Pathol,* 57(2), pp. 152-6.

Bayat, A. (2002). Science, medicine, and the future: Bioinformatics. *BMJ,* 324(7344), pp. 1018-22.

Berthet, N., Descorps-Declere, S., Nkili-Meyong, A.A., Nakoune, E., Gessain, A., Manuguerra, J.C. & Kazanji, M. (2016). Improved assembly procedure of viral RNA genomes amplified with Phi29 polymerase from new generation sequencing data. *Biol Res,* 49(1), p. 39.

Blitvich, B.J. & Firth, A.E. (2015). Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses,* 7(4), pp. 1927-59.

Blomstrom, A.L. (2011). Viral metagenomics as an emerging and powerful tool in veterinary medicine. *Vet Q,* 31(3), pp. 107-14.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A. & Caporaso, J.G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods,* 10(1), pp. 57-9.

Bolling, B.G., Eisen, L., Moore, C.G. & Blair, C.D. (2011). Insect-specific flaviviruses from Culex mosquitoes in Colorado, with evidence of vertical transmission. *Am J Trop Med Hyg,* 85(1), pp. 169-77.

Bolling, B.G., Olea-Popelka, F.J., Eisen, L., Moore, C.G. & Blair, C.D. (2012). Transmission dynamics of an insect-specific flavivirus in a naturally infected Culex pipiens laboratory colony and effects of co-infection on vector competence for West Nile virus. *Virology,* 427(2), pp. 90-7.
Bolling, B.G., Weaver, S.C., Tesh, R.B. & Vasilakis, N. (2015). Insect-Specific Virus Discovery: Significance for the Arbovirus Community. *Viruses,* 7(9), pp. 4911-28.
Bouquet, J., Melgar, M., Swei, A., Delwart, E., Lane, R.S. & Chiu, C.Y. (2017). Metagenomic-based Surveillance of Pacific Coast tick Dermacentor occidentalis Identifies Two Novel Bunyaviruses and an Emerging Human Ricksettsial Pathogen. *Sci Rep,* 7(1), p. 12234.
Brackney, D.E., Beane, J.E. & Ebel, G.D. (2009). RNAi targeting of West Nile virus in mosquito midguts promotes virus diversification. *PLoS Pathog,* 5(7), p. e1000502.
Brackney, D.E., Schirtzinger, E.E., Harrison, T.D., Ebel, G.D. & Hanley, K.A. (2015). Modulation of flavivirus population diversity by RNA interference. *J Virol,* 89(7), pp. 4035-9.
Brand, S.P. & Keeling, M.J. (2017). The impact of temperature changes on vector-borne disease transmission: Culicoides midges and bluetongue virus. *J R Soc Interface,* 14(128).
Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F. & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A,* 99(22), pp. 14250-5.
Brutscher, L.M., McMenamin, A.J. & Flenniken, M.L. (2016). The Buzz about Honey Bee Viruses. *PLoS Pathog,* 12(8), p. e1005757.
Buchfink, B., Xie, C. & Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods,* 12(1), pp. 59-60.
Burt, F.J., Chen, W., Miner, J.J., Lenschow, D.J., Merits, A., Schnettler, E., Kohl, A., Rudd, P.A., Taylor, A., Herrero, L.J., Zaid, A., Ng, L.F.P. & Mahalingam, S. (2017). Chikungunya virus: an update on the biology and pathogenesis of this emerging pathogen. *Lancet Infect Dis,* 17(4), pp. e107-e117.
Bzhalava, D., Johansson, H., Ekstrom, J., Faust, H., Moller, B., Eklund, C., Nordin, P., Stenquist, B., Paoli, J., Persson, B., Forslund, O. & Dillner, J. (2013). Unbiased approach for virus detection in skin lesions. *PLoS One,* 8(6), p. e65953.
Calisher, C.H., Poland, J.D., Calisher, S.B. & Warmoth, L.A. (1985). Diagnosis of Colorado tick fever virus infection by enzyme immunoassays for immunoglobulin M and G antibodies. *J Clin Microbiol,* 22(1), pp. 84-8.
Carver, S., Bestall, A., Jardine, A. & Ostfeld, R.S. (2009). Influence of hosts on the ecology of arboviral transmission: potential mechanisms influencing dengue, Murray Valley encephalitis, and Ross River virus in Australia. *Vector Borne Zoonotic Dis,* 9(1), pp. 51-64.
Charrel, R.N., Gallian, P., Navarro-Mari, J.M., Nicoletti, L., Papa, A., Sanchez-Seco, M.P., Tenorio, A. & de Lamballerie, X. (2005). Emergence of Toscana virus in Europe. *Emerg Infect Dis,* 11(11), pp. 1657-63.
Chen, Y.P. & Siede, R. (2007). Honey bee viruses. *Adv Virus Res,* 70, pp. 33-80.
Cholleti, H., Hayer, J., Abilio, A.P., Mulandane, F.C., Verner-Carlsson, J., Falk, K.I., Fafetine, J.M., Berg, M. & Blomstrom, A.L. (2016). Discovery of Novel Viruses in Mosquitoes from the Zambezi Valley of Mozambique. *PLoS One,* 11(9), p. e0162751.
Clifford, C.M., Thomas, L.A., Hughes, L.E., Kohls, G.M. & Philip, C.B. (1968). Identification and comparison of two viruses isolated from ticks of the genus Ornithodoros. *Am J Trop Med Hyg,* 17(6), pp. 881-5.
Coffey, L.L., Page, B.L., Greninger, A.L., Herring, B.L., Russell, R.C., Doggett, S.L., Haniotis, J., Wang, C., Deng, X. & Delwart, E.L. (2014). Enhanced arbovirus surveillance with deep sequencing: Identification of novel rhabdoviruses and bunyaviruses in Australian mosquitoes. *Virology,* 448, pp. 146-58.
Cook, S., Moureau, G., Harbach, R.E., Mukwaya, L., Goodger, K., Ssenfuka, F., Gould, E., Holmes, E.C. & de Lamballerie, X. (2009). Isolation of a novel species of flavivirus and a new strain of Culex flavivirus (Flaviviridae) from a natural mosquito population in Uganda. *J Gen Virol,* 90(Pt 11), pp. 2669-78.
Cook, S., Moureau, G., Kitchen, A., Gould, E.A., de Lamballerie, X., Holmes, E.C. & Harbach, R.E. (2012). Molecular evolution of the insect-specific flaviviruses. *J Gen Virol,* 93(Pt 2), pp. 223-34.
Crabtree, M.B., Sang, R.C., Stollar, V., Dunster, L.M. & Miller, B.R. (2003). Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus. *Arch Virol,* 148(6), pp. 1095-118.

