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], 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 CB2 1QP, United Kingdom; [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

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

Running Head: Discovery of a novel rhabdovirus in Mexican mosquitoes

Keywords: Rhabdovirus, mosquito, *Culex*, *Ochlerotatus*, Mexico, high throughput sequencing, next generation sequencing

Word count – abstract: 236, text: 4341

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 nucleotides and five major open reading frames that 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 the Genbank database and their identity is only suggested based on their genome position. Using specific RT-PCR assays established from the genome sequence, 27,571 *Cx. 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 (MIR; calculated as the number of positive mosquito pools per 1,000 mosquitoes tested) was 0.9, and similar for both female and male. Screening another 140 pools of 5,484 mosquitoes belonging to four other genera identified positive pools of *Ochlerotatus* spp. mosquitoes, indicating that the host range is not restricted to *Cx. 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.

40 INTRODUCTION

41

42 Rhabdoviruses (family *Rhabdoviridae*, order *Mononegavirales*) are a large and versatile group of viruses
43 that are ubiquitous in nature (Kuzmin et al., 2009). The family consists of 11 genera as well as several
44 viruses that have not yet been assigned to a genus (ICTV, 2015). Virions have a distinctive bullet or cone-
45 shaped morphology or appear bacilliform. Rhabdoviruses have broad host ranges that include humans and
46 other terrestrial mammals, birds, reptiles, fish, insects and plants (Kuzmin et al., 2009, Mann and
47 Dietzgen, 2014, Hoffmann et al., 2005). Many rhabdoviruses are transmitted to vertebrate and plant hosts
48 by insect vectors in which they replicate (Hogenhout et al., 2003, Ammar el et al., 2009).

49

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

function as promotor sequences required for the initiation of genome and anti-genome replication, respectively (Walker et al., 2015, Fu, 2005).

Many novel rhadboviruses have been discovered in recent years due to the advent of unbiased high throughput sequencing (UHTS) (Binger et al., 2015, Ito et al., 2013, Kading et al., 2013, Sakai et al., 2015, Stremlau et al., 2015, Tokarz et al., 2014, Quan et al., 2010), including Bas-Congo virus which was associated with an outbreak of acute hemorrhagic 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 that was discovered by UHTS in mosquitoes in the Yucatan Peninsula of Mexico.

RESULTS

Discovery of Merida virus

UHTS of total RNA from a pool of *Cx. quinquefasciatus* collected in Merida in the Yucatan Peninsula of Mexico generated a ~11 kb sequence corresponding to a putative, novel rhabdovirus tentatively named Merida virus (MERDV). Several other novel virus-like sequences were also discovered and these data will be presented elsewhere. The MERDV genome terminal sequences were identified using a combination of 5' and 3' rapid amplification of cDNA ends (RACE) and 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 24-nt sequence at the distal end of the 5' terminus.

Therefore, our sequence may contain nucleotide errors in the 24 nt primer-binding region at the ultimate 3' end of the genome.

Genomic organization

The MERDV genome consists of 11,798 nucleotides (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., 2011, Kuwata et al., 2011, Gubala et al., 2008, Zhu et al., 2011). The leader and trailer sequences are 59% and 58% A/U-rich, respectively. Each ORF is separated by a noncoding region of 54 to 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 *Cx. 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 the Genbank database. 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 also unlikely 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, a N-terminal signal peptide, a transmembrane domain and a C-terminal short hydrophilic cytoplasmic domain (Walker and Kongsuwan, 1999, Coll, 1995). 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 to 17, a hydrophobic transmembrane domain at residues 469 to 491 and a C-terminal hydrophilic cytoplasmic domain at residue 492 to 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 (Soding 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

131 other rhabdoviruses.

132

133 *Phylogenetic relationship to other rhabdoviruses*

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

143

144 *Prevalence in Cx. quinquefasciatus*

145 A total of 27,571 *Cx. quinquefasciatus*, sorted into 728 pools of up to 50 individuals, were screened by
146 MERDV-specific reverse transcription-polymerase chain reaction (RT-PCR) (Tables 3 and 4). Collections
147 were made in Merida and Tixkokob in 2007-2008 using mosquito magnets and in Merida in 2013 using
148 CDC backpack-mounted aspirators. Mosquitoes obtained in 2007-2008 were tested according to gender
149 whereas those from 2013 were not. Overall, 256 pools (8038 mosquitoes) were composed of females, 195
150 pools (7196 mosquitoes) were composed of males and 277 pools (12,337 mosquitoes) were of mixed
151 gender. Twenty-five mosquito pools were positive for MERDV RNA and the overall minimal infection
152 rate (MIR; calculated as the number of positive mosquito pools per 1,000 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.

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: *Ae. aegypti* (n = 419), *An. albimanus* (n = 727), *An. crucians* (n = 691), *An. vestitipennis* (n = 913), *Oc. taeniorhynchus* (n = 1000), *Oc. trivittatus* (n = 734) and *Ps. 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 *Oc. taeniorhynchus*, and in three pools of *Oc. trivittatus* using both primer pairs while all other species were negative. The MERDV MIRs in *Oc. taeniorhynchus* and *Oc. 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 analyzed by Sanger sequencing. The resulting 114-nt sequences had at least 99.1% nucleotide identity with the corresponding region of the MERDV genome sequence identified in *Cx. quinquefasciatus* (data not shown).

