Am. J. Trop. Med. Hyg., 96(1), 2017, pp. 100–109
doi:10.4269/ajtmh.16-0403

Almendravirus: A Proposed New Genus of Rhabdoviruses Isolated from Mosquitoes in Tropical Regions of the Americas

Maria Angelica Contreras,[1,2]† Gillian Eastwood,[3]† Hilda Guzman,[4] Vsevolod Popov,[4] Chelsea Savit,[5] Sandra Uribe,[2] Laura D. Kramer,[3] Thomas G. Wood,[6] Steven G. Widen,[6] Durland Fish,[7] Robert B. Tesh,[4] Nikos Vasilakis,[4]* and Peter J. Walker[8,9]*

[1]*Programa de Estudio y Control de Enfermedades Tropicales (PECET), Sede de Investigacion Universitaria (SIU), Universidad de Antioquia, Medellin, Colombia;* [2]*Grupo de Investigacion en Sistematica Molecular (GSM), Facultad de Ciencias, Universidad Nacional de Colombia, Medellin, Colombia;* [3]*Griffin Arbovirus Laboratory, Wadsworth Center, New York State Department of Health, Albany, New York;* [4]*Department of Pathology, Center for Biodefense and Emerging Infectious Diseases, Center for Tropical Diseases, Institute for Human Infections and Immunity, University of Texas Medical Branch, Galveston, Texas;* [5]*Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut;* [6]*Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, Texas;* [7]*Yale School of Public Health, New Haven, Connecticut;* [8]*CSIRO Health and Biosecurity, Australian Animal Health Laboratory, Geelong, Australia;* [9]*School of Biological Sciences, University of Queensland, St Lucia, Australia*

Abstract. The *Rhabdoviridae* is a diverse family of negative-sense single-stranded RNA viruses, many of which infect vertebrate hosts and are transmitted by hematophagous arthropods. Others appear to be arthropod specific, circulating only within arthropod populations. Herein, we report the isolation and characterization of three novel viruses from mosquitoes collected from the Americas. Coot Bay virus was isolated from *Anopheles quadrimaculatus* mosquitoes collected in the Everglades National Park, Florida; Rio Chico virus was isolated from *Anopheles triannulatus* mosquitoes collected in Panama; and Balsa virus was isolated from two pools of *Culex erraticus* mosquitoes collected in Colombia. Sequence analysis indicated that the viruses share a similar genome organization to Arboretum virus and Puerto Almendras virus that had previously been isolated from mosquitoes collected in Peru. Each genome features the five canonical rhabdovirus structural protein genes as well as a gene encoding a class 1A viroporin-like protein (U1) located between the G and L genes (3′-N-P-M-G-U1-L-5′). Phylogenetic analysis of complete L protein sequences indicated that all five viruses cluster in a unique clade that is relatively deeply rooted in the ancestry of animal rhabdoviruses. The failure of all viruses in this clade to grow in newborn mice or vertebrate cells in culture suggests that they may be poorly adapted to replication in vertebrates.

INTRODUCTION

The *Rhabdoviridae* is a large and diverse family of negative-sense (–) single-stranded RNA (ssRNA) viruses infecting both animals and plants.[1,2] Many rhabdoviruses replicate in arthropods, some of which (e.g., mosquitoes, sand flies, biting midges, and ticks) may serve as vectors of transmission of infection among humans and other vertebrate hosts.[3] Other rhabdoviruses, such as the drosophila sigmaviruses, are known to be transmitted entirely within arthropod populations.[4] Indeed, a recent metagenomics study of a wide range of terrestrial and marine arthropods has revealed that the genomes of rhabdoviruses and other (–) ssRNA viruses are abundant and diverse, suggesting that arthropods are reservoirs of genetic diversity for these viruses and have played a key role in their long history of evolution.[5] These data also suggest that adaptation of some rhabdoviruses to transmission to vertebrates by hematophagous vectors may be a relatively recent and specialized aspect of their evolution. Distinguishing arthropod-specific viruses from those that may present a serious risk of human infection will become increasingly important as the pace of novel virus discovery escalates.

Herein, we describe the isolation of several novel rhabdoviruses from mosquitoes collected in Colombia, Panama, and the Florida Everglades. We show that these viruses are closely related phylogenetically, and in genome organization, to Arboretum virus (ABTV) and Puerto Almendras virus (PTAMV) that were isolated recently from mosquitoes in Peru.[6] Like ABTV and PTAMV, the novel rhabdoviruses replicated effectively in mosquito cells but failed to show evidence of replication in vertebrate cell cultures or newborn mice, suggesting they may not be well adapted to replication in vertebrates. Our data support the formation of a new genus *Almendravirus* within the family *Rhabdoviridae* to accommodate these novel viruses.

MATERIALS AND METHODS

Description of viruses. Coot Bay virus (CBV; strain EVG 5-53) was isolated from a pool of *Anopheles quadrimaculatus* mosquitoes collected on July 25, 2013, from a mangrove swamp near Coot Bay (25°11′N, 80°55′W) in the Everglades National Park, FL. Rio Chico virus (RCHV; strain GAM-195) was isolated from a pool of unidentified mosquitoes collected in December 2012 in largely old-growth forest near Gamboa (9°07′N, 79°42′W), Central Panama. Balsa virus (BALV; strains CoB 76 and CoB 84) was isolated from two pools of *Culex erraticus* mosquitoes collected between June 9 and June 15, 2013, in an open secondary forest near the beach (La Balsa) in San Bernardo (9°21′N, 75°57′W), Cordoba Department, Colombia. Details of the sources of isolation of ABTV and PTAMV have been described.[6] ABTV (strain LO-121) was isolated on February 22, 2009, from a pool of *Aedes* (*Ochlerotattus*) *fulvus* mosquitoes and PTAMV (strain LO-39) was isolated on March 25, 2009, from *Psorophora albigenu* mosquitoes, each at Puerto Almendras (3°50′S, 73°22′W), Loreto Department, Peru.

Processing of field samples and virus isolation. Pools of field-collected female mosquitoes were homogenized in 1.0 mL of phosphate-buffered saline, pH 7.4, with 20% fetal

*Address correspondence to Peter J. Walker, School of Biological Sciences, University of Queensland, St Lucia, 4072 Australia, E-mail: peter.walker@uq.edu.au or Nikos Vasilakis, University of Texas Medical Branch, Galveston, TX, E-mail: nivasila@utmb.edu.
†These authors contributed equally to this work.