Crochu, S., Cook, S., Attoui, H., Charrel, R.N., De Chesse, R., Belhouchet, M., Lemasson, J.J., de Micco, P. & de Lamballerie, X. (2004). Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. *J Gen Virol,* 85(Pt 7), pp. 1971-80.

Davies, C.R., Jones, L.D., Green, B.M. & Nuttall, P.A. (1987). In vivo reassortment of Thogoto virus (a tick-borne influenza-like virus) following oral infection of Rhipicephalus appendiculatus ticks. *J Gen Virol,* 68 (Pt 9), pp. 2331-8.

Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M. & Lasken, R.S. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A,* 99(8), pp. 5261-6.

Delwart, E.L. (2007). Viral metagenomics. *Rev Med Virol,* 17(2), pp. 115-31.

Djikeng, A., Halpin, R., Kuzmickas, R., Depasse, J., Feldblyum, J., Sengamalay, N., Afonso, C., Zhang, X., Anderson, N.G., Ghedin, E. & Spiro, D.J. (2008). Viral genome sequencing by random priming methods. *BMC Genomics,* 9, p. 5.

Dobler, G. (2010). Zoonotic tick-borne flaviviruses. *Vet Microbiol,* 140(3-4), pp. 221-8.

Drake, J.W. & Holland, J.J. (1999). Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A,* 96(24), pp. 13910-3.

Duffy, M.R., Chen, T.H., Hancock, W.T., Powers, A.M., Kool, J.L., Lanciotti, R.S., Pretrick, M., Marfel, M., Holzbauer, S., Dubray, C., Guillaumot, L., Griggs, A., Bel, M., Lambert, A.J., Laven, J., Kosoy, O., Panella, A., Biggerstaff, B.J., Fischer, M. & Hayes, E.B. (2009). Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med,* 360(24), pp. 2536-43.

Edwards, F.W. (1941). Mosquitoes of the Ehiopian Region. III. Culicine adults and pupae. . *British Museum (Natural Histrory), London*.

Erasmus, J.H., Auguste, A.J., Kaelber, J.T., Luo, H., Rossi, S.L., Fenton, K., Leal, G., Kim, D.Y., Chiu, W., Wang, T., Frolov, I., Nasar, F. & Weaver, S.C. (2017). A chikungunya fever vaccine utilizing an insect-specific virus platform. *Nat Med,* 23(2), pp. 192-199.