Attempted virus isolations

An aliquot of every homogenate positive for MERDV RNA (25 for *Cx. quinquefasciatus*, 3 for *Oc. taeniorhynchus* and 3 for *Oc. 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 *Cx. 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; for example, vertebrate sequences display a strong under-representation of UpA and CpG, while 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, Tulloch et al., 2014, Atkinson 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 NCBI rhabdovirus RefSeqs and MERDV, CTRV had the least under-representation of CpG (observed:expected ratio close to unity) while 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 (Binger et al., 2015, Grard et al., 2012, Ito et al., 2013, Kading et al., 2013, Sakai et al., 2015, Stremlau et al., 2015, Tokarz et al., 2014), including several which were isolated from *Anopheles*, *Culex*, *Ochlerotatus* and *Psorophora*

196 spp. mosquitoes (Vasilakis et al., 2014, Coffey et al., 2014, Quan et al., 2010). Here, we report the
197 discovery of a putative, novel rhabdovirus in *Culex* and *Ochlerotatus* spp. mosquitoes from the Yucatan
198 Peninsula of Mexico.
199
200 Rhabdoviruses are hypothesized to be perhaps in majority hosted by insects (Li et al., 2015) and
201 numerous cyto-, nucleo- and dimarhabdoviruses are transmitted by arthropods to their plant or vertebrate
202 hosts (Bourhy et al., 2005, Hogenhout et al., 2003). This includes the bite of hematophagous arthropods
203 (Comer et al., 1990, Perez De Leon et al., 2006), including *Simulium vittatum* blackflies which can
204 transmit vesicular stomatitis New Jersey virus to cattle under experimental conditions (Mead et al., 2009)
205 and *Phlebotomus argentipes* sandflies that efficiently transmitted Chandipura virus to laboratory mice
206 (Mavale et al., 2007). It is currently unknown whether MERDV has the capacity to replicate in vertebrate
207 hosts but should its host range include vertebrates, it is unlikely that virus persistence in nature is
208 dependent upon the bite of hematophagous arthropods. Male mosquitoes do not feed on blood and thus a
209 significant bias between male and female mosquito MIRs would be expected if that would be that case.
210 Instead, the similar MIRs that we determined for MERDV are more compatible with vertical and venereal
211 transmission. The occurrence of rhabdovirus transmission through these modes, in addition to horizontal
212 transmission, has been demonstrated for example in *Ae. aegypti* mosquitoes for Chandipura virus (Mavale
213 et al., 2005) and in phlebotomine sandflies for vesicular stomatitis Indiana virus (Tesh et al., 1972). A
214 hallmark in the lifecycle of insect-specific sigmaviruses is the exclusively vertical transmission through
215 eggs and sperm (Longdon et al., 2011). Currently no data are available to decide to what extend horizontal
216 or vertical transmission contribute to the maintenance of MERDV in nature. Attempts to isolate MERDV
217 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 (Vasilakis et al., 2014, Coffey et al., 2014, Quan et al., 2010, Kuwata et al., 2011). However, it is not without precedent; Manitoba rhabdovirus from *Cx. 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 CPE 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 (EVEs) analogous to the rhabdovirus-like sequences described previously in various insect hosts (Li et al., 2015, Katzourakis and Gifford, 2010, Nene et al., 2007). However, such EVEs 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 *Cx. tritaeniorhynchus* in Japan (Kuwata et al., 2011)

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

The translated primary sequences of MERDV N, G and L ORFs show significant homology to respective ORFs of other rhabdovirses. 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 the GenBank database 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 to 83% (Table S1). Since 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 Merida virus, 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 programs to gain better insights into arbovirus evolution.

287 METHODS

288

289 Mosquito collections

290 Mosquitoes were collected in five study areas in the Yucatan Peninsula of Mexico: Cozumel Island,
291 Merida, Sian Ka'an, Tixkokob and Tzucacab. Descriptions of these study areas are provided elsewhere
292 (Farfan-Ale et al., 2009, Farfan-Ale et al., 2010). Collections were made in 2007 and 2008 using
293 Mosquito Magnets^{TM} (all five study areas) and in 2013 using backpack-mounted aspirators (Merida only).
294 Mosquito magnets Pro-Liberty (American Biophysics Corp, North Kingstown, RI, USA) were baited with
295 propane and octenol, and placed outdoors. Mosquito magnets were turned on between 16:00 and 18:00
296 and collection nets were replaced the following morning between 06:00 and 09:00. CDC back-pack
297 mounted aspirators were used to collect resting mosquitoes inside private residences. Mosquitoes were
298 transported alive to the Universidad Autonoma de Yucatan (UADY), frozen at -80^{o}C and identified on
299 chill tables according to species and sex using morphological characteristics (Darsie, 1996). Mosquitoes
300 were transported on dry ice from the UADY to Iowa State University by World Courier.

