

An Insect Nidovirus Emerging from a Primary Tropical Rainforest

Florian Zirkel,^{a,b,c} Andreas Kurth,^d Phenix-Lan Quan,^b Thomas Briese,^b Heinz Ellerbrok,^d Georg Pauli,^d Fabian H. Leendertz,^c W. Ian Lipkin,^b John Ziebuhr,^e Christian Drosten,^a and Sandra Junglen^{a,c}

Institute of Virology, University of Bonn Medical Center, Bonn, Germany^a; Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, New York, USA^b; Research Group Emerging Zoonoses^c and Center for Biological Safety-1,^d Robert Koch Institute, Berlin, Germany; and Institute of Medical Virology, Justus Liebig University Gießen, Gießen, Germany^e

ABSTRACT Tropical rainforests show the highest level of terrestrial biodiversity and may be an important contributor to microbial diversity. Exploitation of these ecosystems may foster the emergence of novel pathogens. We report the discovery of the first insect-associated nidovirus, tentatively named Cavally virus (CAVV). CAVV was found with a prevalence of 9.3% during a survey of mosquito-associated viruses along an anthropogenic disturbance gradient in Côte d'Ivoire. Analysis of habitat-specific virus diversity and ancestral state reconstruction demonstrated an origin of CAVV in a pristine rainforest with subsequent spread into agriculture and human settlements. Virus extension from the forest was associated with a decrease in virus diversity ($P < 0.01$) and an increase in virus prevalence ($P < 0.00001$). CAVV is an enveloped virus with large surface projections. The RNA genome comprises 20,108 nucleotides with seven major open reading frames (ORFs). ORF1a and -1b encode two large proteins that share essential features with phylogenetically higher representatives of the order *Nidovirales*, including the families *Coronavirinae* and *Torovirinae*, but also with families in a basal phylogenetic relationship, including the families *Roniviridae* and *Arteriviridae*. Genetic markers uniquely conserved in nidoviruses, such as an endoribonuclease- and helicase-associated zinc-binding domain, are conserved in CAVV. ORF2a and -2b are predicted to code for structural proteins S and N, respectively, while ORF3a and -3b encode proteins with membrane-spanning regions. CAVV produces three subgenomic mRNAs with 5' leader sequences (of different lengths) derived from the 5' end of the genome. This novel cluster of mosquito-associated nidoviruses is likely to represent a novel family within the order *Nidovirales*.

IMPORTANCE Knowledge of microbial diversity in tropical rainforests is sparse, and factors driving the emergence of novel pathogens are poorly understood. We discovered and mapped the spread and genetic evolution of a novel mosquito nidovirus from a pristine rainforest to human settlements. Notably, virus diversity decreased and prevalence increased during the process of spreading into disturbed habitats. The novel virus, tentatively termed Cavally virus, contains genetic features common to members of the order *Nidovirales* (families *Coronaviridae*, *Arteriviridae*, and *Roniviridae*), including conservation of the replicase machinery and expression of subgenomic RNA messages, has a basal phylogenetic relationship to the family *Coronaviridae*, and clearly differs from the established nidovirus families. Inclusion of this putative novel family in the nidovirus phylogeny suggests that nidoviruses may have evolved from arthropods.

Received 8 April 2011 Accepted 13 May 2011 Published 14 June 2011

Citation Zirkel F, et al. 2011. An insect nidovirus emerging from a primary tropical rainforest. *mBio* 2(3):e00077-11. doi:10.1128/mBio.00077-11.

Editor Michael Buchmeier, University of California, Irvine

Copyright © 2011 Zirkel et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Sandra Junglen, junglen@virology-bonn.de.

Tropical rainforests comprise the highest level of terrestrial biodiversity (1–3). Microbes are dependent on their hosts, and hot spots of host biodiversity may also be rich in pathogen diversity. Intensive land use changes in West African rainforests began in the 1960s by industrial deforestation, leaving only remnants of the equatorial forest belt (4–6). Patches of forest are typically surrounded by agricultural land and human settlements. Logging is linked to a profound decline in biodiversity (4, 7–12). Declines in biodiversity are associated with an increased risk of infectious disease emergence in humans and domestic animals (reviewed in reference 13). Nevertheless, links between anthropogenic habitat modification and virus emergence remain to be confirmed (13, 14).