Erasmus, J.H., Needham, J., Raychaudhuri, S., Diamond, M.S., Beasley, D.W., Morkowski, S., Salje, H., Fernandez Salas, I., Kim, D.Y., Frolov, I., Nasar, F. & Weaver, S.C. (2015). Utilization of an Eilat Virus-Based Chimera for Serological Detection of Chikungunya Infection. *PLoS Negl Trop Dis,* 9(10), p. e0004119.

Ewing, B. & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res,* 8(3), pp. 186-94.

Fauver, J.R., Grubaugh, N.D., Krajacich, B.J., Weger-Lucarelli, J., Lakin, S.M., Fakoli, L.S., 3rd, Bolay, F.K., Diclaro, J.W., 2nd, Dabire, K.R., Foy, B.D., Brackney, D.E., Ebel, G.D. & Stenglein, M.D. (2016). West African Anopheles gambiae mosquitoes harbor a taxonomically diverse virome including new insect-specific flaviviruses, mononegaviruses, and totiviruses. *Virology,* 498, pp. 288-299.

Firth, A.E. & Atkins, J.F. (2009). A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1' may derive from ribosomal frameshifting. *Virol J,* 6, p. 14.

Firth, A.E., Blitvich, B.J., Wills, N.M., Miller, C.L. & Atkins, J.F. (2010). Evidence for ribosomal frameshifting and a novel overlapping gene in the genomes of insect-specific flaviviruses. *Virology,* 399(1), pp. 153-66.

Flint, S.J., Racaniello, V.R., Rall, G.F. & Skalka, A.M. (2015). Principles of Virology. 4th edition. , 1.

Forsgren, E., de Miranda, J.R., Isaksson, M., Wei, S. & Fries, I. (2009). Deformed wing virus associated with Tropilaelaps mercedesae infesting European honey bees (Apis mellifera). *Exp Appl Acarol,* 47(2), pp. 87-97.

Fort, P., Albertini, A., Van-Hua, A., Berthomieu, A., Roche, S., Delsuc, F., Pasteur, N., Capy, P., Gaudin, Y. & Weill, M. (2012). Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. *Mol Biol Evol,* 29(1), pp. 381-90.

Francois, S., Filloux, D., Roumagnac, P., Bigot, D., Gayral, P., Martin, D.P., Froissart, R. & Ogliastro, M. (2016). Discovery of parvovirus-related sequences in an unexpected broad range of animals. *Sci Rep,* 6, p. 30880.

Frey, K.G., Biser, T., Hamilton, T., Santos, C.J., Pimentel, G., Mokashi, V.P. & Bishop-Lilly, K.A. (2016). Bioinformatic Characterization of Mosquito Viromes within the Eastern United States and Puerto Rico: Discovery of Novel Viruses. *Evol Bioinform Online,* 12(Suppl 2), pp. 1-12.

