Discovery and sequencing of novel and identified mosquito-associated viruses and genetic determinants of flavivirus host specificity

Jermilia Charles

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Molecular Biology Commons, and the Virology Commons

Recommended Citation
Charles, Jermilia, "Discovery and sequencing of novel and identified mosquito-associated viruses and genetic determinants of flavivirus host specificity" (2019). Graduate Theses and Dissertations. 17421. https://lib.dr.iastate.edu/etd/17421

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Discovery and sequencing of novel and identified mosquito-associated viruses and genetic determinants of flavivirus host specificity

by

Jermilia Charles

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Molecular, Cellular, Developmental Biology

Program of Study Committee:
Bradley Blitvich, Major Professor
   Cathy Miller
   Brett Sponseller
   Phillip Gauger
   Steven Whitham

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

Copyright © Jermilia Charles, 2019. All rights reserved.
DEDICATION

I dedicate this dissertation to my mentor, Dr. Bradley Blitvich, without whom none of this would have been possible. Thank you for your guidance and encouragement throughout my studies. I also dedicate it to my family at home and in Ames.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>CHAPTER 1. GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2. INVESTIGATION OF THE GENETIC DETERMINANTS THAT MODULATE HOST-RESTRICTION OF BAT-ASSOCIATED, NO KNOWN VECTOR FLAVIVIRUSES</td>
<td>37</td>
</tr>
<tr>
<td>CHAPTER 3. DETECTION OF NOVEL AND RECOGNIZED RNA VIRUSES IN MOSQUITOES FROM THE YUCATAN PENINSULA OF MEXICO USING METAGENOMICS AND CHARACTERIZATION OF THEIR IN VITRO HOST RANGES</td>
<td>73</td>
</tr>
<tr>
<td>CHAPTER 4. DISCOVERY OF A NOVEL <em>TYMOVIRIDAE</em>-LIKE VIRUS IN MOSQUITOES FROM MEXICO</td>
<td>103</td>
</tr>
<tr>
<td>CHAPTER 5. COMPLETE GENOME SEQUENCES OF TWO INSECT-SPECIFIC FLAVIVIRUSES</td>
<td>111</td>
</tr>
<tr>
<td>CHAPTER 6. COMPLETE NUCLEOTIDE SEQUENCES OF THE LARGE RNA GENOME SEGMENTS OF MAIN DRAIN AND NORTHWAY VIRUSES (FAMILY <em>PERIBUNYAVIRIDAE</em>)</td>
<td>121</td>
</tr>
<tr>
<td>CHAPTER 7. MERIDA VIRUS, A PUTATIVE NOVEL RHABDOVIRUS DISCOVERED IN <em>CULEX</em> AND <em>OCHLEROTATUS</em> SPP. MOSQUITOES IN THE YUCATAN PENINSULA OF MEXICO</td>
<td>128</td>
</tr>
<tr>
<td>CHAPTER 8. GENERAL CONCLUSIONS</td>
<td>159</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to thank my major professor, Dr. Bradley Blitvich, and my committee members, Dr. Cathy Miller, Dr. Brett Sponseller, Dr. Phillip Gauger, and Dr. Steven Whitham, for their guidance and support throughout the course of this research.

In addition, I would also like to thank my biological and church family, friends and colleagues for your love and support through the years. I want to also offer my appreciation to Vern and Tammy who were always very kind and always willing to help. Last, but not least, God, without whom none of this would be possible.
ABSTRACT

The advent of globalization and the continuing effects of climate change have exposed an even greater percentage of the world to disease agents carried by vectors which were once thought to be restricted to tropical or subtropical areas. Arthropods vectors such as mosquitoes and ticks harbor a number of emerging and re-emerging pathogens of public and veterinary health importance worldwide. These pathogens include viruses from four major families including the Flaviviridae, Bunyaviridae, Togaviridae and Reoviridae families. The Flavivirus genus is one of four genera in the Flaviviridae family, with over 70 species, the great majority of which are maintained in horizontal transmission cycles between hematophagous arthropod vectors and vertebrate hosts. However, within this genus also exists two smaller subsets of viruses which are restricted to either insects (insect-specific flaviviruses) or vertebrates (no known vector flaviviruses). Since it is likely that dual-host flaviviruses may have evolved from single-host precursors, investigative studies into the genetic elements which modulate flavivirus host specificity may elucidate the evolutionary hurdles necessary to make the leap from a single- to a dual-host flavivirus. Such studies could have major implications for efficacious vaccine or antiviral drug development.

In this dissertation, I report the creation and characterization of the first bat-associated flavivirus chimeras, designed to investigate host restriction in no known vector (NKV) flaviviruses at the level of attachment and entry. These chimeras were created using two representative dual-host flaviviruses, yellow fever virus (YFV) and Zika virus (ZIKV) whose attachment and entry proteins (premembrane and envelope) were substituted with that of a bat-associated NKV flavivirus, Rio Bravo virus (RBV). The result was the generation of two different chimeric viruses, both of which were able to replicate in both insect and vertebrate cell
lines, indicating that NKV flavivirus restriction is dictated by a post-attachment/entry event. Additional studies examining restriction at downstream stages in the viral life cycle may prove valuable in determining the genetic elements that modulate host restriction in this group.

In the chapters that follow, we also report the discovery and identification of four novel and two previously identified viruses from mosquitoes collected in the Yucatan Peninsula of Mexico using metagenomics. We also report full sequence data for two insect-specific flaviviruses (ISFs) and for the large segment of two bunyaviruses for which full sequence data was not available.
CHAPTER 1. GENERAL INTRODUCTION

Introduction

The *Flavivirus* genus comprises over 70 different species of viruses, the greater majority of which are maintained in a vertebrate/arthropod vector transmission cycle. A number of these viruses remain a cause for public health and veterinary concern, particularly in tropical and subtropical areas. Some of the most medically important flaviviruses are dengue virus (DENV), yellow fever virus (YFV), Japanese encephalitis virus (JEV), Zika virus (ZIKV) and West Nile virus (WNV). With 390 million new infections per year, DENV is currently endemic in over 125 countries with an estimated 50-60% percent of the world’s population expected to be at risk by 2085 with the current rate of climate change (Murray, Quam et al. 2013). The 2015 ZIKV outbreak in the Americas left half a million to 1.5 million people infected with the virus and resulted in a 20-fold increase in the number of microcephaly cases (Karkhah, Nouri et al. 2018). In 2018, the CDC reported over 2500 West Nile virus infections in 49 of the 50 states with over half of them being classified as neuroinvasive.

The aforementioned viruses all exhibit a dual-host tropism and represent one of three subgroups within the *Flavivirus* genus. The other two subgroups are more restricted in their host ranges and infect either insects or vertebrates (Ochsenreiter, Hofacker et al. 2019). Studies into the genetic elements which govern these differential host ranges in flaviviruses could have important implications for vaccine design and antiviral drug development and could elucidate the evolutionary hurdles that were necessary for viruses to make the leap from single to dual-host viruses and several such studies have been attempted (Charlier, Davidson et al. 2010, Tumban, Maes et al. 2013, Saiyasombat, Carrillo-Tripp et al. 2014, Junglen, Korries et al. 2017, Piyasena, Setoh et al. 2017). In chapter two of this dissertation, I report the creation and characterization of
two chimeras generated between two representative dual-host flaviviruses, ZIKV and YFV, and one bat-associated vertebrate-specific flavivirus, Rio Bravo virus (RBV), designed to assess vertebrate-specific flavivirus host restriction at the level of attachment and entry. These chimeras represent the first chimeras generated between a dual-host flavivirus and a bat-associated vertebrate-specific flavivirus.

The advent of metagenomics has simplified the identification of novel viruses making it faster and easier to identify the components of an organism’s virome or even to sequence pre-existing viruses for which complete sequence data had previously been unavailable. In chapters three to seven of this dissertation, I not only report the discovery and isolation of four novel viruses identified in mosquitoes collected in the Yucatan Peninsula of Mexico, but I also report full sequence data for two insect-specific flaviviruses and the large segment of two bunyaviruses for which full sequence data had currently been unavailable.

**Dissertation Organization**

This dissertation is divided into eight different chapters. The first chapter provides a basic background on major virus taxa relevant to what is later presented in the subsequent chapters. The second chapter is a manuscript which will be submitted for publication. The third through seventh chapters are all manuscripts that have been published in different peer-reviewed scientific journals. The eighth chapter summarizes the major conclusions of each study.
**Flaviviruses**

**Past and present**

Although yellow fever virus (YFV) was recorded as the first virus human virus to be discovered only in 1901, the virus had been wreaking havoc in the Americas since the 1600s and had been around for even longer. The first recorded YFV outbreak in the Americas occurred on the Caribbean island of Barbados in 1647. It continued on to the islands of Guadeloupe, St. Kitts, Cuba and the Yucatan Peninsula of Mexico, as well as the east coast of Central America, killing perhaps 20-30% of the local inhabitants before taking an extended hiatus (McNeill 2004, Norrby 2007, Woolhouse, Scott et al. 2012). A number of other significant outbreaks occurred in both the northern and southern U.S throughout the 18th and 19th centuries, the most notable being in New Orleans, Memphis and Philadelphia where thousands of victims succumbed to the disease (Patterson 1992).

The etiological agent for yellow fever disease was discovered by the Reed commission in the early 1900s and by 1937 Max Theiler had developed a vaccine against the virus. The vaccine proved to be safe and effective and is still in use to this day (Norrby 2007). Although YFV is no longer a threat to developed countries such as the U.S, the World Health Organization reports that several countries in Africa and South America are still either endemic for, or have regions that are endemic for, yellow fever. According to the CDC, 200,000 cases of YFV infection occurs every year with 30,000 of those resulting in death and 90% of infections occurring in Africa. Recent large epidemics occurring in Brazil in 2017 through 2018 and Angola and Democratic Republic of the Congo (DRC) in 2016 through 2017, claimed the lives of 674 (2,037 confirmed cases) and 140 (about 1000 confirmed cases) individuals respectively (Manica, Guzzetta et al. 2019).
Other notable flaviviruses which made their debut after YFV include dengue virus (DENV), the symptoms of which were first described by Benjamin Rush in Philadelphia in 1780 (Beaumier, Garcia et al. 2014), West Nile virus (WNV), which was first isolated in Uganda in 1937 and first introduced to the U.S in 1999 during the New York outbreak (Sejvar 2003), Zika virus (ZIKV), another flavivirus first isolated in Uganda back in 1947 and introduced to the Americas in 2015 when an outbreak in Brazil caught the attention of people worldwide (Musso and Gubler 2016). Today, DENV has the broadest distribution of the vector-borne arboviruses, with an estimated 390 million people becoming infected annually, 25,000 succumbing to illness (Muller, Depelsenaire et al. 2017). According to the CDC preliminary maps and data for 2018, WNV infections in people, birds, or mosquitoes had been reported in 49 states in the United States of America and the District of Columbia. Over 2500 human cases were reported to the CDC, over 60% of which were classified as neuroinvasive.

Over 70 flaviviruses have been discovered to date, most of which are transmitted by an arthropod vector. Many of these viruses are a source of significant public health concern, particularly in tropical areas. Most of these viruses have no effective vaccine or antiviral available for treatment. Molecular virology studies that serve to elucidate the mechanisms of infection and host evasion will therefore prove useful (Krol, Brzuska et al. , García, Padilla et al. 2017, Wang, Liu et al. 2017, Sáez-Álvarez, Arias et al. 2019).

**Genome Organization and Virion Structure**

Flaviviruses are a group of positive-sense, single-stranded, non-segmented RNA viruses with genomes ranging from 9-11kb housed in an enveloped, spherical virion particle of about 50nm in diameter. Their genomes are capped at the 5’ end (m-7GpppAmp) but lack a poly(A)
Genomes encode a long open reading frame (ORF) that is flanked by 5’ and 3’ untranslated regions (UTRs) of approximately 100 and 400-700 nt., respectively. The ORF encodes a long polyprotein which is cleaved by viral and cellular proteases during and after translation generating three structural proteins, designated as the capsid (C), pre-membrane/membrane (prM/M) and envelope (E) proteins, and seven nonstructural (NS) proteins, in the order: 5’UTR-C-prM/M-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5 (García, Padilla et al. 2017, Ng, Soto-Acosta et al. 2017, Zakaria, Carletti et al. 2018).

The 5’ and 3’ UTRs play very important roles in viral translation, replication and viral pathogenesis. Long range interactions between the 5’ and 3’ UTR results in a cyclization of the viral genome which enables the initiation of replication by the polymerase enzyme. Generation of subgenomic flavivirus RNA (sfRNA) occurs through the 3’ UTR and plays a role in evasion of vertebrate host immune response (Villordo and Gamarnik 2009, Ng, Soto-Acosta et al. 2017, de Bernardi Schneider and Wolfinger 2019).

Surrounding the flavivirus RNA genome is the C protein. This protein interacts with the RNA to form the nucleocapsid. Surrounding the nucleocapsid is a lipid bilayer in which the other two structural proteins, the prM and E are embedded. The E proteins mediates attachment to cellular receptors and entry into the cell. Flaviviruses are taken up into the cell by receptor-mediated endocytosis (Jones, Ma et al. 2003, Bhuvanakantham and Ng 2013, Oliveira, Mohana-Borges et al. 2017, Simmonds, Becher et al. 2017, Wilder-Smith, Ooi et al. 2019).

The non-structural proteins play different roles in RNA replication and host response modulation. NS3 is the viral protease and forms a complex with its cofactor NS2B cleaving the viral polyprotein at certain junctions. The NS3 as well as the NS5 protein which encodes the
viral polymerase also play a role in capping of viral mRNAs and RNA synthesis (Leung, Pijlman et al. 2008, Bollati, Alvarez et al. 2010, Li, Phoo et al. 2014, Zhao, Yi et al. 2017). NS2A has been shown to play a role in assembly (Leung, Pijlman et al. 2008) while NS1 has been shown to influence pathogenesis for certain flaviviruses (Puerta-Guardo, Glasner et al. 2019). The NS4 proteins, NS4A and NS4B play a role in the induction of ER membrane rearrangements and antagonization of interferon, respectively (McLean, Wudzinska et al. 2011).

**Flavivirus Host-specificity**

Virus in the *Flavivirus* genus can be further classified into three different groups based on host range. Viruses in the first group are termed *arthropod-borne* viruses (arboviruses) and are maintained in horizontal cycles between insect vectors, such as ticks or mosquitoes, and vertebrate hosts, such as humans or horses (Martín-Acebes, Vázquez-Calvo et al. 2016, Colmant, Hobson-Peters et al. 2017). These viruses are therefore considered dual-host flaviviruses. The other two groups are more restricted in their host range and infect either insects (Bolling, Weaver et al. 2015) or vertebrates (Ochsenreiter, Hofacker et al. 2019). The vertebrate-specific flavivirus group can be separated into two groups. The first group includes viruses which have been isolated from bats, while the second group includes those that have been isolated from rodents. The insect-specific flavivirus group can also be further broken down into two different groups: the classical insect-specific flaviviruses are disconnected phylogenetically from the other viruses in the genus while the dual-host affiliated insect-specific viruses phylogenetically associate with the dual-host flaviviruses (Blitvich and Firth 2015).
Vertebrate-specific flaviviruses

Vertebrate-specific flaviviruses are also known as no known vector flaviviruses (NKV) due to the fact that these viruses have never been isolated from field-collected or experimentally-infected arthropod vectors nor do they replicate in mosquito cell lines (Volkova, Tesh et al. 2012). Rather, they are presumably maintained by means of horizontal transmission between bats or rodents in nature. In an experimental study conducted by Fairbrother and Yuill in 1987, deer mice inoculated with the rodent associated NKV flavivirus, Modoc virus, were housed with naïve deer mice for six weeks. Upon testing, 50% of these naïve mice tested positive for the presence of antibodies to the virus. The mice were also tested for the presence of virus in their lungs but only one mouse tested positive. The authors concluded that the virus may be spread via a direct contact mechanism (Fairbrother and Yuill 1987). Another study conducted by Davis and colleagues in 1974 showed that deer mice intranasally inoculated with Modoc virus not only shed the virus in their body fluids but also had virus that could be detected in many organs of their bodies (Davis, Hardy et al. 1974).

To date, only fourteen established NKV flaviviruses have been discovered, six of those having been isolated from rodents and the remaining eight having been isolated from bats. The bat NKV flaviviruses are further broken down into two groups: the Rio Bravo virus group includes Bukalasa bat virus, Carey Island virus, Dakar bat virus, Montana myotis leukoencephalitis virus, Phnom Penh bat virus and Rio Bravo virus while the Entebbe bat virus group only includes Entebbe bat virus and Yokose virus. The rodent associated NKV flaviviruses are all included in the Modoc virus group and include Apoi virus, Cowbone Ridge virus, Jutiapa virus, Modoc virus, Sal Vieja virus and San Perlita virus.
Rio Bravo virus was so named after its first isolation in 1954 from the salivary glands of a Mexican free-tailed bat at Rio Bravo School in Kern County, California. Although the mode of transmission is not definitive, it has been suggested that the virus may be transmitted through aerosols and/or saliva among bats living in close proximity to each other (Volkova, Tesh et al. 2012). Indeed, a higher incidence of infection has been demonstrated in bats living in poorly ventilated, congested roosts (Constantine and Woodall 1964). The few cases of laboratory acquired infections acquired in humans suggest that Rio Bravo virus infection leads to a systemic infection accompanied by inflammation of the testes or ovaries as well as complications with the central nervous system, while monkeys infected with the virus show febrile symptoms accompanied by drowsiness, transitory weakness of extremities, and nervousness (Sulkin 1961, Constantine and Woodall 1964, Volkova, Tesh et al. 2012). Inspired by the discovery of the novel NKV flavivirus, researchers at the Uganda Virus Research Institute (UVRI) conducted a viral hunt of their own using bats captured in the attic of the institute. Their investigation led to the discovery of Entebbe bat virus (ENTV) which was isolated in 1957 from the salivary glands of a little free-tailed bat (Kading, Kityo et al. 2015).

Modoc virus was so named after Modoc county in California where the virus was first isolated in 1958. As previously mentioned, Modoc virus is a rodent associated virus isolated from the mammary gland tissue of a white-footed deer mouse. One study done by Adams and colleagues showed that the virus caused encephalitis in 12.5% of hamsters and established a persistent infection in surviving hamsters who were found to shed virus in their urine 5 months after infection (Adams, Travassos da Rosa et al. 2013). The virus has also been implicated as the etiological agent of meningitis in a boy in California and antibodies were found in the sera of American Indians in Canada (Fairbrother and Yuill 1987).
In Chapter two of this dissertation, I report the construction and characterization of two chimeras generated between two representative dual-host flaviviruses, YFV and ZIKV and one representative NKV flavivirus, RBV. These chimeras were designed as a tool to investigate the genetic elements that govern flavivirus host specificity, specifically at the attachment and entry level and represent the very first to be constructed or described using a bat associated NKV flavivirus.

**Insect-specific flaviviruses**

The first of the insect-specific flaviviruses to be discovered was cell fusing agent virus (CFAV). The virus was discovered in 1975 by Stollar and colleagues after an observation made in *Aedes albopictus* cells. These cells were infected with a then unidentified agent that induced the formation of syncytia in these mosquito cells whereas vertebrate cells appeared refractory to viral infection (Stollar and Thomas 1975, Öhlund, Lundén et al. 2019).

Almost 30 years went by before another group could report the discovery of another insect-specific flavivirus. In 2003, Sang and colleagues reported the discovery of the second insect-specific flavivirus, Kamiti River virus (KRV). This virus was isolated from *Aedes macintoshi* mosquito larvae and pupae collected in Kenya. Like CFAV, KRV did not replicate in vertebrate cells or in mice, however, unlike CFAV, infection with KRV did not lead to syncytia formation (Crabtree, Sang et al. 2003, Sang, Gichogo et al. 2003).

Hoshino and colleagues reported the discovery of the third insect-specific flavivirus only a few years after the KRV report. This novel virus designated Culex flavivirus (CxFV) was isolated in Japan from Culex species mosquitoes. Mosquito cell lines tested supported replication
of the virus, exhibiting mild signs of cytopathic effects (CPE) while mammalian cell lines were refractory to viral infection (Hoshino, Isawa et al. 2007, Blitvich and Firth 2015).

In 2009, Crabtree and colleagues reported the discovery of yet another insect-specific flavivirus, Quang Binh virus, isolated from Culex tritaeniorhyncus mosquitoes collected in Vietnam. Mosquito cells infected with this virus exhibited signs of CPE while Vero cells failed to support its replication (Crabtree, Nga et al. 2009). Very soon after this virus was reported, Hoshino and colleagues reported the discovery of the fifth insect-specific flavivirus isolated from Aedes albopictus and Aedes flavopictus mosquitoes in Japan. The virus only replicated in one mosquito cell line, exhibiting mild signs of CPE. Only two years later, Tyler and colleagues reported the discovery of the next insect-specific flavivirus isolated from Culex tarsalis mosquitoes in Alberta, Canada. The virus was designated Calbertado virus (CLBOV) due to the fact that it was identified in the states of California and Colorado and the Canadian province of Alberta (Tyler, Bolling et al. 2011). Several other insect-specific flaviviruses were subsequently discovered and are still being discovered, the most recent of which was Sabethes flavivirus (SbFV) in South America, giving these viruses a distribution that spans every inhabited continent (Blitvich and Firth 2015, Gravina, Suzukawa et al. 2019).

Although these viruses do not appear to cause any human or veterinary disease, they are still worth studying for a number of different reasons. Investigations into the genetic elements that govern flaviviral host specificity will elucidate the evolutionary hurdles necessary to make the leap from insect-specific flavivirus to dual-host flavivirus. Two such studies have been conducted using chimeras generated between dual-host flaviviruses and classical insect-specific flaviviruses. By swapping the prM and E proteins of the dual-host flaviviruses with that of the insect-specific flaviviruses, the authors showed that flaviviral host restriction for insect-specific
flaviviruses occurred at multiple different levels (Junglen, Korries et al. 2017, Piyasena, Setoh et al. 2017). The results of such studies could also aid in the development of more effective vaccines against dual-host flaviviruses (Junglen, Korries et al. 2017). In addition to this, co-infection with multiple flaviviruses, for example an insect-specific flavivirus and a dual-host flavivirus, has also been shown to have an effect on viral transmission with some appearing to enhance infection/transmission (Kuwata, Isawa et al. 2015) while others appear to reduce to block infection/ transmission (Goenaga, Kenney et al. 2015).

In Chapter 2, I attempted to construct an ISF chimera by replacing the prM and E genes of a dual-host flavivirus, ZIKV, with that of a dual-host affiliated insect-specific flavivirus, Long Pine Key virus (LPKV). The chimera constructed failed to generate viable virus possibly due to incompatibilities in the genomes of the viruses used. Additionally, in chapter four, I fully sequenced the genomes of two previously described ISFs for which complete sequence data were unavailable.

**Negeviruses**

In 2013 Vasilakis and colleagues reported the discovery of six different novel viruses isolated from mosquitoes and phlebotomine sand flies. These viruses all appeared to exhibit an insect-restricted phenotype with mice and mammalian cells tested being refractory to viral infection. This group of viruses, emerging from five different continents/regions (South America, North America, Africa, Asia and the Middle East), were most closely related to viruses in the Cilevirus genus and formed a new taxon designated the Negevirus taxon (Vasilakis, Forrester et al. 2013). Citrus leprosis virus C, the type species of the Cilevirus genus is an agent of citrus leprosis capable of causing serious damage to infected trees. The virus is transmitted via mite
vectors and circulates in South and Central America causing major agricultural losses. The virus also presented a financial burden in the state of Florida (United States of America) up until the 1960s, after which reports appeared to cease (Roy, Choudhary et al. 2013, Roy, Hartung et al. 2015).

Ever since the initial report by Vasilakis and colleagues in 2013, a number of other Negevirus have been discovered. These include viruses isolated from areas as diverse as the Caribbean island of Trinidad (Auguste, Carrington et al. 2014) to the continent of Australia (O'Brien, McLean et al. 2017). Other negeviruses/negevirus-like viruses have also been isolated from the Philippines, Côte d’Ivoire, Portugal, Japan, Brazil, Colombia, Peru, Panama, USA, Nepal and China giving the Negevirus taxon a host range that covers every continent but Antartica (Kallies, Kopp et al. 2014, Nabeshima, Inoue et al. 2014, Carapeta, do Bem et al. 2015, Kawakami, Kurnia et al. 2016, Fujita, Kuwata et al. 2017, Nunes, Contreras-Gutierrez et al. 2017, Wang, Wu et al. 2018, Wang, Guo et al. 2019).

Viruses in the Negevirus taxon are single-stranded, positive sense RNA viruses with genomes which range from 9-10kb in size (Nunes, Contreras-Gutierrez et al. 2017). Genomes are housed by spherical virion particles 45-55nm in diameter and possess three different reading frames separated by short intergenic regions (Vasilakis, Forrester et al. 2013). Encoded in the first ORF is the putative non-structural replicase proteins possessing four different domains: a viral methyltransferase (vMet), a ribosomal RNA (rRNA) methyltransferase (MTase), an RNA helicase, and an RNA-dependent RNA polymerase (RdRP) domain (Kallies, Kopp et al. 2014, O'Brien, McLean et al. 2017, Wang, Wu et al. 2018, Wang, Guo et al. 2019). Although it remains to be confirmed what proteins are encoded by ORFs 2 and 3, that are predicted to encode the viral glycoprotein and membrane protein, respectively (Carapeta, do Bem et al. 2015,

Negeviruses currently fall into one of two groups based on phylogenetic affiliation. The Nelorpiirus clade of viruses also known as the group I negeviruses include viruses such as Negev (NEGV), Loreto (LORV), and Piura (PIUV) viruses, the first three viruses described in that clade and the viruses that cluster with them (Fujita, Kuwata et al. 2017, O'Brien, McLean et al. 2017, Kondo, Chiba et al. 2019). The second group, the group II negeviruses or sandewaviruses, include Santana (SANV), Dezidougou (DEZV, Wallerfield (WALV) viruses and the viruses that cluster with them (Kallies, Kopp et al. 2014, Carapeta, do Bem et al. 2015, Nunes, Contreras-Gutierrez et al. 2017).

In this dissertation, I report the discovery and characterization of a novel negevirus, isolated from mosquitoes collected in the Yucatan Peninsula of Mexico (Chapter 3).

Hubei Noda-like viruses

Among the newly discovered viruses isolated from mosquitoes collected in Mexico, was one virus that was most closely related to a group of unclassified RNA viruses designated Hubei Noda-like virus 1 to 26. These viruses were recently discovered during a large-scale metagenomics study conducted by Shi and colleagues in 2016. The study led to the discovery of 1445 RNA viruses recovered from over 220 invertebrate species over nine animal phyla. Among the viruses recovered were the unclassified Hubei Noda-like viruses which appeared to be most closely related to the viruses in the Nodaviridae family (Shi, Lin et al. 2016). Members of this family fall into one of two established genera: the Alphanodavirus genus infects insects, while
the Betanodavirus genus infects fish. Alphanodavirus RNA replication has also been shown to be supported by other organisms both in vivo (select mammals) and in vitro (vertebrate, plants and yeast cells) (Bailey, Newman et al. 1975, Ball, Amann et al. 1992, Price, Eckerle et al. 2005, Schuster, Zirkel et al. 2014, Sahul Hameed, Ninawe et al. 2019).