301

302 High throughput sequencing

303 Mosquitoes were homogenized as previously described (Farfan-Ale et al., 2009) and total RNA was
304 extracted using Trizol (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions.
305 Extracts were reverse transcribed using SuperScript III (Thermo Fisher, Waltham, MA, USA) with
306 random hexamers. The complementary DNA (cDNA) was RNase-H treated prior to second strand
307 synthesis with Klenow Fragment (NEB, Ipswich, MA, USA). The generated double stranded cDNA was
308 sheared to an average fragment size of 200 bp using manufacturer's standard settings (Covaris focused-
309 ultrasonicator E210; Woburn, MA, USA). Sheared products were purified (Agencourt Ampure DNA

purification beads, Beckman Coulter, Brea, CA, USA) 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 quantitated by Bioanalyzer (Agilent, Santa Clara, CA, USA) for sequencing. Sequencing on the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA) resulted in an average 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 the GenBank database.

RT-PCR and Sanger sequencing

Total RNA was analyzed by RT-PCR using MERDV-specific primers RHAB-for (5'-CAATCACATCGACTACTCTAAATGGA-3') and RHAB-rev (5'-GATCAGACCTAGCTTGGCTGTTC-3') which target a 490-nt region of the L protein gene, or RHAB-121-for (5'-AACGCCCGACATGACTACTATCG-3') and RHAB-280-rev (5'-TTCCGTACCTCCCATATGAGTGG-3') which target a 160-nt region of the N protein. Complementary DNAs were generated using Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA) and PCRs were performed using *Taq* polymerase (Invitrogen) and the following cycling conditions: 94^{o}C for 3 min then 35 cycles of 94^{o}C for 30 sec, 56^{o}C for 45 sec and 72^{o}C for 1 min followed by a final extension at 72^{o}C for 8 min. RT-PCR products were purified using the Purelink Gel Extraction Kit (Invitrogen). Sanger sequencing was performed using a 3730x1 DNA sequencer (Applied Biosystems, Foster City, CA, USA).

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’-CTCAGAACGGAAGAGGTATACT G-3’). Complementary DNAs were purified by ethanol precipitation and oligo(dC) tails were added to the 3′ ends using 15 units of terminal deoxynucleotidyl transferase (Invitrogen) in 1x tailing buffer (10 mM Tris-HCl [pH 8.4], 25 mM KCl, 1.5 mM $MgCl_2$ and 0.02 mM dCTP). Tailing reactions were performed at 37^{o}C for 30 min and then terminated by heat-inactivation (65^{o}C for 10 min). Oligo dC-tailed cDNAs were purified by ethanol precipitation then PCR amplified using a consensus forward primer specific to the C-tailed termini (5′-GACATCGAAAGGGGGGGGGGG-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 units of poly(A) polymerase (Ambion, Austin, TX, USA) in 1 x reaction buffer (40 mM Tris-HCl [pH 8.0], 10 mM $MgCl_2$, 2.5 mM $MnCl_2$, 250 mM NaCl, 50 µg of bovine serum albumin/ml and 1 mM ATP). Tailing reactions were performed at 37^{o}C for 1 hr and terminated by heat-inactivation (65^{o}C for 10 min). Poly(A)-tailed RNA was reverse transcribed using a poly(A) tail-specific primer (5′-GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTT-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 reverse transcription primer (5′-GGCCACGCGTCGACTAGTAC-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) and kanamycin

(50 μg/ml), 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, Hilden, Germany) 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 the Genbank database by application of tblastn (Altschul et al., 1990). Percent amino acid identities and similarities of select rhabdovirus protein sequences were calculated using ClustalW2 (available at: 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 v. 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 2 ml of Liebovitz's L15 medium (Invitrogen) supplemented with 2% fetal bovine serum, 2 mM L-glutamine, 100 units/ml penicillin, 100 μg/ml streptomycin and 2.5 μg/ml fungizone. Samples were filtered using a 0.22 μm filter and inoculated onto subconfluent monolayers of *Aedes albopictus* C6/36 cells in 75 cm^2 flasks. Cells were incubated for at least 1 hr at room temperature on an orbital shaker. Another 12 ml of L15 maintenance medium was added to each flask, and cells were incubated at 28^oC for 7 days. After two additional blind passages, supernatants were harvested and tested by RT-PCR for the presence of MERDV RNA.

ACKNOWLEDGMENTS

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

386 ABBREVIATIONS

387

388	3’RACE:	3’ rapid amplification of cDNA ends
389	5’RACE:	5’ rapid amplification of cDNA ends
390	CDC:	Centers for Disease Control and Prevention
391	cDNA:	Complementary DNA
392	CTRV:	Culex tritaeniorhynchus rhabdovirus
393	EVE:	Endogenous viral element
394	G:	Glycoprotein
395	L:	RNA-dependent RNA polymerase
396	M:	Matrix
397	MERDV:	Merida virus
398	MIR:	Minimal infection rate
399	N:	Nucleoprotein
400	ORF:	Open reading frame
401	P:	Phosphoprotein
402	PKC:	Protein kinase C
403	RT-PCR:	Reverse transcription-polymerase chain reaction
404	Tyrosine:	TYR
405	UADY:	Universidad Autonoma de Yucatan
406	UHTS:	Unbiased high throughput sequencing

REFERENCES

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

AMMAR EL, D., TSAI, C. W., WHITFIELD, A. E., REDINBAUGH, M. G. & HOGENHOUT, S. A. 2009. Cellular and molecular aspects of rhabdovirus interactions with insect and plant hosts. *Annu Rev Entomol,* 54**,** 447-68.