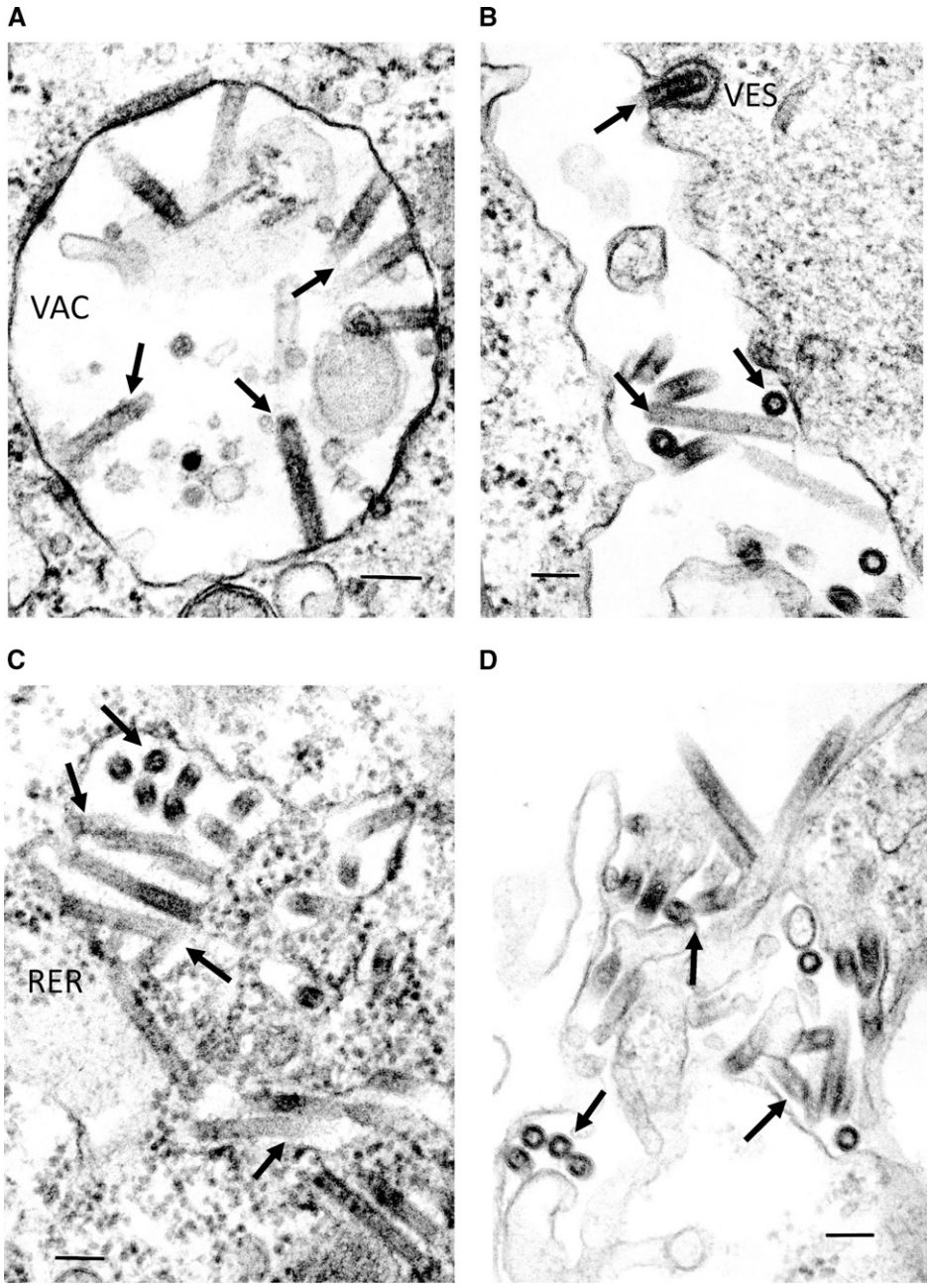

FIGURE 1. Ultrastructure of Balsa virus (BALV) and Coot Bay virus (CBV) virions in C6/36 mosquito cells. (**A**) BALV strain CoB 76 virions forming into a spacious intracellular vacuole (VAC). (**B**) BALV strain CoB 76 virions forming at the cell surface; cross-sections of the virions. One virion is seen in the process of internalization into a clathrin-coated vesicle (VES). (**C**) CBV strain EVG 5-53 virions forming into expanded cisterns of granular endoplasmic reticulum (RER). (**D**) BALV strain CoB 84 virions forming at the cell surface; cross-sections of the virions. Virions are indicated with arrows. Bars = 100 nm.

bovine serum, using a TissueLyser (Qiagen, Valencia, CA) and 3-mm stainless steel beads. After centrifugation at 10,000 rpm in a microcentrifuge for 5 minutes, 100 µL of the supernatant was inoculated into single 12.5-cm^2 flasks with monolayer cultures of C6/36 cells. The cultures were maintained at 28°C for 6–7 days and were examined every 2 days for evidence of viral cytopathic effect (CPE). Cultures showing CPE were harvested for transmission electron microscopy and next-generation sequencing as described below.

Evaluation of growth of viruses in vertebrate cell cultures. Vertebrate cell lines were originally obtained from the American Type Culture Collection (ATCC). Baby hamster kidney cells (BHK-21, ATCC CCL-10), African green monkey kidney cells (Vero E6, ATCC CRL-1586), and duck embryo fibroblasts (ATCC CCL-141) were cultured at 37°C, and toad (*Xenopus laevis*) epithelial cells (XLK-WG, ATCC CRL-2527) were cultured at 28°C. Cells were grown in 25-cm^2 flasks in 5 mL of culture medium as recommended in the ATCC specification sheets. Confluent monolayers were inoculated with 200 µL of virus supernatant obtained following culture in C6/36 mosquito cells and adsorbed for 2 hours. Each flask was then rinsed three times with 5 mL

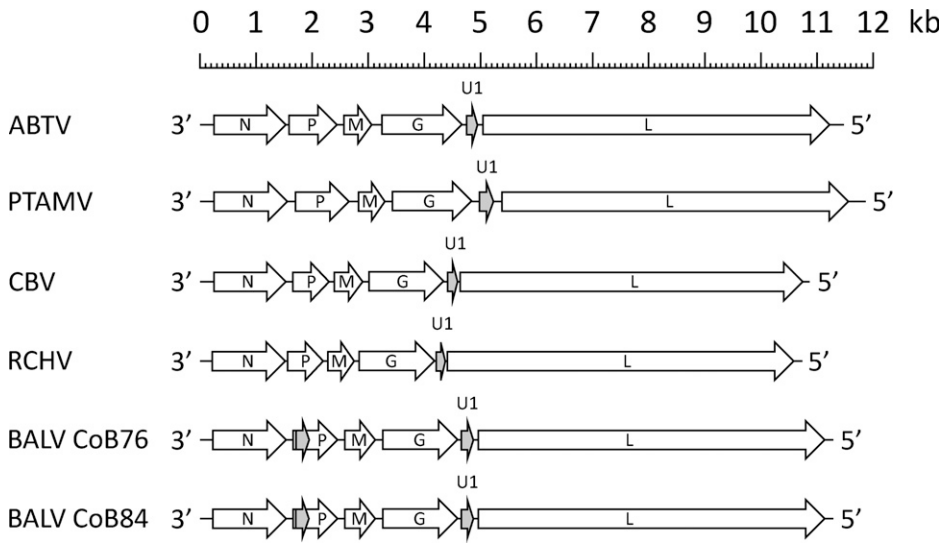

FIGURE 2. Comparisons of the genome organizations of Arboretum virus (ABTV), Puerto Almendras virus (PTAMV), Coot Bay virus (CBV), Rio Chico virus (RCHV), and Balsa virus (BALV) strains CoB 76 and CoB 84. Schematic representation of the genome organizations with the locations of each open reading frame shown approximately to scale.

of maintenance medium with aspiration to remove all remaining medium between washes. The medium was then replaced and the cultures were then examined daily for CPE for 6–7 days. Assays for BALV (strain CoB 76) and CBV (strain EVG 5-53) RNA were conducted on cell culture supernatant fluids that were collected daily and stored at –80°C. After each daily collection, all remaining cell culture fluid was removed by aspiration and fresh medium was added. The samples were assayed by polymerase chain reaction (PCR) using primers targeting the respective L genes (primer sequences and PCR conditions are available upon request).