Nodaviruses are positive sense, single-stranded RNA viruses with bipartite genomes of approximately 3.1 and 1.4kb, respectively. These genomes are encapsidated in a non-enveloped, spherical virion particle of approximately 25-33nm. Members of this family have a 5’ cap but lack a poly(A) tail. The first segment encodes an RdRp and a posttranscriptional gene silencing suppressor protein known as B2. This protein is a non-structural protein expressed from a subgenomic RNA produced from the first fragment during RNA replication. The second fragment encodes the viral capsid protein (Li, Scotti et al. 2007, NaveenKumar, Shekar et al. 2013, Schuster, Zirkel et al. 2014).

Interestingly, phylogenetic analyses conducted on each of the two segments of our newly discovered virus showed that while the first segment analyzed was more closely related to members of the Betanodaviridus genus, the second was not. Rather, this segment, which encodes the capsid protein was more closely affiliated with the alphanodaviruses. This might indicate that this virus is either a reassortant formed from a reassortment event between one alpha- and one betanodavirus or that we recovered segments from two different viruses.

**Mesoniviridae**

The Mesoniviridae family is one of now nine established families in the order Nidovirales which was recently amended by the International Committee on Taxonomy of Viruses (ICTV). Before 2018, this order comprised of solely four different families which
included the two well-studied *Coronaviridae* and *Arteriviridae* families, the *Roniviridae* family and the most recent and somewhat newly established *Mesoniviridae* family (Nga, Parquet Mdel et al. 2011, Warrilow, Watterson et al. 2014, Blanck and Ziebuhr 2016, de Wilde, Pham et al. 2018, Xiao-yu, Yong et al. 2019). Viruses in this order have a wide host range which varies from humans to crustaceans and are united by a number of common features including positive-sense, single-stranded RNA genomes, that possess a 5’ cap on one end and a poly(A) tail on the other end of their genomes (Corman, Eckerle et al. 2016, Wang, Xia et al. 2017).

Families in this order have similar genome organizations but the number of encoded open reading frames (ORFs) varies among nidoviruses. Mesoniviruses, like the other *Nidovirales*, have two overlapping ORFs at the 5’ end of their genomes which code for two different polyproteins. These polyproteins are later processed to generate proteins that play important roles in the regulation of protein expression and viral replication. A “slippery” sequence encoded in the genome allows for ribosomal frameshifting which is necessary for production of the second polyprotein encoded by ORF1b. 3’ to these proteins are a number of smaller ORFs that differ based on the virus (Nga, Parquet Mdel et al. 2011, Zirkel, Roth et al. 2013, Vasilakis, Guzman et al. 2014, Posthuma, te Velthuis et al. 2017).

Virus in the order *Nidovirales* represent the largest RNA viruses and can vary from 12kb (*Arteriviridae*) to over 30kb (*Coronaviridae* and *Roniviridae*). While the OF1b gene of all *Nidovirales* have been found to possess several important domains such as an RdRp, a superfamily 1 helicase (HEL1), putative multinuclear zinc-binding domain (ZBD) and uridylate-specific endoribonuclease (NendoU) domain, the latter two being restricted to *Nidovirales*, only the larger ones have been shown to encode a 3’-to-5’ exoribonuclease (ExoN) in the HEL1 region. It is assumed that the acquisition of this ExoN enabled the larger nidoviruses to evolve
from their smaller precursors. Indeed, the relatively recent discovery of the first mesonivirus allowed researchers to bridge the gap between the small and large nidoviruses in existence at the time. This family of viruses which averages approximately 20kb, also possesses an ExoN domain (Gorbalenya, Enjuanes et al. 2006, Nga, Parquet Mdel et al. 2011, Lauber, Ziebuhr et al. 2012, Sadeghi, Popov et al. 2017).

The first two mesoniviruses to be described were discovered relatively recently in apparent quick succession in mosquitoes collected in Vietnam and Côte d’Ivoire respectively. The first virus, Nam Dinh virus, became the prototypical member of a new family in Nidovirales order and also became the first known member of the order to infect insects (Zirkel, Kurth et al. 2011, Thuy, Huy et al. 2013, Zhou, Jin et al. 2017). Both as these viruses, like the other novel mesoniviruses that were later discovered, exhibited a mosquito-specific phenotype and do not appear to cause disease in vertebrates. Despite their apparent insect-restricted phenotype, these viruses are still worth studying not only due to their phylogenetic relationship with other more dangerous viruses in that very same order, but also because it has been shown that infection with insect-specific viruses may have an effect on the transmission of arthropod-borne viruses whose replication is supported by vertebrates (Vasilakis, Guzman et al. 2014, Goenaga, Kenney et al. 2015, Öhlund, Lundén et al. 2019).

In chapter 3 of this dissertation, I report the isolation of a previously described mesonivirus, Houston virus, which was discovered in the United States of America (Vasilakis, Guzman et al. 2014, Cigarroa-Toledo, Baak-Baak et al. 2018). This virus was isolated from mosquitoes collected in the Yucatan Peninsula of Mexico. Further characterization of the virus was done using eleven different cell lines to investigate the in vitro host range after sequencing
was completed. The information generated from our studies served to provide further insights into the nature of this new family of viruses.

**Tymoviruses**

The *Tymovirus* genus is one of three established genera in the *Tymoviridae* family of viruses which also includes the marafiviruses and the maculaviruses (Igori, Lim et al. 2017). Members of this family have traditionally been recognized as plant viruses vectored by coleopteran and hemipteran insects (Wang, Lv et al. 2012, Li, Lin et al. 2016). Recently, however, our group reported the discovery and isolation of a novel *tymoviridae*-like virus which was recovered from mosquitoes collected in the Yucatan Peninsula of Mexico (Charles, Tangudu et al. 2019). This was the second report of a *tymoviridae*-like virus being isolated from mosquitoes, the first being by a group in China back in 2012 (Wang, Lv et al. 2012). Members of the *Tymoviridae* family are non-enveloped, positive-sense, single stranded RNA viruses with genomes that range from 6.0-7.5kb with isometric virion particles that are approximately 30 nm in diameter (2012, Li, Lin et al. 2016, Thekke-Veetil, Ho et al. 2018).

One unusual feature of tymoviruses is the high cytidine content of their genomes which can range from 32-50% (Hull 2002, Martelli, Sabanadzovic et al. 2002, Edwards and Weiland 2010, 2012). The typical tymovirus genome possesses a cap at the 5’ end (\(^{m7}GpppG\)) followed by two overlapping open reading frames (ORFs) which encode the viral replicase and movement protein (MP), respectively (Segwagwe, Putnam et al. 2008, de Oliveira, Nagata et al. 2013). The first ORF is the largest of three and possesses an RdRp, papain-like protease, (MTase) and helicase domains (Hull 2002). The MP encoded by the second ORF has been implicated in a number of possible roles including movement from cell to cell, suppression of RNA silencing
and as a genetic element that governs viral host range (Segwagwe, Putnam et al. 2008, Nicolini, Pio-Ribeiro et al. 2012). The third ORF is located at the 3’ end of the genome and encodes a subgenomic mRNA which is translated to produce the coat protein. Directly upstream to this ORF is a 16-nucleotide region which appears to be conserved in all tymoviruses. This region is referred to as the tymobox and is assumed to form part of the promoter region for ORF3 (Ding, Howe et al. 1990, Hull 2002, Segwagwe, Putnam et al. 2008). Members of the marafivirus genus also share a similar sequence designated the “marafi-box”, although interestingly no such region has been found in the maculaviruses (Martelli, Sabanadzovic et al. 2002). The genome terminates with a 3’ untranslated region (UTR), part of which forms a tRNA-like structure (TLS) which has been shown to enhance translation (Nicolini, Pio-Ribeiro et al. 2012). The 3’ terminus of the genome is also considered to have important implications for replication of the viral genome (Segwagwe, Putnam et al. 2008). Unlike the two other two genera in the Tymoviridae family, tymoviruses do not possess a poly(A) tail (Kreuze, Koenig et al. 2013).

Tymoviruses unlike most other members of the Tymoviridae family can be transmitted mechanically with a relatively unlimited tissue tropism in their hosts. Marafi- and maculaviruses however, are restricted to the phloem of the plant (2012). Members of the marafivirus genus mainly infect monocotyledonous plants whereas tymovirus and maculavirus genera infect mainly dicotyledonous plants. Viruses in the Tymoviridae family have a widespread distribution in general but distribution of individual members may vary from individual to widespread (Martelli, Sabanadzovic et al. 2002, 2012). Some signs of infection in plants may include the presence of isomeric particles in the virion, cytoplasm and vacuole, the presence of membrane-derived vesicles on the periphery of mitochondria and/or chloroplasts and disruption in the shape of certain organelles. More genus-specific diseases symptoms include flecks in the leaves of
plants infected with maculaviruses, dwarfing, chlorotic stripes and vein clearing in the
marafiviruses and bright yellow mosaic or mottling in the tymoviruses and Poinsettia mosaic
virus, an unassigned member of the family (PnMV) (Crestani 1986, Alexandre, Duarte et al.

In chapter 4 of this dissertation, we report the discovery of the second tymoviridae-like
virus to be recovered in mosquitoes.

Rhabdoviridae

The Rhabdoviridae family is one of now eight families in the order Mononegavirales.
Viruses in this order have single stranded, negative sense, non-segmented RNA genomes that are
housed in a lipid envelope. Over 130 established rhabdovirus species assigned to 18 different
genera have been discovered thus far. These viruses have a very diverse host range which differs
according to assigned genus. The Lyssavirus, Novirhabdovirus, Perhabdovirus, Sprivirus,
Tupavirus genera have a host range that appears to be restricted to vertebrates, while the
Curiovirus, Ephemerovirus, Hapavirus, Ledantevirus, Sripuvirus, Tibrovirus, and Vesiculovirus
genera infect vertebrates via the assistance of an arthropod vector. Members of the
Almendravirus and Sigmavirus genus are restricted to arthropod hosts whereas the
Cytorhabdoviruses, Dichorhaviruses, Nucleorhabdoviruses, Varicosaviruses have all been
classified as plant viruses (Contreras, Eastwood et al. 2017, Amarasinghe, Arechiga Ceballos et
al. 2018, Walker, Blasdell et al. 2018). Dichorhaviruses and Varicosaviruses were added to the
Rhabdoviridae family only recently and unlike the other viruses in this family, the genomes of
these viruses are divided into two separate segments (Koloniuk, Fránová et al. 2018).
Of the 18 established genera, the *Lyssavirus* and *Vessiculovirus* genera are the two most well-studied. The type species of the *Lyssavirus* genus is the very well-known rabies virus (RABV) which has an almost 100% fatality rate in humans. This virus claims almost 60,000 lives yearly, most of which reside in Asia or Africa. The type species of the *Vesiculovirus* genus is vesicular stomatitis virus (VSV) which causes disease in cattle, pigs and horses (Albertini, Baquero et al. 2012, de Souza and Madhusudana 2014, Apanga, Awoonor-Williams et al. 2016, 2017). Other examples of rhaboviruses include but are not limited to infectious hematopoietic necrosis virus, viral hemorrhagic septicemia virus and spring viremia of carp virus, which cause viral disease and mortality in fish (Purcell, Laing et al. 2012), potato yellow dwarf virus and lettuce necrotic yellows virus, which infect plants (Redinbaugh and Hogenhout 2005) and Adelaide River virus and bovine ephemeral fever virus which both infect cattle (Wang, Cowley et al. 1995, Trinidad, Blasdell et al. 2014).

The *Rhabdoviridae*, as the name suggests, is so named based on their rod-like shape; these viruses take on a sort of bullet or cone shape in animals and bacilliform shape in plants (Martinez, Ribeiro et al. 2013, Dietzgen, Kondo et al. 2017). The virion particle is 75nm in diameter and 100-300nm in length. Their genomes range from 11-16kb and like the other viruses in the order *Mononegavirales* encode five proteins in the order 3′-nucleo-(or nucleocapsid) protein - phosphoprotein- matrix protein - glycoprotein - polymerase-5′ (3′-N-P-M-G-L-5′). These proteins and flanked by complementary 3′ leader and 5′ trailer sequences. Rhabdoviruses may also encode additional accessory proteins (Blondel, Maarifi et al. 2015, Dietzgen, Kondo et al. 2017, Whitfield, Huot et al. 2018, Xu, Cantalupo et al. 2018, Guo, Duan et al. 2019).

The nucleocapsid protein interacts with the viral genome to form the ribonucleoprotein complex (RNP). This complex serves as the template for transcription and replication after being
released into the cell (Luo, Green et al. 2007). The L protein encodes the viral polymerase and together with the P protein forms the polymerase complex which acts to transcribe the RNP (Morin, Liang et al. 2016). The M protein acts as a bridge between the RNP and the cytoplasmic domain of the G protein and has been implicated in a number of other roles including modulation of transcription, translation, apoptosis and the promotion of budding (Koser, McGettigan et al. 2004, Graham, Assenberg et al. 2008). The G protein which protrudes from the viral surface mediates attachment to cell surface receptors and entry of the virus into host cells (Albertini, Baquero et al. 2012). The 3’ leader sequence serves as a transcriptional entry site and has been shown to possess sequences which are inversely complementary to the 5’ trailer sequence which terminates transcription and replication (Hinzman, Barr et al. 2002, Albertini, Ruigrok et al. 2011).

In chapter 7 of this dissertation, I report the discovery of a novel rhabdovirus designated Merida virus which was discovered in mosquitoes collected in Mexico. This virus has not yet been assigned to a genus and due to our inability to obtain a viral isolate, we were not able to conduct host range studies to determine tropism.

**Peribunyaviridae**

Viruses in the recently established *Peribunyaviridae* family are tripartite, single-stranded, negative sense or ambisense RNA viruses that range in size from 11-19kb in total (2012, Briese, Calisher et al. 2013, Maes, Adkins et al. 2019). This family is home to four different genera including the *Orthobunyavirus* genus, which is the largest of the RNA viruses, having had over 170 viruses assigned across 18 different serogroups (Travassos da Rosa, de Souza et al. 2017, Maes, Adkins et al. 2019). Over the past two years, several changes have been made to this
family including establishment of the *Peribunyaviridae* family as the *Bunyaviridae* family was elevated in rank to the *Bunyavirales* order, the dissolution of the *Hantavirus, Nairovirus* and *Phlebovirus* genera from the family and the addition of two additional genera, the *Herbevirus* and *Shangavirus* genera (Roundy, Azar et al. 2017, Maes, Adkins et al. 2019).

Members of the *Tospovirus* genus, unlike the other genera, uniquely infect plants and include viruses such as the type species tomato spotted wilt virus which has a wide distribution and host range and impatiens necrotic spot virus which infects vegetables, decorative plants and tomatoes leading to major agricultural losses (Oliver and Whitfield 2016, Smirnova, Shneider et al. 2016). The *Shangavirus* genus has but one species member designated Shuāngào insect virus 1 while the *Herbevirus* genus has been split into three member species (Maes, Adkins et al. 2019).

The largest genus, the *Orthobunyavirus* genus is the most medically significant of the *Peribunyaviridae* causing serious disease in both human and animal hosts. Infection with La Crosse virus, a California serogroup *Orthobunyavirus* vectored by mosquitoes, causes encephalitis in humans with most cases being reported in children under 15 years of age (Shope 1996, Byrd 2016). Jamestown Canyon virus is another California serogroup virus, which, like the other members of that serogroup, cause febrile illness and neurologic disease. This virus is believed to be amplified in white-tailed deer (Webster, Dimitrova et al. 2017). Oropouche virus is another *Orthobunyavirus* which is transmitted mainly by midges and leads to febrile symptoms in humans (Romero-Alvarez and Escobar 2018).

The tripartite genome exists in the form of small (S), medium (M) and large (L) segments housed in a spherical or pleomorphic virion averaging 80–120 nm (2012, Yadav, Chaubal et al. 2017). The S segment encodes the nucleoprotein which interacts with viral RNA and the viral
polymerase (encoded by the L segment) to form the ribonucleoproteins. The S protein of certain bunyaviruses also encodes non-structural proteins. The M segment encodes a polyprotein that gets cleaved into two glycoproteins Gn and Gc which are both embedded in the lipid envelope, forming the spike proteins. The M segment of viruses in the Orthobunyavirus genus also encodes a non-structural protein NSm between these two proteins (Överby, Pettersson et al. 2008, 2012, Hellert, Aebischer et al. 2019).

In chapter 6 of this dissertation, I report sequencing of the large genome segment of two Orthobunyaviruses, Northway virus and Main Drain virus, viruses for which complete sequence data had previously been unavailable. Main Drain virus has been implicated as a source of encephalomyelitis in horses (W Emmons, D Woodie et al. 1983) while Northway virus appears to prefer leporids (Brabb and Di Giacomo 2012).

References


Krol, E., G. Brzuska and B. Szewczyk "Production and Biomedical Application of *Flavivirus*-like Particles." *Trends in Biotechnology*.


CHAPTER 2. INVESTIGATION OF THE GENETIC DETERMINANTS THAT MODULATE HOST-RESTRICTION OF BAT-ASSOCIATED, NO KNOWN VECTOR FLAVIVIRUSES

Modified from a manuscript to be submitted Journal General Virology

Jermilia Charles, Chandra Tangudu, Bradley J. Blitvich

Department of Veterinary Microbiology and Preventive Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa

Author for Correspondence: Bradley J. Blitvich, PhD; 2116 Veterinary Medicine, Iowa State University, Ames, Iowa, 50011; E-mail: blitvich@iastate.edu; Telephone: 515-294-9861; Fax: 515-294-8500

Abstract

The Flavivirus genus is one of three genera in the Flaviviridae family which harbors a number of emerging and re-emerging pathogens of public and veterinary health importance worldwide. Most members of this genus are transmitted via the bite of an infected arthropod vector to a vertebrate host and are termed arboviruses, whereas others appear to exhibit a more restricted host range, displaying host tropisms that appear to be confined to either insects (insect-specific flaviviruses) or vertebrates (no known vector flaviviruses). Investigative studies into the possible genetic elements that govern flaviviral host specificity could elucidate the evolutionary hurdles that these host-restricted flaviviruses had to overcome in order to expand their host ranges. Information gathered from such studies could have positive implications for efficacious vaccine as well as drug development. For this study, we conducted an investigation into no known vector (NKV) flaviviral host restriction at the level of attachment and entry. This was examined using two different chimeras generated using dual-host flaviviruses, yellow fever virus (YFV) and Zika.
virus (ZIKV), and a representative NKV flavivirus, Rio Bravo virus (RBV). Gibson assembly was used to assemble fragments containing all but two protein-coding genes (prM and E genes) from either dual-host flavivirus, the prM and E genes of RBV, a cytomegalovirus (CMV) promoter and a hepatitis delta virus ribozyme (HDR) followed by the simian virus 40 (SV40) polyadenylation sequence (poly(A)). The assembled DNA products were transfected into baby hamster kidney (BHK) cells and then later passaged on to Vero and *Aedes albopictus* (C6/36) cells and monitored for signs of cytopathic effects and tested for virus by RT-PCR and plaque assay. A more extensive *in vitro* host range study was also conducted on the Zika/Rio Bravo virus chimera (Zikrio). Results of our analyses showed that both chimeras exhibited a dual-host phenotype even though the attachment and entry protein genes had been switched with that of an NKV flavivirus. This would imply that NKV flavivirus host restriction occurs at a post-attachment/entry stage for RBV and possibly other NKV flaviviruses. We also employed this strategy to conduct an investigation into the genetic determinants of insect-specific flavivirus (ISF) host specificity. Attempts made at generating chimeras between YFV and an ISF, Long Pine key virus (LPKV) were however unsuccessful.

**Introduction**

Globalization has had far-reaching impacts, opening for us doors that we never even knew to exist in centuries past; however, it also brings with it a wave of new challenges. With the world now being smaller than it has ever been, humans (and animals) are exposed to increasing pathogenic threats such as those posed by vector-borne diseases. Climate change accompanied by the advent of globalization and increased international travel has facilitated the spread of insect vectors such as the *Aedes* species mosquitoes which support the replication of several medically significant flaviviruses such as Zika (ZIKV), yellow fever (YFV) and dengue

Flaviviruses are a group of positive-sense, non-segmented, enveloped RNA viruses in the 
*Flaviviridae* family (Lopez-Denman and Mackenzie 2017, Ng, Soto-Acosta et al. 2017). Their genomes average approximately 11kb and encode a long polyprotein which is cleaved by host and viral proteases co- and post-translationally to generate at least ten proteins. These proteins include the structural capsid (C), pre-membrane/membrane (prM/M) and envelope (E) proteins and seven nonstructural proteins designated NS1-NS5 in the order: 5’UTR-C-prM(M)-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3’UTR. The C protein interacts with the viral RNA genome to form the nucleocapsid, while the prM protein interacts with the E to mediate viral attachment and entry into host cells. The nonstructural proteins play roles in viral replication, assembly and evasion of the immune response (Brand, Bisaillon et al. 2017, García, Padilla et al. 2017, Lopez-Denman and Mackenzie 2017, Ng, Soto-Acosta et al. 2017, Zhang, Jia et al. 2017, Kellman, Offerdahl et al. 2018).

Over seventy flaviviruses have been discovered thus far, most of which exhibit a dual-host phenotype, cycling between arthropod (i.e. tick or mosquito) vectors and vertebrate hosts (Piyasena, Setoh et al. 2017). However, two smaller subsets of viruses in this genus appear to exhibit a more restricted host phenotype replicating in only insects (insect-specific flaviviruses) (Piyasena, Setoh et al. 2017) or only vertebrates (no known vector flaviviruses) (Blitvich and Firth 2017). No known vector (NKV) flaviviruses are divided into two different groups; bat-associated NKV flaviviruses and rodent-associated NKV flaviviruses. These viruses have never been isolated from, nor do they replicate in field collected or experimentally infected arthropods but are presumably maintained by means of horizontal transmission (Fairbrother and Yuill 1987,
Volkova, Tesh et al. 2012). Insect specific flaviviruses (ISFs) can also be further divided into two different groups: dual-host affiliated ISFs, which despite their apparent insect-restricted phenotype, phylogenetically affiliate with the dual-host flaviviruses, and the classical ISFs (Blitvich and Firth 2015). As it is likely that dual-host flaviviruses originated from single-host ancestors, data generated from studies investigating the genetic elements which dictate these differential host ranges may prove valuable (Halbach, Junglen et al. 2017). The flaviviral genome encodes ten different proteins, many of which either form complexes or interact with each other to mediate different stages in the viral life cycle. It is therefore very likely that these evolutionary jumps from single to dual-host tropisms, relied on major changes in multiple genes or genetic elements in the viral genome.

Studies done using the rodent-associated NKV flavivirus, Modoc virus (MODV), has shown that host restriction occurs at a stage post attachment and entry and is not mediated by specific elements in the 3’UTR (Charlier, Davidson et al. 2010, Tumban, Maes et al. 2013, Saiyasombat, Carrillo-Tripp et al. 2014). Similar studies done using ISF chimeric viruses showed that ISFs faced opposition at multiple different levels in the viral life cycle upon attempted infection of a vertebrate cell (Junglen, Korries et al. 2017, Piyasena, Setoh et al. 2017). Indeed, for one study, the researchers failed to recover any virus even when the attachment and entry step, mediated by prM and E proteins, was bypassed using electroporation (Junglen, Korries et al. 2017).

Here we report the generation of two chimeric viruses designed to investigate whether NKV flaviviruses face opposition in insect cells at the attachment and entry level. These chimeric viruses are the first of their kind, having been generated from dual-host flaviviruses and a bat-associated NKV flavivirus. The two chimeras, Yelrio and Zikrio, were generated by
swapping out the prM and E proteins of either YFV or ZIKV with that of a bat-associated NKV flavivirus- Rio Bravo virus (RBV). The results were two chimeric viruses that appear to replicate in both insect and vertebrate cells proving that unlike the ISFs, host restriction in this subgroup most likely occurs at a downstream post-attachment/entry stage. Attempts were also made at generating chimeric viruses between YFV and an ISF, Long Pine Key virus (LPKV), however, they proved unsuccessful.

**Materials and Methods**

**Cell Culture**

Cell lines used in this study are as follows: *Aedes albopictus* (C6/36), African green monkey kidney (Vero), *Anopheles gambiae* (Sua 4.0), baby hamster kidney 21 (BHK-21), baby hamster kidney-derived cells that constitutively express T7 RNA polymerase (BSR-T7/5) (Buchholz, Finke et al. 1999), *Culex tarsalis* (CT), duck embryonic fibroblast (DEF), human epithelial (HeLa), murine microglia, rhesus macaque (LLC-MK2), *Spodoptera frugiperda* (Sf9) and *Trichoplusia ni* (High FiveTM) cells. Cells were cultured in Dulbecco’s modified Eagle medium (all mammalian cells except BHK-21 cells which were cultured in minimum essential medium and murine microglia which were cultured in DMEM/F12), Eagle’s minimum essential medium (EMEM) (DEF cells), Express Five medium (High Five cells), Liebovitz L15 medium (C6/36 cells), Schneider’s Drosophila medium (CT and Sua 4.0 cells) and SF-900 medium (Sf9 cells) (all cell culture media from Thermo Fisher Scientific, Massachusetts, USA except EMEM which was from the American Type Cell Collection). All media were supplemented with 10% fetal bovine serum (FBS) (except for CT cells which were supplements with 7% FBS), 2 mM L-glutamine, 100 units/ml of 17 penicillin and 100 μg/ml of streptomycin, with the exception of the Express Five and SF-900 media which were serum-free. In addition to this, murine microglia
cells were also supplemented with sodium pyruvate. Vertebrate cells were cultured at 37°C with 5% CO2 whereas invertebrate cells were cultured at 28°C.

**Viruses**

RNA extracted from YFV (strain 17D), ZIKV (strain PRVABC59) and RBV (strain M64) was used to generate complementary DNAs (cDNAs) in a reverse transcription reaction (RT) which were then used as templates for polymerase chain reactions (PCRs).

**Construction of chimeric viruses**

Two chimeric flaviviruses were generated employing the strategy illustrated in Figure 1. The first chimera was created using all but the prM and E genes of the Zika virus genome which was replaced with the corresponding genes of the Rio Bravo virus genome. The second chimera was generated using a similar strategy, where the prM and E genes of yellow fever virus were swapped out for the corresponding genes of Rio Bravo virus.