ARTSOB, H., DOANE, F., SEKLA, L., STACKIW, W. & BRUST, R. 1991. Manitoba virus, a new rhabdovirus isolated from Culex tarsalis mosquitoes collected in Manitoba, Canada. *Can J Microbiol,* 37**,** 329-32.

ATKINSON, N. J., WITTEVELDT, J., EVANS, D. J. & SIMMONDS, P. 2014. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res,* 42**,** 4527-45.

BINGER, T., ANNAN, A., DREXLER, J. F., MULLER, M. A., KALLIES, R., ADANKWAH, E., WOLLNY, R., KOPP, A., HEIDEMANN, H., DEI, D., AGYA-YAO, F. C., JUNGLEN, S., FELDT, T., KURTH, A., OPPONG, S., ADU-SARKODIE, Y. & DROSTEN, C. 2015. A novel rhabdovirus isolated from the straw-colored fruit bat Eidolon helvum, with signs of antibodies in swine and humans. *J Virol,* 89**,** 4588-97.

BOURHY, H., COWLEY, J. A., LARROUS, F., HOLMES, E. C. & WALKER, P. J. 2005. Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene. *J Gen Virol,* 86**,** 2849-58.

CASTRESANA, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol,* 17**,** 540-52.

COFFEY, L. L., PAGE, B. L., GRENINGER, A. L., HERRING, B. L., RUSSELL, R. C., DOGGETT, S. L., HANIOTIS, J., WANG, C., DENG, X. & DELWART, E. L. 2014. Enhanced arbovirus surveillance with deep sequencing: Identification of novel rhabdoviruses and bunyaviruses in Australian mosquitoes. *Virology,* 448**,** 146-58.

COLL, J. M. 1995. The Glycoprotein-G of Rhabdoviruses. *Archives of Virology,* 140**,** 827-851.

COLLINS, P. L., OLMSTED, R. A., SPRIGGS, M. K., JOHNSON, P. R. & BUCKLER-WHITE, A. J. 1987. Gene overlap and site-specific attenuation of transcription of the viral polymerase L gene of human respiratory syncytial virus. *Proc Natl Acad Sci U S A,* 84**,** 5134-8.

COMER, J. A., TESH, R. B., MODI, G. B., CORN, J. L. & NETTLES, V. F. 1990. Vesicular stomatitis virus, New Jersey serotype: replication in and transmission by Lutzomyia shannoni (Diptera: Psychodidae). *Am J Trop Med Hyg,* 42**,** 483-90.

DARSIE, R. F., JR. 1996. A survey and bibliography of the mosquito fauna of Mexico (Diptera: Culicidae). *Journal of the American Mosquito Control Association,* 12**,** 298-306.

EDGAR, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics,* 5**,** 113.

FARFAN-ALE, J. A., LORONO-PINO, M. A., GARCIA-REJON, J. E., HOVAV, E., POWERS, A. M., LIN, M., DORMAN, K. S., PLATT, K. B., BARTHOLOMAY, L. C., SOTO, V., BEATY, B. J., LANCIOTTI, R. S. & BLITVICH, B. J. 2009. Detection of RNA from a novel West Nile-like virus and high prevalence of an insect-specific flavivirus in mosquitoes in the Yucatan Peninsula of Mexico. *Am J Trop Med Hyg,* 80**,** 85-95.

FARFAN-ALE, J. A., LORONO-PINO, M. A., GARCIA-REJON, J. E., SOTO, V., LIN, M., STALEY, M., DORMAN, K. S., BARTHOLOMAY, L. C., HOVAV, E. & BLITVICH, B. J. 2010. Detection of flaviviruses and orthobunyaviruses in mosquitoes in the Yucatan Peninsula of Mexico in 2008. *Vector Borne Zoonotic Dis,* 10**,** 777-83.

FU, Z. F. 2005. Genetic comparison of the rhabdoviruses from animals and plants. *Curr Top Microbiol Immunol,* 292**,** 1-24.

GILLICH, N., KUWATA, R., ISAWA, H. & HORIE, M. 2015. Persistent natural infection of a Culex tritaeniorhynchus cell line with a novel Culex tritaeniorhynchus rhabdovirus strain. *Microbiol Immunol*.

GOLOVKO, G., KHANIPOV, K., ROJAS, M., MARTINEZ-ALCANTARA, A., HOWARD, J. J., BALLESTEROS, E., GUPTA, S., WIDGER, W. & FOFANOV, Y. 2012. Slim-Filter: an interactive windows-based application for illumina genome analyzer data assessment and manipulation. *Bmc Bioinformatics,* 13.