We recently conducted a survey of mosquito-borne viruses within and around a primary forest habitat in Côte d'Ivoire, West

Africa (14). Findings of highly divergent viruses suggested an extension of the diversity of several virus taxa in primary forest habitats (15, 16). We also identified short sequence fragments of an unusual virus in insects that was distantly related to coronaviruses (14). Here we report the genomic sequence and organization of this first insect nidovirus and its evolutionary divergence along an anthropogenic disturbance gradient extending from primary forest into human settlements.

The order *Nidovirales* comprises the families *Coronaviridae* (subfamilies *Coronavirinae* [CoV] and *Torovirinae* [ToV]), as well as the monogeneric families *Arteriviridae* (ArV, genus *Arterivirus*) and *Roniviridae* (RoV, genus *Okavirus*) (17). Nidoviruses have a wide range of hosts, including crustaceans (RoV) (18), fish (ToV) (19), birds (CoV) (20), and a variety of mammals (ArV, CoV, and ToV) (17, 21–23). With plus-stranded genomes between 26 and

TABLE 1 CAVV isolation from pools of female mosquitoes

Sampling site	Pool	Mosquito species	No. of mosquitoes	CPE ^a	% CAVV isolation (no. positive/total)	Genetic distance ^b (%)
Camps	A4	<i>Culex</i> spp.	20	3 dpi	7.5 (5/67)	0.1
	A21	<i>Culex (Eumelanomyia)</i> spp.	10	Weak		
	A26	<i>C. nebulosus</i>	10	3 dpi		
	A28	<i>C. nebulosus</i>	22	7 dpi		
	A31	<i>Culex</i> spp.	9	2nd P, 7 dpi		
Primary forest	B14	Female, ND ^c	17	1st P, 3 dpi	4.1 (4/98)	1.4
	B15	<i>C. nebulosus</i>	1	2nd P, 7 dpi		
	B40	Female, ND	2	3 dpi		
	B42	<i>C. nebulosus, C. cinerellus</i>	9	3 dpi		
	C45	<i>C. nebulosus</i>	16	3 dpi		
Secondary forest	C64	<i>Culex (Eumelanomyia)</i> spp.	10	7 dpi	5.1 (5/99)	0.5
	C65	<i>Culex (Eumelanomyia)</i> spp.	8	3 dpi		
	C79	<i>Aedes harrisoni</i>	6	1st P, 7 dpi		
	C82	<i>Anopheles</i> spp.	20	2 dpi		
	D27	<i>Aedes harrisoni</i>	2	7 dpi		
Plantations	D28	<i>Anopheles</i> spp.	2	3 dpi	9 (9/100)	0.1
	D42	<i>C. decens</i>	20	3 dpi		
	D50	<i>C. nebulosus</i>	20	3 dpi		
	D51	<i>C. nebulosus</i>	20	3 dpi		
	D60	Female, ND	15	1st P, 5 dpi		
	D61	Female, ND	11	1st P, 5 dpi		
	E9	<i>Uranotaenia chorleyi</i>	7	3 dpi		
	F2	Female, ND	23	Weak		
	F3	<i>Aedes harrisoni</i>	9	1st P, 5 dpi		
	F24	<i>C. nebulosus</i>	30	3 dpi		
Human settlements	F25	<i>C. nebulosus</i>	21	3 dpi	26.5 (18/68)	0.3
	F26	<i>C. nebulosus</i>	50	3 dpi		
	F27	<i>C. nebulosus</i>	40	1st P, 5 dpi		
	F28	<i>C. nebulosus</i>	20	3 dpi		
	F30	<i>C. nebulosus</i>	20	3 dpi		
	F32	<i>C. nebulosus</i>	20	3 dpi		
	F33	<i>C. nebulosus</i>	12	3 dpi		
	F35	<i>C. nebulosus</i>	16	3 dpi		
	F47	<i>Culicinae</i> spp.	10	8 dpi		
	F52	<i>C. quinquefasciatus</i>	20	3 dpi		
	F53	<i>C. quinquefasciatus</i>	8	1st P, 3 dpi		
	F54	<i>C. antennatus</i>	20	3 dpi		
	F55	<i>C. antennatus</i>	9	8 dpi		
	F56	<i>Culex</i> spp.	18	2nd P, 8 dpi		
	F63	<i>Anopheles</i> spp.	21	2nd P, 8 dpi		

^a Shown is the time point when the first CPE was visible. P, passage; dpi, days postinfection.