For the creation of the first chimera, Zikrio, six pairs of primers were designed. Each internal primer was designed such that the 5’ end of one fragment would overlap with the 3’ end of the adjacent fragment. In some cases, this required the introduction of overhangs (sequences from another virus, CMV or ribozyme sequence). Overlaps ranging from 15-26 nucleotides (Zikrio) and 18-28 nucleotides (Yelrio) were designed such that the melting temperature always exceeded 50°C, a requirement for efficient Gibson assembly. cDNAs were generated using Superscript III reverse transcriptase (Invitrogen) and PCR was done using Phusion high fidelity DNA polymerase (New England Biolabs).

The six fragments generated from PCR reactions included the following: 5’ CMV promoter- 5’UTR+ Capsid of ZIKV- prME of RBV-Fragment 4 of Zika- Fragment 5 of Zika-
ribozyme followed by polyadenylation signal 3’. The primer sequences and binding locations are included in the supplementary materials.

All fragments were purified using the PureLink PCR purification kit (Thermo Fisher) before proceeding to Gibson assembly. DNA concentrations were measured using a Nanodrop (Thermo Fisher). Fragments were assembled using the Gibson assembly master mix (New England Biolabs) according to the manufacturer’s directions. A similar strategy was employed for generation of Yelrio chimera.

**Transfections and virus recovery**

The entire Gibson assembly reaction was used for transfection reactions. 25 cm$^2$ flasks of either BHK 21(Zikrio) or BSR-T7/5 (Yelrio) cells were seeded the day prior to transfection. Cells were transfected with Gibson reaction mix containing 1.67 µg of full-length chimeric DNA product mixed with 900 µl of serum-free Opti-MEM (Invitrogen) and 12 µl of Lipofectamine 2000 reagent (Thermo Fisher) and incubated for approximately 20 minutes before adding to cells. Cells were left overnight, and fresh media was added the next day. Transfected cells were incubated for 4-6 days after which they were frozen and thawed and aliquots collected and inoculated onto subconfluent monolayers of Vero and BSR-T7/5 (Yelrio) cells. Cells were monitored daily for signs of CPE and left for up to 7-10 days if no CPE was observed and frozen if they exhibited 50-70% CPE. Additional passages were performed in the same cell type or, where specified, an alternate cell line before testing each passage for the presence of each virus by RT-PCR.
**Plaque assays**

Viruses were subjected to serial tenfold dilutions, inoculated onto confluent monolayers of Vero cells in 35-mm culture dishes then incubated at 37°C for 60 min. Three milliliters of neutral red-deficient minimum essential medium (Invitrogen) supplemented with 2% FBS, antibiotics and 1% agar were added to each well. Plates were incubated at 37°C for 5 days for Zika, Rio Bravo, Zikrio and Yelrio virus plaque assays and 6 days for yellow fever virus plaque assays. Another 3 ml of the same medium containing 0.22% neutral red was then added to each well, and plaques were counted 24 h later. Viral titers were expressed as plaque-forming units per milliliter (pfu/ml).

**Plaque size comparisons**

Viruses were inoculated onto confluent monolayers of Vero cells in 35-mm culture dishes then incubated at 37°C for 60 min. Three milliliters of neutral red deficient minimum essential medium (Invitrogen) supplemented with 2% FBS, antibiotics and 1% agar were added to each well, and plates were incubated at 37°C for 3, 5 or 7 days. To fix the cells, 2 ml of 10% formaldehyde was added directly onto each agar overlay and the plates were incubated at 37°C for at least 60 min. Agar overlays were gently removed, and 0.5 ml of 0.25% crystal violet (w/v) in 20% methanol was added to each well. Once the desired intensity was reached, plates were rinsed several times with tap water and photographed.

**Reverse transcription-polymerase chain reaction**

Total RNA was extracted from supernatants using the QIAamp viral RNA mini kit (Qiagen, Valencia, CA). Complementary DNAs were generated using Superscript III reverse
transcriptase (Thermo Fisher). PCRs were performed using Taq DNA polymerase (Thermo Fisher) and primers specific to each chimera, with the forward and reverse chimera spanning junctions between the dual-host flavivirus and RBV. For the sequencing of the genomes, primers spanning the entire length of the genomes were designed.

**In vitro host range studies**

The *in vitro* host ranges of ZIKV, RBV and Zikrio were assessed using 11 cell lines. Briefly, cell monolayers approaching confluency in 25 cm² flasks were inoculated with 50 μl of virus-containing supernatant and incubated for 7 days or until 40-60% of the cell monolayer displayed CPE. A total of five passages were performed, with supernatants collected after each passage and tested by RT-PCR for viral RNA. A cell line was considered to support virus replication if RT-PCR products were detected after every passage.

**Results**

Before a flavivirus can even begin its life cycle, two very crucial hurdles or barriers to infection must be overcome: viral attachment and entry. To determine whether NKV flaviviruses face opposition in insect cells at the level of attachment and entry, we generated two different full-length flavivirus chimeras. The first chimera was generated using overlapping DNA fragments spanning all but the prM and E regions of the ZIKV genome, the prM and E genes of RBV, CMV promoter and HDR followed by SV40 poly(A) signal. The strategy employed to generate the full-length chimeras is illustrated in a schematic in figure 1 using Zikrio as an example. Fragments were assembled by Gibson assembly before being transfected into BHK21 cells (Zikrio) or BSR-T7/5 (Yelrio). For the Yelrio experiments, we switched to using BSR-T7/5 cells not because the promoter in the chimera was changed to a T7 promoter, but because we realized that these cells were able to live for almost twice as long sometimes after being
transfected or infected. Lipofectamine complexes were left on transfected cells overnight as that seemed to produce better results than the shorter incubations. After about 5 days the cells were frozen and thawed and aliquots of each supernatant was collected and inoculated onto Vero cells (Zikrio) and BHKT7 cells (Yelrio). After about a week, supernatant was again collected and used to inoculate fresh Vero cells. CPE was observed within a few days. Both chimeras were also inoculated onto C6/36 cells. Samples were assessed by plaque assay, RT-PCR and Sanger sequencing to confirm viral production.

**Host range**

Replication of both chimeras generated appeared to be supported by both vertebrate and mosquito cells. Vero cells infected with either Yelrio or Zikrio viruses exhibited severe cytopathic effects (CPE). While this was not the case for C6/36 cells infected with these viruses-these cells exhibited no signs of CPE up to 7 days post infection-data generated from RT-PCR and plaque assay analysis confirmed that these cells do indeed support the replication of both chimeras (Table 2). Supernatant from infected C6/36 cells was also used to inoculate Vero cells in which CPE was indeed observed. More extensive *in vitro* host range studies were also conducted using the first of the two generated chimeras, Zikrio, as well as both of its parental viruses ZIKV and RBV. For these studies, eleven different cell lines of insect (mosquito and lepidopteran) and vertebrate (avian, hamster, human, monkey and murine) origin were used. The findings of this study are summarized in Table 1. Zikrio was found to replicate in four of the eleven cell lines tested: Vero and LLC-MK2, which are both monkey cells, DEFs which are of avian origin and C6/36s which are of mosquito origin. RBV replication was also supported by four out of the eleven cell lines tested, all of which were of vertebrate origin and included the DEF avian cell line, the BHK hamster cell line and the two monkey cell lines Vero and LLC-
MK2. This virus, as it has been shown before, did not demonstrate the capacity to replicate in any of the insect cell lines tested. The other parental virus, ZIKV, replicated in five of the eleven cell lines tested including both monkey cell lines Vero and LLC-MK2 and the avian cell line DEF, three cell lines that also support the replication of both Zikrio and RBV. ZIKV also replicated in BSR-T7/5 and C6/36 mosquito cells.

**Genome sequencing**

The full genomes of Yelrio and Zikrio were sequenced (except the final approximately 50-100 nucleotides in the 5’ and 3’ regions for which sequence data was not taken to be reliable) and was found to harbor seven and eight mutations, respectively. In the Yelrio virus genome, six of the seven mutations were silent while the eighth, located in the prME region of the genome was found to be a missense mutation which converted leucine to proline. Half of the eight mutations in the Zikrio genome were found to be silent while the other four were found to be missense mutations located in the prME, NS1 and NS5 regions.

**Growth kinetics**

Vero and C6/36 cells were infected in triplicate with Zikrio, Yelrio, and the three parental viruses with aliquots being collected at 24-hour intervals for 7 days. Viral titers were then measured using plaque assay analysis. The results are shown in Table 3. In Vero cells, Zikrio, like ZIKV, reached peak titers by day 2 but, unlike ZIKV, which demonstrated a sharp decline from that point, Zikrio demonstrated a steadier decline as the week progressed. RBV, although demonstrating comparable titers by that point, did not reach its peak until the 5th day. Titers also remained high in Vero cells infected with RBV up until day 7. Yelrio did not reach its peak titer until day 4, demonstrating growth kinetics more similar to that of YFV which had the lowest
peak titer and reached its peak on day 3. Like Zikrio and RBV, this virus also appeared to

demonstrate a more modest, gradual decline from the point at which it reached its peak to the 7th
day.

Both chimeras demonstrated the ability to replicate in C6/36 cells, however the
replication kinetics differed drastically from that which was seen in Vero cells. Yelrio reached a
mean peak titer of 6 log_{10} pfu/ml in Vero cells whereas C6/36 cells only saw a mean peak titer of
3 log_{10} pfu/ml. Zikrio reached an even higher mean peak titer of 7 log_{10} pfu/ml in Vero cells
whereas C6/36 cells again showed a far more reduced mean peak titer of 4 log_{10} pfu/ml. RBV
was not included in C6/36 growth curve experiments as it has previously been shown that the
virus does not replicate in insect cells. The replication kinetics of the both chimeras were more
similar to the dual-host flavivirus from which their backbones were derived as compared to
RBV, the NKV flavivirus from which their prM and E proteins were derived. That is, the
replication kinetics of Yelrio was more similar to YFV as compared to RBV and the replication
kinetics of Zikrio was more similar to ZIKV as compared to RBV.

**Plaque size**

To compare plaque sizes between the parent viruses (RBV, ZIKV and YFV) and the
chimeras generated, Vero cells were plated in 6-well plates and infected with each virus. Cells
were fixed at 3 different intervals- 3, 5 and 7 days- before staining with crystal violet. On day 3
(results not shown) none of the plaques appeared to be visible, even for the viruses producing the
largest plaques; however, by day 5, four of the five viruses produced plaques that could now be
clearly seen with the naked eye (figure 2). On day 5, RBV produced the largest plaques with a
mean diameter of 2.77 mm (sample std dev of 0.70 mm) followed by Yelrio with a mean
diameter of 2.3 mm (std dev 0.65 mm). The Zikrio plaques more closely mimicked that of the
ZIKV plaques which were the third largest plaques. ZIKV plaques had a mean diameter of 1.69 mm (std dev 0.52 mm) while Zikrio had a mean diameter of 1.29 mm (std dev 0.39 mm). YFV plaques were not yet visible at this time point. By day 7, the RBV plaques had more than doubled in size and had a mean diameter of 7.60 mm (std dev 1.46 mm). Yelrio virus again produced the second largest plaques at this time point, appearing to exhibit more of an intermediate plaque phenotype when compared to the two parent viruses: RBV, from which the prM and E genes were derived, and YFV, from which the rest of the genome was derived. This virus had a mean diameter of 4.66 mm (std dev 0.87 mm). By this day 7 time point, the YFV plaques had now become visible, averaging a diameter of 1.77 mm (std dev 0.73 mm). ZIKV plaques had a mean diameter of 2.61 mm (std dev 1.10 mm) while Zikrio plaques averaged 2.35 mm (std dev 0.47 mm). In summary, Yelrio plaques were more similar in size to RBV plaques than to YFV plaques on day 5, but by day 7 exhibited a more intermediate plaque size. Zikrio plaques, however, were more similar in size to ZIKV plaques than to RBV plaques on either day.

Discussion

Here we report the findings of a study that was conducted to investigate the genetic elements that govern flaviviral host specificity via characterization of chimeric viruses generated between representative dual-host flaviviruses and one representative bat-associated NKV flavivirus. We originally sought to investigate host restriction in both insects and vertebrates at the attachment and entry level, however, despite multiple attempts, we were not able to generate functional chimeras between the insect-specific and arthropod-borne viruses used.

In spite of the obstacles faced with our insect-specific flavivirus chimeras, we were able to generate two functional chimeras encoding the prM and E genes of an NKV flavivirus in a dual-host flavivirus backbone. Prior to our study, four other studies had been conducted wherein
the authors had successfully generated chimeric viruses using a number of different dual-host and NKV flaviviruses. However, those studies were all done using a rodent-associated NKV, MODV. Until now, no such studies had been conducted using a bat-associated NKV chimeric flavivirus. Charlier and colleagues successfully generated viral chimeras using YFV and DENV backbones carrying the surface proteins of MODV. The results of their study led the authors to conclude that host-restriction in mosquito cells was caused by intracellular host factors as opposed to an attachment/entry event mediated by the prM and E proteins (Charlier, Molenkamp et al. 2004, Charlier, Davidson et al. 2010).

Another study conducted by our research group utilized a reciprocal chimera which was generated between MODV and West Nile viruses (WNV). This chimera contained the prM and E genes of WNV embedded in a MODV backbone. The authors were able to successfully recover the virus from vertebrate cells infected, however mosquito cells appeared to be refractory to viral infection (Saiyasombat, Carrillo-Tripp et al. 2014).

The two other chimeras were generated by replacing the conserved pentanucleotide sequence and variable region (respectively) of DENV, with MODV sequences. Both of these elements are located in the 3’ untranslated region of the flavivirus genome which plays a role in the evasion of host anti-viral responses, genome packaging, replication and translation (Schuessler, Funk et al. 2012, Tumban, Maes et al. 2013, Pijlman 2014, Roby, Pijlman et al. 2014, Mazeaud, Freppel et al. 2018). The pentanucleotide sequence motif in the 3’UTR (CACAG) has been shown to be very well conserved among dual-host flaviviruses while exhibiting minor changes in NKV flaviviruses (Charlier, Leyssen et al. 2002, Khromykh, Kondratieva et al. 2003, Tilgner, Deas et al. 2005, Silva, Molenkamp et al. 2007). In their study however, Tumban and colleagues concluded that neither of the two elements swapped were necessary for expansion of host range.
as the changes did not appear to affect efficiency of infection in mosquitoes (Tumban, Maes et al. 2013).

We, too, attempted the creation of three different chimeras or mutants. The first two were created by swapping out the conserved pentanucleotide sequence of YFV with that of either Apoi virus (APOIV) or RBV. The third was a mutant which was created by deleting one of the conserved sequences (CS1) in the 3’untranslated region of YFV (Hahn, Hahn et al. 1987). Unfortunately, both chimeric viruses-APOIV and RBV- reverted back to wild type while the delta cs1 mutant failed to produce virus.

In the end, we decided to proceed with the two functional chimeras generated. In vitro host range studies done with the first chimera, Zikrio, showed permissive replication in cells of both vertebrate and insect origin, namely mosquito, duck and monkey. The fact that the viral replication was supported by monkey and duck cells did not come as a surprise as RBV replication is also supported by all three cell lines. RBV replication is, however, not supported by C6/36 mosquito cells (or other insect cells in general), a cell line which did support the growth of Zikrio (Hendricks, Hardy et al. 1983, Hendricks, Patick et al. 1988, Billoir, de Chesse et al. 2000, Volkova, Tesh et al. 2012). As the prM and E genes of Zika virus, genes that mediate attachment and entry, were swapped out for those of RBV, we conclude that host restriction for NKV flaviviruses, in this case RBV, occurs at a downstream stage post-attachment/entry. These findings are similar to those reported by Charlier and colleagues in 2010. However, this group used the prM and E genes of MODV which is a rodent-associated NKV flavivirus, whereas our chimeras were created with RBV, a bat-associated NKV flavivirus. Our chimeric viruses represented the first to be created using a bat-associated NKV.
One surprising finding that we did note was the fact that Zikrio did not appear to replicate in either BHK-21 or BSR-T7/5 cells. This was indeed very surprising as BHK-21 cells were the cells used in transfection experiments. It was also surprising because BSR-T7/5 cells also support the replication of both of the parental viruses. Upon sequencing of the Zikrio genome, there was one missense mutation in the prME region of the genome, one in NS1 and two in NS5. This mutation in the prME could have been enough to alter the host range of the virus. A study done by Goo and colleagues in 2017 showed that a single mutation in the flaviviral E protein was enough to have an effect on viral stability, pathogenesis and antigenicity (Goo, VanBlargan et al. 2017). Yet another study done by Monath and colleagues in 2002 showed that one mutation in the hinge region of this protein led to decreased viscerotropism for monkeys (Monath, Arroyo et al. 2002). Other studies have also shown expansion or restriction of host range resulting from single amino acid changes in surface proteins (Aytay and Schulze 1991, Rainey, Natanson et al. 2003, Martinez, Ndungo et al. 2013). Another less likely possibility could be that the mutations present in the NS1 or NS5 proteins presented a level of host restriction. The NS5 protein plays a role in viral immune evasion, replication and methylation of viral RNAs (Zhou, Ray et al. 2007, Davidson 2009, Issur, Geiss et al. 2009, Zou, Chen et al. 2011, Best 2017). The NS1 protein also plays a role in replication, pathogenesis and immune evasion (Muller and Young 2013, Rastogi, Sharma et al. 2016). It is therefore possible, though unlikely, that mutations in either one or both of these proteins is interfering with replication or inhibiting the antiviral activity of this virus. A final possibility is that this virus faces opposition at the level of attachment and entry but not at any of the other downstream stages in the viral life cycle. Transfection of the chimera into the cells would allow it to bypass that requirement, however, it would again encounter opposition after egress, when attempting to infect new cells.
A number of arbovirus chimeras have been described, many of which were created for vaccine purposes (Pletnev, Bray et al. 2001, Komar, Langevin et al. 2009, Chattopadhyay, Wang et al. 2013, Suzuki, Ishikawa et al. 2014, Erasmus, Needham et al. 2015, Erasmus, Seymour et al. 2018, Lei, Takeda et al. 2018). Differences in viral replication patterns, when compared to parental viruses, vary by study and would suggest that multiple different elements contribute to growth kinetics. A study done by Junglen and colleagues in 2017 led to the generation of a chimera which replicated less efficiently than the parental virus from which most of its genes were derived, reaching lower peak titers in the cell line tested even when both parental viruses replicated in this cell line (Junglen, Korries et al. 2017). Another study done by Yang and colleagues in 2018 generated a chimera which also replicated less efficiently than its parental viruses even when replication of both parental viruses was supported in the cell line used (Yang, Yang et al. 2018). Conversely, a study done with a chimera constructed using the prME of ZIKV virus in a YFV-17D backbone actually resulted in a chimera which produced higher amounts of viral genetic material up to 5 days post infection in Vero cells (Touret, Gilles et al. 2018).

Another chimera generated between two different alphaviruses, Eilat (EILV) and Chikungunya (CHIKV) viruses, showed similar results reaching higher peak titers than those previously reported for EILV from which its non-structural genes were derived (Erasmus, Auguste et al. 2017). Yet another study conducted in 2015 showed that a chimera generated between two other alphaviruses, Sindbis virus (SINV) and EILV, exhibited replication kinetics that were similar to that of their parent viruses, although the authors did see a decline in viral titers after 72 hours (Nasar, Gorchakov et al. 2014). In our studies using Vero cells, Zikrio peaked on the same day as ZIKV at an average of 3.75 log7 while RBV peaked much later but with a higher average of 8.6 log7. After peaking, ZIKV titers exhibited a rather sharp decline in titer as opposed to Zikrio
which exhibited a more steady decline. Yelrio exhibited a similar replication pattern peaking on day 4 at a titer higher than, although closer to, YFV but lower than RBV which peaked a day later. Both chimeras had peak titers that were slightly higher than, although closer to, the parental virus from which the non-structural genes were derived, but lower than that from which the prM and E genes were derived.

Plaque assay analysis was also done with four of our five viruses which had been shown to replicate in mosquito, specifically C6/36 cells. Although these cells seemed to support the replication of both chimeras, the replication efficiency was greatly reduced. This may be due to a reduced binding affinity given the fact that the prM and E proteins supplied here have been specialized for vertebrate and not mosquito cell entry.

Plaque sizes also differed among parental viruses and the two chimeras generated. Interestingly, Yelrio exhibited a plaque size more similar to that of RBV than YFV (day 5) whereas Zikrio exhibited a plaque size more similar to that of ZIKV than that of RBV even if the prM and E genes for both chimeras were derived from RBV. By day 7, however, Yelrio exhibited more of an intermediate plaque size when compared to both parental viruses. This would suggest that genetic elements both within and outside of the prM and E proteins dictate plaque size. Studies done with other chimeras have also shown that genetic elements both within and outside the prME influence plaque size (Bray and Lai 1991, Maharaj, Anishchenko et al. 2012, McAuley, Torres et al. 2016, Amarilla, Setoh et al. 2017, Li, Dong et al. 2018, Yang, Yang et al. 2018). A study done by Amarilla and colleagues in 2017 utilizing reciprocal viral chimeras generated by swapping the prM and E genes of one out for the other, showed that while one chimera produced plaques that were larger than both parental viruses, the other produced plaques that were smaller than one parental virus and similar in size to the other (Amarilla, Setoh et al.}
2017). Taken together, we see that the genetic elements that govern plaque size differ from virus to virus and may exist both within or outside the envelope/surface proteins.

As previously mentioned, we also attempted the generation of insect-specific chimeric viruses to study host restriction in insect-specific flaviviruses. To date, only two such studies have been conducted using representative insect-specific flaviviruses and dual-host flaviviruses. The first study was conducted by Junglen and colleagues in 2017 where the authors generated chimeric viruses between a newly discovered insect-specific flavivirus, Niénokoué virus (NIEV) and a dual-host flavivirus, YFV. Based on their findings, the authors concluded that insect-specific viruses faced opposition at multiple different stages in their life cycle including attachment, entry, assembly and release (Junglen, Korries et al. 2017). Another study which was published only months later, reported similar findings with a chimera generated between WNV, a dual-host arbovirus and Palm Creek virus (PCV), an insect-specific flavivirus (Piyasena, Setoh et al. 2017). Insect-specific flaviviruses, however, can be further divided into two different subgroups based on phylogenetic affiliation. The first group, designated classical insect-specific flaviviruses, form a completely distinct clade, while the second, the dual-host affiliated insect-specific flaviviruses, in spite of their apparent insect-restricted phenotype, group with the dual-host flaviviruses. Both of the insect-specific flaviviruses used in the aforementioned studies are classical insect-specific flaviviruses (Blitvich and Firth 2015, Calzolari, Ze-Ze et al. 2016, Guzman, Contreras-Gutierrez et al. 2018). Given the fact that dual-host affiliated insect-specific flaviviruses are phylogenetically distinct from the classical insect-specific flaviviruses, it is possible that they may face opposition at fewer/different levels. With that in mind, we attempted the creation of chimeras between a representative dual-host affiliated flavivirus, Long Pine Key virus and a dual-host flavivirus, YFV. We were able to recover virus from neither mosquito
(C6/36) nor vertebrate (BHK) cells transfected. This may be due to incompatibilities between the different proteins used.

In summary, we report the creation and recovery of two flavivirus chimeras generated by replacing the prM and E proteins of two respective dual-host flaviviruses with the corresponding genes of a bat-associated NKV flavivirus, RBV. These chimeras were designed to study host restriction of NKV flaviviruses in insect cells at the attachment and entry level. Both chimeric viruses demonstrated the ability to replicate in vertebrate as well as mosquito cells, suggesting that host restriction in NKV flaviviruses occurs at a post-attachment/entry stage. Further studies should be conducted using chimeras constructed using different combinations of flaviviral gene swaps to determine possible stages where these viruses face opposition.

References


Table 1. *In vitro* host range of Zikrio. *In vitro* host range was assessed using eleven different cell lines of insect and vertebrate origin.