GRARD, G., FAIR, J. N., LEE, D., SLIKAS, E., STEFFEN, I., MUYEMBE, J. J., SITTLER, T., VEERARAGHAVAN, N., RUBY, J. G., WANG, C., MAKUWA, M., MULEMBAKANI, P., TESH, R. B., MAZET, J., RIMOIN, A. W., TAYLOR, T., SCHNEIDER, B. S., SIMMONS, G., DELWART, E., WOLFE, N. D., CHIU, C. Y. & LEROY, E. M. 2012. A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. *PLoS Pathog,* 8**,** e1002924.

GREENBAUM, B. D., LEVINE, A. J., BHANOT, G. & RABADAN, R. 2008. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. *PLoS Pathog,* 4**,** e1000079.

GUBALA, A., DAVIS, S., WEIR, R., MELVILLE, L., COWLED, C. & BOYLE, D. 2011. Tibrogargan and Coastal Plains rhabdoviruses: genomic characterization, evolution of novel genes and seroprevalence in Australian livestock. *J Gen Virol,* 92**,** 2160-70.

GUBALA, A. J., PROLL, D. F., BARNARD, R. T., COWLED, C. J., CRAMERI, S. G., HYATT, A. D. & BOYLE, D. B. 2008. Genomic characterisation of Wongabel virus reveals novel genes within the Rhabdoviridae. *Virology,* 376**,** 13-23.

HOFFMANN, B., BEER, M., SCHUTZE, H. & METTENLEITER, T. C. 2005. Fish rhabdoviruses: molecular epidemiology and evolution. *Curr Top Microbiol Immunol,* 292**,** 81-117.

HOGENHOUT, S. A., REDINBAUGH, M. G. & AMMAR EL, D. 2003. Plant and animal rhabdovirus host range: a bug's view. *Trends Microbiol,* 11**,** 264-71.

ICTV 2015. International Committee on Taxonomy of Viruses. Virus taxonomy: 2013 release. http://www.ictvonline.org/virusTaxonomy.asp. .

ITO, T., SUZAKI, K. & NAKANO, M. 2013. Genetic characterization of novel putative rhabdovirus and dsRNA virus from Japanese persimmon. *J Gen Virol,* 94**,** 1917-21.

KADING, R. C., GILBERT, A. T., MOSSEL, E. C., CRABTREE, M. B., KUZMIN, I. V., NIEZGODA, M., AGWANDA, B., MARKOTTER, W., WEIL, M. R., MONTGOMERY, J. M., RUPPRECHT, C. E. & MILLER, B. R. 2013. Isolation and molecular characterization of Fikirini rhabdovirus, a novel virus from a Kenyan bat. *J Gen Virol,* 94**,** 2393-8.

KAPOOR, A., SIMMONDS, P., LIPKIN, W. I., ZAIDI, S. & DELWART, E. 2010. Use of nucleotide composition analysis to infer hosts for three novel picorna-like viruses. *J Virol,* 84**,** 10322-8.

KATZOURAKIS, A. & GIFFORD, R. J. 2010. Endogenous viral elements in animal genomes. *PLoS Genet,* 6**,** e1001191.

KUWATA, R., ISAWA, H., HOSHINO, K., TSUDA, Y., YANASE, T., SASAKI, T., KOBAYASHI, M. & SAWABE, K. 2011. RNA splicing in a new rhabdovirus from Culex mosquitoes. *J Virol,* 85**,** 6185-96.

KUZMIN, I. V., NOVELLA, I. S., DIETZGEN, R. G., PADHI, A. & RUPPRECHT, C. E. 2009. The rhabdoviruses: biodiversity, phylogenetics, and evolution. *Infect Genet Evol,* 9**,** 541-53.

LI, C. X., SHI, M., TIAN, J. H., LIN, X. D., KANG, Y. J., CHEN, L. J., QIN, X. C., XU, J., HOLMES, E. C. & ZHANG, Y. Z. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife,* 4.

LONGDON, B., WILFERT, L., OBBARD, D. J. & JIGGINS, F. M. 2011. Rhabdoviruses in two species of Drosophila: vertical transmission and a recent sweep. *Genetics,* 188**,** 141-50.

MANN, K. S. & DIETZGEN, R. G. 2014. Plant rhabdoviruses: new insights and research needs in the interplay of negative-strand RNA viruses with plant and insect hosts. *Arch Virol,* 159**,** 1889-900.

MAVALE, M. S., FULMALI, P. V., GHODKE, Y. S., MISHRA, A. C., KANOJIA, P. & GEEVARGHESE, G. 2007. Experimental transmission of Chandipura virus by Phlebotomus argentipes (diptera: psychodidae). *Am J Trop Med Hyg,* 76**,** 307-9.

MAVALE, M. S., GEEVARGHESE, G., GHODKE, Y. S., FULMALI, P. V., SINGH, A. & MISHRA, A. C. 2005. Vertical and venereal transmission of Chandipura virus (Rhabdoviridae) by Aedes aegypti (Diptera: Culicidae). *J Med Entomol,* 42**,** 909-11.

MEAD, D. G., LOVETT, K. R., MURPHY, M. D., PAUSZEK, S. J., SMOLIGA, G., GRAY, E. W., NOBLET, R., OVERMYER, J. & RODRIGUEZ, L. L. 2009. Experimental transmission of vesicular stomatitis New Jersey virus from Simulium vittatum to cattle: clinical outcome is influenced by site of insect feeding. *J Med Entomol,* 46**,** 866-72.