^b Maximal pairwise nucleotide distance with a 603-bp RT-PCR fragment in the ORF1b gene.

^c ND, not determined.

32 kb, CoV, ToV, and RoV have the largest known RNA genomes and are referred to as “large” nidoviruses (17). Genomes of the “small” nidoviruses (ArV) comprise 13 to 16 kb. All nidoviruses encode two replicase polyproteins, pp1a and pp1ab, encoded by open reading frame 1a (ORF1a) and ORF1b located at the 5' end of the genome and followed by genes encoding structural proteins and, in most cases, several accessory (nonessential) proteins. A distinctive feature of nidoviruses is their transcription strategy. Genes downstream of the replicase polyprotein gene are expressed from a nested set of 3'-coterminally subgenomic mRNAs (in Latin, nidus means nest) (24–27). In this study, analyses of subgenomic mRNAs and major features of genome organization and phylogenetic analysis were employed for a taxonomic classification of a novel insect nidovirus. The summation of results suggests the discovery of a previously unrecognized family within the order *Nidovirales*.

RESULTS AND DISCUSSION

Prevalence and divergence of CAVV along a gradient of habitat modifications. Between February and June 2004, 7,067 mosquitoes were trapped along an anthropogenic disturbance gradient in the area of the Taï National Park, Côte d'Ivoire. The gradient comprised sampling sites in the primary (pristine) forest, in secondary (modified) forest, in agriculturally exploited forest edge areas, and in adjacent human settlements (14–16). Initial analyses by cell culture and electron microscopy yielded viral particles with CoV-like morphology in supernatants of C6/36 cells infected with mosquito homogenates. A short genome fragment with low but significant identity to ORF1b of CoV was identified (14). This virus was tentatively named Cavally virus (CAVV), after a river near the sites where mosquitoes were caught.

Extensive attempts at virus isolation from adult female mosquitoes pooled in small numbers (up to 23 mosquitoes per pool)

yielded virus in 40 (9.3%) of the 432 pools tested. As shown in Table 1, virus was most frequently isolated from *Culex* mosquitoes, especially *Culex nebulosus*. Mosquitoes of the genera *Aedes*, *Anopheles*, and *Uranotaenia* were also found to be infected but at lower rates. CAVV was present in all sampled habitat types but with the highest prevalence in human settlements (Table 1). Analysis of variance showed the observed virus isolation rate in human settlements (and only there) to be significantly different from the mean isolation rates (F test, $P < 0.00001$) and from all isolation rates in any other habitat. In order to gain insight into the genetic diversity of CAVV, a 603-nucleotide (nt) genome fragment representing the 1b polymerase gene region was amplified and sequenced from all positive pools. The pairwise genetic distance between the isolates was up to 15% at the nucleotide level and up to 9.9% at the amino acid level. In particular, the nucleotide distance of isolate CAVV/A4/CI/2004 from all other isolates ranged between 13.5 and 15%, whereas the range of distances was below 1.8% for all other CAVV isolates. This suggested the presence of a diversified virus population comprising two different clusters whose distance is compatible with different species.

CAVV is an insect-associated virus whose hosts were encountered along a gradient of anthropogenic habitat modifications. To examine possible influences of habitat modification on virus diversity, average nucleotide distances over sequence pairs were determined within the amplified 603-nt fragments from all pools and were assigned to one of the habitat types from where the respective isolates originated, i.e., (i) primary forest or research camps within a primary forest (9 isolates), (ii) secondary forest (5 isolates), (iii) plantations (6 isolates), and (iv) villages (19 isolates). The mean numbers of nucleotide exchanges in pairwise sequence comparisons were determined within the samples pertaining to each habitat type. CAVV/A4/CI/2004 was excluded to prevent bias from a virus that may be evolutionarily disconnected from the main clade (putative second species). In habitat types i to iv, the mean pairwise exchange rates within the 603-nt fragment were 4.35, 2.8, 0.6, and 2, respectively. Even though isolates from villages predominated numerically in the data set, the level of diversification was highest in primary forest (chi-square test on exchange rates, $P < 0.01$).