Froussard, P. (1992). A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. *Nucleic Acids Res,* 20(11), p. 2900.
Froussard, P. (1993). rPCR: a powerful tool for random amplification of whole RNA sequences. *PCR Methods Appl,* 2(3), pp. 185-90.
Gish, W. & States, D.J. (1993). Identification of protein coding regions by database similarity search. *Nat Genet,* 3(3), pp. 266-72.
Goodwin, S., McPherson, J.D. & McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet,* 17(6), pp. 333-51.
Gratz, N.G. (1999). Emerging and resurging vector-borne diseases. *Annu Rev Entomol,* 44, pp. 51-75.
Gray, S.M. & Banerjee, N. (1999). Mechanisms of arthropod transmission of plant and animal viruses. *Microbiol Mol Biol Rev,* 63(1), pp. 128-48.
Grobbelaar, A.A., Weyer, J., Leman, P.A., Kemp, A., Paweska, J.T. & Swanepoel, R. (2011). Molecular epidemiology of Rift Valley fever virus. *Emerg Infect Dis,* 17(12), pp. 2270-6.
Gubler, D.J. (2001). Human arbovirus infections worldwide. *Ann N Y Acad Sci,* 951, pp. 13-24.
Gubler, D.J. (2002). The global emergence/resurgence of arboviral diseases as public health problems. *Arch Med Res,* 33(4), pp. 330-42.
Gubler, D.J., Kuno, G., Sather, G.E., Velez, M. & Oliver, A. (1984). Mosquito cell cultures and specific monoclonal antibodies in surveillance for dengue viruses. *Am J Trop Med Hyg,* 33(1), pp. 158-65.
Hahn, C.S., Lustig, S., Strauss, E.G. & Strauss, J.H. (1988). Western equine encephalitis virus is a recombinant virus. *Proc Natl Acad Sci U S A,* 85(16), pp. 5997-6001.
Hall, R.J., Wang, J., Todd, A.K., Bissielo, A.B., Yen, S., Strydom, H., Moore, N.E., Ren, X., Huang, Q.S., Carter, P.E. & Peacey, M. (2014). Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J Virol Methods,* 195, pp. 194-204.
He, C.Q., Ding, N.Z., He, M., Li, S.N., Wang, X.M., He, H.B., Liu, X.F. & Guo, H.S. (2010). Intragenic recombination as a mechanism of genetic diversity in bluetongue virus. *J Virol,* 84(21), pp. 11487-95.
Hobson-Peters, J., Warrilow, D., McLean, B.J., Watterson, D., Colmant, A.M., van den Hurk, A.F., Hall-Mendelin, S., Hastie, M.L., Gorman, J.J., Harrison, J.J., Prow, N.A., Barnard, R.T., Allcock, R., Johansen, C.A. & Hall, R.A. (2016). Discovery and characterisation of a new insect-specific bunyavirus from Culex mosquitoes captured in northern Australia. *Virology,* 489, pp. 269-81.
Hobson-Peters, J., Yam, A.W., Lu, J.W., Setoh, Y.X., May, F.J., Kurucz, N., Walsh, S., Prow, N.A., Davis, S.S., Weir, R., Melville, L., Hunt, N., Webb, R.I., Blitvich, B.J., Whelan, P. & Hall, R.A. (2013). A new insect-specific flavivirus from northern Australia suppresses replication of West Nile virus and Murray Valley encephalitis virus in co-infected mosquito cells. *PLoS One,* 8(2), p. e56534.
Hoogstraal, H. (1979). The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe, and Africa. *J Med Entomol,* 15(4), pp. 307-417.
Hopkins, G.H.E. (1952). Mosquitoes of Ethiopian Region, vol. I: Larval Bionomics of Mosquitoes and Taxonomy of Culicine Larvae *British Museum (Natural Histrory), London*.
Innis, B.L., Nisalak, A., Nimmannitya, S., Kusalerdchariya, S., Chongswasdi, V., Suntayakorn, S., Puttisri, P. & Hoke, C.H. (1989). An enzyme-linked immunosorbent assay to characterize dengue infections where dengue and Japanese encephalitis co-circulate. *Am J Trop Med Hyg,* 40(4), pp. 418-27.
Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L. & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature,* 451(7181), pp. 990-3.
Junglen, S., Korries, M., Grasse, W., Wieseler, J., Kopp, A., Hermanns, K., Leon-Juarez, M., Drosten, C. & Kummerer, B.M. (2017). Host Range Restriction of Insect-Specific Flaviviruses Occurs at Several Levels of the Viral Life Cycle. *mSphere,* 2(1).
Karlsson, O.E., Belak, S. & Granberg, F. (2013). The effect of preprocessing by sequence-independent, single-primer amplification (SISPA) on metagenomic detection of viruses. *Biosecur Bioterror,* 11 Suppl 1, pp. S227-34.
Kenney, J.L., Solberg, O.D., Langevin, S.A. & Brault, A.C. (2014). Characterization of a novel insect-specific flavivirus from Brazil: potential for inhibition of infection of arthropod cells with medically important flaviviruses. *J Gen Virol,* 95(Pt 12), pp. 2796-808.
Kessell, A., Hyatt, A., Lehmann, D., Shan, S., Crameri, S., Holmes, C., Marsh, G., Williams, C., Tachedjian, M., Yu, M., Bingham, J., Payne, J., Lowther, S., Wang, J., Wang, L.F. & Smith,