Evaluation of growth in suckling mice. One litter of 2-day old ICR mouse pups (N = 10) was inoculated intracranially with approximately 15 µL of culture fluid from C6/36 cultures infected with BALV (strain CoB 76) or CBV (strain EVG 5-53). After inoculation, the pups were returned to their dams and were examined daily for 14 days for signs of illness or death. The mice were purchased from Harlan Sprague-Dawley (Indianapolis, IN); animal work at University of Texas Medical Branch was carried out under an Institutional Animal Care and Use Committee-approved protocol (no. 9505045).

Transmission electron microscopy. For ultrastructural analysis, infected C6/36 cells were fixed for at least 1 hour in a mixture of 2.5% formaldehyde prepared from paraformaldehyde powder and 0.1% glutaraldehyde in 0.05 M cacodylate buffer (pH 7.3), to which 0.03% picric acid and 0.03% $CaCl_2$ were added. The monolayers were washed in 0.1 M cacodylate buffer, and cells were scraped off and processed further as a pellet. The pellets were postfixed in 1% OsO_4 in 0.1 M cacodylate buffer (pH 7.3) for 1 hour, washed with distilled water, and stained en block with 2% aqueous uranyl acetate for 20 minutes at 60°C. The pellets were dehydrated in ethanol, processed through propylene oxide, and embedded in Poly/Bed 812 (Polysciences, Warrington, PA). Ultrathin sections were cut on a Leica EM UC7 µL tramicrotome (Leica Microsystems, Buffalo Grove, IL), stained with lead citrate, and examined in a Phillips 201 transmission electron microscope (FEI Phillips, Hillsboro, OR) at 60 kV.

Extraction of viral RNA. C6/36 cells grown to 90% confluence in 25-cm^2 culture flasks were infected with respective viruses. The virus harvest and isolation of viral RNA for next-generation genome sequencing was processed as described previously.[7]

Next-generation sequencing. Viral RNA (~0.9 µg) was fragmented by incubation at 94°C for 8 minutes in 19.5 µL of fragmentation buffer (Illumina 15016648, Illumina, San Diego, CA). Sequencing libraries were prepared from the sample RNAs using an Illumina TruSeq RNA v2 kit following the manufacturer's protocol. The samples were sequenced on a HiSeq 1500 using the Rapid-Run 2 × 50 paired-end protocol. Reads in FASTQ format were quality filtered, and any adapter sequences were removed, using Trimmomatic software.[8] The de novo assembly program ABySS[9] (Michael Smith Genome Sciences Centre, Vancouver, Canada) was used to assemble the reads into contigs, using several different sets of reads, and k values from 20 to 40. Contigs covering nearly the full length of the viruses in each sample were obtained from 150,000 reads (BALV strains CoB 76 and CoB 84) and 1,000,000 reads for CBV. Reads were mapped back to the contigs using bowtie2,[10] and visualized with the Integrated Genomics Viewer[11] to verify that the assembled contigs were correct. There were 11.3, 13.6, and 26 million read-pairs in samples of BALV CoB 76, BALV CoB 84, and CBV, respectively. Read-pairs mapping to the virus in each sample were ~288,000 (2.1%), 403,000 (3.6%), and 468,000 (1.8%), respectively.

Bioinformatic analysis. Translation of genome sequences and structural analysis of proteins was conducted using bioinformatics resources accessed through the ExPASy portal (http://www.expasy.org/proteomics), including Translate (nucleotide sequence translation), Compute PI/MW (pI and molecular weight prediction), TMHMM and TMPred (transmembrane protein prediction), SignalP (signal peptidase cleavage sites), and NetNGlyc (N-linked glycosylation sites).

Phylogenetic analysis. Sequence alignments of complete L protein sequences were created using MUSCLE (MEGA

TABLE 1
Putative transcriptional regulatory sequences of almendraviruses

Gene	Location	ABTV	PTAMV	CBV	RCHV	BALV
N	TI*	AACAATCCTAA	AACAACCCTAA	AACAAACAAAA	AACAAACCTAA	AACAAACCTAA
				AACAAGCAAAG		
				AACAAACTAAA		
	TTP†	TATTAAAAAAACTCAA	AATGAAAAAAACTCAA	ATATAAAAAACTCAA	ATATAAAAAACTCAA	TTATAAAAAACTCAA
P	TI	AACAAACCTAA	AACAACCCTAA	AACGATTAATT	AACAAACCTAA	AACAAACCTAA
	TTP	TTCGAAAAAAACTCAA	TATCAAAAAAACTCAA	ATATAAAAAACTCAA	TTATAAAAAACTCAA	ATATAAAAAACTCAA
M	TI	AACAAACCTAC	AACAAGCCTAA	AACATATAAAT	AACAAACTTAA	AACAAACCTAA
	TTP	GTTGAAAAAAACTCAA	TTAGAAAAAAACTCAA	ATATAAAAAACTCAA	ATTTAAAAAACACAA	ATATAAAAAACTCAG
G	TI	AACAAACCTAG	AACAACCCTAA	AACATAAAAGA	AACAATCCTAA	AACAAACCTAA
	TTP	TATCAAAAAAACTCAA	TATCAAAAAAACTCAG	ATATAAAAAACTCAA	ATATAAAAAACTCAA	TTATAAAAAACTCAA
U1	TI	AACAAACCTAC	AACAAACCTAC	AAGAAATAAAA	AATAATCCTAC	TAAAAACCTAA
	TTP	TATGAAAAAAACTCAA	TATGAAAAAAACTCAA	ATATAAAAAACTCAA	ATTTAAAAAACTCAA	TTTTAAAAAACTCAA
L	TI	AACAAACCTAG	AACAAACCTAG	AAGAAATACAA	AACAAACCTAA	AACAAACCTAA
	TTP	TATGAAAAAAACTCAT	GATGAAAAAAACTCAA	ATATAAAAAACTCAC	ATATAAAAAACTCAT	TTATAAAAAACTCAA
Con	TI	AACAAACCTAA	AACAAACCTAA	Cryptic	AACAAACCTAA	AACAAACCTAA
	TTP	TATGAAAAAAACTCAA	TATGAAAAAAACTCAA	ATATAAAAAACTCAA	ATATAAAAAACTCAA	TTATAAAAAACTCAA

ABTV = Arboretum virus; BALV = Balsa virus; CBV = Coot Bay virus; PTAMV = Puerto Almendras virus; RCHV = Rio Chico virus.
*Transcription initiation sequence.
†Transcription termination-polyadenylation sequence.