NENE, V., WORTMAN, J. R., LAWSON, D., HAAS, B., KODIRA, C., TU, Z. J., LOFTUS, B., XI, Z., MEGY, K., GRABHERR, M., REN, Q., ZDOBNOV, E. M., LOBO, N. F., CAMPBELL, K. S., BROWN, S. E., BONALDO, M. F., ZHU, J., SINKINS, S. P., HOGENKAMP, D. G., AMEDEO, P., ARENSBURGER, P., ATKINSON, P. W., BIDWELL, S., BIEDLER, J., BIRNEY, E., BRUGGNER, R. V., COSTAS, J., COY, M. R., CRABTREE, J., CRAWFORD, M., DEBRUYN, B., DECAPRIO, D., EIGLMEIER, K., EISENSTADT, E., EL-DORRY, H., GELBART, W. M., GOMES, S. L., HAMMOND, M., HANNICK, L. I., HOGAN, J. R., HOLMES, M. H., JAFFE, D., JOHNSTON, J. S., KENNEDY, R. C., KOO, H., KRAVITZ, S., KRIVENTSEVA, E. V., KULP, D., LABUTTI, K., LEE, E., LI, S., LOVIN, D. D., MAO, C., MAUCELI, E., MENCK, C. F., MILLER, J. R., MONTGOMERY, P., MORI, A., NASCIMENTO, A. L., NAVEIRA, H. F., NUSBAUM, C., O'LEARY, S., ORVIS, J., PERTEA, M., QUESNEVILLE, H., REIDENBACH, K. R., ROGERS, Y. H., ROTH, C. W., SCHNEIDER, J. R., SCHATZ, M., SHUMWAY, M., STANKE, M., STINSON, E. O., TUBIO, J. M., VANZEE, J. P., VERJOVSKI-ALMEIDA, S., WERNER, D., WHITE, O., WYDER, S., ZENG, Q., ZHAO, Q., ZHAO, Y., HILL, C. A., RAIKHEL, A. S., SOARES, M. B., KNUDSON, D. L., LEE, N. H., GALAGAN, J., SALZBERG, S. L., PAULSEN, I. T., DIMOPOULOS, G., COLLINS, F. H., BIRREN, B., FRASER-LIGGETT, C. M. & SEVERSON, D. W. 2007. Genome sequence of Aedes aegypti, a major arbovirus vector. *Science,* 316**,** 1718-23.

PEREZ DE LEON, A. A., O'TOOLE, D. & TABACHNICK, W. J. 2006. Infection of guinea pigs with vesicular stomatitis New Jersey virus Transmitted by Culicoides sonorensis (Diptera: Ceratopogonidae). *J Med Entomol,* 43**,** 568-73.

QUAN, P. L., JUNGLEN, S., TASHMUKHAMEDOVA, A., CONLAN, S., HUTCHISON, S. K., KURTH, A., ELLERBROK, H., EGHOLM, M., BRIESE, T., LEENDERTZ, F. H. & LIPKIN, W. I. 2010. Moussa virus: a new member of the Rhabdoviridae family isolated from Culex decens mosquitoes in Cote d'Ivoire. *Virus Res,* 147**,** 17-24.

QUAN, P. L., WILLIAMS, D. T., JOHANSEN, C. A., JAIN, K., PETROSOV, A., DIVINEY, S. M., TASHMUKHAMEDOVA, A., HUTCHISON, S. K., TESH, R. B., MACKENZIE, J. S., BRIESE, T. & LIPKIN, W. I. 2011. Genetic characterization of K13965, a strain of Oak Vale virus from Western Australia. *Virus Research,* 160**,** 206-213.

RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HOHNA, S., LARGET, B., LIU, L., SUCHARD, M. A. & HUELSENBECK, J. P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol,* 61**,** 539-42.

SAKAI, K., HAGIWARA, K., OMATSU, T., HAMASAKI, C., KUWATA, R., SHIMODA, H., SUZUKI, K., ENDOH, D., NAGATA, N., NAGAI, M., KATAYAMA, Y., OBA, M., KURANE, I., SAIJO, M., MORIKAWA, S., MIZUTANI, T. & MAEDA, K. 2015. Isolation and characterization of a novel Rhabdovirus from a wild boar (Sus scrofa) in Japan. *Vet Microbiol*.

SCHNEEMANN, A., SCHNEIDER, P. A., KIM, S. & LIPKIN, W. I. 1994. Identification of signal sequences that control transcription of borna disease virus, a nonsegmented, negative-strand RNA virus. *Journal of virology,* 68**,** 6514-22.

SCHNEIDER, P. A., SCHNEEMANN, A. & LIPKIN, W. I. 1994. RNA splicing in Borna disease virus, a nonsegmented, negative-strand RNA virus. *Journal of virology,* 68**,** 5007-12.

SHI, C., LIU, Y., HU, X., XIONG, J., ZHANG, B. & YUAN, Z. 2015. A metagenomic survey of viral abundance and diversity in mosquitoes from hubei province. *PLoS One,* 10**,** e0129845.