I. (2012). Cygnet River virus, a novel orthomyxovirus from ducks, Australia. *Emerg Infect Dis,* 18(12), pp. 2044-6.
Kobayashi, D., Isawa, H., Fujita, R., Murota, K., Itokawa, K., Higa, Y., Katayama, Y., Sasaki, T., Mizutani, T., Iwanaga, S., Ohta, N., Garcia-Bertuso, A. & Sawabe, K. (2017). Isolation and characterization of a new iflavirus from Armigeres spp. mosquitoes in the Philippines. *J Gen Virol,* 98(11), pp. 2876-2881.
Koonin, E.V. & Dolja, V.V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol,* 28(5), pp. 375-430.
Kurkela, S., Ratti, O., Huhtamo, E., Uzcategui, N.Y., Nuorti, J.P., Laakkonen, J., Manni, T., Helle, P., Vaheri, A. & Vapalahti, O. (2008). Sindbis virus infection in resident birds, migratory birds, and humans, Finland. *Emerg Infect Dis,* 14(1), pp. 41-7.
Kuwata, R., Isawa, H., Hoshino, K., Sasaki, T., Kobayashi, M., Maeda, K. & Sawabe, K. (2015). Analysis of Mosquito-Borne Flavivirus Superinfection in Culex tritaeniorhynchus (Diptera: Culicidae) Cells Persistently Infected with Culex Flavivirus (Flaviviridae). *J Med Entomol,* 52(2), pp. 222-9.
Kuwata, R., Isawa, H., Hoshino, K., Tsuda, Y., Yanase, T., Sasaki, T., Kobayashi, M. & Sawabe, K. (2011). RNA splicing in a new rhabdovirus from Culex mosquitoes. *J Virol,* 85(13), pp. 6185-96.
Lambrechts, L. & Scott, T.W. (2009). Mode of transmission and the evolution of arbovirus virulence in mosquito vectors. *Proc Biol Sci,* 276(1660), pp. 1369-78.
Lanciotti, R.S. & Kerst, A.J. (2001). Nucleic acid sequence-based amplification assays for rapid detection of West Nile and St. Louis encephalitis viruses. *J Clin Microbiol,* 39(12), pp. 4506-13.
Lanciotti, R.S., Kerst, A.J., Nasci, R.S., Godsey, M.S., Mitchell, C.J., Savage, H.M., Komar, N., Panella, N.A., Allen, B.C., Volpe, K.E., Davis, B.S. & Roehrig, J.T. (2000). Rapid detection of west nile virus from human clinical specimens, field-collected mosquitoes, and avian samples by a TaqMan reverse transcriptase-PCR assay. *J Clin Microbiol,* 38(11), pp. 4066-71.
Lanciotti, R.S., Roehrig, J.T., Deubel, V., Smith, J., Parker, M., Steele, K., Crise, B., Volpe, K.E., Crabtree, M.B., Scherret, J.H., Hall, R.A., MacKenzie, J.S., Cropp, C.B., Panigrahy, B., Ostlund, E., Schmitt, B., Malkinson, M., Banet, C., Weissman, J., Komar, N., Savage, H.M., Stone, W., McNamara, T. & Gubler, D.J. (1999). Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science,* 286(5448), pp. 2333-7.
Langmead, B. & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9(4), pp. 357-9.
Le Gall, O., Christian, P., Fauquet, C.M., King, A.M., Knowles, N.J., Nakashima, N., Stanway, G. & Gorbalenya, A.E. (2008). Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T = 3 virion architecture. *Arch Virol,* 153(4), pp. 715-27.
Leland, D.S. & Ginocchio, C.C. (2007). Role of cell culture for virus detection in the age of technology. *Clin Microbiol Rev,* 20(1), pp. 49-78.
Li, C.X., Shi, M., Tian, J.H., Lin, X.D., Kang, Y.J., Chen, L.J., Qin, X.C., Xu, J., Holmes, E.C. & Zhang, Y.Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife,* 4.
Li, H. & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics,* 26(5), pp. 589-95.
Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K. & Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics,* 25(15), pp. 1966-7.
Liu, S., Vijayendran, D. & Bonning, B.C. (2011). Next generation sequencing technologies for insect virus discovery. *Viruses,* 3(10), pp. 1849-69.
Longdon, B., Murray, G.G., Palmer, W.J., Day, J.P., Parker, D.J., Welch, J.J., Obbard, D.J. & Jiggins, F.M. (2015). The evolution, diversity, and host associations of rhabdoviruses. *Virus Evol,* 1(1), p. vev014.
Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W. & Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience,* 1(1), p. 18.