6.0)[12,13] and ambiguously aligned regions were removed using Gblocks.[14] The resulting alignment comprising 921 amino acids was used to infer phylogenetic relationships using the maximum-likelihood (ML) method in MEGA 6.0 employing the WAG model of amino acid substitution, earest neighbor interchange (NNI) and subtree pruning and regrafting (SPR) branch-swapping. The phylogenetic robustness of each node was determined using 100 bootstrap replicates and NNI branch-swapping. Trees were annotated using Figtree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree).

RESULTS AND DISCUSSION

Each of the four virus isolates was initially isolated in cultures of C6/36 cells. On the first passage, CBV (EVG 5-53) produced marked CPE in the mosquito cells within 48 hours. The two BALV virus strains (CoB 76 and CoB 84) also caused massive CPE within 2 days. The original mosquito homogenate yielding RCHV (GAM195) produced CPE in the C6/36 cell culture 4 days after inoculation, but on the second C6/36 passage, CPE was observed after 2 days. However, subsequent next generation sequencing of the GAM195 isolate indicated that it contained two distinct viruses, RCHV and Wallerfield virus, a negevirus, so we are uncertain which virus caused the CPE.

After one or two passages in C6/36 cells, the four rhabdovirus isolates were inoculated into cultures of Vero E6 and BHK-21 cells maintained at 37°C. No CPE was observed in the mammalian cells after 14 days. Samples (0.15 µL) of each virus suspension were also inoculated intracranially into single litters of 1- to 2-day-old, outbred (ICR) mice. Pups were examined daily for 14 days, but no illness or deaths were observed in the animals. To further assess the potential for virus replication in vertebrate cells, one isolate of BALV (CoB 76) and the CBV (EVG 5-53) isolate were each inoculated onto confluent monolayers of Vero cells, duck embryo cells, and toad (*X. laevis*) cells, and the culture supernatants were tested daily by PCR for the presence of viral RNA. In each cell line, BALV RNA was detected as a strong band in the inoculum and as weaker bands of decreasing intensity in supernatants collected on day 0 (triple-washed cells) and from day 1 to day 4, but was not detected in supernatants collected on day 5 or day 6. CBV was also detected as a moderately intense band in the inoculum, but was not detected in the supernatants collected from any of the cell lines from day 0 (postinoculation triple-washed cells) to day 6. There was no evidence of CPE in any of the cell lines inoculated with BALV or CBV. Taken together, these results suggested that these viruses, as proposed previously for ABTV and PTAMV, do not replicate in vertebrate cells and may be insect-specific viruses.[6]

Three of the virus isolates were characterized initially by electron microscopy. Ultrathin sections of infected C6/36 cells were processed as described previously[15] and examined by transmission electron microscopy. BALV strain CoB 76 virions (50 –55 nm in diameter and 180 –310 nm in length) were observed budding into intracytoplasmic vacuoles or from the plasmalemma into extracellular space (Figure 1A and B). A virion was also observed in the process of internalization into a clathrin-coated vesicle (Figure 1B). BALV strain

CoB 84 virions (~50 nm in diameter and 140 –290 nm in length) were found mostly inside intracellular vacuoles (Figure 1C). CBV virions (~50 nm in diameter and 130–320 nm in length) were forming either from the cell surface or into expanded cisterns of granular endoplasmic reticulum (Figure 1D). In each case, the variations in the length of virions resulted in morphology that varied from the bullet-like shapes to long rod-like shapes, sometimes with two rounded ends. These variations were also observed previously for ABTV and PTAMV,[6] and may be indicative of the packaging of full-length genomes and smaller defective-interfering particles.

Complete genome sequences were determined for BALV strain CoB 76 and BALV CoB 84 (11,287 nucleotides and 11,286 nucleotides, respectively). For CBV and RCHV, complete coding regions were sequenced (10,869 nucleotides and 10,735 nucleotides, respectively) with only the extreme genome termini not determined. BLAST searches (blastx)[16] of the National Center for Biotechnology Information non-redundant protein sequences (nr) database indicated that each of the genomes was most closely related to ABTV (GenBank KC994644) and PTAMV (GenBank KF534749) which had been isolated from mosquito pools collected in Brazil by passage in C6/36 cells and which had also failed to grow in mammalian cell cultures or in newborn mice.[6] Translation of the genome sequences indicated that BALV, CBV, and RCHV each have similar organizations to ABTV and PTAMV, comprising the five canonical rhabdovirus structural protein genes (N, P, M, G, and L) and a gene encoding a small hydrophobic protein (U1 or SH) located between the G and L genes (Figure 2). Each gene is flanked by transcription regulatory sequences which conform to the consensus sequences 5′-UUGUUUGGAUU (transcription initiation) and [U]7GAGUU-3′ (transcription termination), except for CBV in which transcription initiation sequences appear to be highly variable and somewhat cryptic (Table 1). The novel genome sequences have been deposited in GenBank under accession numbers KX228196–KX228199.

The structural characteristics of polypeptides encoded in each of the genes of BALV, CBV, and RCHV were compared with those of ABTV and PTAMV. The N gene of each virus contains a single long open reading frame (ORF) encoding a polypeptide with the structural characteristics of rhabdovirus nucleoproteins. The N proteins range in size from 429 to 438 amino acids (predicted molecular weights 48.2–49.8 kDa) and share amino acid sequence identity (p-distance; MEGA 6.0) ranging from 28.4% (ABTV and CBV) to 61.9% (ABTV and PTAMV). Alignment of the sequences with the vesicular stomatitis Indiana virus (VSIV) N protein indicated conservation of eight basic residues that are located in the RNA-binding cavity, six of which coordinate with phosphate groups in the ribonucleoprotein (RNP) with VSIV genomic RNA, and conservation of two large aromatic residues (F, W, Y) that form a hydrophobic patch on the C-terminal lobe of the VSIV RNA-binding cavity (Supplemental Figure 1).[17,18] Interestingly, one of the six RNA-coordinating residues in VSIV N (R-146) is not conserved in the CBV N protein, aligning to a polar residue (S-155) in stretch of amino acids that lacks basic residues. Variations in phosphate-coordinating residues have also been reported between VSIV and rabies virus RNPs.[18]

The P genes of each virus contain long ORFs encoding polypeptides that vary significantly in size (210–320 amino acids; predicted 24.4–36.5 kDa) and display little overall sequence identity (> 25%). However, the proteins are similar in net charge (pI = 5.07–5.89) and, like the P proteins of other rhabdoviruses, they are rich in charged amino acids (D, E, K, R, H) which, in VSIV, form extensive acidic and basic domains on the protein surface.[19] Structural variability is typical of rhabdovirus P proteins which are intrinsically disordered.[20] Uniquely among this set of viruses, the P genes of each strain of BALV contains an alternative ORF (Px) commencing 35 nucleotides downstream of the P ORF initiation codon. The putative Px proteins (81 amino acids; 9.8 kDa) are highly basic (pI = 10.66) and share a high level of sequence identity (97.5%) (Figure 3A). Small basic proteins are commonly encoded in alternative ORFs in the P genes of rhabdoviruses[21] including VSIV for which two small basic carboxy-coterminal proteins (C and C′) have been shown to be expressed in infected cells.[22,23]