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Zika virus</th>
<th>Rio Bravo virus</th>
<th>Zikrio virus</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mosquito cells</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6/36</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>CT</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sua 4.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>Lepidopteran cells</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Five</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sf9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>Vertebrate cells</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSR-T7/5</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>DEF</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>HeLa</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LLC-MK2</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Murine microglia</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Vero</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Table 2a. Comparison of the replication kinetics of ZIKV, Zikrio, YFV, Yelrio and RBV in Vero cells.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Virus titer in pfu/mL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1</td>
</tr>
<tr>
<td>ZIKV</td>
<td>1.96log5</td>
</tr>
<tr>
<td>Zikrio</td>
<td>7.50log4</td>
</tr>
<tr>
<td>YFV</td>
<td>1.25log4</td>
</tr>
<tr>
<td>Yelrio</td>
<td>1.67log2</td>
</tr>
<tr>
<td>RBV</td>
<td>1.81log5</td>
</tr>
</tbody>
</table>
Table 2b. Comparison of the replication kinetics of ZIKV, Zikrio, YFV and Yelrio in C6/36 cells.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 6</th>
<th>Day 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZIKV</td>
<td>5.16log4</td>
<td>9.4log4</td>
<td>1.22log5</td>
<td>6.25log5</td>
<td>3.15log5</td>
<td>2.59log5</td>
<td>2.42log5</td>
</tr>
<tr>
<td>Zikrio</td>
<td>9.3log3</td>
<td>1.28 log4</td>
<td>6.5log3</td>
<td>1.82log4</td>
<td>2.35log3</td>
<td>5.25log3</td>
<td>1.32log4</td>
</tr>
<tr>
<td>YFV</td>
<td>1.4log4</td>
<td>5.7log4</td>
<td>2.33log5</td>
<td>2.45log5</td>
<td>4.25log5</td>
<td>3.57log5</td>
<td>3.32log5</td>
</tr>
<tr>
<td>Yelrio</td>
<td>4.0log2</td>
<td>3.13log2</td>
<td>1.83log2</td>
<td>3.95log2</td>
<td>1.05log3</td>
<td>4.4log2</td>
<td>5.8log2</td>
</tr>
</tbody>
</table>
Table 3. Comparison of the plaque sizes of Zikrio, Yelrio and the parental viruses on Vero cells. Plaque diameter was measured at 5- and 7-days post infection using image J software.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Size of plaques ± sample standard deviation (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 5</td>
</tr>
<tr>
<td>ZIKV</td>
<td>1.69 ± 0.52</td>
</tr>
<tr>
<td>Zikrio</td>
<td>1.29 ± 0.39</td>
</tr>
<tr>
<td>YFV</td>
<td>0</td>
</tr>
<tr>
<td>Yelrio</td>
<td>2.3 ± 0.65</td>
</tr>
<tr>
<td>RBV</td>
<td>2.77 ± 0.70</td>
</tr>
</tbody>
</table>
Table 4. Mutations accrued in the genome of Zikrio and Yelrio virus chimera.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Nucleotide position</th>
<th>Amino Acid position</th>
<th>Nucleotide change</th>
<th>Amino acid change</th>
<th>Protein affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zikrio</td>
<td>1552</td>
<td>489</td>
<td>A-&gt;C</td>
<td>Asp-&gt;Ala</td>
<td>Envelope</td>
</tr>
<tr>
<td></td>
<td>2390</td>
<td>768</td>
<td>A-&gt;G</td>
<td>Silent</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3192</td>
<td>1036</td>
<td>A-&gt;G</td>
<td>Lys-&gt;Glu</td>
<td>NS1</td>
</tr>
<tr>
<td></td>
<td>3737</td>
<td>1217</td>
<td>C-&gt;T</td>
<td>Silent</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>7373</td>
<td>2429</td>
<td>C-&gt;T</td>
<td>Silent</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>8193</td>
<td>2703</td>
<td>T-&gt;C</td>
<td>Silent</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>8899</td>
<td>2938</td>
<td>A-&gt;G</td>
<td>Glu-Gly</td>
<td>NS5</td>
</tr>
<tr>
<td></td>
<td>9610</td>
<td>3175</td>
<td>A-&gt;G</td>
<td>His-&gt;Arg</td>
<td>NS5</td>
</tr>
<tr>
<td>Yelrio</td>
<td>142</td>
<td>8</td>
<td>G-&gt;A</td>
<td>Silent</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1134</td>
<td>339</td>
<td>T-&gt;C</td>
<td>Leu-Pro</td>
<td>Envelope</td>
</tr>
<tr>
<td></td>
<td>4030</td>
<td>1304</td>
<td>T-&gt;C</td>
<td>Silent</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>6001</td>
<td>1961</td>
<td>T-&gt;C</td>
<td>Silent</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>6256</td>
<td>2046</td>
<td>G-&gt;C</td>
<td>Silent</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>9886</td>
<td>3256</td>
<td>C-&gt;T</td>
<td>Silent</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>10,120</td>
<td>3334</td>
<td>G-&gt;A</td>
<td>Silent</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 1. Schematic illustrating strategy used to generate chimeras
Figure 2. Comparison of the plaque sizes of A) Yelrio and B) Zikrio and the parental viruses. Confluent monolayers of Vero cells were infected with individual viruses and then stained and imaged 5- and 7-days post infection.
Figure 2 continued…

Abbreviations:

RBV- Rio Bravo virus; YRV- Yelrio virus; YFV- Yellow fever virus

ZRV- Zikrio virus; ZIKV- Zika Virus
### Supplementary material

#### Table 1a. Primers used during the construction of full-length Yelrio DNAs

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Polarity</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>YRCMVFor</td>
<td>sense</td>
<td>GGGGTTCCGCGCACATTTCC</td>
</tr>
<tr>
<td>YRCMVRev</td>
<td>antisense</td>
<td>CACACAGGATTATTACTCGGTTCACTAAACGAG</td>
</tr>
<tr>
<td>YRCapFor</td>
<td>sense</td>
<td>GTTTAGTGAACCCGAGTAAAAATCCTGTGTC</td>
</tr>
<tr>
<td>YRCapRev</td>
<td>antisense</td>
<td>CATAGCCATCAATCCACCCCTGTCAACACAGC</td>
</tr>
<tr>
<td>YRprMEFor</td>
<td>sense</td>
<td>CGGGTGGATTTGATGGCTATGCAAGGTGATCCC</td>
</tr>
<tr>
<td>YRprMERev</td>
<td>antisense</td>
<td>GCATCCTTGATCTCCCCATGACTCCAGTTGTAATC</td>
</tr>
<tr>
<td>YRF4For</td>
<td>sense</td>
<td>GAGTCATGGGAGATCAAGGATCGCCCATCAAACCTTTGG</td>
</tr>
<tr>
<td>YRF4Rev</td>
<td>antisense</td>
<td>CTGTTATTGAATTCCAGCCACTGACC</td>
</tr>
<tr>
<td>YRF5For</td>
<td>sense</td>
<td>GGTCAGTGGCTGGAATTTCAATAACAG</td>
</tr>
<tr>
<td>YRF5Rev</td>
<td>antisense</td>
<td>TGCCGGCCAGAGGTTTTTGTGT</td>
</tr>
<tr>
<td>YRRiboFor</td>
<td>sense</td>
<td>CAAAAACCTCTGGCCCGGATGGTC</td>
</tr>
<tr>
<td>YRRiboRev</td>
<td>antisense</td>
<td>GCTGCGCAACTTGGGGGAAG</td>
</tr>
</tbody>
</table>

#### Table 1b. PCR products generated for construction of Yelrio DNAs

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Primers</th>
<th>Position in genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMV</td>
<td>ZRCMVFor + ZRCMVRev</td>
<td>1-781</td>
</tr>
<tr>
<td>5’UTR to capsid</td>
<td>ZRCapFor + ZRCapRev</td>
<td>1-473</td>
</tr>
<tr>
<td>prME</td>
<td>ZRprMEFor + ZRprMERev</td>
<td>423-2369</td>
</tr>
<tr>
<td>Fragment 4</td>
<td>ZRF4For + ZRF4Rev</td>
<td>2490-7077</td>
</tr>
<tr>
<td>Fragment 5</td>
<td>ZRF5For + ZRF5Rev</td>
<td>7052-10,807</td>
</tr>
<tr>
<td>Ribo+poly(A)</td>
<td>ZRRiboFor + ZRRiboRev</td>
<td>1-192</td>
</tr>
</tbody>
</table>
Table 2a. Primers used during the construction of full-length Zikrio DNAs

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Polarity</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZRCMVFor</td>
<td>sense</td>
<td>GAATAAGGGCCGACACGGAAATGTC</td>
</tr>
<tr>
<td>ZRCMVRev</td>
<td>antisense</td>
<td>CAACAACTCGGTTCACTAAACGAGCTCT</td>
</tr>
<tr>
<td>ZRCapFor</td>
<td>sense</td>
<td>CGTTTACGTGACCCGAGTTTGGTCTGTGTGAATC</td>
</tr>
<tr>
<td>ZRCapRev</td>
<td>antisense</td>
<td>GCCATCAATGCCCATTAGCTGTGGTCAGC</td>
</tr>
<tr>
<td>ZRprMEF</td>
<td>sense</td>
<td>CACAGCTATGGCATTGGCTATGCAGG</td>
</tr>
<tr>
<td>ZRPrMEREv</td>
<td>antisense</td>
<td>CCCACATCTCCCATGACTCCAGTTGAAATC</td>
</tr>
<tr>
<td>ZRF4For</td>
<td>sense</td>
<td>GTCATGGGAGATGTGGGGTGCTC</td>
</tr>
<tr>
<td>ZRF4Rev</td>
<td>antisense</td>
<td>GGACGGCTGCTGGCTTGGATAATGAAATGTC</td>
</tr>
<tr>
<td>ZRF5For</td>
<td>sense</td>
<td>GACAACCTTTTCAATCCCCGACGCTCC</td>
</tr>
<tr>
<td>ZRF5Rev</td>
<td>antisense</td>
<td>TGCCGGCCAGACCATGGATTTC</td>
</tr>
<tr>
<td>ZRRiboFor</td>
<td>sense</td>
<td>TGGGTCTGCGCCGCGATGGTC</td>
</tr>
<tr>
<td>ZRRiboRev</td>
<td>antisense</td>
<td>CTCAGGGTGCAATGCCAGCGCTTAATTTC</td>
</tr>
</tbody>
</table>

Table 2b. PCR products generated for construction of Zikrio DNAs

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Primers</th>
<th>Position in genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMV</td>
<td>YRCMVFor + YRCMVRev</td>
<td>1-858</td>
</tr>
<tr>
<td>5'UTR to</td>
<td>YRCapFor + YRCapRev</td>
<td>1-481</td>
</tr>
<tr>
<td>capsid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prME</td>
<td>YRprMEFor + YRprMEREv</td>
<td>423-2369</td>
</tr>
<tr>
<td>Fragment 4</td>
<td>YRF4For + YRF4Rev</td>
<td>2453-7202</td>
</tr>
<tr>
<td>Fragment 5</td>
<td>YRF5For + YRF5Rev</td>
<td>7177-10,862</td>
</tr>
<tr>
<td>Ribo+poly(A)</td>
<td>YRRiboFor + YRRiboRev</td>
<td>1-379</td>
</tr>
</tbody>
</table>
CHAPTER 3. DETECTION OF NOVEL AND RECOGNIZED RNA VIRUSES IN MOSQUITOES FROM THE YUCATAN PENINSULA OF MEXICO USING METAGENOMICS AND CHARACTERIZATION OF THEIR IN VITRO HOST RANGES

Modified from a manuscript published in Journal of General Virology 2018;99:1729–1738

Jermilia Charles,1 Chandra S. Tangudu,1 Stefanie L. Hurt,1 Charlotte Tumescheit,2 Andrew E. Firth,2 Julian E. Garcia-Rejon,3 Carlos Machain-Williams3 and Bradley J. Blitvich1

1Department of Veterinary Microbiology and Preventive Medicine, College of Veterinary Medicine, Iowa State University, Ames, IA 50011.

2 Department of Pathology, University of Cambridge, Cambridge, UK

3 Laboratorio de Arbovirologia, Centro de Investigaciones Regionales ‘Dr Hideyo Noguchi’, Universidad Autónoma de Yucatán, Mérida, Yucatán, Mexico

Corresponding author: 2116 Veterinary Medicine, Iowa State University, Ames, IA 50011 (e-mail: blitvich@iastate.edu).

Abstract

A metagenomics approach was used to detect novel and recognized RNA viruses in mosquitoes from the Yucatan Peninsula of Mexico. A total of 1359 mosquitoes of 7 species and 5 genera (Aedes, Anopheles, Culex, Mansonia and Psorophora) were sorted into 37 pools, homogenized and inoculated onto monolayers of Aedes albopictus (C6/36) cells. A second blind passage was performed and then total RNA was extracted and analysed by RNA-seq. Two novel viruses, designated Uxmal virus and Mayapan virus, were identified. Uxmal virus was isolated from three pools of Aedes (Ochlerotatus) taeniorhynchus and phylogenetic data indicate that it should be classified within the recently proposed taxon Negevirus. Mayapan virus was recovered
from two pools of *Psorophora ferox* and is most closely related to unclassified *Nodaviridae*-like viruses. Two recognized viruses were also detected: Culex flavivirus (family *Flaviviridae*) and Houston virus (family *Mesoniviridae*), with one and two isolates being recovered, respectively. The *in vitro* host ranges of all four viruses were determined by assessing their replicative abilities in cell lines of avian, human, monkey, hamster, murine, lepidopteran and mosquito (*Aedes, Anopheles* and *Culex*) origin, revealing that all viruses possess vertebrate replication-incompetent phenotypes. In conclusion, we report the isolation of both novel and recognized RNA viruses from mosquitoes collected in Mexico, and add to the growing plethora of viruses discovered recently through the use of metagenomics.

**Introduction**

The advent of next-generation sequencing (NGS) has revolutionized virus discovery [1–5]. This technology allows for the rapid and inexpensive detection of nucleic acids in diverse biological and environmental samples and offers distinct advantages over conventional virus detection techniques. Unfortunately, NGS is not always complemented by virus isolation experiments and therefore many newly discovered viruses are known solely from sequence data [6, 7]. Although many viruses cannot be easily propagated in cell culture, the importance of virus isolation cannot be understated because a newly discovered virus cannot be phenotypically characterized if an isolate is unavailable. Information on host range, transmissibility, pathogenesis, antigenicity and epidemiology are essential for the development and implementation of effective virus surveillance, diagnosis, control and prevention strategies.

A rapidly increasing number of novel RNA viruses have been identified in mosquitoes by NGS in recent years [8–13]. One of the first mosquito-associated RNA viruses to be discovered
using this technology was Negev virus, the prototype virus of a newly proposed taxon, designated Negevirus, which also includes Dezidougou virus (DEZV) and Wallerfield virus (WALV) [8, 12]. Negeviruses have been isolated exclusively from mosquitoes and phlebotomine sandflies and are assumed to have insect-restricted host ranges [8, 12, 14–21]. Phylogenetic analyses have shown that negeviruses separate into two monophyletic clades (Nelorpivirus and Sandewavirus) and the amount of genetic diversity between viruses of different clades is sufficient to allow each clade to be elevated to the level of genus [20].

Negeviruses have a close phylogenetic relationship with plant viruses of the genera Cilevirus, Higreivirus and Blunervirus, some of which are arthropod-borne, and therefore it has been postulated that negeviruses possess the capacity to replicate in plants, although direct experimental evidence is lacking [12, 19, 20, 22].

Shi and colleagues recently discovered over a thousand novel putative RNA viruses in the largest NGS investigation of viruses ever performed on invertebrates [23]. This groundbreaking study provides valuable insights into virus biodiversity and evolution. More than 220 species of terrestrial and marine invertebrates from 9 metazoan phyla collected in China were examined. Among the numerous viruses discovered were multiple unclassified RNA viruses designated as Hubei noda-like virus 1 to 26. Phylogenetic studies revealed that Hubei noda-like viruses are closely related to viruses in the family Nodaviridae, a group of non-enveloped, bipartite RNA viruses that are classified according to their host associations. The family contains two established genera, Alphanodavirus (insect-associated) and Betanodavirus (fish-associated), and one proposed genus, Gammanodavirus (decapod-associated), besides a large number of other divergent but unclassified proposed members [23–25].
In the present study, mosquitoes of multiple species from the Yucatan Peninsula of Mexico were tested for the presence of novel and recognized RNA viruses. Mosquitoes were assayed by virus isolation in mosquito cell culture and then total RNA was extracted from all cultures, regardless of whether cytopathic effect (CPE) occurred, and examined for viral sequences using metagenomics. The replicative abilities of each novel and recognized virus were assessed using multiple vertebrate and invertebrate cell lines to determine their *in vitro* host ranges.

**Results**

**Mosquito collections and virus discovery**

This study was performed using 1359 mosquitoes of 7 species and 5 genera collected in the Yucatan Peninsula of Mexico in 2007 and 2008 (Table 1). The mosquitoes were sorted into 37 pools and homogenized, and then an aliquot of each homogenate underwent 2 blind passages in *Aedes albopictus* (C6/36) cells. Eight homogenates caused cell death and seven others caused mild to severe clumping or rounding of cells. Total RNA was extracted from all second passage cultures, regardless of whether CPE was observed, and analysed by RNA-seq for viral sequences. Two novel viruses, designated Uxmal virus (UXMV) and Mayapan virus (MYPV), were discovered (Table 2). The viruses were named after archaeological ruins in the Yucatan Peninsula of Mexico. Two recognized viruses were also isolated: Culex flavivirus (CxHV; genus *Flavivirus*, family *Flaviviridae*) and Houston virus (HOUV; a variant of the species Alphamesonivirus-1 of the genus *Alphamesonivirus*, family *Mesoniviridae*). An acronym has not been previously assigned to Houston virus, but HOUV is used in this paper. Virus sequences were not recovered from several samples that produced CPE, indicating that the metagenomics analysis did not identify all viruses.
Uxmal virus

UXMV was isolated from three pools of female Aedes (Ochlerotatus) taeniorhynchus collected in Merida (Table 2). All second-passage C6/36 cell cultures displayed 40–60% CPE at 3 to 4 days post-inoculation, as indicated by the presence of elongated cells, clumping of cells into a chain-like formation and cell death. The in vitro host range experiments confirmed that C6/36 cells are permissive to UXMV replication and revealed that Culex tarsalis (CT) cells also support replication (Table 3). Unlike for the C6/36 cells, CPE was not observed in the CT cells. Anopheles, lepidopteran and vertebrate cells did not support virus replication. Homogenates prepared from 268 male Ae. taeniorhynchus in 15 pools were also tested by reverse transcription/polymerase chain reaction (RT-PCR) for UXMV, and all were negative.

The genome of one isolate (designated UXMV-M985) was fully sequenced (GenBank accession no. MH719095). The sequence consists of 9227 nt, excluding the polyadenylate tail, and is most closely related to the genomes of WALV (69.5% nt identity with 67.4% coverage) and DEZV (66.9% nt identity with 70.0% coverage). The genomic organization of UXMV is characteristic of viruses in the taxon Negevirus, with the genome predicted to encode three major open reading frames (ORFs) that are separated by short (22 to 119 nt) intergenic regions and flanked by 5’ and 3’ untranslated regions (UTRs) of 235 nt and 286 nt, respectively (Fig. 1a).

ORF1 encodes a putative replicase protein of 2239 residues that is most closely related to the corresponding translation products of WALV (60.2% identity and 77.6% similarity, with 100% coverage) and DEZV (60.8% identity and 77.2% similarity, with 98.6% coverage). The putative protein contains predicted viral methyltransferase (pfam01660), 2’-O-methyltransferase (pdb3R24), helicase (pfam01443) and RNA-dependent RNA polymerase (RdRp; pfam00978).
domains at residues 82 to 413, 775 to 965, 1312 to 1579 and 1783 to 2220, respectively. These domains are highly conserved among the replicase proteins of negeviruses [12, 19].

ORF 2 is predicted to encode a membrane-associated protein of 413 residues. The putative protein is most closely related to the corresponding translation products of WALV (44.9% identity and 64.7% similarity, with 100% coverage) and DEZV (35.3% identity and 54.4% similarity, with 98.3% coverage). The putative protein is predicted to contain an amino-terminal signal peptide (residues 1 to 16), in addition to three carboxy-terminal transmembrane domains (residues 337 to 356, 368 to 389 and 395 to 410) and two carboxy-terminal cytoplasmic domains (residues 357 to 367 and 411 to 413), which is consistent with the ORF 2-encoded translation products of other negeviruses [16, 17, 26]. Three potential N-linked glycosylation sites are also present (residues 178, 244 and 269).

ORF 3 is predicted to encode a membrane-associated protein of 200 residues. The putative protein is most closely related to the corresponding translation products of WALV (63.0% identity and 82.5% similarity, with 100% coverage) and DEZV (62.1% identity and 75.6% similarity, with 99.5% coverage). The putative protein is predicted to contain four transmembrane domains (residues 49 to 71, 101 to 121, 133 to 158 and 165 to 183) and two cytoplasmic domains (residues 72 to 100 and 159 to 164). Two potential N-linked glycosylation sites are also present (residues 42 and 72). An additional ORF, which we have called ORF3a, spans nucleotide positions 8236 to 8448. The ORF overlaps the 5’ end of ORF3 and is conserved in sandewaviruses (except the divergent Hubei virga-like virus 7 and Lodeiro virus) and encodes a cysteine-rich putative protein of 70 residues (Fig. 2a). ORF3a mostly fills a previously presumed non-coding gap between ORFs 2 and 3 (Fig. 1a).
A phylogenetic tree was constructed with Bayesian methods using the deduced amino acid sequence of the RdRp of UXMV and the corresponding regions of related sequences (Fig. 3). UXMV is most closely related phylogenetically to WALV and DEZV, and the posterior support for this topological arrangement is 1.0. Most of the viruses used in the analysis belong to one of two previously established clades: *Nelorpivirus* (group I) and *Sandewavirus* (group II). UXMV is a group II negavirus.

**Mayapan virus**

MYPV was isolated from two pools of female *Psorophora ferox* collected in Tzucacab (Table 2). CPE was not observed in the first-or second-passage C6/36 cell cultures. The *in vitro* host range experiments revealed that MYPV has a narrow host range because only C6/36 cells were permissive to virus infection (Table 3). Male *Ps. ferox* were not tested for MYPV because none were available in our archived collections.

The genomic organization of MYPV has similarities with that of viruses in the family *Nodaviridae*. Viruses classified within this family have positive-sense, single-stranded, bipartite RNA genomes and the two segments are denoted as RNA1 and RNA2 [27]. RNA1 (3.1 kb) encodes the RdRp and RNA2 (1.4 kb) encodes the capsid protein. A subgenomic RNA (sgRNA3; 0.4 kb) is produced from RNA1 and encodes one or two small proteins, B1 and B2. The function of B1, which is identical to the C-terminus of the RdRp, is unknown, and B2 is a suppressor of RNA interference (RNAi) [28]. An additional subgenomic RNA (0.3 kb) is uniquely produced from RNA1 of mosinovirus, a nodavirus isolated from *Culicidae* spp. mosquitoes in Cote d’Ivoire [29].

Two non-overlapping MYPV sequences were recovered, indicating that the virus has a bipartite genome (GenBank accession nos MH719096 and MH719097; Fig. 1b). A less likely
explanation is that we recovered genome segments from two distinct but related viruses. RNA1 consists of 3045 nt and is most closely related to the corresponding region of Hubei noda-like virus 5 (82.3% identity with 98.6% coverage), an unclassified Nodaviridae-like virus detected in spiders and insects in China [23]. The detection of Hubei noda-like virus 5 in spiders suggests that the virus naturally infects spiders or was acquired from recently eaten insects. RNA1 of MYPV is predicted to encode a large ORF flanked by 5′ and 3′ UTRs of 39 and 171 nt, respectively. The putative protein consists of 944 residues and is most closely related to the corresponding translation product of Hubei noda-like virus 5 (90.8% identity and 95.2% similarity, with 97.6% coverage), followed by mosinovirus (43.2% identity and 60.7% similarity, with 96.3% coverage). Domains characteristic of RdRps (cd01699) and dephospho-coenzyme A kinases (cd02022) are present at residues 469 to 685 and 283 to 349, respectively. A short ORF (nt 2591 to 2932) overlapping the 3′ end of the RdRp ORF is also conserved in Hubei noda-like virus 5 and might encode a B2 protein of 113 aa (Figs 1b and 2b).

RNA2 of MYPV consists of 1446 nt and is most closely related to the corresponding region of Hubei noda-like virus 11 (69.2% identity and 12.1% coverage), an unclassified Nodaviridae-like virus detected in spiders in China [23]. The sequence is predicted to encode a single ORF flanked by 5′ and 3′ UTRs of 71 and 142 nt, respectively. The putative protein consists of 410 residues and is most closely related to a corresponding translation product of Hubei noda-like virus 11 (44.5% identity and 58.8% similarity, with 89.6% coverage). A domain homologous to the alphanoda-like capsid protein is present at residues 55 to 364.

Phylogenetic trees were constructed with Bayesian methods using the deduced amino acid sequences of the RdRp and capsid protein of MYPV and the corresponding regions of related viruses (Fig. 4). In the analysis performed using RdRp sequences, it was demonstrated
that MYPV is most closely related to Hubei noda-like virus 5. The posterior support for this
topological arrangement is 1.0. MYPV and Hubei noda-like virus 5 belong to a clade that
includes selected other unclassified Nodaviridae-like viruses. A second major clade consists of
additional unclassified Nodaviridae-like viruses, along with viruses in the genus Betanodavirus.
Notably, the RdRp of MYPV appeared to be more closely related to RdRps of betanodaviruses
than to RdRps of alphanodaviruses such as Flock House, black beetle and Nodamura viruses.
Alphanodaviruses were not included in the tree because their RdRps have limited (about 25%)
amino acid identity (with 70–80% coverage) to that of MYPV. In contrast, in the analysis
performed using capsid protein sequences, MYPV was positioned basally to viruses in the
genus Alphanodavirus. The alphanodavirus capsid protein is structurally very different from the
betanodavirus capsid protein [30], which is why betanodaviruses do not appear in the capsid
protein tree.

**Recognized viruses**

CxFV was isolated from a pool of male *Cx. quinquefasciatus* collected in Tixkokob 2008
( Table 2 ). C6/36 cells were permissive to CxFV replication, but the virus was unable to
replicate in any other cell line ( Table 3 ). The genome of the CxFV isolate (designated CxFV-
T1123) was fully sequenced (GenBank accession no. MH719098) and shown to consist of 10837
nt. blast analysis revealed that CxFV-T1123 has the greatest (99.8%) nucleotide identity with
CxFV-Mex07, an isolate of CxFV recovered from *Cx. quinquefasciatus* in Tixkokob in 2007
[31]. HOUV was isolated from two pools of female mosquitoes collected in Merida ( Table 2 ).
One isolate was recovered from *Cx. quinquefasciatus* and the other from *Ae. taeniorhynchus.*
HOUV replicated in C6/36 and CT cells, with CPE only being observed in C6/36 cells ( Table
3). The entire genome of one isolate (designated HOUV-M742) was fully sequenced and shown
82
to consist of 20129 nt, excluding the polyadenylate tail (GenBank accession number
MH719099). blast analysis revealed that HOUV-M742 has the greatest (99.2%) nucleotide
identity with V3872, an isolate of HOUV recovered from Cx. quinquefasciatus in Houston in
2004 [32].

Discussion

We report the isolation and identification of two novel viruses and two recognized
viruses in mosquitoes from the Yucatan Peninsula of Mexico. Sequences corresponding to a
putative novel rhabdovirus (Merida virus) were identified in mosquitoes in the same area in an
earlier metagenomics study, but an isolate was not recovered by virus isolation in C6/36 cells,
even though the cell line supports the replication of numerous mosquito-associated viruses [33].
Other research groups have also discovered novel virus genomes in mosquitoes by NGS but
could not recover isolates from inoculated C6/36 cells [13, 34]. The inability to culture some
mosquito-associated viruses in C6/36 cells could be because these viruses are species-specific or
have only adapted to a few mosquito species and lack the capacity to replicate in Ae.
albopictus cells. The experimental approach used in this study was designed to avoid the
situation in which novel virus-like sequences were identified but isolates were not recovered. To
this end, the metagenomics analysis was preceded by virus isolation experiments and therefore
an isolate was already on hand for every virus that was identified. Our experimental approach
also allowed for the detection of both cytopathic and non-cytopathic viruses because NGS was
not restricted to cultures that exhibited CPE.

Our data demonstrate that UXMV should be classified in the taxon Negevirus. This
newly proposed monophyletic taxon is composed of viruses detected exclusively in mosquitoes
and sandflies, with most being discovered through the use of metagenomics [8, 12, 14–20].
Negeviruses are assumed to have vertebrate-incompetent replication phenotypes because they cannot replicate in suckling mice or any vertebrate cell lines that have been tested. The most comprehensive \textit{in vitro} host range experiments were performed by Vasilakis and colleagues, who assessed the replicative abilities of six negeviruses in three vertebrate cell lines [baby hamster kidney (BHK-21), human embryonic kidney (HEK-293) and Vero cells] and five mosquito cell lines \cite{alam2021negevirus} [\textit{Ae. albopictus} (C6/36 and C7/10), \textit{Anopheles albimanus}, \textit{An. gambiae} and \textit{Cx. tarsalis} cells], in addition to \textit{Drosophila melanogaster} and \textit{Phlebotomus papatasi} cells \cite{alam2021negevirus}. All viruses replicated in every mosquito cell line, but none replicated in any other cell line. We provide additional evidence that negeviruses lack the capacity to replicate in vertebrate cells because UXMV could not replicate in any of the six vertebrate cell lines tested. A notable difference between the two studies is that all previously tested negeviruses replicated in \textit{Anopheles} cells, while UXMV could not, suggesting that the mosquito host range of UXMV is not as broad as that of other negeviruses.

Negeviruses have a ubiquitous geographical distribution, having been reported in every continent with the exception of Antarctica \cite{alam2021negevirus}. The first negevirus to be identified in Mexico was Piura virus after its isolation from \textit{Aedes}, \textit{Culex}, \textit{Mansonia}, \textit{Psorophora} and \textit{Wyeomyia} spp. mosquitoes in the southern state of Chiapas \cite{alam2021negevirus}. Piura virus belongs to the \textit{Nelorpivirus} clade. UXMV is the only \textit{Sandewavirus} clade virus known to occur in Mexico.