Lutomiah, J.J., Mwandawiro, C., Magambo, J. & Sang, R.C. (2007). Infection and vertical transmission of Kamiti river virus in laboratory bred Aedes aegypti mosquitoes. *J Insect Sci,* 7, pp. 1-7.
Maan, S., Maan, N.S., Guimera, M., Nomikou, K., Singh, K.P., Pullinger, G., Belaganahalli, M.N. & Mertens, P.P. (2012). Genome sequence of a reassortant strain of bluetongue virus serotype 23 from western India. *J Virol,* 86(12), pp. 7011-2.
Mansfield, K.L., Jizhou, L., Phipps, L.P. & Johnson, N. (2017). Emerging Tick-Borne Viruses in the Twenty-First Century. *Front Cell Infect Microbiol,* 7, p. 298.
Marklewitz, M., Zirkel, F., Kurth, A., Drosten, C. & Junglen, S. (2015). Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. *Proc Natl Acad Sci U S A,* 112(24), pp. 7536-41.
Markoff, L. (2003). 5'- and 3'-noncoding regions in flavivirus RNA. *Adv Virus Res,* 59, pp. 177-228.
McGinnis, S. & Madden, T.L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res,* 32(Web Server issue), pp. W20-5.
McMullan, L.K., Folk, S.M., Kelly, A.J., MacNeil, A., Goldsmith, C.S., Metcalfe, M.G., Batten, B.C., Albarino, C.G., Zaki, S.R., Rollin, P.E., Nicholson, W.L. & Nichol, S.T. (2012). A new phlebovirus associated with severe febrile illness in Missouri. *N Engl J Med,* 367(9), pp. 834-41.
Melian, E.B., Hinzman, E., Nagasaki, T., Firth, A.E., Wills, N.M., Nouwens, A.S., Blitvich, B.J., Leung, J., Funk, A., Atkins, J.F., Hall, R. & Khromykh, A.A. (2010). NS1' of flaviviruses in the Japanese encephalitis virus serogroup is a product of ribosomal frameshifting and plays a role in viral neuroinvasiveness. *J Virol,* 84(3), pp. 1641-7.
Morens, D.M., Folkers, G.K. & Fauci, A.S. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature,* 430(6996), pp. 242-9.
Murray, K.O., Walker, C. & Gould, E. (2011). The virology, epidemiology, and clinical impact of West Nile virus: a decade of advancements in research since its introduction into the Western Hemisphere. *Epidemiol Infect,* 139(6), pp. 807-17.
Musso, D., Nilles, E.J. & Cao-Lormeau, V.M. (2014). Rapid spread of emerging Zika virus in the Pacific area. *Clin Microbiol Infect,* 20(10), pp. O595-6.
Myrmel, M., Oma, V., Khatri, M., Hansen, H.H., Stokstad, M., Berg, M. & Blomstrom, A.L. (2017). Single primer isothermal amplification (SPIA) combined with next generation sequencing provides complete bovine coronavirus genome coverage and higher sequence depth compared to sequence-independent single primer amplification (SISPA). *PLoS One,* 12(11), p. e0187780.
Nasar, F., Erasmus, J.H., Haddow, A.D., Tesh, R.B. & Weaver, S.C. (2015a). Eilat virus induces both homologous and heterologous interference. *Virology,* 484, pp. 51-8.
Nasar, F., Gorchakov, R.V., Tesh, R.B. & Weaver, S.C. (2015b). Eilat virus host range restriction is present at multiple levels of the virus life cycle. *J Virol,* 89(2), pp. 1404-18.
Nasar, F., Palacios, G., Gorchakov, R.V., Guzman, H., Da Rosa, A.P., Savji, N., Popov, V.L., Sherman, M.B., Lipkin, W.I., Tesh, R.B. & Weaver, S.C. (2012). Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. *Proc Natl Acad Sci U S A,* 109(36), pp. 14622-7.
Ongus, J.R., Peters, D., Bonmatin, J.M., Bengsch, E., Vlak, J.M. & van Oers, M.M. (2004). Complete sequence of a picorna-like virus of the genus Iflavirus replicating in the mite Varroa destructor. *J Gen Virol,* 85(Pt 12), pp. 3747-55.
Patel, R.K. & Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One,* 7(2), p. e30619.
Pepin, M., Bouloy, M., Bird, B.H., Kemp, A. & Paweska, J. (2010). Rift Valley fever virus(Bunyaviridae: Phlebovirus): an update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Vet Res,* 41(6), p. 61.
Pettersson, J.H., Shi, M., Bohlin, J., Eldholm, V., Brynildsrud, O.B., Paulsen, K.M., Andreassen, A. & Holmes, E.C. (2017). Characterizing the virome of Ixodes ricinus ticks from northern Europe. *Sci Rep,* 7(1), p. 10870.
Pfeffer, M. & Dobler, G. (2010). Emergence of zoonotic arboviruses by animal trade and migration. *Parasit Vectors,* 3(1), p. 35.
Presti, R.M., Zhao, G., Beatty, W.L., Mihindukulasuriya, K.A., da Rosa, A.P., Popov, V.L., Tesh, R.B., Virgin, H.W. & Wang, D. (2009). Quaranfil, Johnston Atoll, and Lake Chad viruses are novel members of the family Orthomyxoviridae. *J Virol,* 83(22), pp. 11599-606.
Quan, P.L., Junglen, S., Tashmukhamedova, A., Conlan, S., Hutchison, S.K., Kurth, A., Ellerbrok, H., Egholm, M., Briese, T., Leendertz, F.H. & Lipkin, W.I. (2010). Moussa virus: a new