SIMMEN, M. W. 2008. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. *Genomics,* 92**,** 33-40.

SODING, J., BIEGERT, A. & LUPAS, A. N. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research,* 33**,** W244-W248.

STREMLAU, M. H., ANDERSEN, K. G., FOLARIN, O. A., GROVE, J. N., ODIA, I., EHIANE, P. E., OMONIWA, O., OMOREGIE, O., JIANG, P. P., YOZWIAK, N. L., MATRANGA, C. B., YANG, X., GIRE, S. K., WINNICKI, S., TARIYAL, R., SCHAFFNER, S. F., OKOKHERE, P. O., OKOGBENIN, S., AKPEDE, G. O., ASOGUN, D. A., AGBONLAHOR, D. E., WALKER, P. J., TESH, R. B., LEVIN, J. Z., GARRY, R. F., SABETI, P. C. & HAPPI, C. T. 2015. Discovery of novel rhabdoviruses in the blood of healthy individuals from West Africa. *PLoS Negl Trop Dis,* 9**,** e0003631.

TESH, R. B., CHANIOTIS, B. N. & JOHNSON, K. M. 1972. Vesicular stomatitis virus (Indiana serotype): transovarial transmission by phlebotomine sandlies. *Science,* 175**,** 1477-9.

TOKARZ, R., SAMEROFF, S., LEON, M. S., JAIN, K. & LIPKIN, W. I. 2014. Genome characterization of Long Island tick rhabdovirus, a new virus identified in Amblyomma americanum ticks. *Virol J,* 11**,** 26.

TULLOCH, F., ATKINSON, N. J., EVANS, D. J., RYAN, M. D. & SIMMONDS, P. 2014. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *Elife,* 3**,** e04531.

VASILAKIS, N., CASTRO-LLANOS, F., WIDEN, S. G., AGUILAR, P. V., GUZMAN, H., GUEVARA, C., FERNANDEZ, R., AUGUSTE, A. J., WOOD, T. G., POPOV, V., MUNDAL, K., GHEDIN, E., KOCHEL, T. J., HOLMES, E. C., WALKER, P. J. & TESH, R. B. 2014. Arboretum and Puerto Almendras viruses: two novel rhabdoviruses isolated from mosquitoes in Peru. *J Gen Virol,* 95**,** 787-92.

VASILAKIS, N., WIDEN, S., DA ROSA, A. P. A. T., WOOD, T. G., WALKER, P. J., HOLMES, E. C. & TESH, R. B. 2013. Malpais spring virus is a new species in the genus vesiculovirus. *Virology Journal,* 10.

WALKER, P. J., DIETZGEN, R. G., JOUBERT, D. A. & BLASDELL, K. R. 2011. Rhabdovirus accessory genes. *Virus Res,* 162**,** 110-25.

WALKER, P. J., FIRTH, C., WIDEN, S. G., BLASDELL, K. R., GUZMAN, H., WOOD, T. G., PARADKAR, P. N., HOLMES, E. C., TESH, R. B. & VASILAKIS, N. 2015. Evolution of genome size and complexity in the rhabdoviridae. *PLoS Pathog,* 11**,** e1004664.

WALKER, P. J. & KONGSUWAN, K. 1999. Deduced structural model for animal rhabdovirus glycoproteins. *Journal of General Virology,* 80**,** 1211-1220.

ZHU, R. L., LEI, X. Y., KE, F., YUAN, X. P. & ZHANG, Q. Y. 2011. Genome of turbot rhabdovirus exhibits unusual non-coding regions and an additional ORF that could be expressed in fish cell. *Virus Res,* 155**,** 495-505.

FIGURE LEGENDS

FIGURE 1. Coding capacity of Merida virus sequence. Main open reading frames (ORFs), as well as minor ORFs unlikely to be expressed, are indicated along a scaled representation of the antigenomic strand. The size of minor ORFs is indicated by their codon number. Blue, frame 1; green, frame 2; red, frame 3.

FIGURE 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 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.

FIGURE 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, fXpY, were expressed relative to their expected frequencies, fX x fY, 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