MYPV is a \textit{Nodaviridae}-like virus that was recovered from \textit{Ps. ferox}. Two pools of female \textit{Ps. ferox} were tested and both yielded isolates, indicating that MYPV commonly infects this species in the study area. However, the sample size was small and additional females need to be tested to accurately determine the prevalence of MYPV in \textit{Ps. ferox}. Adult males and
immatures should also be tested to provide some indication of whether MYPV is maintained in nature by vertical transmission, but none are available in our archived collections. All vertebrate cell lines were refractory to virus replication, indicating that MYPV has a vertebrate-incompetent replication phenotype. The ability of Aedes, but not Anopheles or Culex, cells to support MYPV replication indicates that the virus has a narrow mosquito host range. C6/36 cells have a dysfunctional RNAi response and it remains to be determined whether MYPV can establish an infection in Aedes cells that possess a functional RNAi pathway [35, 36]. Recent studies have also revealed that CT cells possess a dysfunctional RNAi response [37].

The isolation of CxFV from Cx. quinquefasciatus was not unexpected because we have previously reported a high prevalence of CxFV in mosquitoes of this species in the Yucatan Peninsula of Mexico [31, 38, 39]. CxFV is assumed to have an insect-restricted host range because it has been isolated exclusively from mosquitoes and cannot replicate in any vertebrate cell lines that have been tested [40–42]. Our study provides additional evidence that the virus is insect-specific, with several previously untested vertebrate cell lines being refractory to infection. The inability of CT cells to support CxFV infection was not unexpected because the cell line is persistently infected with Calbertado virus, a closely related insect-specific flavivirus, which is potentially suppressing the replication of CxFV by superinfection exclusion [43].

HOUV was originally isolated from Ae. albopictus and Cx. quinquefasciatus in the United States and was then isolated from Cx. quinquefasciatus in Xkaladzonot, a rural town in the Yucatan Peninsula of Mexico [32, 44]. HOUV is a variant of the species Alphamesonivirus-1, which also includes Cavally virus from Cote d’Ivoire, Nam Dinh virus from Vietnam and China, and Ngewotan virus from Indonesia and Australia [45–49]. Alphamesonivirus-1 belongs to the recently established family Mesoniviridae, which consists exclusively of viruses assumed
to have insect-restricted host ranges [32, 50]. The findings from our *in vitro* host range experiments support this assumption and indicate that the mosquito host range of HOUV is restricted to *Culicinae* mosquitoes because it replicated in *Aedes* and *Culex*, but not *Anopheles*, cell cultures.

In summary, we report the isolation and sequence characterization of both novel and recognized RNA viruses from mosquitoes collected in the Yucatan Peninsula of Mexico. None of the viruses could replicate in any of the vertebrate cell lines tested, suggesting that they all possess insect-restricted host ranges. Despite their apparent vertebrate-incompetent replication phenotypes, further investigation into these viruses is warranted because coinfection experiments have revealed that insect-specific viruses can alter the replication and transmission of pathogenic viruses in mosquitoes [51–54]. Viral sequences were not recovered from some C6/36 cell cultures that exhibited CPE, indicating that the metagenomics analysis failed to detect some viruses. One explanation for this observation is that these viruses are highly divergent from all known viruses and their genomes have no significant identity with any sequences in the GenBank database.

**Methods**

**Cell culture**

The cell lines used in this study were as follows: *Aedes albopictus* (C6/36), African green monkey kidney (Vero), *Anopheles gambiae* (Sua 4.0), baby hamster kidney (BSR-T7/5), *Culex tarsalis* (CT), duck embryonic fibroblast (DEF), human epithelial (HeLa), murine microglia, rhesus macaque (LLC-MK2), *Spodoptera frugiperda* (Sf9) and *Trichoplusia ni* (High Five) cells. Cells were cultured in Dulbecco’s modified Eagle’s medium (all mammalian cells), Eagle’s minimum essential medium (DEF cells), Express Five medium (High Five cells), Liebovitz L15
medium (C6/36 cells), Schneider’s *Drosophila* medium (CT and Sua 4.0 cells) and SF-900 medium (Sf9 cells) (all cell culture media were obtained from Thermo Fisher Scientific, Massachusetts, USA). All media were supplemented with 10% foetal bovine serum, 2 mM l-glutamine, 100 units ml$^{-1}$ of penicillin and 100 µg ml$^{-1}$ of streptomycin, with the exception of the Express Five and SF-900 media, which were serum-free. Vertebrate cells were cultured at 37 °C with 5% CO$_2$ and invertebrate cells were cultured at 28 °C.

**Mosquitoes**

Mosquitoes were collected at study sites in Merida, Tixkokob and Tzucacab in the Yucatan Peninsula of Mexico. Detailed descriptions of the study sites and the protocols used for the collection, identification and homogenization of mosquitoes have been provided elsewhere [31, 38, 55]. Briefly, collections were made in 2007 and 2008 using Mosquito Magnets Pro-Liberty (American Biophysics Corp., Rhode Island, USA) baited with propane and octenol. Mosquitoes were transported alive to the laboratory, frozen at −80 °C and sorted into pools of up to 50 according to species, sex, date of collection and study site. Mosquitoes were placed in polypropylene, round-bottom 5 ml tubes with 1.8 ml of CO$_2$-independent cell culture medium (Thermo Fisher Scientific) and four 4.5mm diameter copper-clad steel beads (BB-calibre airgun shot) and then homogenized by vortexing for 30 s. The homogenates were centrifuged (12000 g, 10 min, 4 °C) and supernatants were collected.

**Virus isolation in C6/36 cell culture**

An aliquot (100 µl) of each supernatant was added to 2 ml of Liebovitz L15 medium supplemented with 2% foetal bovine serum, l-glutamine, penicillin and streptomycin. Samples were filtered and inoculated onto subconfluent monolayers of C6/36 cells in 25 cm$^2$ flasks. Another 4 ml of the aforementioned medium was added to each flask and the cells were
incubated at 28 °C and monitored regularly for CPE. Supernatants were collected when 40–60% of the cell monolayer displayed CPE. If CPE was not observed, supernatants were collected at 7 days post-inoculation. A second blind passage was performed and then the cell monolayers and supernatants were harvested.

**Library preparation**

Total RNA was extracted from all second-passage C6/36 cell cultures using Trizol Reagent (Thermo Fisher Scientific), fragmented using RNase III (Thermo Fisher Scientific) and assessed for quality using an Agilent 2100 Bioanlyser (Agilent, CA, USA). Libraries were constructed using the Ion Total RNA-Seq Kit v2 (Thermo Fisher Scientific) and barcoded using the Ion Xpress RNA-Seq Barcode 1–16 kit (Thermo Fisher Scientific). Libraries were assessed for quality and analysed at the Genomic Technologies Facility at Iowa State University using an Ion Proton Sequencer (Thermo Fisher Scientific).

**Bioinformatics**

Ion-proton reads were analysed using the FastX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) to remove bar-codes and low-quality ends (Phred quality score ≥33). Duplicate reads were identified and removed using Cdhit-454 (http://weizhongli-lab.org/cd-hit/). Host sequences were depleted by mapping the remaining reads to *Ae. aegypti*, *Ae. albopictus*, *An. gambiae* and/or *Cx. quinquefasciatus* genomes using Bowtie 2 [56]. Unmapped reads were analysed using the sortMeRNA program to remove ribosomal RNA-related reads [57]. The remaining reads were subjected to *de novo* SPAdes assembly (version 3.5.0) [58]. Contigs were aligned by blastn, blastx and tblastx to the NCBI nucleotide database (downloaded August 2017) using an e-value of <10⁻⁵. Unaligned contigs were translated into all six reading frames and aligned to the HMM-FRAME, pfam, CDD and
TIGRFAM databases (downloaded August 2017). The data were transformed by Python programming (https://www.python.org/).

**RT-PCR and Sanger sequencing**

RT-PCR and Sanger sequencing were performed to verify the virus sequences identified by NGS and to close gaps between contigs. RT-PCR was also used to test cell cultures for the presence of viral RNA in the host range experiments. Complementary DNAs were generated using Superscript III reverse transcriptase (Thermo Fisher Scientific) and PCRs were performed using high-fidelity Taq polymerase (Thermo Fisher Scientific) in accordance with the manufacturer’s instructions. Primers were designed from the sequences generated by NGS and are available upon request. RT-PCR products were purified using the PureLink gel extraction kit (Thermo Fisher Scientific) and sequenced using a 3730×1 DNA sequencer (Applied Biosystems, CA, USA).

**5′ and 3′ rapid amplification of cDNA ends**

The 5′ and 3′ ends of virus genomes were identified using 5′ and 3′ rapid amplification of cDNA ends, respectively. Briefly, a DNA adaptor (5′-rApp/TGGAATTCTCGGGTGCCAAGGT/ddC-3′) was ligated to the viral genomic and anti-genomic RNAs using T4 RNA ligase (New England BioLabs, MA, USA). Complementary cDNAs were created using SuperScript III (Thermo Fisher Scientific) and an adapter-specific primer. PCRs were performed using adapter- and gene-specific primers, and amplicons were purified and subjected to Sanger sequencing.

**Sequence alignments and identification of conserved domains**

The nucleotide and deduced amino acid sequences of each virus were compared to other sequences in the GenBank database by the application of blastn and blastp, respectively [59].
Percentage amino acid identities and similarities were calculated using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). Conserved domains were identified using blastp, InterProScan (http://www.ebi.ac.uk/interpro/search/sequence-search) and HHpred (PMID 15980461). Potential N-linked glycosylation sites were identified using the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) with the consensus sequence defined as Asn-X-Ser/Thr (where X is not Pro) in the context of specific surrounding sequences.

In vitro host range experiments

The in vitro host ranges of all novel and recognized viruses were assessed using 11 cell lines. Briefly, cell monolayers approaching confluency in 25 cm$^2$ flasks were inoculated with 50 µl of virus-containing supernatant and incubated for 7 days or until 40–60% of the cell monolayer displayed CPE. A total of five passages were performed, with supernatants being collected after each passage and tested by RT-PCR for viral RNA. A cell line was considered to support virus replication if RT-PCR products were detected after every passage.

Funding

Mosquito collections were made possible by grant 5R21AI067281 from the National Institutes of Health. The metagenomics analysis was supported by an intramural grant from Iowa State University. Host range experiments were supported by grant R01AI114720 from the National Institutes of Health. A.E.F. and C.T. are supported by Wellcome Trust grant 106207 and European Research Council grant 646891 to A.E.F.
Acknowledgments

The authors thank Kristina Larsen and Sean Johnston for technical assistance.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Ethical Statement

No human subjects or vertebrate animals were used in this study.

References


44. Cigarroa-Toledo N, Baak-Baak CM, Cetina-Trejo RC, Cordova-Fletes C, Martinez-Nuñez MA et al. Complete genome sequence of houston virus, a newly discovered mosquito-specific virus isolated from *Culex quinquefasciatus* in Mexico. *Microbiol Resour Announc* 2018;7:


49. Zirkel F, Kurth A, Quan PL, Briese T, Ellerbrok H et al. An insect nidovirus emerging from a primary tropical rainforest. MBio 2011;2:e0007711


60. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004;5:113

Fig. 1. Predicted genomic organization of (a) Uxmal virus and (b) Mayapan virus. The predicted viral methyltransferase, 2'-O-methyltransferase, helicase and RNA-dependent RNA polymerase domains located within ORF 1 of Uxmal virus are denoted as MTase, 2'OMTase, Hel and RdRp, respectively.
**Fig. 2.** Alignment of (a) the translation products potentially encoded by ORF3a of Uxmal virus and selected related viruses and (b) deduced amino acid sequences of the putative B2 proteins of Mayapan virus and Hubei noda-like virus 5.

(a)  
<table>
<thead>
<tr>
<th>Virus Type</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanay virus (KF425261)</td>
<td>MNSTEICIF--------------------------VCLTVVALIVINKLFYFLEHLIVENGSKTHD-SR</td>
</tr>
<tr>
<td>Goutanap virus (KF588035)</td>
<td>MCFVSVIIIVLFLFGF---------------------LSSLITDCN------------------EFDCTAYERYG</td>
</tr>
<tr>
<td>Biggievirus (MF281708)</td>
<td>MFYCLILL---------------------------VFIVVYLLY-PSKPSYV----------------PAKEK-YG</td>
</tr>
<tr>
<td>Culex negev-like virus (MF176277)</td>
<td>MFWYSSL---------------------------IGFVVYLL-YSKPDYV----------------SAEKE-YG</td>
</tr>
<tr>
<td><em>Uxmal virus</em></td>
<td>MNQLTTLDLAY--------------------------LALLYYVSSVFNLIHLLE-----------RKNGSAYW-YG</td>
</tr>
<tr>
<td>Wallerfield virus (KF042857)</td>
<td>MNIVDCVTIILWSHF-----------------------LIRLHYLSRKFYAFVYI--------NYGPAAWY-YG</td>
</tr>
<tr>
<td>Dezidouguou virus (JQ675604)</td>
<td>MITCICVVWL--------------------------FLYLSVDCYHTYL------------RKNGYASWY-YG</td>
</tr>
<tr>
<td>Santana virus (JQ675606)</td>
<td>MSIVILPPLYTLYLYHPRFvbYIDTITILYISILKTIFDISINHNLTRDNGSASYW-HG</td>
</tr>
<tr>
<td>Bustos virus (LC103139)</td>
<td>MFPLAASVLQGES-....YILQIFGRICLLFYIVDVFNRQLFLL---------RYHGPQAYW-YG</td>
</tr>
<tr>
<td>Biratnagar virus (KX518758)</td>
<td>MYPLILAVLLQSEF-....PLYQFGVRCLLCICIDIIERYLFLL--------RHHGPQAYR-YG</td>
</tr>
</tbody>
</table>

(b)  
<table>
<thead>
<tr>
<th>Virus Type</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanay virus (KF425261)</td>
<td>RSQGDLFQNRSSRESN----------------------RDSRYSEENRYQDQWQD------</td>
</tr>
<tr>
<td>Goutanap virus (KF588035)</td>
<td>CTSLT-YGSEDCCCPCPWPWLLQD-----------------SASTCPARSGTSTSQT--</td>
</tr>
<tr>
<td>Biggievirus (MF281708)</td>
<td>CTALE-YGAECCICCPSTAWLLQQLADAESHNYGNDRSAATSSAETS</td>
</tr>
<tr>
<td>Culex negev-like virus (MF176277)</td>
<td>CPAET-YGSEGCLCPSTAWLLQQLADAESHYDCHGDCPATTSSTTT</td>
</tr>
<tr>
<td><em>Uxmal virus</em></td>
<td>CEAH-YGLPECTCP------------------------QEAYQYSLGSDP----------</td>
</tr>
<tr>
<td>Wallerfield virus (KF042857)</td>
<td>CEASC-CCSPDLCPC------------------------PHSCGEADQPS6GIES--</td>
</tr>
<tr>
<td>Dezidouguou virus (JQ675604)</td>
<td>CEAKN-YGDEGVCPSN------------------------KDPIC------------------</td>
</tr>
<tr>
<td>Santana virus (JQ675606)</td>
<td>CTAEEF-YEAEDCVCP------------------------KAEEISLRHGIE--</td>
</tr>
<tr>
<td>Bustos virus (LC103139)</td>
<td>CSACFDCPSGCYCPP------------------------REKSGETSERIRE------</td>
</tr>
<tr>
<td>Biratnagar virus (KX518758)</td>
<td>CSPHDVSSCCCCCP------------------------FEKCGEGFEFRE--------</td>
</tr>
</tbody>
</table>

Hubei noda-like virus 5 (KX883080)  
| Hubei noda-like virus 5 (KX883161)              | MSLTKQSLSELAVAVQRQRQNYKTAEARLRLVTLRGLKVLGKRLAKSLKPLAPDHYKKN |
| **Mayapan virus**                               | MSLTKQSLSELAVAVQRQRQNYKTAEARLRLVTLRGLKVLGKRLAKSLKPLAPDHYKKN |
  
Hubei noda-like virus 5 (KX883080)  
| Hubei noda-like virus 5 (KX883161)              | ERPWVRMVMNMTELLTVHPDLVPGDAQQPAPVQYDKCDPNWDPKPSGPPVPHE |
| **Mayapan virus**                               | ERPWVRMVMNMTELLTVHPDLVPGDAQQPAPVQYDKCDPNWDPKPSGPPVPHE |

**Mayapan virus**

\[
\begin{array}{llllllllllllllll}
\text{DRSVRVDKMTCLMTVHPDLVPGDAQQPAPVQYDRCMDNWDPKPSGSPTPH-} \\
\vdots \text{*:**:*::*:**:******** *** ******** ******** ** **} \\
\end{array}
\]
Fig. 3. Phylogenetic tree for Uxmal virus and selected related reference and non-reference sequences. The amino acid sequences of the ORF encoding the RdRp were aligned using muscle [60]. A maximum-likelihood phylogenetic tree was estimated using the Bayesian Markov chain Monte Carlo method implemented in MrBayes version 3.2.6 [61], sampling across the default set of fixed amino acid rate matrices with one million generations and discarding the first 25% as burn-in. The figure was produced using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The tree is midpoint-rooted and selected nodes are labelled with posterior probability values. GenBank accession numbers are indicated next to virus names.
Fig. 4. Phylogenetic tree for Mayapan virus and selected related reference and non-reference sequences. Amino acid sequences of the ORF encoding the (a) RdRp and (b) capsid protein were aligned using muscle [60]. A maximum-likelihood phylogenetic tree was estimated using the Bayesian Markov chain Monte Carlo method implemented in MrBayes version 3.2.6 [61], sampling across the default set of fixed amino acid rate matrices with one million generations and discarding the first 25% as burn-in. The figure was produced using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The tree is midpoint-rooted and selected nodes are labelled with posterior probability values. GenBank accession numbers are indicated next to virus names.
Table 1. Mosquitoes tested by virus isolation in cell culture and viral metagenomics

<table>
<thead>
<tr>
<th>Species</th>
<th>No. of mosquitoes</th>
<th>No. of pools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aedes aegypti</td>
<td>62</td>
<td>4</td>
</tr>
<tr>
<td>Aedes taeniorhynchus</td>
<td>764</td>
<td>16</td>
</tr>
<tr>
<td>Aedes trivittatus</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>Anopheles vestitipennis</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>Culex quinquefasciatus</td>
<td>305</td>
<td>10</td>
</tr>
<tr>
<td>Mansonia titillans</td>
<td>89</td>
<td>2</td>
</tr>
<tr>
<td>Psorophora ferox</td>
<td>93</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>1359</td>
<td>37</td>
</tr>
</tbody>
</table>
Table 2. Novel and recognized RNA viruses isolated from mosquitoes collected in the Yucatan Peninsula of Mexico

<table>
<thead>
<tr>
<th>Virus</th>
<th>No. of isolations</th>
<th>Isolate name(s)</th>
<th>Taxonomic classification</th>
<th>Study site</th>
<th>Mosquito species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culex flavivirus</td>
<td>1</td>
<td>CxFV-T1123</td>
<td><em>Flavivirus (Flaviviridae)</em></td>
<td>Tixkokob</td>
<td><em>Cx. quinquefasciatus</em></td>
</tr>
<tr>
<td>Houston virus</td>
<td>2</td>
<td>HOUV-M742, HOUV-M2668</td>
<td><em>Alphamesonivirus(Mesoniviridae)</em></td>
<td>Merida</td>
<td><em>Cx. quinquefasciatus, Ae. taeniorhynchus</em></td>
</tr>
<tr>
<td>Mayapan virus</td>
<td>2</td>
<td>MYPV-H44, MYPV-H56</td>
<td><em>a Nodaviridae?</em></td>
<td>Tzucacab</td>
<td><em>Ps. ferox</em></td>
</tr>
<tr>
<td>Uxmal virus</td>
<td>3</td>
<td>UXMV-M985, UXMV-M1000, UXMV-M2038</td>
<td><em>Negevirus</em></td>
<td>Merida</td>
<td><em>Ae. taeniorhynchus</em></td>
</tr>
</tbody>
</table>

*The closest known relatives of Mayapan virus are unclassified viruses that phylogenetically group with viruses in the family *Nodaviridae.*
Table 3. *In vitro* host ranges of novel and recognized viruses isolated from mosquitoes in the Yucatan Peninsula of Mexico

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Virus</th>
<th>Culex flavivirus</th>
<th>Houston virus</th>
<th>Mayapan virus</th>
<th>Uxmal virus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mosquito cells</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6/36</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>a+</td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>Sua 4.0</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td><strong>Lepidopteran cells</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Five</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>Sf9</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td><strong>Vertebrate cells</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSR-T7/5</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>DEF</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>HeLa</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>LLC-MK2</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>Murine microglia</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>Vero</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
</tbody>
</table>

* a, CPE was observed.
CHAPTER 4. DISCOVERY OF A NOVEL TYMOVIRIDAE-LIKE VIRUS IN MOSQUITOES FROM MEXICO

Modified from a manuscript published in Archives of Virology 164: 649-652 (2019)

Jermilia Charles¹, Chandra S. Tangudu¹, Stefanie L. Hurt¹, Charlotte Tumescheit², Andrew E.Firth², Julian E. Garcia-Rejon³, Carlos Machain-Williams³, Bradley J. Blitvich¹

¹Department of Veterinary Microbiology and Preventive Medicine, College of Veterinary Medicine, Iowa State University, Ames, IA, USA
²Department of Pathology, University of Cambridge, Cambridge UK
³Laboratorio de Arbovirologia, Centro de Investigaciones Regionales “Dr. Hideyo Noguchi”Universidad Autónoma de Yucatán, Mérida, Mexico

Corresponding author: Bradley J. Blitvich; 2116 Veterinary Medicine, Iowa State University, Ames, IA 50011, USA, e-mail: blitvich@iastate.edu

Abstract

A novel Tymoviridae-like virus, designated Ek Balam virus, was isolated from male Culex quinquefasciatus mosquitoes collected in Yucatan, Mexico. The genome was fully sequenced and shown to have no more than 69% nt sequence identity to its closest known relative. Mosquito cells were permissive to Ek Balam virus replication, but mammalian and avian cells were refractory, suggesting that vertebrates are not involved in the maintenance of the virus in nature.

All viruses currently classified as members of the family Tymoviridae are plant viruses, and some are major agricultural pathogens [1, 2]. The family includes three genera: Maculavirus, Marafivirus and Tymovirus. Viruses in the genera Marafivirus and Tymovirus are transmitted by leafhoppers and beetles, respectively, but no vectors have been identified for viruses in the genus Maculavirus. Family members contain a positive-sense, single-stranded RNA genome of 6.0 to 7.5 kb. Genomes are capped at the 5’ end, often possess a tRNA structure at the 3’ end and, except for viruses in the genus Tymovirus, are
polyadenylated. The number of open reading frames (ORFs) varies according to genus and species. Viruses in the genus Maculavirus have four ORFs, and these encode a replicase, coat protein and two small proline-rich proteins, p31 and p16 [2, 3].

Mosquitoes are not recognized vectors of viruses in the family Tymoviridae, although a novel Tymoviridae-like virus, designated Culex-originated Tymoviridae-like virus (CuTLV), was recently isolated from Culex spp. mosquitoes in China [4]. The genomic organization of CuTLV is similar to that of viruses in the genus Maculavirus, although notable differences were also described, including the absence of a polyadenylated tail and an ORF encoding p32. CuTLV has not yet been classified by the International Committee on Taxonomy of Viruses, and it remains to be seen whether it will be assigned to the genus Maculavirus or to a new genus or family. We report the discovery of a second novel Tymoviridae-like virus that naturally infects mosquitoes.

The novel virus, designated Ek Balam virus (EkBV), was detected in male Culex quinquefasciatus collected in 2007 in the rural town of Tixkokob in Yucatan, Mexico. The virus was named after an archeological ruin in Yucatan. Descriptions of study sites and protocols for the collection, identification and homogenization of mosquitoes are provided elsewhere [5, 6]. EkBV was isolated after two blind passages in Aedes albopictus (C6/36) cells, and its genome was fully sequenced by Ion Proton sequencing, reverse transcription polymerase chain reaction, Sanger sequencing, and 5’ and 3’ rapid amplification of cDNA ends as described previously [7].

The genome of EkBV consists of 6516 nt (GenBank accession no. MH822863). The sequence is most closely related to the genome sequence of CuTLV (68.3% nt sequence identity with 29.6% coverage), followed by bat tymo-like virus (65.8% nt sequence identity with 41.5% coverage), an unclassified Tymoviridae-like virus detected in guano of Myotis yumanensis (GenBank accession nos. NC_030844.1 and KX580887.1; Alex Greninger,
University of Washington, Seattle, WA, USA, personal communication). It is not known whether bat tymo-like virus replicates in bats or was acquired through the consumption of virus-infected material (i.e., insects), but the latter appears more likely because the virus is most closely related to EkBV and CuTLV. The next-closest relatives to EkBV are viruses in the genus *Maculavirus*. The genome of EkBV has a high cytosine and low guanine content (A, 23.0%; C, 31.7%; G; 18.6%; T, 26.7%), as has been reported for CuTLV and recognized members of the family *Tymoviridae* [1, 3, 4, 8]. The genome lacks a polyadenylate tail, and its 5’ and 3’ untranslated regions consist of 144 nt and 433 nt, respectively. We could not provide any convincing evidence of a tRNA-like structure in the 3’ UTR (data not shown). The genome does not possess a classical “tymobox” or “marafibox”, conserved subgenomic promoter sequences characteristic of tymoviruses and marafiviruses, respectively [1, 2].

The genome of EkBV is predicted to contain at least two major ORFs (Fig. 1). The first ORF encodes a putative replicase of 1724 residues that is most closely related to the corresponding translation product of CuTLV (50.9% identity and 64.1% similarity, with 100% coverage). Predicted viral methyltransferase (pfam01660), tymovirus endopeptidase (pfam05381), helicase (pfam01443) and RdRp (pfam00978) domains are present at residues 38 to 318, 683 to 770, 870 to 1102 and 1400 to 1591, respectively. The second ORF encodes a putative coat protein of 243 residues that is most closely related to the corresponding translation product of CuTLV (61.3% identity and 75.4% similarity, with 100% coverage). A predicted tymovirus coat protein domain (pfam00983) is present at residues 113 to 196. The ORFs are separated by a 32-nt intergenic region.

A phylogenetic tree was constructed using Bayesian methods based on the deduced amino acid sequence of the replicase of EkBV and the corresponding regions of some related
sequences (Fig. 2). EkBV is most closely related phylogenetically to CuTLV, and the posterior support for this topological arrangement is 0.97. EkBV, CuTLV and bat tymo-like virus belong to a clade that is basal to recognized members of the family Tymoviridae. Viruses in this family separate into three distinct genus-specific clades.