member of the Rhabdoviridae family isolated from Culex decens mosquitoes in Cote d'Ivoire. *Virus Res,* 147(1), pp. 17-24.
Reuter, G., Boros, A., Pal, J., Kapusinszky, B., Delwart, E. & Pankovics, P. (2016). Detection and genome analysis of a novel (dima)rhabdovirus (Riverside virus) from Ochlerotatus sp. mosquitoes in Central Europe. *Infect Genet Evol,* 39, pp. 336-341.
Reyes, G.R. & Kim, J.P. (1991). Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol Cell Probes,* 5(6), pp. 473-81.
Rosseel, T., Ozhelvaci, O., Freimanis, G. & Van Borm, S. (2015). Evaluation of convenient pretreatment protocols for RNA virus metagenomics in serum and tissue samples. *J Virol Methods,* 222, pp. 72-80.
Rosseel, T., Pardon, B., De Clercq, K., Ozhelvaci, O. & Van Borm, S. (2014). False-positive results in metagenomic virus discovery: a strong case for follow-up diagnosis. *Transbound Emerg Dis,* 61(4), pp. 293-9.
Ryan, S.J., Carlson, C.J., Stewart-Ibarra, A.M., Borbor-Cordova, M.J., Romero, M.M., Cox, S.A., Mahon, R., Trotman, A., St Ville, S. & Ahmed, S. (2017). Outbreak of Zika Virus Infections, Dominica, 2016. *Emerg Infect Dis,* 23(11), pp. 1926-1927.
Saiyasombat, R., Bolling, B.G., Brault, A.C., Bartholomay, L.C. & Blitvich, B.J. (2011). Evidence of efficient transovarial transmission of Culex flavivirus by Culex pipiens (Diptera: Culicidae). *J Med Entomol,* 48(5), pp. 1031-8.
Shafiq, M., Minakshi, P., Bhateja, A., Ranjan, K. & Prasad, G. (2013). Evidence of genetic reassortment between Indian isolate of bluetongue virus serotype 21 (BTV-21) and bluetongue virus serotype 16 (BTV-16). *Virus Res,* 173(2), pp. 336-43.
Shahhosseini, N., Luhken, R., Jost, H., Jansen, S., Borstler, J., Rieger, T., Kruger, A., Yadouleton, A., de Mendonca Campos, R., Cirne-Santos, C.C., Ferreira, D.F., Garms, R., Becker, N., Tannich, E., Cadar, D. & Schmidt-Chanasit, J. (2017). Detection and characterization of a novel rhabdovirus in Aedes cantans mosquitoes and evidence for a mosquito-associated new genus in the family Rhabdoviridae. *Infect Genet Evol,* 55, pp. 260-268.
Shi, C., Liu, Y., Hu, X., Xiong, J., Zhang, B. & Yuan, Z. (2015). A metagenomic survey of viral abundance and diversity in mosquitoes from Hubei province. *PLoS One,* 10(6), p. e0129845.
Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., Qin, X.C., Li, J., Cao, J.P., Eden, J.S., Buchmann, J., Wang, W., Xu, J., Holmes, E.C. & Zhang, Y.Z. (2016). Redefining the invertebrate RNA virosphere. *Nature*.
Simon, F., Savini, H. & Parola, P. (2008). Chikungunya: a paradigm of emergence and globalization of vector-borne diseases. *Med Clin North Am,* 92(6), pp. 1323-43, ix.
Smith, K.F., Goldberg, M., Rosenthal, S., Carlson, L., Chen, J., Chen, C. & Ramachandran, S. (2014). Global rise in human infectious disease outbreaks. *J R Soc Interface,* 11(101), p. 20140950.
Stollar, V. & Thomas, V.L. (1975). An agent in the Aedes aegypti cell line (Peleg) which causes fusion of Aedes albopictus cells. *Virology,* 64(2), pp. 367-77.
Sun, Q., Zhao, Q., An, X., Guo, X., Zuo, S., Zhang, X., Pei, G., Liu, W., Cheng, S., Wang, Y., Shu, P., Mi, Z., Huang, Y., Zhang, Z., Tong, Y., Zhou, H. & Zhang, J. (2017). Complete genome sequence of Menghai rhabdovirus, a novel mosquito-borne rhabdovirus from China. *Arch Virol,* 162(4), pp. 1103-1106.
Suttle, C. (2005). Crystal ball. The viriosphere: the greatest biological diversity on Earth and driver of global processes. *Environ Microbiol,* 7(4), pp. 481-2.
Suzuki, S., Ishida, T., Ohue, M., Kakuta, M. & Akiyama, Y. (2017a). GHOSTX: A Fast Sequence Homology Search Tool for Functional Annotation of Metagenomic Data. *Methods Mol Biol,* 1611, pp. 15-25.
Suzuki, Y., Frangeul, L., Dickson, L.B., Blanc, H., Verdier, Y., Vinh, J., Lambrechts, L. & Saleh, M.C. (2017b). Uncovering the Repertoire of Endogenous Flaviviral Elements in Aedes Mosquito Genomes. *J Virol,* 91(15).
Tanaka, M. (1993). Rapid identification of flavivirus using the polymerase chain reaction. *J Virol Methods,* 41(3), pp. 311-22.
Tantely, L.M., Boyer, S. & Fontenille, D. (2015). A review of mosquitoes associated with Rift Valley fever virus in Madagascar. *Am J Trop Med Hyg,* 92(4), pp. 722-9.
Taylor, R.M., Hurlbut, H.S., Work, T.H., Kingston, J.R. & Hoogstraal, H. (1966). Arboviruses isolated from ARGAS TICKS IN Egypt: Quaranfil, Chenuda, and Nyamanini. *Am J Trop Med Hyg,* 15(1), pp. 76-86.