610 acid sequence and the original total numbers of each of the 61 codons were maintained, but
611 synonymous codons were randomly shuffled between the different sites where the corresponding
612 amino acid is used in the original sequence. Next, the UpA and CpG frequencies in the original
613 sequence were expressed relative to their mean frequencies in the codon-shuffled sequences.
614 Because codon usage is factored out, the UpA and CpG relative frequencies tend to be less
615 extreme in (B) compared to (A). Each point represents a single rhabdovirus sequence. Solid
616 points correspond to species within defined genera, color-coded by genus (see key). Annotated
617 open circles correspond to species that are currently unassigned at genus level, color-coded by
618 host (or presumed host) taxa. Asterisks in the key indicate clades with uncertain host taxa:
619 viruses in the unclassified "arthropod-infecting" clades (yellow open circles) have been isolated
620 from arthropods but not from vertebrates; the sole representative of genus *Tupavirus* has been
621 isolated from mammals but not from arthropods, although its phylogenetic position suggests that
622 it may be arthropod-borne; the presence of viruses derived from vertebrates and viruses derived
623 from arthropods in each of the unclassified "vertebrate (arthropod-borne)" clades (brick-red open
624 circles) suggests that all of these viruses are likely arboviruses. GenBank accession numbers of
625 sequences used: NC_000855, NC_000903, NC_001542, NC_001560, NC_001615, NC_001652,
626 NC_002251, NC_002526, NC_002803, NC_003243, NC_003746, NC_005093, NC_005974,
627 NC_005975, NC_006429, NC_006942, NC_007020, NC_007642, NC_008514, NC_009527,
628 NC_009528, NC_011532, NC_011542, NC_011639, NC_013135, NC_013955, NC_016136,
629 NC_017685, NC_017714, NC_018381, NC_018629, NC_020803, NC_020804, NC_020805,
630 NC_020806, NC_020807, NC_020808, NC_020809, NC_020810, NC_022580, NC_022581,
631 NC_022755, NC_024473, NC_025251, NC_025253, NC_025255, NC_025340, NC_025341,
632 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 1. Noncoding sequences in the Merida virus genome

Gene	Noncoding sequences at the start of the gene	Noncoding sequences at the end of the gene	*IGR
N	*ACGAAAACAAAAAATCCCCACTCAACAGTCAGAATCCATGTTGTCGTTG GAGGTCTATGGAAATCCTT***AACAU**AACUAGUAUUAAUUAACUCUAA UAUUUGACACUUUUGGAUUUUCUGGAACGCCCGAC	GCCCUCACCUCGAGGACCCUCGGAUGCCCAGCAGGUUA**CAUGAA AAAAA**	ACUCC
P	**AACAU**AACUAACUCGACCUCGGAAUCCGAUCAUUCACA	**CAUGAAAAAAA**	CUCC
M	**AACAU**CACUCACCUGAGACUCAUUCCCAGGUUAUUCUUGCCAAC	AUCCCCCUUUAUAGACUUGGACCUUGUUAUUCCACAAUAAAGA CAUAACAUAACUAGAAA**CUUGAAAAAAA**	†N/A
G	**AACAU**AACUAGAAACUUGAAAAAAAGUUCCUCUGUGAAUUCCAG GUAGACGGGCCGAAAAAG	AGACCGUUAUGUAUCCCGCCUGUUCUUGGUUUGUCUGUGACCU GAAUUCAAUUUUGCCGUACUAUUGGAUAAUUCUUUUCCUCCUC UGAUUAUAUGUCUGUAAACUUUUAA**CAUGAAAAAAA**	†N/A
L	**AACAU**GAAAAAAAUCAACAAAACUCAACGGGUAUCAUAUCAAAA	GACCAGAUCAAAGAGGGAAAAGAGACAGAAGAGAGAA**CAUGAA AAAAA***CATGATCCCGATTCCTATAGTCTGATAAGGACCTCTAGGAGTA TGCTTGTTGAGTGGGGATTTTTTGTTTTCGT*	

*Intergenic region located immediately downstream of the gene of interest that is neither translated nor transcribed to mRNA; †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

Protein	ORF Genomic location	ORF length (nt)	Protein length (aa)	Protein mass (kDa)
N	131-1567	1437	478	54.2
P	1660-2862	1203	400	43.8
M	2917-3477	564	187	21.0
G	3586-5112	1530	509	56.8
L	5266-11676	6411	2136	241.2

Table 3. Minimal infection rates for Merida virus in *Culex quinquefasciatus* in the Yucatan Peninsula of Mexico, 2007-2008 and 2013

Study site	Date	No. mosquitoes tested	No. pools tested (positive)	[‡]MIR
[*]Tixkokob	2007-2008	9071	247 (2)	0.2
[*]Merida	2007-2008	6163	204 (14)	2.3
[†]Merida	2013	12,337	277 (9)	0.7
Total	-	27,571	728 (25)	0.9

[*]Mosquitoes were collected outdoors using mosquito magnets; [†]Mosquitoes were collected inside private residences using CDC backpack mounted aspirators; [‡]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

Study site	[*]No. mosquitoes tested			No. pools tested (positive)			[†]MIR		
	Females	Males	Total	Females	Males	Total	Females	Males	Total
[c]Merida	3018	3145	6163	112 (7)	92 (7)	204 (14)	2.3	2.2	2.3
[‡]Tixkokob	5020	4051	9071	144 (2)	103 (0)	247 (2)	0.4	0.0	0.2
Total	8038	7196	15,234	256 (9)	195 (7)	451 (16)	1.1	1.0	1.1

[*]*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

Species	[*]No. mosquitoes tested	No. pools tested (positive)	[†]MIR	[‡]Study site where positive pools were collected
Ae. aegypti	419	20 (0)	0	
An. albimanus	727	20 (0)	0	
An. crucians	691	20 (0)	0	
An. vestitipennis	913	20 (0)	0	
Oc. taeniorhynchus	1000	20 (3)	3.0	Cozumel Island
Oc. trivittatus	734	20 (3)	4.1	Cozumel Island, Merida and Tzucacab
Ps. cyanescens	1000	20 (3)	0	
Total	5484	140 (6)	-	

[*]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