The M genes contain ORFs encoding mildly basic proteins (pI = 8.61–9.40) of similar size (159–182 amino acids; predicted 19.1–21.4 kDa). They share identifiable levels of sequence identity across the data set, but poor homology with the M proteins of VSIV and other rhabdoviruses (Supplemental Figure 2). Rhabdovirus M proteins are a structural component of virions and play roles in maturation and

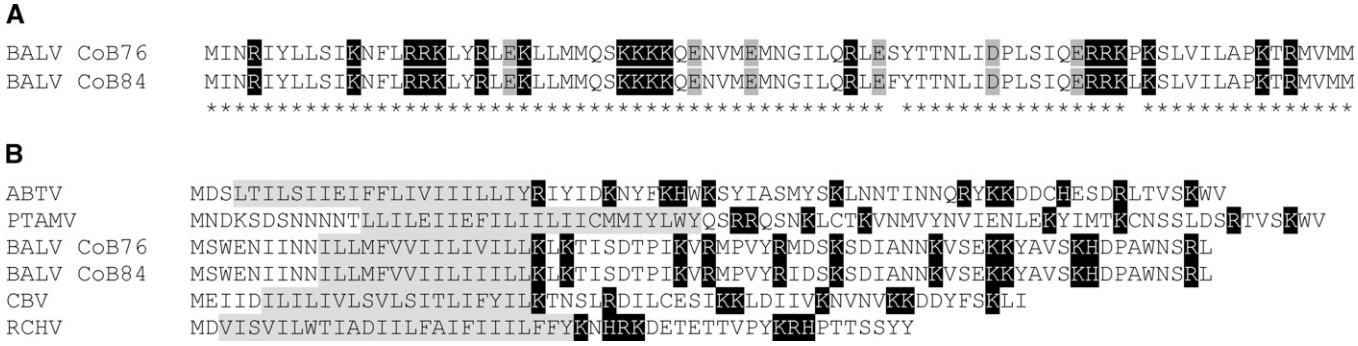

FIGURE 3. (**A**) ClustalX alignment of the amino acid sequences of the Px proteins of Balsa virus (BALV) strains CoB 76 and CoB 84 showing the high level of sequence identity and the high proportion of basic amino acid residues. Basic amino acids (K, R) are shaded in black and acidic amino acids (D, E) are shaded in grey. (**B**) Comparison of the amino acid sequences of the U1 proteins of each virus illustrating the similar structures including the predicted transmembrane domains (shaded grey) and relatively high proportions of basic amino acid residues (K, R, H) in the C-terminal domains.

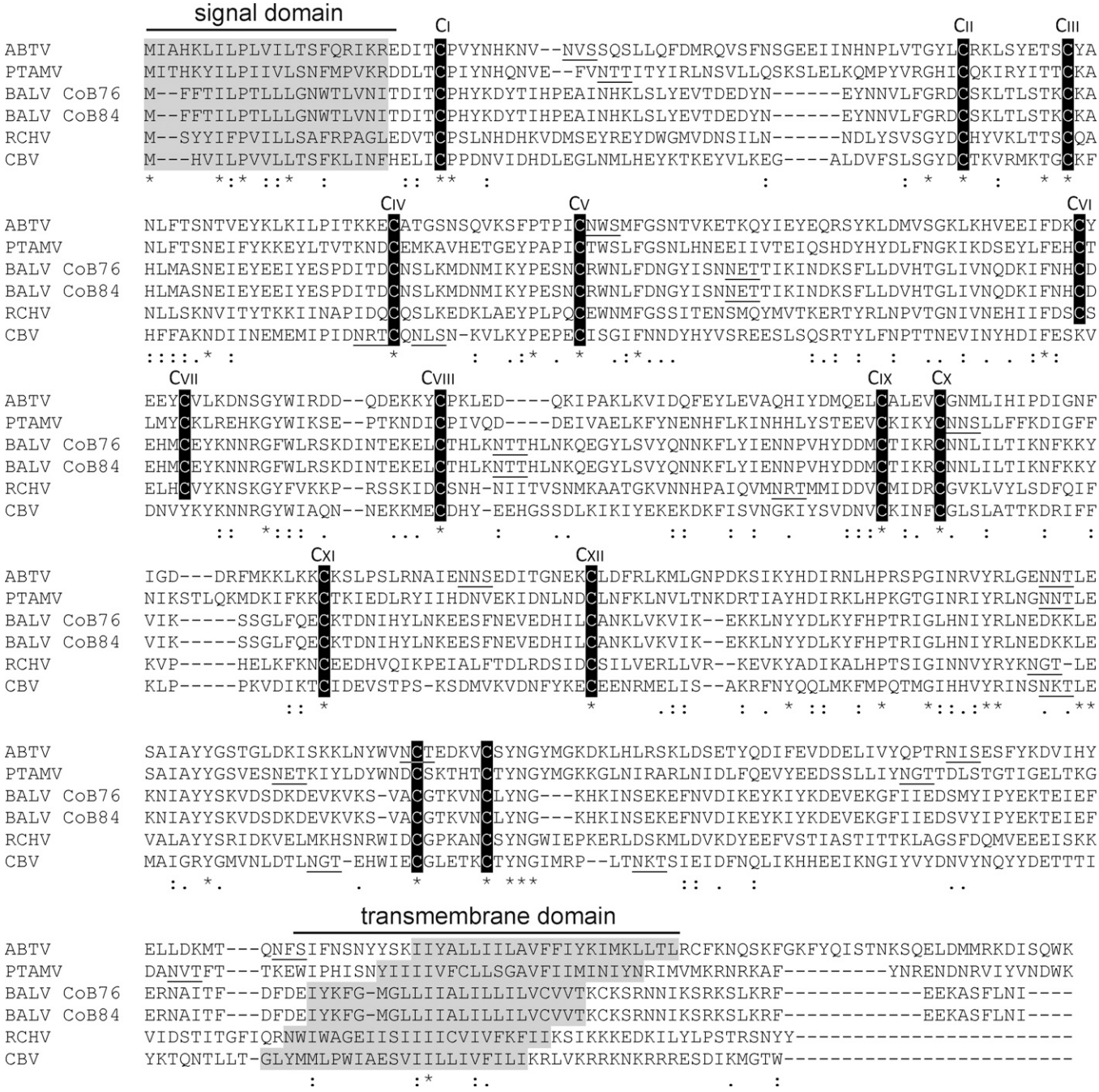

FIGURE 4. ClustalX alignment of the amino acid sequences of the G proteins of each virus. The signal domains and transmembrane domains are shaded grey. Cysteine residues in the ectodomain are shaded black. Cysteine residues that form conserved disulfide bridges in animal rhabdoviruses (C_I–C_{XII}) are numbered according to the scheme of Walker and Kongsuwan.[26] Predicted N-glycosylation sites are underlined. Fully conserved (*), strongly conserved (:), or weakly conserved (.) amino acids are indicated below the alignment.

modulating intracellular processes during virus replication.[24] The N-terminal domain of the VSIV M protein contains a basic region that is important in membrane localization and late budding domain motifs (PPPY and PSAP) that are required for membrane fusion and release of virions from infected cells.[25] Similar clusters of basic residues are evident in the N-terminal domains of the M proteins of each of the viruses in this data set (Supplemental Figure 2), and sequences similar to recognized late budding motifs have been identified previously in ABTV and PTAMV.[6] However, we could find no evidence of recognized late budding motifs in the CBV, RCHV, or BALV M proteins.