Tsetsarkin, K.A., Vanlandingham, D.L., McGee, C.E. & Higgs, S. (2007). A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog,* 3(12), p. e201.

Tsetsarkin, K.A. & Weaver, S.C. (2011). Sequential adaptive mutations enhance efficient vector switching by Chikungunya virus and its epidemic emergence. *PLoS Pathog,* 7(12), p. e1002412.

van der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J. & Berkhout, B. (2004). Identification of a new human coronavirus. *Nat Med,* 10(4), pp. 368-73.

van Oers, M. (2010). Genomics and biology of Iflaviruses. *Insect Virology, Norfolk: Caister Academic Press*, pp. 231-250.

Vasilakis, N. & Tesh, R.B. (2015). Insect-specific viruses and their potential impact on arbovirus transmission. *Curr Opin Virol,* 15, pp. 69-74.

Walker, A.R., Bounattour, A., Camicas, J.L., Estrada-Pena, A., Horak, I.G., Latif, A., Pegram, R.G. & Preston, P.M. (2003). Ticks of domestic animals in Africa: Aguide to identification of species. . *Biosience Reports*.

Weaver, S.C. & Barrett, A.D. (2004). Transmission cycles, host range, evolution and emergence of arboviral disease. *Nat Rev Microbiol,* 2(10), pp. 789-801.

Weaver, S.C., Kang, W., Shirako, Y., Rumenapf, T., Strauss, E.G. & Strauss, J.H. (1997). Recombinational history and molecular evolution of western equine encephalomyelitis complex alphaviruses. *J Virol,* 71(1), pp. 613-23.

Weaver, S.C. & Reisen, W.K. (2010). Present and future arboviral threats. *Antiviral Res,* 85(2), pp. 328-45.

WHO (2014). A global brief on vector-borne diseases. *Geneva: World Health Organization*.

Xia, H., Hu, C., Zhang, D., Tang, S., Zhang, Z., Kou, Z., Fan, Z., Bente, D., Zeng, C. & Li, T. (2015). Metagenomic profile of the viral communities in Rhipicephalus spp. ticks from Yunnan, China. *PLoS One,* 10(3), p. e0121609.

Yanase, T., Kato, T., Aizawa, M., Shuto, Y., Shirafuji, H., Yamakawa, M. & Tsuda, T. (2012). Genetic reassortment between Sathuperi and Shamonda viruses of the genus Orthobunyavirus in nature: implications for their genetic relationship to Schmallenberg virus. *Arch Virol,* 157(8), pp. 1611-6.

Yang, X., Charlebois, P., Gnerre, S., Coole, M.G., Lennon, N.J., Levin, J.Z., Qu, J., Ryan, E.M., Zody, M.C. & Henn, M.R. (2012). De novo assembly of highly diverse viral populations. *BMC Genomics,* 13, p. 475.

Zerbino, D.R. (2010). Using the Velvet de novo assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics,* Chapter 11, p. Unit 11 5.

Zhao, Y., Tang, H. & Ye, Y. (2012). RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics,* 28(1), pp. 125-6.