To estimate the habitat association of the putative common ancestor of the 39 virus isolates, sequence fragments were subjected to phylogenetic analysis using a minimum evolution model, considering either the numbers of pairwise synonymous nucleotide exchanges or the percentages of overall (synonymous and nonsynonymous) pairwise nucleotide exchanges. CAVV/A4/CI/2004 was used as an outgroup. Phylogenies were subjected to ancestral state reconstruction under a parsimony assumption. These simple models of phylogeny and trait evolution were chosen because the ordered and sparse pattern of nucleotide exchange across the alignment suggested the absence of multiple or saturating exchanges. As shown in Fig. 1, ancestral state reconstruction suggested the most recent common ancestor linked to the outgroup, as well as the common ancestor of the ingroup, to have existed in a primary forest habitat. This matches the general concept wherein novel viruses are being transferred to areas of human settlement in the course of agricultural exploitation of primary habitats.

We tested 269 pools of 1,716 adult male mosquitoes collected during this survey for CAVV infection by specific real-time reverse transcription (RT)-PCR (14). CAVV was not detected in male

mosquitoes, suggesting that CAVV is likely dependent on amplifying vertebrate hosts.

Viral growth and morphology. Virus growth kinetics and morphological presentation in insect cells were studied. Cells showed strong cytopathic effects (CPE) manifesting in aggregation of cells at 48 h postinfection (hpi) (Fig. 2A and B). Virus replication was measured by real-time RT-PCR every 3 h for 2 days. Maximal RNA concentrations were reached at 15 to 18 hpi, indicating a fast replication cycle (Fig. 2C). Enveloped, spherical CoV-like virions with a mean diameter of 120 nm and large, club-shaped surface projections were detected in cell culture supernatants at 48 hpi by negative staining electron microscopy (Fig. 3). In ultrathin sections of fixed insect cells analyzed by transmission electron microscopy at 48 hpi, vesicles containing spherical, potentially enveloped particles were observed in the cytoplasm of infected cells. These were 50 to 60 nm in diameter and lacked surface projections (Fig. 3A and B). A role for these vesicles filled with virions resembling steps during virus maturation, as shown for other plus-strand RNA viruses, remains to be investigated (28–30). Furthermore, tubular structures likely of viral origin were detected in the cytoplasm of infected cells (Fig. 3A). Separation or adsorption of putative virions on cell membranes is shown in Fig. 3C. These particles were morphologically indistinguishable from typical particles encountered in cleared cell culture supernatant (Fig. 3D).

Full genome sequencing of isolate CAVV/C79/CI/2004 was achieved by a combination of adaptor-based random RT-PCR and ultradeep sequencing. The entire CAVV genome comprised 20,108 nt, excluding the 3' poly(A) tail (GenBank accession number HM746600), a size intermediate between that of arteriviruses (13 to 16 kb) and CoV or RoV (26 to 32 kb) (17). The genome contained seven major ORFs, as well as untranslated regions of 362 and 570 nt at the 5' and 3' ends, respectively (Table 2; Fig. 4A). In an effort to identify potential functional domains, the seven ORFs were compared by psiBLAST to a database restricted to nidoviruses (Table 2).

Putative replicase polyprotein genes. The replicase genes of nidoviruses share several domains and features conserved across the families *Arteriviridae*, *Coronaviridae*, and *Roniviridae* (25). These are (from the N to the C terminus) transmembrane domain 1 (TMD1), TMD2, a 3C-like protease (3CL^{PRO}) (31), TMD3, a ribosomal frameshift site (RFS) (32, 33), an RNA-dependent RNA polymerase (RdRp) (34), a zinc-binding domain (ZBD), an RNA helicase (HEL), and a uridylyate-specific endoribonuclease (NendoU) (35, 36). CoV, ToV, and RoV also share a 3'-5' exonuclease (ExoN) upstream of NendoU (37) and a ribose-2'-O-methyltransferase (MT) at the C terminus (38, 39). CoV and ToV also encode an ADP-ribose 1-phosphatase upstream of TMD1 (19, 40, 41).

Using TMHMM v2.0, three hydrophobic regions comprising putative multiple membrane-spanning domains (TMD1_{25L-L47}, TMD2_{1128I-Y1272}, and TMD3_{1727Y-M1780}) were identified within the first predicted ORF of the CAVV genome. Three TMDs are also found in ORF1a of CoV and ToV, whereas RoV has four TMDs (18). However, the position of CAVV TMD1 was similar to that in RoV (18). Between TMD2 and TMD3, a putative 3CL^{PRO} domain was identified. Comparative sequence analysis suggested that the CAVV 3CL^{PRO} domain is a cysteine protease with a Cys-His-Asp catalytic triad (see Fig. S1a in the supplemental material).