The G genes each contain a single long ORF encoding a class 1 transmembrane glycoprotein with an N-terminal signal domain and a near-C-terminal transmembrane domain (Figure 4). The unprocessed G proteins vary in size (442–476 amino acids; predicted 51.5–55.8 kDa) and in the number (2–6) of predicted N-glycosylation sites (NetNGlyc 1.0 server), the locations of which are not generally conserved. Each virus contains the 12 highly conserved cysteine residues (C_I–C_{XII})

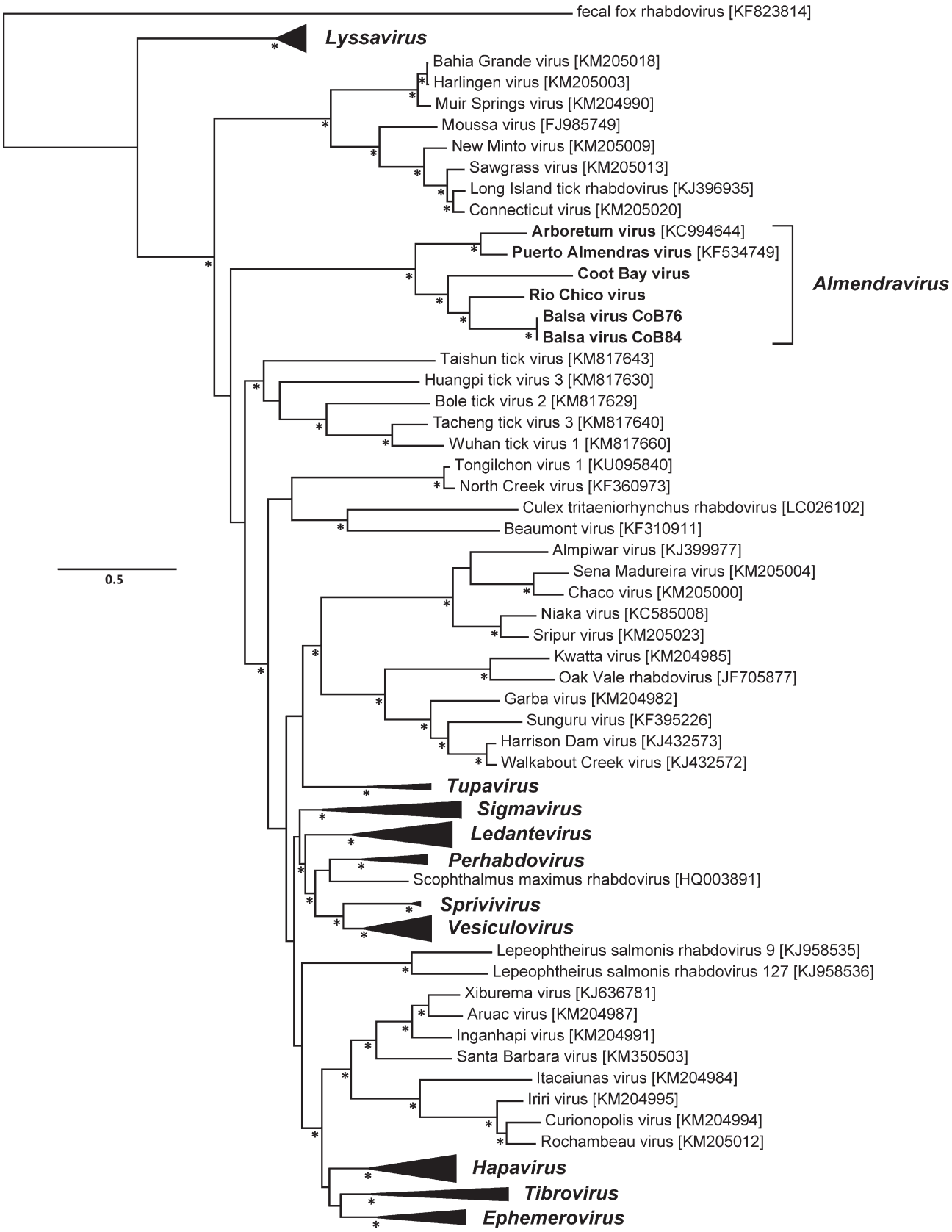

FIGURE 5. A maximum-likelihood phylogenetic tree inferred from a MUSCLE alignment of the L protein sequences of 137 animal rhabdoviruses. Clades containing viruses assigned to eight existing genera (*Lyssavirus*, *Tupavirus*, *Sigmavirus*, *Perhabdovirus*, *Sprivivirus*, *Vesiculovirus*, *Tibrovirus*, and *Ephemerovirus*) and two proposed new genera (*Hapavirus*, *Ledantevirus*) have been collapsed to reduce complexity. Asterisks (*) indicate well-supported nodes in the tree (bootstrap proportion ≥ 75%). Horizontal branch lengths are drawn to a scale of amino acid substitutions/site, the scale bar indicating a value of 0.5.

that are characteristic of rhabdovirus G proteins,[26] except for CBV which lacks cysteines C_{VI} and C_{VII}. These 12 cysteine residues form six disulfide bridges that stabilize the folded secondary structures of the trimerization, fusion and plextrin homology domains of the G protein.[27] In most animal rhabdoviruses, cysteines C_{VI} and C_{VII} form a disulfide bridge that stabilizes an exposed loop in the fusion domain.[27] Interestingly, this disulfide bridge is also absent from the G proteins of sigmaviruses which are known to be insect specific.[4,26] As reported previously for ABTV and PTAMV,[6] each virus also contains two additional conserved cysteine residues that are likely to form a unique seventh disulfide bridge in the lateral domain.

The U1 genes each encode small hydrophobic proteins (51–80 amino acids; predicted 6.1–9.5 kDa) which are predicted to feature a short N-terminal luminal domain, a central transmembrane domain, and a longer C-terminal cytoplasmic domain that is rich in basic residues (TMpred; http://embnet.vital-it.ch/software/TMPRED_form.html) (Figure 3B). This structure is characteristic of class 1A viroporins.[28] ORFs encoding viroporin-like proteins have been reported in a wide range of animal rhabdoviruses, including ABTV and PTAMV, and are usually located immediately following the G gene.[6,21,29] The α1 viroporin of the rhabdovirus, bovine ephemeral fever virus, has been shown to localize in the Golgi complex, increase cellular permeability, and interact with importins to disrupt nuclear trafficking.[30]

The L genes of each virus encode the large multifunctional RNA-dependent RNA polymerase (2,035–2,062 amino acids; predicted 234.5–241.1 kDa) that contains each of the conserved regions (CR_I–CR_{VI}) and functional domains (RdRp, capping, connector, methyltransferase, and CTD) that are characteristic of the L proteins of all nonsegmented (–) ssRNA viruses. Amino acid sequence identity between the L proteins of viruses in the data set ranges from 42.4% (ABTV and CBV) to 65.1% (ABTV and PTAMV).