The in vitro host range of EkBV was determined by assessing its replicative ability in cell lines of avian, mammalian and mosquito origin. Six cell lines were tested: African green monkey kidney (Vero), Anopheles gambiae (Sua 4.0), baby hamster kidney (BSR-T7/5), C6/36, Culex tarsalis (CT) and duck embryonic fibroblast (DEF) cells. A cell line was considered to support EkBV replication if viral RNA was detected by RT-PCR after five passages. All mosquito cell lines were permissive to replication, indicating that the virus has a broad mosquito host range. A cytopathic effect was not observed in any mosquito cell cultures. All vertebrate cell lines were refractory to virus infection, consistent with a vertebrate replication-incompetent phenotype. CuTLV is also assumed to have a vertebrate-replication-incompetent phenotype because it cannot replicate in Vero or baby hamster kidney (BHK-21) cells [4].

The inability of EkBV to replicate in vertebrate cells and its detection in male mosquitoes suggests that the virus is not maintained in nature in a mosquito-vertebrate transmission cycle. Our findings indicate that EkBV is more likely to be maintained by vertical (i.e., transovarial) or horizontal (i.e., venereal) transmission among insect hosts. Alternatively, the virus could be maintained by mechanical (i.e., per os) transmission. Experimental infections are needed to determine whether EkBV is maintained in nature by direct mosquito-to-mosquito transmission and whether it is capable of infecting plants. The detection of EkBV in field-collected mosquitoes and C6/36 cells after five passages suggests that the virus replicates in mosquitoes.
and is not a surface or gut contaminant acquired after coming in contact with or feeding upon infectious material.

In conclusion, we report the isolation of a novel *Tymoviridae*-like virus from mosquitoes in Mexico and add to the rapidly increasing number of newly discovered viruses. Based on the genetic distance between EkBV, its closest known relative (CuTLV) and recognized members of the family *Tymoviridae*, we propose that EkBV and CuTLV are members of two distinct species that should be assigned to the same, yet-to-be-established genus. It remains to be determined whether these viruses will be classified as members of the family *Tymoviridae* or of a new family.

**References**


Fig. 1. Genomic organization of (A) Ek Balam virus, (B) Culex originated *Tymoviridae*-like virus and (C) bat tymo-like virus.
Fig. 2. Phylogenetic tree for EkBV and selected related reference and non-reference sequences. Amino acid sequences of the open reading frame encoding the replicase were aligned using MUSCLE [9]. A maximum-likelihood phylogenetic tree was constructed using the Bayesian Markov chain Monte Carlo method implemented in MrBayes version 3.2.6 [10], sampling across the default set of fixed amino acid rate matrices with 1,000,000 generations, discarding the first 25% as burn-in. The figure was produced using FigTree (available at http://tree.bio.ed.ac.uk/software/figtree/). The tree is midpoint-rooted, and selected nodes are labelled with posterior probability values. GenBank accession numbers are shown next to virus names.
CHAPTER 5. COMPLETE GENOME SEQUENCES OF TWO INSECT-SPECIFIC FLAVIVIRUSES

Modified from a manuscript published in Archives of Virology 2017; 162(12): 3913–3917.

Jermilia Charles,1 Chandra S. Tangudu,1 Andrew E. Firth,2 and Bradley J. Blitvich1

1Department of Veterinary Microbiology and Preventive Medicine, College of Veterinary Medicine, Iowa State University, Ames, IA, USA
2Department of Pathology, University of Cambridge, Cambridge UK
Corresponding author: 2116 Veterinary Medicine, Iowa State University, Ames, IA 50011 (e-mail: blitvich@iastate.edu)

Abstract

We determined the complete genomic sequences of two previously discovered insect-specific flaviviruses, Marisma mosquito virus (MMV) and Nanay virus (NANV), using a combination of high-throughput sequencing, reverse transcription-polymerase chain reaction, 5′ and 3′ rapid amplification of cDNA ends and Sanger sequencing. Complete polyprotein amino acid sequence alignments revealed that the closest known relatives of MMV and NANV are Donggang virus (89% identity, 95% similarity) and Nounané virus (53% identity, 70% similarity), respectively. Phylogenetic inference is in agreement with these findings. Potential programmed −1 ribosomal frameshifting sites were bioinformatically identified in the genomes of both viruses.

Viruses in the genus Flavivirus (family Flaviviridae) can be divided into three distinct groups based on their host ranges and mode of transmission [2]. Dual-host flaviviruses are transmitted horizontally between hematophagous arthropods (i.e. mosquitoes and ticks) and vertebrate hosts. Viruses in the other two groups possess vertebrate-specific or insect-specific host ranges. Insect-specific flaviviruses (ISFs) are further divided into classical ISFs (cISFs) and dual-host affiliated flaviviruses (dISFs). Viruses in the dISF group phylogenetically affiliate with dual-host flaviviruses despite their apparent insect-restricted phenotypes. Examples of dISFs
include Marisma mosquito virus (MMV) and Nanay virus (NANV) in addition to Barkedji virus (BJV), Chaoyang virus (CHAOV), Donggang virus (DONV), Ilomantsi virus (ILOV), Lammi virus (LAMV), Nhumirim virus (NHUV) and Nounané virus (NOUV). MMV was originally isolated from Ochlerotatus caspius in Italy in 2001 to 2007 [17]. NANV was originally isolated from Culex (Melanoconion) ochssa in Peru in 2009 [6]. All these dISFs have had most, if not all, of their genomes fully sequenced aside from MMV and NANV for which partial envelope and/or NS5 gene sequence data are available. The objective of this study was to fully sequence the genomes of MMV and NANV.

MMV (isolate HU4528/07) and NANV (isolate PRD316/PER/09) were obtained from the World Reference Center for Emerging Viruses and Arboviruses at the University of Texas Medical Branch in Galveston, TX. MMV and NANV had been passaged four and at least five times, respectively in C6/36 (Aedes albopictus) cells prior to receipt and each virus underwent an additional passage in C6/36 cells at Iowa State University. Total RNA was extracted using Trizol Reagent (Invitrogen, Carlsbad, CA) and RNA was fragmented using RNase III and assessed for quality using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). Libraries were constructed using the Ion Total RNA-Seq Kit v2 (ThermoFisher, Carlsbad, CA) and barcoded using the Ion Xpress™ RNA-Seq Barcode 1-16 Kit (ThermoFisher). Libraries were assessed for quality and analyzed at the Genomic Technologies Facility at Iowa State University using an Ion Proton Sequencer (ThermoFisher). All ion-Torrent reads were mapped to Ae. albopictus and Ae. aegypti genomes using Bowtie 2 [13]. Unmapped reads were analyzed using the sortMeRNA program [12] to remove rRNA-related reads. All remaining reads with Phred values ≥33 were subjected to de novo SPAdes (ver 3.5.0) assembly [1]. Contigs were mapped to a reference flavivirus genome using LASTZ [9]. Alignment files were manually verified on TABLET [16].
Reverse transcription-polymerase chain reaction and Sanger sequencing were used to close gaps and confirm the accuracy of assembled contigs. The 5′ and 3′ ends of each genome were identified using 5′ and 3′ rapid amplification of cDNA ends, respectively. Briefly, a DNA adaptor (5′-rApp/TGGAATTCTCGGGTGCCAAGGT/ddC-3′) was ligated to the viral genomic and anti-genomic RNAs using T4 RNA ligase (New England BioLabs, Ipswich, MA). Complementary cDNAs were created using SuperScript III (Invitrogen) and an adapter-specific primer. PCRs were performed using adapter- and gene-specific primers, and amplicons were subjected to Sanger sequencing.

The complete genome of MMV consists of 10,848 nt. (Genbank Accession No. MF139576) and contains a 5′ untranslated region (UTR) of 119 nt., a long open reading frame (ORF) of 10,353 nt., and a 3′ UTR of 376 nt (Figure 1a). The predicted amino acid sequence of the MMV polyprotein was aligned to other amino acid sequences in the Genbank database revealing that the closest known relatives of MMV are DONV (89% identity, 95% similarity) and ILOV (71% identity, 83% similarity). The complete genome of NANV is slightly smaller (10,804 nt; Genbank Accession No. MF139575) and contains a 5′ UTR of 106 nt., a long ORF of 10,299 nt., and a 3′ UTR of 399 nt (Figure 1b). Amino acid sequence alignments revealed that the closest known relatives of NANV are NOUV (53% identity, 70% similarity) and NHUV (52% identity, 69% similarity).

The phylogenetic placements of MMV and NANV, relative to other members of the Flavivirus genus, were assessed using the Bayesian Markov chain Monte Carlo based method implemented in MrBayes [15]. Complete polyprotein amino acid sequences were aligned using MUSCLE [5] and a phylogenetic tree was constructed using MrBayes (Figure 2). MMV is most closely related phylogenetically to DONV. Both viruses belong to a distinct clade that also
includes CHAOV, LAMV and ILOV. NANV is most closely related phylogenetically to NOUV and these two viruses belong to a distinct clade that also includes BJV and NHUV.

The sequences of MMV and NANV were inspected for potential −1 ribosomal frameshifting (−1 PRF) motifs because several groups of flaviviruses, including dISFs, appear to utilize −1 PRF during translation of their genomic RNA [2, 7, 10, 14]. Such frameshifting occurs at specific sites which normally comprise a slippery heptanucleotide sequence and a 3′-adjacent RNA structure. In eukaryotes, the consensus motif for the slippery heptanucleotide is X_XXY_YYZ, where XXX represents any three identical nucleotides although a number of exceptions occur (such as GGA), YYY represents AAA or UUU, Z is A, C or U, and underscores represent codons in the original reading frame. The 3′ RNA structure is normally a stem-loop or pseudoknot and is separated from the shift site by a spacer region of 5–9 nt. PRF has been predicted to occur in the NS2B region of CHAOV, DONV, LAMV, and ILOV [2, 10] and the ability of the identified motifs to stimulate −1 PRF has been verified in reporter constructs [8]. MMV contains a conserved G_GAU_UUU shift site sequence followed by a predicted RNA stem-loop structure in the NS2B region, as previously described for CHAOV, DONV and LAMV (Figure 3a). As noted previously, the −1 frame ORF varies considerably in length from six codons in DONV to 107 codons in CHAOV and LAMV. Similarly to DONV, MMV has a six-codon ORF.

The genome of NANV harbors a U_UUU_UUU potential shift site that aligns with the U_UUU_UUA shift site previously proposed in NOUV (Figure 3b), and tested in reporter constructs [8]. While the NOUV shift site is followed by a compact 13-bp stem-loop, the potential shift site in NANV is followed by a more extended potential stem-loop. Nonetheless, the stem-loop has a stable base (seven consecutive Watson-Crick pairs, four being G:Cs) and is
within the critical 5–9 nt. separation from the shift site. While the simple presence of a shift site and potential RNA structure should be viewed with caution (as they may occur spuriously, and not all RNA structures have the correct geometry to efficiently stimulate frameshifting [4]), the conservation of such features at a similar genomic location between related species lends weight to –1 PRF predictions. Thus, it seems plausible that NOUV and NANV may represent another group of flaviviruses that utilize –1 PRF.

In summary, we report the complete genome sequences of two previously discovered ISFs, MMV and NANV, and provide bioinformatic evidence that both viruses utilize –1 PRF. It remains to be proven whether dISFs evolved from dual-host flaviviruses or are themselves the precursors but the former theory has been favored [11]. A rapidly growing number of ISFs have been discovered in recent years, and the availability of complete genome sequence data will allow for more robust comparative genomic studies between dual- and single-host flaviviruses and could, ultimately, provide novel insight into the evolutionary mechanisms that condition their differential host ranges and transmissibilities.

Acknowledgements
This study was supported by a grant from the National Institutes of Health (R01AI114720). A.E.F. is supported by a Wellcome Trust grant (106207) and a European Research Council (ERC) European Union’s Horizon 2020 research and innovation programme grant (646891). The authors thank Robert Tesh for providing isolates of MMV and NANV.

Conflict of Interest
The authors declare that they have no conflict of interest.
**Ethical Approval**
The research reported here did not involve the use of human subjects or vertebrate animals.

**Ethical Statement**
This study represents original work that has not been submitted to any other journal for publication. The authors have no conflict of interest to declare. No human or animal ethics approval was required for the completion of this study.

**References**


Fig. 1. Schematic representation of each flavivirus genome and polyprotein. Genomic organization of (A) Marisma mosquito virus and (B) Nanay virus. Lengths of the 5' and 3' untranslated regions as well as the structural and nonstructural protein genes are drawn to scale.
Fig. 2. Phylogenetic tree for genus Flavivirus. Complete polyprotein amino acid sequences were aligned using MUSCLE [5]. Regions of ambiguous alignment were excised using Gblocks [3] with default parameters, after which 1604 amino acid positions were retained. A maximum likelihood phylogenetic tree was estimated using the Bayesian Markov chain Monte Carlo method implemented in MrBayes version 3.2.3 [15] sampling across the default set of fixed amino acid rate matrices, with one million generations, discarding the first 25% as burn-in. The figure was produced using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The tree is midpoint-rooted, and nodes are labelled with posterior probability values where different from 1.00. Species names are color-coded as follows: cISFs—blue; dISFs—green; no known vector (NKV) flaviviruses—red; mosquito/vertebrate flaviviruses—purple; tick/vertebrate flaviviruses—black
Fig. 3. Predicted sites of ribosomal frameshifting. (a) Marisma mosquito virus and other members of the Chaoyang/Lammi/Donggang/Ilomantsi clade have a conserved G_GAU_UUU slippery heptanucleotide (orange highlight) followed by a predicted RNA stem-loop structure (blue letters) in the NS2B region. Compensatory substitutions (i.e. paired substitutions that preserve the predicted base-pairings) are indicated in pink. Parentheses indicate the predicted base-pairings. Numbers in the last column show the length of the −1 frame ORF. (b) A potential shift site (orange highlight) and adjacent RNA stem-loop structure (blue letters), conserved in location but not sequence, in the NS2A region of Nanay virus and Nounane virus.
CHAPTER 6. COMPLETE NUCLEOTIDE SEQUENCES OF THE LARGE RNA GENOME SEGMENTS OF MAIN DRAIN AND NORTHWAY VIRUSES (FAMILY PERIBUNYAVIRIDAE)

Modified from a manuscript published in Archives of Virology 2018; 163(8): 2253–2255.

Jermilia Charles, Chandra S. Tangudu, and Bradley J. Blitvich

1Department of Veterinary Microbiology and Preventive Medicine, College of Veterinary Medicine, Iowa State University, Ames, IA, USA
Corresponding author: 2116 Veterinary Medicine, Iowa State University, Ames, IA 50011 (e-mail: blitvich@iastate.edu)

Abstract

The large RNA genome segments of Main Drain virus (MDV) and Northway virus (NORV) were fully sequenced and shown to consist of 6860 and 6875 nucleotides, respectively. Sequence alignments revealed that the large RNA segment of MDV is most closely related to the corresponding region of NORV, with 76.8% nucleotide sequence identity, and the large RNA segment of NORV is most closely related to the corresponding region of Maguari virus, with 79.1% identity.

Main Drain virus (MDV) and Northway virus (NORV) are members of the genus Orthobunyavirus (family Peribunyaviridae). MDV was originally isolated from Culicoides variipennis in California in 1964, and additional isolations have since been made elsewhere in the western United States [2]. Serologic evidence suggests that MDV also occurs in central and eastern regions of the United States [9]. MDV is not a recognized human pathogen, although it has been isolated from the brain of a horse that died of encephalitis [7] and experimental infection studies have shown that it is a cause of severe musculoskeletal and nervous system malformations in ovine fetuses [6]. NORV was originally isolated from Aedes and Culiseta spp. mosquitoes in Alaska in 1970 and has since been reported in
California and northwestern Canada [2, 3, 8]. NORV is not a recognized pathogen of humans or vertebrate animals, but antibodies to the virus have been identified in humans, livestock and wild animals [9, 11].

All viruses in the genus *Orthobunyavirus* possess a tripartite, single-stranded, negative-sense RNA genome [10]. The three RNA segments are designated as small, (S), medium (M) and large (L) and are approximately 1.0, 4.5 and 6.9 kb in length, respectively. The L RNA segment codes for the L protein, which is an RNA-dependent RNA polymerase. All three RNA segments of the virus have the same complementary nucleotide sequences at their 3’ and 5’ termini. Base pairing of these terminal sequences results in the formation of panhandle structures that are required for RNA synthesis. The S and M RNA segments of MDV and NORV have been fully sequenced, but less than 600 nt of L RNA sequence data are available for each virus [1, 5]. The objective of this study was to fully sequence the L RNA segments and verify the S and M segment sequences of MDV and NORV, thereby allowing complete genome sequences to be available.

MDV (strain BFS 5015) and NORV (strain 0234) were obtained from the World Reference Center for Emerging Viruses and Arboviruses at the University of Texas Medical Branch in Galveston, TX. MDV had been passaged six times in suckling mouse brains prior to receipt and underwent an additional passage in African Green Monkey kidney (Vero) cells at Iowa State University (ISU). NORV had been passaged four times in suckling mouse brains and once each in baby hamster kidney (BHK-21) and Vero cells prior to receipt and underwent an additional passage in Vero cells at ISU. The L RNA segment of each virus was sequenced by Ion-Torrent sequencing, reverse transcription polymerase chain reaction, 5’ and 3’ rapid amplification of cDNA ends, and Sanger sequencing as described previously [4]. Contigs were
mapped to the L segment of Cache Valley virus (prototype strain) and previously reported S and M segment sequences of MDV and NORV.

The L RNA genome segment of MDV consists of 6860 nt (GenBank accession no. MG652604). The open reading frame (ORF) is 6711 nt long and is flanked by 5’ and 3’ untranslated regions (UTRs) of 49 and 100 nt, respectively. Twenty-seven of the 28 terminal nucleotides at the 5’ and 3’ ends of the genome segment are complementary. The exception is the A and C at nucleotide positions 9 and 6852, respectively (nucleotide numbers correspond to the positive-sense RNA). Application of blastn followed by Clustal Omega (accessed at https://www.ebi.ac.uk/Tools/msa/clustalo/) revealed that the L RNA segment of MDV is most closely related to the corresponding region of NORV, with 76.8% nt sequence identity (Table 1). The predicted translation product consists of 2236 amino acids and has 86.6% identity and 94.5% similarity to the corresponding region of NORV.

Our MDV S segment sequence is identical to the previously reported sequence, while an extra 20 and 14 nt were discovered at the 5’ and 3’ end, respectively, of the M segment (GenBank accession no. MH102379). The S and M RNA segment sequences of MDV have 85.3% and 72.6% nt sequence identity, respectively, to the corresponding regions of their closest known relatives (Table 1). The first 11 nucleotides at each terminus are conserved between all three genomic segments of MDV.

The L RNA genome segment of NORV consists of 6875 nt (GenBank accession no. MG544835). The ORF is 6714 nt long and is flanked by 5’ and 3’ UTRs of 51 and 110 nt, respectively. Twenty-nine of the 30 terminal nucleotides at the 5’ and 3’ ends of the genome segment are complementary. The exception is the A and C at nucleotide positions 9 and 6861, respectively. Sequence alignments revealed that the L RNA segment of NORV is most closely
related to the corresponding region of Maguari virus (MAGV; isolate BeAr 7272), with 79.6% nt sequence identity (Table 1). The predicted translation product consists of 2237 amino acids and has 91.5% identity and 97.1% similarity to the corresponding regions of MAGV.

Our NORV S segment sequence contains one nonsynonymous substitution and two nucleotide substitutions in the 3’ UTR when compared to the previously reported sequence (GenBank accession no. MH015350). The M segment sequence contains an extra 20 and 14 nt at the 5’ and 3’ end, respectively, in addition to one synonymous substitution (GenBank accession no. MH102378). The S and M RNA segment sequences have 86.3% and 76.4% nt sequence identity, respectively, to the corresponding regions of their closest known relatives (Table 1). The first 11 nucleotides at each terminus are conserved between all three genomic segments of NORV.

In conclusion, the L RNA genome segments of MDV and NORV were fully sequenced, and their S and M RNA genome segments were verified, and thus, complete genome sequences are available for both viruses. The availability of complete genome sequence data is needed for the accurate identification and classification of viruses, especially viruses with segmented genomes, because partial genome sequencing will not always identify reassortants. Complete genome sequence data are not available for many orthobunyaviruses. Often, the S and M RNA segments are fully sequenced but the L RNA segment is not. Future studies should address this issue.
Acknowledgements

This study was supported by an intramural grant from Iowa State University. The authors thank Robert Tesh for providing isolates of MDV and NORV.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

The research reported here did not involve the use of human subjects or vertebrate animals.

Ethical statement

This study represents original work that has not been submitted to any other journal for publication. The authors have no conflict of interest to declare. No human or animal ethics approval was required for the completion of this study.

References


### Table 1. Genetic relatedness of the small, medium and large genome segments of Main Drain and Northway viruses and selected other orthobunyaviruses

<table>
<thead>
<tr>
<th>Virus</th>
<th>Small Main Drain</th>
<th>Medium Main Drain</th>
<th>Large Main Drain</th>
<th>Small Northway</th>
<th>Medium Northway</th>
<th>Large Northway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bunyamwera virus</td>
<td>78.6</td>
<td>58.5</td>
<td>73.5</td>
<td>79.8</td>
<td>63.2</td>
<td>73.6</td>
</tr>
<tr>
<td>Batai virus</td>
<td>78.4</td>
<td>59.0</td>
<td>73.1</td>
<td>81.0</td>
<td>70.4</td>
<td>74.3</td>
</tr>
<tr>
<td>Birao virus</td>
<td>78.2</td>
<td>8-</td>
<td>-</td>
<td>78.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bozo virus</td>
<td>75.3</td>
<td>-</td>
<td>-</td>
<td>76.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cache Valley virus</td>
<td>81.9</td>
<td>59.7</td>
<td>76.1</td>
<td>86.2</td>
<td>74.1</td>
<td>77.5</td>
</tr>
<tr>
<td>Fort Sherman virus</td>
<td>80.2</td>
<td>59.8</td>
<td>76.3</td>
<td>84.1</td>
<td>74.3</td>
<td>78.8</td>
</tr>
<tr>
<td>Germiston virus</td>
<td>69.4</td>
<td>57.0</td>
<td>-</td>
<td>71.2</td>
<td>59.3</td>
<td>-</td>
</tr>
<tr>
<td>Ilesha virus</td>
<td>76.6</td>
<td>61.3</td>
<td>73.3</td>
<td>78.3</td>
<td>67.9</td>
<td>73.5</td>
</tr>
<tr>
<td>Main Drain virus</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>82.5</td>
<td>60.6</td>
<td>76.7</td>
</tr>
<tr>
<td>Maguari virus</td>
<td>81.6</td>
<td>59.7</td>
<td>76.2</td>
<td>85.2</td>
<td>76.4</td>
<td>79.1</td>
</tr>
<tr>
<td>Mboke virus</td>
<td>75.3</td>
<td>-</td>
<td>-</td>
<td>77.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ngari virus</td>
<td>85.3</td>
<td>59.5</td>
<td>73.9</td>
<td>85.9</td>
<td>71.3</td>
<td>73.8</td>
</tr>
<tr>
<td>Northway virus</td>
<td>82.5</td>
<td>60.6</td>
<td>76.8</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Playas virus</td>
<td>81.7</td>
<td>59.7</td>
<td>76.0</td>
<td>86.0</td>
<td>74.1</td>
<td>77.4</td>
</tr>
<tr>
<td>Potosi virus</td>
<td>82.3</td>
<td>72.6</td>
<td>75.5</td>
<td>86.3</td>
<td>60.4</td>
<td>77.1</td>
</tr>
<tr>
<td>Shokwe virus</td>
<td>79.4</td>
<td>-</td>
<td>-</td>
<td>80.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tensaw virus</td>
<td>81.3</td>
<td>60.6</td>
<td>76.5</td>
<td>86.0</td>
<td>73.1</td>
<td>78.6</td>
</tr>
<tr>
<td>Tlacotalpan virus</td>
<td>81.4</td>
<td>60.1</td>
<td>76.0</td>
<td>84.8</td>
<td>74.6</td>
<td>77.6</td>
</tr>
<tr>
<td>Xingu virus</td>
<td>75.8</td>
<td>-</td>
<td>-</td>
<td>76.8</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Sequence data not available

Genbank Accession Nos. used for the alignments are as follows: Bunyamwera virus (NC_001927.1, M11852.1, X14383.1), Batai virus (KU746869.1, KU746870.1, KU746871.1), Birao virus (AM711131.1), Bozo virus (AM711132.1), Cache Valley virus (KX100133.1, KX100134.1, KX100135.1), Fort Sherman virus (KX100130.1, KX100131.1, KX100132.1), Germiston virus (M19420.1, M21951.1), Ilesha virus (KF234073.1, KF234074.1, KF234075.1), Main Drain virus (X73469.1, EU004187.1, MG652604), Maguari virus (KY910431.1, KY910430.1, KY910429.1), Mboke virus (AY593727.1), Ngari virus (KM507341.1, KM514677.1, KM507334.1), Northway virus (X73470.1, EU004188.1, MG544835), Playas virus (KX100121.1, KX100122.1, KX100123.1), Potosi virus (MF066370.1, MF066369.1, MF066368.1), Shokwe virus (EU564831.1), Tensaw virus (FJ943505.1, FJ943506.1, FJ943509.1), Tlacotalpan virus (KX100118.1, KX100119.1, KX100120.1) and Xingu virus (EU564830.1).
CHAPTER 7. MERIDA VIRUS, A PUTATIVE NOVEL RHABDOVIRUS DISCOVERED IN CULEX AND OCHLEROTATUS SPP. MOSQUITOES IN THE YUCATAN PENINSULA OF MEXICO


Jermilia Charles, 1 Andrew E. Firth, 2 Maria A. Loroño-Pino, 3 Julian E. Garcia-Rejon, 3 Jose A. Farfan-Ale, 3 W. Ian Lipkin, 4 Bradley J. Blitvich, 1 and Thomas Briese 4

1 Department of Veterinary Microbiology and Preventive Medicine, College of Veterinary Medicine, Iowa State University, Ames, IA, USA
2 Department of Pathology, University of Cambridge, Cambridge, UK
3 Laboratorio de Arbovirología, Centro de Investigaciones Regionales Dr Hideyo Noguchi, Universidad Autónoma de Yucatán, Mérida, Yucatán, México
4 Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, NY, USA
Corresponding author: Bradley J. Blitvich; 2116 Veterinary Medicine, Iowa State University, Ames, IA 50011, USA, e-mail: blitvich@iastate.edu

Abstract

Sequences corresponding to a putative, novel rhabdovirus [designated Merida virus (MERDV)] were initially detected in a pool of Culex quinquefasciatus collected in the Yucatan Peninsula of Mexico. The entire genome was sequenced, revealing 11 798 nt and five major ORFs, which encode the nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and RNA-dependent RNA polymerase (L). The deduced amino acid sequences of the N, G and L proteins have no more than 24, 38 and 43 % identity, respectively, to the corresponding sequences of all other known rhabdoviruses, whereas those of the P and M proteins have no significant identity with any sequences in GenBank and their identity is only suggested based on their genome position. Using specific reverse transcription-PCR assays established from the genome sequence, 27 571 C. quinquefasciatus which had been sorted in 728 pools were screened to assess the
prevalence of MERDV in nature and 25 pools were found positive. The minimal infection rate (calculated as the number of positive mosquito pools per 1000 mosquitoes tested) was 0.9, and similar for both females and males. Screening another 140 pools of 5484 mosquitoes belonging to four other genera identified positive pools of *Ochlerotatus* spp. mosquitoes, indicating that the host range is not restricted to *C. quinquefasciatus*. Attempts to isolate MERDV in C6/36 and Vero cells were unsuccessful. In summary, we provide evidence that a previously undescribed rhabdovirus occurs in mosquitoes in Mexico.