Putative functional motifs with similarity to CoV and ToV in

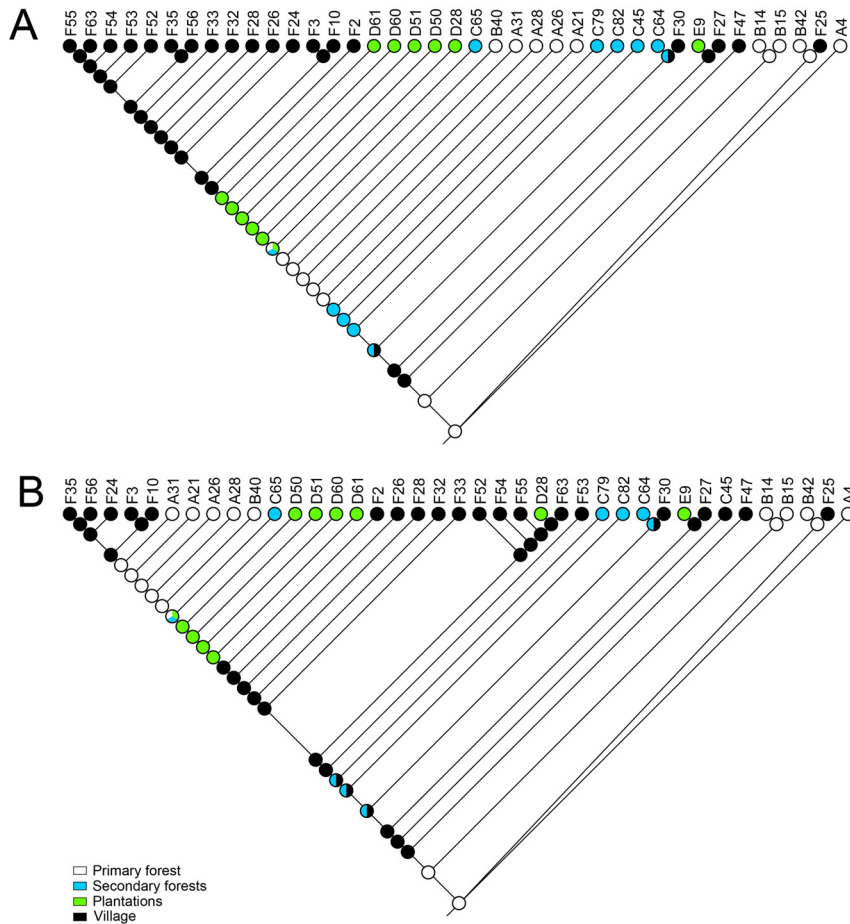


FIG 1 CAVV ancestral state reconstruction based on a maximum-parsimony algorithm. Minimum evolution phylogenies based on the numbers of synonymous exchanges (A) and the percent nucleotide distance (B) were calculated in Mega 5 (66). CAVV strain A4 served as an outgroup. Tip traits in Mesquite ([http://mesquiteproject.org/mesquite.html](http://mesquiteproject.org/mesquite/mesquite.html)) were defined by the habitats from which the virus isolates were taken (identified by color code at the bottom left of panel B). The circles at each root point identify by colored segments the likelihood of the root point taxon having existed in the habitat shown.

the second ORF included an RdRp domain (₆₃₁Y-V₇₈₇, cd01699), a ZBD (₁₀₅₁C-L₁₁₁₃, UPF1, pfam09416, Fig. S1b), and a HEL domain (₁₄₆₆K-I₁₇₁₇, superfamily I DNA and RNA helicases, COG1112). Phylogenies of RdRp and HEL domains indicated equidistant basal relationships to CoV and RoV (see below). Furthermore, a putative NendoU was identified. Sequence alignments suggest that the active site of the CAVV NendoU involves residues His₄₆₇₀, His₄₆₈₅, and Lys₄₇₂₅ (Fig. S1c). Although short regions with low similarity suggest conservation of the ExoN and MT domains, reliable alignments could not be generated and biochemical evidence remains to be obtained to confirm the functionality of these domains.