An ML phylogenetic tree was inferred using the complete L protein sequences of CBV, RCHV, BALV, and 134 other animal rhabdoviruses, including all members of existing genera, proposed new genera, and many rhabdoviruses that are currently unassigned (Figure 5). Novirhabdoviruses (infecting fish) and several rhabdovirus genomes detected by metagenomics analysis of arthropods were excluded as the L protein sequences are too divergent, limiting the length of the Gblocks-pruned alignment and thus reducing phylogenetic resolution. Fecal fox rhabdovirus, which was detected by metagenomics analysis of fecal samples from red foxes (*Vulpes vulpes*) in Spain,[31] was the most divergent sequence and was used as the outgroup in the tree. In this analysis, CBV, RCHV, and BALV clustered with ABTV and PTAMV as a distinct monophyletic group with strong bootstrap support (bootstrap proportion = 100%). Although several of the deeply rooted nodes in the tree were unresolved, the almendravirus clade was more deeply rooted than clades representing all other genera except lyssaviruses.

Previous studies have identified ABTV and PTAMV as novel and phylogenetically distinct rhabdoviruses and proposed the creation of a new genus *Almedravirus* to which the viruses should be assigned.[6,21] Our data support this proposal and indicate that the genus should include five new species (*Arboretum almendravirus*, *Puerto Almendras almendravirus*, *Coot Bay almendravirus*, *Rio Chico almendravirus*, and *Balsa almendravirus*). Each of these viruses has been isolated from mosquitoes, they failed to grow in mammalian cell cultures or in suckling mice, and they have a genome organization featuring a gene between the G and L genes that encodes a small hydrophobic class 1A viroporin-like protein. They also have a similar a similar arrangement of cysteine residues in the G proteins, supporting previous observations of the genus-specificity of disulfide bridges.[26]

Failure of these viruses to grow in newborn mice or in vertebrate cells in vitro may be indicative of poor adaptation to replication in vertebrates. An increasing number of RNA viruses isolated from mosquitoes in recent years have been characterized as insect specific, replicating only in mosquitoes with no apparent vertebrate host. These include positive-sense ssRNA viruses (flaviviruses, togaviruses, negeviruses, mesoniviruses, nodaviruses, tymoviruses), double-stranded RNA viruses (reoviruses), and (–) ssRNA viruses (rhabdoviruses).[32] Indeed, recent metagenomics surveys of various arthropods have revealed a vast and deeply rooted diversity of both positive-sense and negative-sense RNA viruses, suggesting that they may be the ancestral hosts from which vertebrate-infecting viruses have evolved.[5,33] Our analysis indicated that, although several deep nodes remained unresolved, the almendravirus clade is clearly basal to most recognized families of arthropod-borne rhabdoviruses infecting vertebrates. However, it was also evident that the clade is not more deeply rooted than lyssaviruses (infecting only mammals) or several other viruses that were isolated from mosquitoes that have a broad host range in mammals and reptiles (Bahia Grande virus, Reed Ranch virus, and Muir Springs virus).[34] We suggest, therefore, that further experimental infections and epidemiological studies are required to determine whether the almendraviruses are truly insect specific.

Poor resolution of deeper roots in the phylogeny also suggests that, unless major ancestral lineages are now extinct, there may be many as yet undiscovered rhabdoviruses that would help clarify the phylogenetic position of almendraviruses and add further insights into the evolutionary history of the family.

Received May 18, 2016. Accepted for publication September 27, 2016.

Published online October 31, 2016.

Note: Supplemental figures appear at www.ajtmh.org.

Acknowledgments: We thank Richard Hoyos and Juan David Suaza for help for fieldwork and entomological processing of the samples in Colombia. We thank Matt Aliota for his assistance in screening samples at NYSDOH.

Financial support: This work was supported in part by contract HHSN272201000040I/HHSN2700004/D4 from the National Institutes of Health. Laboratory work in the United States was funded in part by NIH grant R24 AI120942. Maria Angelica Contreras was supported by Programa de Doctorados Nacionales–Colciencias (Convacatoria 567) PhD Fellowship from Colombia. Gillian Eastwood, and collections in Panama, were supported by a Robert E. Shope International Fellowship from the American Society of Tropical Medicine and Hygiene in 2012; kindly assisted by Jose R. Loaiza, and the Smithsonian Tropical Research Institute. We also acknowledge support from Yale Institute for Biospheric Studies, Center for Ecoepidemiology, and the U.S. National Park Service (Collecting Permit no. EVER-2013-SCI-0032). The field collection in Colombia was funded in part by Colciencias grant 111549326198.

Disclaimer: Animal use in this work was conducted under protocol no. 9505045, approved by the IACUC at the University of Texas Medical Branch.

Authors' addresses: Maria Angelica Contreras, Programa de Estudio y Control de Enfermedades Tropicales (PECET), Sede de Investigacion Universitaria (SIU), Universidad de Antioquia, Medellin, CO, and Grupo de Investigacion en Sistematica Molecular (GSM), Facultad de Ciencias, Universidad Nacional de Colombia, Medellin, CO, E-mail: maria.contreras@pecet-colombia.org. Gillian Eastwood, Griffin Arbovirus Laboratory, Wadsworth Center, New York State Department of Health, Slingerlands, NY, E-mail: gill2g@hotmail.com. Hilda Guzman, Department of Pathology, University of Texas Medical Branch, Houston, TX, E-mail: hguzman@utmb.edu. Vsevolod Popov and Robert B. Tesh, Department of Pathology, Center for Biodefense and Emerging Infectious Diseases, Center for Tropical Diseases, Institute for Human Infections and Immunity, University of Texas Medical Branch, Galveston, TX, E-mails: vpopov@utmb.edu and rtesh@utmb.edu. Chelsea Savit, Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, E-mail: savvysavit@gmail.com. Sandra Uribe, Sede de Investigacion Universitaria (SIU), Universidad de Antioquia, Medellin, CO, E-mail: siuribesoto@gmail.com. Laura D. Kramer, Zoonotic Diseases, New York State Department of Health, Slingerlands, NY, and [3]Griffin Arbovirus Laboratory, Wadsworth Center, New York State Department of Health, Albany, NY, E-mail: laura.kramer@health.ny.gov. Thomas G. Wood and Steven G. Widen, Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, E-mails: tgwood@utmb.edu and sgwiden@utmb.edu. Durland Fish, Epidemiology and Public Health, Yale University, New Haven, CT, E-mail: durland.fish@yale.edu. Nikos Vasilakis, Department of Pathology, University of Texas Medical Branch, Galveston, TX, and Center for Biodefense and Emerging Infectious Diseases, University of Texas Medical Branch, Galveston, TX, E-mail: nivasila@utmb.edu. Peter J. Walker, CSIRO Health and Biosecurity, Australian Animal Health Laboratory, Geelong, Australia, and School of Biological Sciences, University of Queensland, St Lucia, Australia, E-mail: peter.walker@uq.edu.au.