**Introduction**

The family *Rhabdoviridae* (order *Mononegavirales*) is composed of a large and versatile group of viruses that are ubiquitous in nature (Kuzmin *et al.*, 2009). The family consists of 11 genera as well as several viruses that have not yet been assigned to a genus (http://www.ictvonline.org/virusTaxonomy.asp). Virions have a distinctive bullet or cone-shaped morphology or appear bacilliform. Rhabdoviruses have broad host ranges that include humans and other terrestrial mammals, birds, reptiles, fish, insects and plants (Hoffmann *et al.*, 2005; Kuzmin *et al.*, 2009; Mann & Dietzgen, 2014). Many rhabdoviruses are transmitted to vertebrate and plant hosts by insect vectors in which they replicate (Hogenhout *et al.*, 2003; Ammar *et al.*, 2009).

Rhabdoviruses have a negative-sense ssRNA genome of ~11–16 kb. A universal feature of the rhabdovirus genome is the presence of at least five genes that code in 3’ → 5’ order for the structural proteins: nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and RNA-dependent RNA polymerase (L) (Fu, 2005; Walker *et al.*, 2015). Each ORF is flanked by relatively conserved *cis*-acting transcription initiation and transcription termination/polyadenylation signals that regulate mRNA expression. Rhabdovirus genomes are
often interspersed with smaller ORFs that encode accessory proteins, most of which are of unknown function (Walker et al., 2011). ORFs that encode accessory proteins can occur as alternative or overlapping ORFs within the major structural protein genes or as independent ORFs in the intergenic regions that separate the structural protein genes. Leader and trailer sequences are located at the 3′ and 5′ termini of the rhabdovirus genome, respectively. These sequences are non-coding, A/U-rich and usually 50–100 nt in length. The first 10–20 nt of the leader and trailer sequences commonly exhibit partial complementary, and function as promoter sequences required for the initiation of genome and anti-genome replication, respectively (Fu, 2005; Walker et al., 2015).

Many novel rhabdoviruses have been discovered in recent years due to the advent of unbiased high-throughput sequencing (UHTS) (Quan et al., 2010; Ito et al., 2013; Kading et al., 2013; Tokarz et al., 2014; Binger et al., 2015; Sakai et al., 2015; Stremlau et al., 2015), including Bas-Congo virus which was associated with an outbreak of acute haemorrhagic fever in humans in the Democratic Republic of Congo (Grard et al., 2012). Here, we report the genomic organization and prevalence of an apparently novel rhabdovirus tentatively named Merida virus (MERDV) that was discovered by UHTS in mosquitoes in the Yucatan Peninsula of Mexico.

**Results**

**Discovery of MERDV**

UHTS of total RNA from a pool of *Culex quinquefasciatus* collected in Merida in the Yucatan Peninsula of Mexico generated a ~11 kb sequence corresponding to a putative, novel rhabdovirus tentatively named MERDV. Several other novel virus-like sequences were also discovered and
the data will be presented elsewhere. The MERDV genome terminal sequences were identified using a combination of 5′ and 3′ RACE and Sanger sequencing. Difficulties were encountered during the 3′ RACE because the reverse primer bound preferentially to an A-rich region located slightly upstream of the 3′ terminus. The 3′ end of the genome was eventually identified by taking advantage of the partial complementarity that exists between the 5′ and 3′ termini of the rhabdovirus genome. The 3′ end of the MERDV genome was amplified and sequenced using a reverse primer designed from the inverse complement of the 24nt sequence at the distal end of the 5′ terminus. Therefore, our sequence may contain nucleotide errors in the 24nt primer-binding region at the ultimate 3′ end of the genome.

**Genomic organization**

The MERDV genome consists of 11,798 nt (GenBank accession number KU194360) and its organization is consistent with that of the classical rhabdovirus genome: short leader and trailer sequences (68 and 74 nt, respectively) flank five structural protein genes in the order 3′-N–P–M–G–L-5′ (Fig. 1). The terminal nucleotides at the 5′ end of the MERDV genome are 5′-ACG-3′ and these same trinucleotides are located at the 5′ termini of other, related rhabdovirus genomes (Gubala *et al.*, 2008, 2011; Kuwata *et al.*, 2011; Zhu *et al.*, 2011). The leader and trailer sequences are 59 and 58 % A/U-rich, respectively. Each ORF is separated by a non-coding region of 54–153 nt that contains transcription initiation and termination sequences identified as AACAU and CAUG[A]7, respectively, through sequence alignment of conserved nucleotides. The only exception to these consensus motifs is the CUUG[A]7 transcription termination sequence that regulates M mRNA expression (Table 1).

The genomic location and length of each predicted ORF is shown in Table 2. The ORF that encodes the N protein consists of 1437 nt. The predicted translation product is most closely
related to the corresponding protein of *Culex tritaeniorhynchus* rhabdovirus (CTRV; 24% identity and 45% similarity) which was recently discovered in *C. tritaeniorhynchus* in Japan (Kuwata *et al.*, 2011) and Yongjia tick virus 2 (also 24% identity and 45% similarity) from *Haemaphysalis hystricis* ticks in China (Li *et al.*, 2015). The next two ORFs encode translation products that have no significant identity with any other sequences in GenBank. These two ORFs are assumed to encode the P and M proteins based on their positions in the genome. The ORF that encodes the G protein consists of 1530 nt and the predicted translation product is most closely related to the corresponding protein of CTRV (38% identity and 58% similarity). The next closest match is to a tandem rhabdovirus-like glycoprotein domain repeat sequence identified by the *Aedes aegypti* sequencing consortium (Nene *et al.*, 2007). The largest ORF in the MERDV genome consists of 6411 nt and encodes the L protein. The predicted translation product is also most closely related to the corresponding region of CTRV (44% identity and 65% similarity). Minor ORFs of 102, 62, 83 and 75 codons overlap the N, P, L and L genes, respectively, but are not obviously accessible via ribosomal scanning (Fig. 1). The AUG of a fifth ORF (75 codons) that overlaps the 5′ end of the P ORF is apparently upstream of the P mRNA transcription start site, so is also unlikely to be accessible for translation.

**Predicted domains and post-translational modifications**

The G proteins of rhabdoviruses have several common characteristics, including the presence of two to six potential N-linked glycosylation sites, 12 well-conserved cysteine residues, an N-terminal signal peptide, a transmembrane domain and a C-terminal short hydrophilic cytoplasmic domain (Coll, 1995; Walker & Kongsuwan, 1999). The G protein of MERDV is predicted to contain four potential N-linked glycosylation sites (one less than the G protein of CTRV) (Kuwata *et al.*, 2011). All 12 conserved cysteines are present with two
additional cysteines located at residues 12 and 485 (in CTRV one additional cysteine is present) (Kuwata et al., 2011). The G protein of MERDV is predicted to contain a signal peptide at residues 1–17, a hydrophobic transmembrane domain at residues 469–491 and a C-terminal hydrophilic cytoplasmic domain at residues 492–508, as common for rhabdoviral G proteins. Multiple protein kinase C (PKC) and tyrosine (TYR) phosphorylation sites are present in the N, P and M proteins of MERDV, consistent with many other rhabdoviruses including CTRV (Kuwata et al., 2011). Analysis with HHpred (Söding et al., 2005) revealed homology between the putative M protein of MERDV and Pfam family PF06326 (‘vesiculovirus matrix proteins’) indicating that it is indeed homologous to the M proteins of other rhabdoviruses.

**Phylogenetic relationship to the other rhabdoviruses**

The L protein sequences from MERDV and 73 other rhabdovirus species were aligned using MUSCLE (Edgar, 2004) and a phylogenetic tree was reconstructed using MrBayes (Ronquist et al., 2012) (Fig. 2). MERDV is most closely related to CTRV, consistent with the amino acid sequence alignments. More distantly, MERDV is related to North Creek virus which was identified in Culex sitiens in Australia (Coffey et al., 2014). Our analysis also indicated that MERDV cannot be assigned to one of the currently established rhabdovirus genera. Note that analyses of different genome regions (e.g. N) or of alignments with poorly aligning regions removed (e.g. with GBLOCKS; Castresana, 2000) provide different topologies in some of the deeper branches, but the clustering of MERDV with CTRV is consistent (data not shown).

**Prevalence in C. quinquefasciatus**

A total of 27,571 C. quinquefasciatus, sorted into 728 pools of up to 50 individuals, were screened by MERDV-specific reverse transcription (RT)-PCR. Collections were made in Merida and Tixkokob in 2007–2008 using mosquito magnets, and in Merida in 2013 using Centers for
Disease Control and Prevention (CDC) backpack-mounted aspirators. Mosquitoes obtained in 2007–2008 were tested according to gender, whereas those from 2013 were not. Overall, 256 pools (8038 mosquitoes) were composed of females, 195 pools (7196 mosquitoes) were composed of males and 277 pools (12 337 mosquitoes) were of mixed gender. Twenty-five mosquito pools were positive for MERDV RNA and the overall minimal infection rate (MIR; calculated as the number of positive mosquito pools per 1000 mosquitoes tested) was 0.9. The MIRs for female and male mosquitoes were similar (1.1 and 1.0, respectively). Evidence of MERDV infection was detected in mosquitoes collected in both study areas and during both time periods. See Tables 3 and 4.

**Detection of MERDV sequence in other mosquito species**

Another 5484 mosquitoes belonging to seven species were tested by RT-PCR using MERDV-specific primers RHAB-for and RHAB-rev as well as RHAB-121-for and RHAB-280-rev. Mosquito species tested were as follows: *Aedes aegypti* (*n* = 419), *Anopheles albimanus* (*n* = 727), *Anopheles crucians* (*n* = 691), *Anopheles vestitipennis* (*n* = 913), *Ochlerotatus taeniorhynchus* (*n* = 1000), *Ochlerotatus trivittatus* (*n* = 734) and *Psorophora cyanescens* (*n* = 1000) (Table 5). Collections were made using mosquito magnets at five study sites (Cozumel Island, Merida, Sian Ka'an, Tixkokob and Tzucacab) in 2007–2008. Mosquitoes had been sorted into 140 pools (20 pools per species) and all were female. MERDV RNA was detected in three pools of *O. taeniorhynchus* and in three pools of *O. trivittatus* using both primer pairs, whilst all other species were negative. The MERDV MIRs in *O. taeniorhynchus* and *O. trivittatus* were calculated as 3.0 and 4.1, respectively. All six PCR products generated using primers RHAB-121-for and RHAB-280-rev were analysed by Sanger sequencing. The resulting 114nt sequences had at least 99.1 %
nucleotide identity with the corresponding region of the MERDV genome sequence identified in *C. quinquefasciatus* (data not shown).

**Attempted virus isolations**

An aliquot of every homogenate positive for MERDV RNA (25 for *C. quinquefasciatus*, three for *O. taeniorhynchus* and three for *O. trivittatus*) was tested by inoculation of C6/36 cells. Cytopathic effects were not observed in any cultures and a faint RT-PCR signal was only occasionally observed in supernatants or cell lysates harvested from the initially inoculated C6/36 cell monolayers; no RT-PCR signal was obtained after any of the second or third blind passages. Three positive homogenates from *C. quinquefasciatus* were also tested by virus isolation in Vero cells, but all were negative.

**Dinucleotide usage preferences of MERDV**

Vertebrate, invertebrate and plant virus hosts preferentially have certain codon and dinucleotide usage biases, e.g. vertebrate sequences display a strong under-representation of UpA and CpG, whilst insect sequences display a strong under-representation of UpA but not of CpG (Simmen, 2008). RNA virus sequences often have preferences that mimic those of their native hosts (Greenbaum *et al.*, 2008; Atkinson *et al.*, 2014; Tulloch *et al.*, 2014). Thus, analysis of dinucleotide frequencies in virus genomes may be used to infer host taxa (Kapoor *et al.*, 2010). In a comparison of UpA and CpG usage in the L protein ORF of 80 National Center for Biotechnology Information rhabdovirus RefSeqs and MERDV, CTRV had the least under-representation of CpG (observed/expected ratio close to unity) whilst MERDV ranked second or third depending on the randomization protocol utilized (Fig. 3), suggesting that MERDV, as well as CTRV, are not well adapted to vertebrate hosts.
Discussion

The advent of UHTS has resulted in the discovery of many novel rhabdoviruses (Grard et al., 2012; Ito et al., 2013; Kading et al., 2013; Tokarz et al., 2014; Binger et al., 2015; Sakai et al., 2015; Stremlau et al., 2015), including several which were isolated from Anopheles, Culex, Ochlerotatus and Psorophora spp. mosquitoes (Quan et al., 2010; Coffey et al., 2014; Vasilakis et al., 2014). Here, we report the discovery of a putative, novel rhabdovirus in Culex and Ochlerotatus spp. mosquitoes from the Yucatan Peninsula of Mexico.

It is hypothesized that the majority of Rhabdoviruses are hosted by insects (Li et al., 2015), and numerous cyto-, nucleo- and dimarhabdoviruses are transmitted by arthropods to their plant or vertebrate hosts (Hogenhout et al., 2003; Bourhy et al., 2005). This includes the bite of haematophagous arthropods (Comer et al., 1990; Pérez De León et al., 2006), including Simulium vittatum blackflies, which can transmit vesicular stomatitis New Jersey virus to cattle under experimental conditions (Mead et al., 2009), and Phlebotomus argentipes sandflies, which efficiently transmitted Chandipura virus to laboratory mice (Mavale et al., 2007). It is currently unknown whether MERDV has the capacity to replicate in vertebrate hosts, but should its host range include vertebrates, it is unlikely that virus persistence in nature is dependent upon the bite of haematophagous arthropods. Male mosquitoes do not feed on blood, and thus a significant bias between male and female mosquito MIRs would be expected if that would be that case. Instead, the similar MIRs that we determined for MERDV are more compatible with vertical and venereal transmission. The occurrence of rhabdovirus transmission through these modes, in addition to horizontal transmission, has been demonstrated, for example, in Aedes aegypti mosquitoes for Chandipura virus (Mavale et al., 2005) and in
phlebotomine sandflies for vesicular stomatitis Indiana virus (Tesh et al., 1972). A hallmark in
the life cycle of insect-specific sigmaviruses is the exclusively vertical transmission through eggs
and sperm (Longdon et al., 2011). Currently, no data are available to decide to what extent
horizontal or vertical transmission contribute to the maintenance of MERDV in nature. Attempts
to isolate MERDV by inoculation of Vero cells were unsuccessful, and together with the
determined CpG and UpA dinucleotide usage biases, suggest that MERDV is likely not to infect
vertebrates in nature. However, attempts to isolate MERDV in C6/36 cells also were
unsuccessful thus far.

The inability to recover an isolate of MERDV in C6/36 cells is unexpected because this
cell line supports the replication of a diverse range of mosquito-associated viruses, including
several rhabdoviruses recently identified in Culex and Ochlerotatus spp. mosquitoes (Quan et al.,
2010; Kuwata et al., 2011; Coffey et al., 2014; Vasilakis et al., 2014). However, it is not without
precedent; Manitoba rhabdovirus from Culex tarsalis is reported to not propagate in C6/36 cells,
although it does replicate in Vero, primary chick embryo and mouse neuroblastoma cells
(Artsob et al., 1991). Shortcomings in sample handling and possible failures in the cold-chain
during transport are unlikely for the failure to isolate MERDV because one-quarter of our
mosquito homogenates induced virus-like cytopathic effect when inoculated onto C6/36 cells,
indicating that other, undetermined viruses did successfully propagate. Another possibility is that
MERDV does not actively replicate in mosquitoes. It cannot be excluded that some of the field-
collected mosquitoes passively carried MERDV without the virus being capable of replicating in
the mosquito, despite the detection over multiple years, several locations and specific species.
Alternatively, we may have discovered another example of endogenous viral elements analogous
to the rhabdovirus-like sequences described previously in various insect hosts (Nene et al.,
2007; Katzourakis & Gifford, 2010; Li et al., 2015). However, such endogenous viral elements have thus far been reported to concern only partial sequences, at best covering one gene, but never what appears to be a complete, functional genome as we have found for MERDV.

Amino acid sequence alignments and phylogenetic analyses indicated CTRV as the closest known relative of MERDV. CTRV was isolated in C6/36 cells from *C. tritaeniorhynchus* in Japan (Kuwata et al., 2011), and later detected in *Culex, Aedes, Anopheles* and *Armigeres* spp. mosquitoes in China (Li et al., 2015; Shi et al., 2015). Recent studies have also shown that NIID-CTR cells, which were established from *C. tritaeniorhynchus* embryos, are persistently infected with CTRV (Gillich et al., 2015). CTRV establishes a non-cytolytic infection and, similar to sigmaviruses, employs vertical transmission. However, in contrast to sigmaviruses, CTRV replicates in the nucleus of the infected cell similar to nucleorhabdoviruses and it is the only known rhabdovirus that requires the cellular splicing machinery for its mRNA maturation. The coding region for the L protein of CTRV is interrupted by a 76nt intron (Kuwata et al., 2011). Inspection of the MERDV sequence provided no evidence for the use of splicing similar to CTRV, suggesting that MERDV may not require a nuclear phase. Our proposed transcriptional signals match those confirmed for CTRV. Whilst conservation of the termination signal sequence is seen in comparison to other rhabdoviruses, including the *Drosophila*-specific sigmaviruses and Moussa virus, a potentially mosquito-specific rhabdovirus from *Culex decans* mosquitoes (Quan et al., 2010), the initiation signal sequence differs from the two other viruses. Additionally, whereas sigmaviruses and Moussa virus do not show overlap of genes, the G and L genes of CTRV and both the M/G and G/L genes of MERDV show overlap. Other rhabdoviruses also possess overlapping transcription termination and transcription initiation sequences in their genomes including two more recently
discovered mosquito-associated rhabdoviruses: Malpais Spring virus and Oak Vale virus (Quan et al., 2011; Vasilakis et al., 2013). Indeed, the positioning of the initiation signal of the downstream gene in front of the termination signal of the preceding gene or the use of splicing are not unprecedented in mononegaviruses, e.g. in human metapneumovirus or in bornaviruses, where these mechanisms have been hypothesized to adjust transcription levels possibly in conjunction with persistent infection, or attenuate gene expression in addition to the $3'\rightarrow 5'$ transcriptional gradient characteristic for mononegaviruses (Collins et al., 1987; Schneemann et al., 1994; Schneider et al., 1994).

The translated primary sequences of MERDV N, G and L ORFs show significant homology to the respective ORFs of other rhabdoviruses. However, as also observed for CTRV, sigmaviruses and Moussa viruses, the P and M ORFs are far more diverse and do not exhibit homology to any other sequences in GenBank or to each other (except for the M proteins of sigmaviruses, which have a very distant relationship to the corresponding protein of Flanders virus). In addition, sigmaviruses contain an additional ORF (designated ORF X) between the P and G ORFs.

Although CTRV is the closest known relative of MERDV, these two viruses exhibit considerable sequence dissimilarity. The L proteins of MERDV and CTRV, which represent the most conserved genome regions of the rhabdoviruses, show an amino acid divergence of 56 %. This is reaching the divergence observed between rhabdoviruses belonging to different genera, which is commonly in a range of 47–83 % (Table S1, available in the online Supplementary Material). As species and genus demarcations for rhabdoviruses also include factors such as biological characteristics (e.g. host range) and serological cross-reactivity, additional work will
be needed to accurately determine the taxonomic status of MERDV within the family *Rhabdoviridae*.

In summary, we provide evidence that a novel rhabdovirus occurs in mosquitoes in the Yucatan Peninsula of Mexico. This apparent virus, provisionally named MERDV, is most closely related to CTRV, although it shows considerable sequence and biological divergence. Our findings underscore the vast diversity of this virus family, highlight the power of next-generation sequence technology in the discovery of novel viruses and provide the basis for improved surveillance programmes to gain better insights into arbovirus evolution.

**Methods**

**Mosquito collections**

Mosquitoes were collected in five study areas in the Yucatan Peninsula of Mexico: Cozumel Island, Merida, Sian Ka'an, Tixkokob and Tzucacab. Descriptions of these study areas are provided elsewhere (Farfan-Ale *et al.*, 2009, 2010). Collections were made in 2007 and 2008 using Mosquito magnets (all five study areas) and in 2013 using CDC backpack-mounted aspirators (Merida only). Mosquito magnets Pro-Liberty (American Biophysics) were baited with propane and octenol, and placed outdoors. Mosquito magnets were turned on between 16:00 and 18:00 and collection nets were replaced the following morning between 06:00 and 09:00. CDC back-pack mounted aspirators were used to collect resting mosquitoes inside private residences. Mosquitoes were transported alive to the Universidad Autonoma de Yucatan, frozen at −80 °C, and identified on chill tables according to species and sex using morphological characteristics (Darsie, 1996). Mosquitoes were transported on dry ice from the Universidad Autonoma de Yucatan to Iowa State University by World Courier.
High-throughput sequencing

Mosquitoes were homogenized as described previously (Farfan-Ale et al., 2009) and total RNA was extracted using TRIzol (Invitrogen) following the manufacturer's instructions. Extracts were reverse transcribed using SuperScript III (Thermo Fisher) with random hexamers. The cDNA was RNase-H-treated prior to second-strand synthesis with Klenow fragment (NEB). The generated double-stranded cDNA was sheared to a mean fragment size of 200bp using a Covaris focused-ultrasonicator E210 and the manufacturer's standard settings. Sheared products were purified (Agencourt Ampure DNA purification beads; Beckman Coulter) and libraries constructed. Sheared nucleic acid was end-repaired, dA-tailed, ligated to sequencing adapters (NEBNext modules; NEB), PCR-amplified (Phusion High-Fidelity DNA polymerase; NEB) and quantified by an Agilent Bioanalyzer for sequencing. Sequencing on the Illumina HiSeq 2500 platform (Illumina) resulted in a mean of 180 million reads per lane. Samples were de-multiplexed using Illumina software and FastQ files generated. Data were quality-filtered and trimmed (Slim-Filter), and de novo assembled using Dwight assembler at custom settings (Golovko et al., 2012). The generated contiguous sequences (contigs) and unique singleton reads were subjected to homology search using BLASTN and BLASTX against GenBank.

RT-PCR and Sanger sequencing

Total RNA was analysed by RT-PCR using MERDV-specific primers RHAB-for (5’-CAATCACATCGACTACTCTAAATGGA-3’) and RHAB-rev (5’-GATCAGACCTAGCTTGCTGTTC-3’), which target a 490nt region of the L protein gene, or RHAB-121-for (5’-AACGCCCGACATGACTACTATCG-3’) and RHAB-280-rev (5’-TTCCGTACCTCCCATATGAGTGG-3’), which target a 160nt region of the N protein. cDNAs were generated using SuperScript III reverse transcriptase (Invitrogen), and PCRs were performed using Taq polymerase (Invitrogen) and the following cycling conditions: 94 °C for
3 min, then 35 cycles of 94 °C for 30 s, 56 °C for 45 s and 72 °C for 1 min followed by a final extension at 72 °C for 8 min. RT-PCR products were purified using the Purelink Gel Extraction kit (Invitrogen). Sanger sequencing was performed using a 3730 × 1 DNA sequencer (Applied Biosystems).

5′ and 3′ RACE

The extreme 5′ and 3′ ends of the MERDV genome were determine by 5′ and 3′ RACE, respectively. In the 5′ RACE reactions, total RNA was reversed transcribed using a MERDV-specific primer (5′-CTCAGAACGGAAGAGGTATACTG-3′). cDNAs were purified by ethanol precipitation and oligo(dC) tails were added to the 3′ ends using 15 U terminal deoxynucleotidyltransferase (Invitrogen) in 1 × tailing buffer (10mM Tris/HCl, pH 8.4, 25mM KCl, 1.5mM MgCl₂ and 0.02mM dCTP). Tailing reactions were performed at 37 °C for 30 min and then terminated by heat inactivation (65 °C for 10 min). Oligo dC-tailed cDNAs were purified by ethanol precipitation, and then PCR-amplified using a consensus forward primer specific to the C-tailed termini (5′-GACATCGAAAGGGGGGGGGG-3′) and a reverse primer specific to the MERDV cDNA sequence (5′-TTCCGTACCTCCCATATGAGTGG-3′). In the 3′ RACE reactions, polyadenylate [poly(A)] tails were added to the 3′ ends of the genomic RNA using 6 U poly(A) polymerase (Ambion) in 1 × reaction buffer (40mM Tris/HCl, pH 8.0, 10mM MgCl₂, 2.5mM MnCl₂, 250mM NaCl, 50 μg BSA ml⁻¹ and 1mM ATP). Tailing reactions were performed at 37 °C for 1 h and terminated by heat inactivation (65 °C for 10 min). Poly(A)-tailed RNA was reverse transcribed using a poly(A) tail-specific primer (5′-GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTT-3′). Complementary DNAs were PCR amplified using a forward primer specific to the MERDV cDNA sequence (5′-
AAGAACATCGGGTATTGATCCGG-3′) and a reverse primer that matched the 5′ half of the poly(A)-specific RT primer (5′-GGCCACCGCTCGACTAGTAC-3′).

PCR products generated from the 5′ and 3′ RACE reactions were inserted into the pCR4-TOPO cloning vector (Invitrogen), and ligated plasmids were transformed into competent TOPO10 *Escherichia coli* cells (Invitrogen). Cells were grown on Luria–Bertani agar containing ampicillin (50 μg ml⁻¹) and kanamycin (50 μg ml⁻¹), and colonies were screened for inserts by PCR amplification. An aliquot of each PCR product was examined by 1% agarose gel electrophoresis, and selected PCR products were purified by QIAquick spin column (Qiagen) and sequenced using a 3730x1 DNA sequencer.

**Amino acid sequence alignments and prediction algorithms**

The predicted amino acid sequences of MERDV were compared to all other sequences in GenBank by application of TBLASTN (Altschul *et al*., 1990). Per cent amino acid identities and similarities of select rhabdovirus protein sequences were calculated using CLUSTAL W (http://simgene.com/ClustalW). The following prediction algorithms were used for the amino acid sequence analysis: NetNGlyc 1.0 server (for the identification of potential N-linked glycosylation sites), SignalP 4.1 server (for the identification of potential signal peptides), TMHMM server 2.0 (for the identification of potential transmembrane domains and cytoplasmic domains) and NetPhos 2.0 Server (for the identification of potential PKC and TYR phosphorylation sites).