REFERENCES

1. Dietzgen RG, Calisher CH, Kurath G, Kuzman IV, Rodriguez LL, Stone DM, Tesh RB, Tordo N, Walker PJ, Wetzel T, Whitfield AE, 2012. *Rhabdoviridae*. King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, eds. *Virus Taxonomy*, Ninth Report of the International Committee on Taxonomy of Viruses. San Diego, CA: Elsevier, 654–681.
2. Kuzmin IV, Novella IS, Dietzgen RG, Padhi A, Rupprecht CE, 2009. The rhabdoviruses: biodiversity, phylogenetics, and evolution. *Infect Genet Evol 9:* 541–553.
3. Kuzmin IV, Walker PJ, 2016. Vector-borne rhabdoviruses. Vasilakis N, Gubler DJ, eds. *Arboviruses: Molecular Biology, Evolution and Control*. Norfolk, VA: Caister Academic Press, 389.
4. Longdon B, Obbard DJ, Jiggins FM, 2010. Sigma viruses from three species of *Drosophila* form a major new clade in the rhabdovirus phylogeny. *Proc Biol Sci 277:* 35–44.
5. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, Qin XC, Xu J, Holmes EC, Zhang YZ, 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife 4:* e05378.
6. Vasilakis N, Castro-Llanos F, Widen SG, Aguilar PV, Guzman H, Guevara C, Fernandez R, Auguste AJ, Wood TG, Popov V, Mundal K, Ghedin E, Kochel TJ, Holmes EC, Walker PJ, Tesh RB, 2014. Arboretum and Puerto Almendras viruses: two novel rhabdoviruses isolated from mosquitoes in Peru. *J Gen Virol 95:* 787–792.
7. Vasilakis N, Forrester NL, Palacios G, Nasar F, Savji N, Rossi SL, Guzman H, Wood TG, Popov V, Gorchakov R, Gonzalez AV, Haddow AD, Watts DM, da Rosa AP, Weaver SC, Lipkin WI, Tesh RB, 2013. Negevirus: a proposed new taxon of insect-specific viruses with wide geographic distribution. *J Virol 87:* 2475–2488.
8. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B, 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res 40:* W622–W627.
9. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I, 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res 19:* 1117–1123.
10. Langmead B, Salzberg SL, 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods 9:* 357–359.
11. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP, 2011. Integrative genomics viewer. *Nat Biotechnol 29:* 24–26.
12. Edgar RC, 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics 5:* 113.
13. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol 30:* 2725–2729.
14. Talavera G, Castresana J, 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol 56:* 564–577.
15. Vasilakis N, Widen S, Mayer SV, Seymour R, Wood TG, Popov V, Guzman H, Travassos da Rosa AP, Ghedin E, Holmes EC, Walker PJ, Tesh RB, 2013. Niakha virus: a novel member of the family *Rhabdoviridae* isolated from phlebotomine sandflies in Senegal. *Virology 444:* 80–89.
16. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res 25:* 3389–3402.
17. Green TJ, Zhang X, Wertz GW, Luo M, 2006. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science 313:* 357–360.
18. Luo M, Green TJ, Zhang X, Tsao J, Qiu S, 2007. Conserved characteristics of the rhabdovirus nucleoprotein. *Virus Res 129:* 246–251.
19. Ribeiro EA, Favier A, Gerard FCA, Leyrat C, Brutscher B, Blondel D, Ruigrok RWH, Blackledge M, Jamin M, 2008. Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J Mol Biol 382:* 525–538.
20. Karlin D, Ferron F, Canard B, Longhi S, 2003. Structural disorder and modular organization in *Paramyxovirinae* N and P. *J Gen Virol 84:* 3239–3252.
21. Walker PJ, Firth C, Widen SG, Blasdell KR, Guzman H, Wood TG, Paradkar PN, Holmes EC, Tesh RB, Vasilakis N, 2015. Evolution of genome size and complexity in the *Rhabdoviridae*. *PLoS Pathog 11:* e1004664.
22. Spiropoulou CF, Nichol ST, 1993. A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus. *J Virol 67:* 3103–3110.
23. Peluso RW, Richardson JC, Talon J, Lock M, 1996. Identification of a set of proteins (C′ and C) encoded by the bicistronic P gene of the Indiana serotype of vesicular stomatitis virus and analysis of their effect on transcription by the viral RNA polymerase. *Virology 218:* 335–342.
24. Gubala AJ, Proll DF, Barnard RT, Cowled CJ, Crameri SG, Hyatt AD, Boyle DB, 2008. Genomic characterisation of Wongabel virus reveals novel genes within the *Rhabdoviridae*. *Virology 376:* 13–23.
25. Freed EO, 2002. Viral late domains. *J Virol 76:* 4679–4687.
26. Walker PJ, Kongsuwan K, 1999. Deduced structural model for animal rhabdovirus glycoproteins. *J Gen Virol 80:* 1211–1220.
27. Roche S, Bressanelli S, Rey FA, Gaudin Y, 2006. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science 313:* 187–191.
28. Nieva JL, Madan V, Carrasco L, 2012. Viroporins: structure and biological functions. *Nat Rev Microbiol 10:* 563–574.
29. Walker PJ, Dietzgen RG, Joubert DA, Blasdell KR, 2011. Rhabdovirus accessory genes. *Virus Res 162:* 110–125.
30. Joubert DA, Blasdell KR, Audsley MD, Trinidad L, Monaghan P, Dave KA, Lieu K, Amos-Ritchie R, Jans DA, Moseley GW, Gorman JJ, Walker PJ, 2014. Bovine ephemeral fever rhabdovirus a1 protein has viroporin-like properties and binds importin b1 and importin 7. *J Virol 88:* 1591–1603.
31. Bodewes R, Ruiz-Gonzalez A, Schurch AC, Osterhaus AD, Smits SL, 2014. Novel divergent rhabdovirus in feces of red fox, Spain. *Emerg Infect Dis 20:* 2172–2174.

32. Bolling BG, Weaver SC, Tesh RB, Vasilakis N, 2015. Insect-specific virus discovery: significance for the arbovirus community. *Viruses 7:* 4911–4928.
33. Shi M, Lin XD, Vasilakis N, Tian JH, Li CX, Chen LJ, Eastwood G, Diao XN, Chen MH, Chen X, Qin XC, Widen SG, Wood TG, Tesh RB, Xu J, Holmes EC, Zhang YZ, 2015. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the *Flaviviridae* and related viruses. *J Virol 90:* 659–669.
34. Kerschner JH, Calisher CH, Vorndam AV, Francy DB, 1986. Identification and characterization of Bahia Grande, Reed Ranch and Muir Springs viruses, related members of the family *Rhabdoviridae* with widespread distribution in the United States. *J Gen Virol 67:* 1081–1089.