**Virus isolation**

An aliquot (200 μl) of each supernatant that tested positive for MERDV RNA was added to 2ml Liebovitz's L15 medium (Invitrogen) supplemented with 2 % FBS, 2 mM L-glutamine, 100U ml⁻¹ penicillin, 100 μg ml⁻¹ streptomycin and 2.5 μg fungizone ml⁻¹. Samples were
filtered using a 0.22 μm filter and inoculated onto subconfluent monolayers of *Aedes albopictus* C6/36 cells in 75 cm² flasks. Cells were incubated for at least 1 h at room temperature on an orbital shaker. Another 12 ml L15 maintenance medium was added to each flask and cells were incubated at 28 °C for 7 days. After two additional blind passages, supernatants were harvested and tested by RT-PCR for the presence of MERDV RNA.

**Acknowledgements**

The authors thank Valeria Bussetti for expert technical assistance. This study was supported by the National Institutes of Health (5R21AI067281, AI057158, 5R21AI067281 and AI088647;), the US Department of Defense and an intramural grant from Iowa State University. A. E. F. is supported by a grant from the Wellcome Trust (106207).

**References**


of Long Island tick rhabdovirus, a new virus identified in Amblyomma americanum ticks. Virol J
11, 26.

attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide
frequencies. eLife 3, e04531.

10, 69.

Vasilakis, N., Castro-Llanos, F., Widen, S. G., Aguilar, P. V., Guzman, H., Guevara, C.,
Almendras viruses: two novel rhabdoviruses isolated from mosquitoes in Peru. J Gen Virol
95, 787–792.


Walker, P. J., Firth, C., Widen, S. G., Blasdell, K. R., Guzman, H., Wood, T. G., Paradkar, P.

rhabdovirus exhibits unusual non-coding regions and an additional ORF that could be expressed
in fish cell. Virus Res 155, 495–505.
Fig. 1. **Coding capacity of MERDV sequence.** Main ORFs, as well as minor ORFs unlikely to be expressed, are indicated along a scaled representation of the anti-genomic strand. The size of minor ORFs is indicated by their codon number: blue, frame 1; green, frame 2; red, frame 3.
Fig. 2. Phylogenetic tree for MERDV and selected other rhabdovirus sequences. L protein amino acid sequences were aligned using muscle (Edgar, 2004). A maximum-likelihood phylogenetic tree was estimated using the Bayesian Markov chain Monte Carlo method implemented in MrBayes version 3.2.3 (Ronquist et al., 2012) sampling across the default set of fixed amino acid rate matrices with 10 million generations, discarding the first 25% as burn-in. The original figure was produced using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The tree is midpoint-rooted and selected nodes are labelled with posterior probability values. Rhabdovirus genera, where defined, are labelled on the far right. GenBank accession numbers are indicated next to virus names. Bar indicates amino acid substitutions per site.
Fig. 3. Relative UpA and CpG frequencies in the L protein ORF of different rhabdovirus species. UpA and CpG frequencies were calculated in two different ways. (a) In each sequence, the numbers of UpA and CpG dinucleotides, and A, C, G and U mononucleotides, were counted. Dinucleotide frequencies, $f_{XY}$, were expressed relative to their expected frequencies, $f_X \times f_Y$, in the absence of selection. (b) To factor-out codon and amino acid usage, 1000 shuffled ORF sequences were generated for each virus sequence. In each shuffled sequence, the original amino acid sequence and the original total numbers of each of the 61 codons were maintained, but synonymous codons were randomly shuffled between the different sites where the corresponding amino acid is used in the original sequence. Next, the UpA and CpG frequencies in the original sequence were expressed relative to their mean frequencies in the codon-shuffled sequences. As codon usage is factored-out, the UpA and CpG relative frequencies tend to be less extreme in (b) compared with (a). Each point represents a single rhabdovirus sequence. Solid points correspond to species within defined genera, colour coded by genus (see key). Annotated open circles correspond to species that are currently unassigned at genus level, colour coded by host (or presumed host) taxa. Asterisks in the key indicate clades with uncertain host taxa: viruses in the unclassified ‘arthropod-infecting’ clades (yellow open circles) were isolated from arthropods but
not from vertebrates; the sole representative of the genus Tupavirus was isolated from mammals but not from arthropods, although its phylogenetic position suggests that it may be arthropod-borne; the presence of viruses derived from vertebrates and viruses derived from arthropods in each of the unclassified ‘vertebrate (arthropod-borne)’ clades (brick-red open circles) suggests that all of these viruses are likely arboviruses. GenBank accession numbers of sequences used: NC_000855, NC_000903, NC_001542, NC_001560, NC_001615, NC_001652, NC_002251, NC_002526, NC_002803, NC_003243, NC_003746, NC_005093, NC_005974, NC_005975, NC_006429, NC_006942, NC_007020, NC_007642, NC_008514, NC_009527, NC_009528, NC_011532, NC_011542, NC_011639, NC_013135, NC_013955, NC_016136, NC_017685, NC_017714, NC_018381, NC_018629, NC_020803, NC_020804, NC_020805, NC_020806, NC_020807, NC_020808, NC_020809, NC_020810, NC_022580, NC_022581, NC_022755, NC_024473, NC_025251, NC_025253, NC_025255, NC_025340, NC_025341, NC_025342, NC_025353, NC_025354, NC_025356, NC_025358, NC_025359, NC_025362, NC_025364, NC_025365, NC_025371, NC_025376, NC_025377, NC_025378, NC_025382, NC_025384, NC_025385, NC_025387, NC_025389, NC_025391, NC_025392, NC_025393, NC_025394, NC_025395, NC_025396, NC_025397, NC_025398, NC_025399, NC_025400, NC_025401, NC_025405, NC_025406, NC_025408.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Noncoding sequences at the start of the gene</th>
<th>Noncoding sequences at the end of the gene</th>
<th>^I ^G ^( \text{GR} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>\textit{AACAUAAACUACGACCUGGAAUCCGAUCA}CA</td>
<td>\textit{CAUGAAAAAAAA}</td>
<td>A</td>
</tr>
<tr>
<td>P</td>
<td>\textit{AACAUAAACUACGACCUGGAAUCCGAUCA}CA</td>
<td>\textit{CAUGAAAAAAAA}</td>
<td>A</td>
</tr>
<tr>
<td>M</td>
<td>\textit{AACAUAAACUACGACCUGGAAUCCGAUCA}CA</td>
<td>\textit{CAUGAAAAAAAA}</td>
<td>A</td>
</tr>
<tr>
<td>G</td>
<td>\textit{AACAUAAACUACGACCUGGAAUCCGAUCA}CA</td>
<td>\textit{CAUGAAAAAAAA}</td>
<td>A</td>
</tr>
<tr>
<td>L</td>
<td>\textit{AACAUAAACUACGACCUGGAAUCCGAUCA}CA</td>
<td>\textit{CAUGAAAAAAAA}</td>
<td>A</td>
</tr>
</tbody>
</table>

\*Intergenic region located immediately downstream of the gene of interest that is neither translated nor transcribed to mRNA; \^\( \text{N/ A} \)No IGR present due to gene overlap by 25 nt between the M and G genes and 13 nt between the G and L genes; solid underline indicates sequence overlap. Bolded sequence indicates the predicted transcription start/stop sites. The three nucleotides indicated by dashed underline (AAC) and the AU of the start codon immediately downstream (not shown) are not predicted to serve as a transcription start site because they overlap the M coding sequence and are not in close proximity to any of the other coding sequences. Leader and trailer sequences are italicized.
Table 2. Predicted locations and lengths the open reading frames encoded by the Merida virus genome

<table>
<thead>
<tr>
<th>Protein</th>
<th>ORF Genomic location</th>
<th>ORF length (nt)</th>
<th>Protein length (aa)</th>
<th>Protein mass (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>131-1567</td>
<td>1437</td>
<td>478</td>
<td>54.2</td>
</tr>
<tr>
<td>P</td>
<td>1660-2862</td>
<td>1203</td>
<td>400</td>
<td>43.8</td>
</tr>
<tr>
<td>M</td>
<td>2917-3477</td>
<td>564</td>
<td>187</td>
<td>21.0</td>
</tr>
<tr>
<td>G</td>
<td>3586-5112</td>
<td>1530</td>
<td>509</td>
<td>56.8</td>
</tr>
<tr>
<td>L</td>
<td>5266-11676</td>
<td>6411</td>
<td>2136</td>
<td>241.2</td>
</tr>
</tbody>
</table>
Table 3. Minimal infection rates for Merida virus in *Culex quinquefasciatus* in the Yucatan Peninsula of Mexico, 2007-2008 and 2013

<table>
<thead>
<tr>
<th>Study site</th>
<th>Date</th>
<th>No. mosquitoes tested</th>
<th>No. pools tested (positive)</th>
<th>(^{\ddagger})MIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(^{*})Tixkokob</td>
<td>2007-2008</td>
<td>9071</td>
<td>247 (2)</td>
<td>0.2</td>
</tr>
<tr>
<td>(^{*})Merida</td>
<td>2007-2008</td>
<td>6163</td>
<td>204 (14)</td>
<td>2.3</td>
</tr>
<tr>
<td>(^{*})Merida</td>
<td>2013</td>
<td>12,337</td>
<td>277 (9)</td>
<td>0.7</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>27,571</td>
<td>728 (25)</td>
<td>0.9</td>
</tr>
</tbody>
</table>

*Mosquitoes were collected outdoors using mosquito magnets; \(^{\ddagger}\)Mosquitoes were collected inside private residences using CDC backpack mounted aspirators; \(^{\ddagger}\)Minimal infection rates are expressed as the number of positive mosquito pools per 1,000 mosquitoes tested.
Table 4. Comparison of minimal infection rates for Merida virus in female and male *Culex quinquefasciatus* in the Yucatan Peninsula of Mexico, 2007-2008

<table>
<thead>
<tr>
<th>Study site</th>
<th>No. mosquitoes tested</th>
<th>No. pools tested (positive)</th>
<th>†MIR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Females</td>
<td>Males</td>
<td>Total</td>
</tr>
<tr>
<td><strong>Merida</strong></td>
<td>3018</td>
<td>3145</td>
<td>6163</td>
</tr>
<tr>
<td></td>
<td>(7)</td>
<td>(7)</td>
<td>(14)</td>
</tr>
<tr>
<td><strong>Tixkokob</strong></td>
<td>5020</td>
<td>4051</td>
<td>9071</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>(0)</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>8038</td>
<td>7196</td>
<td>15,234</td>
</tr>
<tr>
<td></td>
<td>(9)</td>
<td>(7)</td>
<td>(16)</td>
</tr>
</tbody>
</table>

*Cx. quinquefasciatus* collected in Merida in 2013 are not listed because males and females were not tested separately; †MIRs are expressed as the number of positive mosquito pools per 1,000 mosquitoes tested; ‡Mosquitoes were collected outdoors using mosquito magnets.
Table 5. Minimal infection rates for Merida virus in selected *Aedes, Anopheles, Ochlerotatus* and *Psorophora* spp. mosquitoes

<table>
<thead>
<tr>
<th>Species</th>
<th><em>No. mosquitoes tested</em></th>
<th>No. pools tested (positive)</th>
<th>†MIR</th>
<th>‡Study site where positive pools were collected</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ae. aegypti</em></td>
<td>419</td>
<td>20 (0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><em>An. albimanus</em></td>
<td>727</td>
<td>20 (0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><em>An. crucians</em></td>
<td>691</td>
<td>20 (0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><em>An. vestitipennis</em></td>
<td>913</td>
<td>20 (0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><em>Oc. taeniorhynchus</em></td>
<td>1000</td>
<td>20 (3)</td>
<td>3.0</td>
<td>Cozumel Island</td>
</tr>
<tr>
<td><em>Oc. trivittatus</em></td>
<td>734</td>
<td>20 (3)</td>
<td>4.1</td>
<td>Cozumel Island, Merida and Tzucacab</td>
</tr>
<tr>
<td><em>Ps. cyanescens</em></td>
<td>1000</td>
<td>20 (3)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>5484</td>
<td>140 (6)</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

*All mosquitoes were female; †Minimal infection rates are expressed as the number of positive mosquito pools per 1,000 mosquitoes tested, ‡Mosquitoes were collected using mosquito magnets at five study sites (Tixkokob, Merida, Cozumel Island, Sian Ka’an and Tzucacab) in 2007-2008*
Supplemental Data

Table S1. Pairwise amino acid sequence identity analysis using L protein sequences of representative members of the 11 currently defined rhabdovirus genera.

<table>
<thead>
<tr>
<th></th>
<th>BEFV</th>
<th>DMelSV</th>
<th>DURV</th>
<th>IHNV</th>
<th>LNYV</th>
<th>PRV</th>
<th>PYDV</th>
<th>RABV</th>
<th>SVCV</th>
<th>TIBV</th>
<th>VSIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEFV</td>
<td></td>
<td>-</td>
<td>37.8</td>
<td>39.6</td>
<td>18.6</td>
<td>20.0</td>
<td>41.3</td>
<td>19.0</td>
<td>31.7</td>
<td>42.5</td>
<td>43.8</td>
</tr>
<tr>
<td>DMelSV</td>
<td>37.8</td>
<td>-</td>
<td>39.6</td>
<td>18.0</td>
<td>19.9</td>
<td>39.5</td>
<td>17.1</td>
<td>33.3</td>
<td>39.3</td>
<td>35.3</td>
<td>39.4</td>
</tr>
<tr>
<td>DURV</td>
<td>39.6</td>
<td>38.6</td>
<td></td>
<td>18.8</td>
<td>20.2</td>
<td>41.5</td>
<td>18.5</td>
<td>32.4</td>
<td>41.3</td>
<td>37.8</td>
<td>41.7</td>
</tr>
<tr>
<td>IHNV</td>
<td>18.6</td>
<td>18.0</td>
<td>18.8</td>
<td></td>
<td>17.4</td>
<td>18.7</td>
<td>17.6</td>
<td>18.9</td>
<td>18.0</td>
<td>18.6</td>
<td>17.6</td>
</tr>
<tr>
<td>LNYV</td>
<td>20.0</td>
<td>19.9</td>
<td>20.2</td>
<td>17.4</td>
<td></td>
<td>20.0</td>
<td>25.7</td>
<td>21.2</td>
<td>20.0</td>
<td>19.9</td>
<td>20.0</td>
</tr>
<tr>
<td>PRV</td>
<td>41.3</td>
<td>39.5</td>
<td>41.5</td>
<td>18.7</td>
<td>20.0</td>
<td>-</td>
<td>18.8</td>
<td>33.0</td>
<td>46.6</td>
<td>39.0</td>
<td>44.9</td>
</tr>
<tr>
<td>PYDV</td>
<td>19.0</td>
<td>17.1</td>
<td>18.5</td>
<td>17.6</td>
<td>25.7</td>
<td>18.8</td>
<td>-</td>
<td>20.4</td>
<td>18.4</td>
<td>18.5</td>
<td>18.4</td>
</tr>
<tr>
<td>RABV</td>
<td>31.7</td>
<td>33.3</td>
<td>32.4</td>
<td>18.9</td>
<td>21.2</td>
<td>33.0</td>
<td>20.4</td>
<td>-</td>
<td>34.1</td>
<td>31.4</td>
<td>33.8</td>
</tr>
<tr>
<td>SVCV</td>
<td>42.5</td>
<td>39.3</td>
<td>41.3</td>
<td>18.0</td>
<td>20.0</td>
<td>46.6</td>
<td>18.4</td>
<td>34.1</td>
<td>-</td>
<td>40.3</td>
<td>52.6</td>
</tr>
<tr>
<td>TIBV</td>
<td>43.8</td>
<td>35.3</td>
<td>37.8</td>
<td>18.6</td>
<td>19.9</td>
<td>39.0</td>
<td>18.5</td>
<td>31.4</td>
<td>40.3</td>
<td>-</td>
<td>39.8</td>
</tr>
<tr>
<td>VSIV</td>
<td>41.8</td>
<td>39.4</td>
<td>41.7</td>
<td>17.6</td>
<td>20.0</td>
<td>44.9</td>
<td>18.4</td>
<td>33.8</td>
<td>52.6</td>
<td>39.8</td>
<td>-</td>
</tr>
</tbody>
</table>

Pairwise sequence alignments were performed using the following viruses (with abbreviation, genus name and Genbank Accession No. denoted in parenthesis): bovine ephemeral fever virus (BEFV, Ephemerovirus, AAG10420), Drosophila melanogaster sigmavirus (DMelSV, Sigmavirus, CBA18272), Durham virus (DURV, Tupavirus, ADB88761, partial sequence), infectious hematopoietic necrosis virus (IHNV, Novirhabdovirus, AAC42155), lettuce necrotic yellows virus (LNYV, Cytorhabdovirus, CAI30426), perch rhabdovirus (PRV, Perhabdovirus, AFX72892), Potato yellow dwarf virus (PYDV, Nucleorhabdovirus, ADE45274), rabies virus (RABV, Lyssavirus, AAA47219), spring viraemia of carp virus (SVCV, Sprivivirus, AAK60424), Tibrogargan virus (TIBV, Tibrovirus, ADG86355) and vesicular stomatitis virus (VSIV, Vesiculovirus, AAA48442). These viruses are the type virus species of their respective genera.
CHAPTER 8. GENERAL CONCLUSIONS

Air travel has indeed made the world a smaller place, with the year 2013 marking a century since its launch. Although globalization has opened for us several doors granting us exposure to new cultures and increased opportunity, it also carries with it its own unique wave of challenges. Along with climate change, the advent of air travel has exposed us to diseases transmitted by vectors that were once thought to be restricted to tropical or subtropical areas. The last two decades and even the last five years witnessed the first cases of major medically relevant arboviruses in the Western Hemisphere. West Nile virus (WNV) was first introduced to the United States two decades ago as it made its debut in the New York epidemic of 1999 (Petersen and Hayes 2008). More recently, Chikungunya virus (CHIKV) (2013), an alphavirus and Zika virus (ZIKV) (2015), a flavivirus, were introduced to the Americas both resulting in newsworthy outbreaks (Morrison 2014, White, Wollebo et al. 2016).

Although the flavivirus group of arboviruses harbors a number of medically relevant viruses, two smaller subsets within this genus also exist, constituting viruses that exhibit a more limited host range. The insect-specific flaviviruses (ISFs) have been isolated from insects collected on every inhabited continent but lack the capacity to replicate in vertebrate cells tested or mice inoculated with virus (Moureau, Cook et al. 2015). The no known vector flaviviruses (NKVs), on the other hand, have been isolated from bats and rodents but do not appear to be vectored by arthropods. The fact that these viruses have failed to replicate in mosquito cell lines or experimentally infected mosquitoes would suggest that they are reliant on other means for propagation and transmission in nature. Indeed, there is evidence to support the possibility of horizontal transmission among bats or rodents nesting together in close-knit communities (Constantine and Woodall 1964, Fairbrother and Yuill 1987, Volkova, Tesh et al. 2012).
Over the years, the ISFs and NKVs have failed to receive the level of attention given to the arboviral flaviviruses due to the fact that these viruses did not appear to be medically relevant. However, studies have shown that prior infection with certain viruses can actually interfere with infection or transmission of a secondary virus. A study done by Newman and colleagues in 2011 showed a positive correlation between WNV and Culex flavivirus (CxFV) infection. The authors reported a fourfold increase in the occurrence of WNV infection in mosquitoes coinfected with CxFV as compared to spatiotemporally matched pools which tested negative for WNV (Newman, Cerutti et al. 2011). Another study done by Hobson-Peters and colleagues in 2013 reported suppression of WNV and Murray Valley encephalitis virus (MVEV) by an ISF, Palm Creek virus (PCV), a phenomenon known as superinfection exclusion (Hobson-Peters, Yam et al. 2013). Taking both of these possibilities into account, ISFs could be used to serve as a means of controlling the transmission and spread of medically relevant flaviviruses but this would require thorough research and would have to be addressed on a case by case basis.

Studies investigating the genetic elements governing flaviviral host specificity could have major implications for vaccine design as well as the discovery of novel flaviviral drug targets. Such studies would also provide insights into the compulsory evolutionary hurdles necessary to make the leap from ISF or NKV flavivirus to dual-host flavivirus. In addition to this, such data would also provide us with tools which would assist in identifying regions in newly discovered viruses that would make them a threat to public or veterinary health.

Studies both done in 2017 by Junglen and colleagues and Piyasena and colleagues, respectively, showed that classical ISF host restriction occurs at multiple levels including the initial stages of attachment and entry (Junglen, Korries et al. 2017, Piyasena, Setoh et al. 2017). No such studies have been done, however, using dual-host affiliated ISFs which are a group of
flaviviruses which, despite their apparent insect-specific host phenotype, cluster phylogenetically with the dual-host flaviviruses. We therefore attempted the generation of chimeras using a representative dual-host flavivirus whose premembrane and envelope genes was swapped with the corresponding genes of a representative dual-host affiliated ISFs as a means of studying host restriction at the level of attachment and entry. We were however unable to generate any functional chimeras, perhaps due to genetic incompatibilities present between the two different genomes.

NKV flavivirus chimeras have also been designed and characterized by a few groups, the first of them being two chimeras generated between two different dual-host flaviviruses, dengue (DENV) and yellow fever virus (YFV) and a representative rodent-associated NKV flavivirus, Modoc virus (MODV), in 2010. The authors studied host-restriction at the level of attachment and entry and concluded the host-restriction in MODV was determined by a stage post attachment/entry (Charlier, Davidson et al. 2010). Tumban and colleagues also generated two chimeras where the conserved pentanucleotide sequence and variable regions in the 3’ untranslated region of DENV, a dual-host flavivirus, was replaced with the corresponding regions of MODV. The authors concluded that these regions did not affect host tropism (Tumban, Maes et al. 2013). We too generated two functional NKV chimeras between two representative dual-host flaviviruses- YFV and ZIKV- and a representative bat-associated NKV flavivirus-Rio Bravo virus (RBV). These chimeras were the first chimeras to be generated using a bat-associated NKV flavivirus. Like Charlier and colleagues, both of our chimeras showed the capacity to replicate in both vertebrate and insect cells suggesting that host-restriction in both bat- and rodent-associated NKV flaviviruses occur at a downstream stage in the viral life cycle.
Additional studies utilizing chimeras generated by swapping additional proteins between dual-host and NKV flaviviruses may provide greater insight into host-restriction among these viruses.

High throughput sequencing technology has significantly simplified the discovery and identification of novel and recognized pathogens including viruses. In this dissertation, I reported the discovery of four novel viruses—a Tymoviridae-like virus, designated Ek Balam virus, a negevirus, designated Uxmal virus, a Nodaviridae-like virus, designated Mayapan virus and a rhabdovirus, designated Merida virus. Two recognized viruses were also identified and isolated during one of our metagenomics studies, one of them being an ISF, Culex flavivirus and the other being an alphamesonvirus, Houston virus. These viruses were all identified in pools of mosquitoes collected from the Yucatan Peninsula of Mexico.

The first of these viruses to be discovered was Merida virus, named after the location in which the virus was discovered. This virus was found to be most closely related to Culex tritaeniorhynchus rhabdovirus (CTRV), a rhabdovirus isolated from Culex tritaeniorhynchus mosquitoes in Japan (Kuwata, Isawa et al. 2011). Unfortunately, despite repeated attempts, we were not able to obtain an isolate in Aedes albopictus (C6/36) or Vero cells. The inability of this virus to replicate in C6/36 cells was unexpected considering the fact that the virus was isolated from mosquitoes. C6/36 cells are also very susceptible to viral infection and the preferred cell line for preparation of reagents for arboviruses (White 1987). It is unlikely that the observed growth restriction occurred due to an interruption in the cold chain while the mosquitoes were being transported as CPE was observed in C6/36 cells inoculated with mosquito homogenates from other pools transported with our Merida virus pool. Though unlikely, it may be possible that positive mosquito pools were not actually infected with Merida virus but were rather passive carriers. Another possibility, though also unlikely, may be that
sequences corresponding to our putative virus may have actually been endogenous viral elements analogous to the rhabdovirus-like sequences described previously in various insect hosts. Although we were not able to perform *in vitro* host range studies on Merida virus, dinucleotide usage preference analyses showed that this virus is not well adapted to and would therefore most likely not propagate in vertebrate hosts.

To avoid the challenges faced with Merida virus isolation, viruses for the remaining metagenomics studies were isolated before being sequenced, thereby allowing us to have isolates for the other three novel and two recognized viruses. *In vitro* host range studies were conducted with these viruses utilizing cell lines of both vertebrate and insect origin. All of the viruses, novel and recognized exhibited an insect-restricted phenotype with vertebrate cell lines all being refractory to infection.

Given the fact that the negevirus taxon is closely related to *Cilevirus* genus of plant viruses, additional studies launching an investigation into plant host permissibility and susceptibility can be explored. The same should be considered for Ek Balam virus, whose closest relatives are members of the *Tymoviridae* family of plant viruses. Although this virus is reported here as a *Tymoviridae*-like virus, it phylogenetically affiliates with a clade in the putative family which also houses two other animal viruses. The first virus and its closest relative, Culex originated *Tymoviridae*-like virus (CTLV), was the first *Tymoviridae*-like virus isolated from mosquitoes (Wang, Lv et al. 2012), Ek Balam virus being the second. The second virus, bat tymo-like virus was isolated from bat faeces. Since the virus was isolated from bat faeces, it is possible that this virus is also an insect-virus ingested by the bat and excreted through its faeces. It is currently uncertain whether this new clade of viruses will form a new genus with a
differential host range in the *Tymoviridae* family or whether they will be reclassified by the ICTV and included in another taxonomic group.

In Chapters 5 and 6, we reported the full sequence data for the genomes of two insect-specific flaviviruses, Marisma mosquito virus (MMV) and Nanay virus (NANV) and for the large segment of two bunyaviruses, Main Drain virus (MDV) and Northway virus (NORV), for which complete sequence data had not previously been available. Data gathered from these sequencing experiments add to the wealth of genomic data currently available and contribute to more accurate and comprehensive comparative genomics studies as we seek to better understand genetic trends associated with specific viral taxa.

In summary, the results of an investigation into the genetic elements which govern flaviviral host specificity showed that host restriction in bat-associated no known vector flaviviruses occurred at a post-entry stage in the viral life cycle. This novel finding will allow future research to focus on the involvement of different genetic elements mediating post-entry steps. I also reported sequence data for several novel and recognized viruses. The availability of such data would allow for the more accurate classification, identification and molecular characterization of these viruses. This is especially important for the identification of segmented viruses, such as bunyaviruses, which have the ability to undergo reassortment. I was also able to assess the *in vitro* host ranges of some of these viruses, giving us an idea of how they are maintained in nature. Our metagenomics studies pave the way for future research aimed at further characterizing viruses which may be of plant, human or veterinary health significance.
References