Putative RFS. Expression of the second major ORF of nidoviruses involves a programmed ribosomal frameshift into the -1 reading frame, occurring just upstream of the ORF1a stop codon (32, 33). The overlap region of ORF1a and -1b typically contains a slippery heptanucleotide sequence and a downstream RNA pseudoknot structure that together promote ribosomal frameshifting (32, 33). With minor variations, previously identified nidovirus slippery sequences conform to the XXXYYYZ consensus sequence conserved in many ribosomal slip sites on viral RNAs (for a review, see reference 42). Thus, for example, ArV, CoV, and ToV use

5'-(U/G)UUAAAC and RoV uses 5'-AAAUUUU as a slippery sequence (18, 42). The short (~35-nt) ORF1a/1b overlap region in CAVV does not contain heptanucleotide sequences related to those of other nidoviruses. The only XXXYYYZ-like sequence identified in this region of the CAVV genome is ₇₈₂₉CCCUUUG. However, previous systematic mutagenesis studies of ribosomal slip sites (43) revealed that the CCCUUUG heptanucleotide sequence does not promote efficient -1 ribosomal frameshifting *in vitro* and, to our knowledge, has not been reported to mediate efficient ribosomal frameshifting in viral or cellular systems. Further analyses revealed ₇₈₃₅GGAUUUU as a further candidate slip site. The GGAUUUU sequence conforms to the simultaneous-slippage model introduced by Jacks et al. (44). More importantly, data obtained for red clover necrotic mosaic (diantho) virus (RCNMV) (45, 46) have shown that this sequence mediates efficient -1 ribosomal frameshifting and thus expression of the downstream p57 polymerase ORF of RCNMV. Furthermore, the CAVV ₇₈₃₅GGAUUUU sequence is located 5 nt upstream of an energetically favorable RNA secondary structure. Both the length of the spacer element and the presence of a putative stem-loop structure adjacent to the proposed frameshift site support the idea that CAVV ORF1b expression is mediated by -1 ribosomal frame-

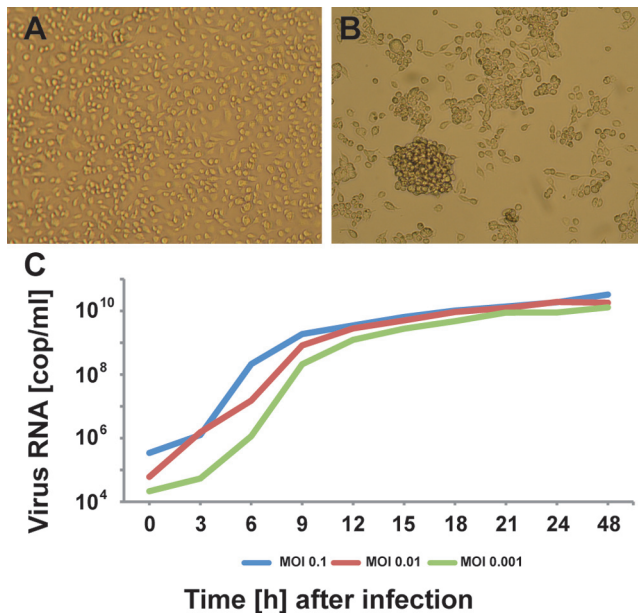


FIG 2 CAVV growth on insect cells. CPE in C6/36 cells infected with CAVV was observed at 48 hpi (B) compared to mock-infected C6/36 cells (A). Numbers of CAVV genome copies per milliliter in cell culture supernatant of C6/36 cells infected with CAVV at MOIs of 0.1, 0.01, and 0.001 were measured by RT-PCR at 0 to 48 hpi (C).

shifting at this site. Even though direct experimental evidence would be desirable to further corroborate this prediction, it seems reasonable to suggest that ORF1a/1b-encoded sequences are fused at a ₂₄₉₁Leu-Asp-Phe-Ser junction site.

Regulation of transcription. Although all nidovirus subgenomic mRNAs are 3' coterminal, different families and genera have different subgenomic mRNA 5' ends. ArV and CoV subgenomic mRNAs contain a common 5' leader sequence derived from the 5' terminus of the genomic RNA. The leader template is fused to nascent minus strand mRNA templates via a copy choice-related template switching process called discontinuous extension of minus strands or discontinuous transcription (DT) (47, 48). Copy choice occurs at transcription-regulating sequences (TRS), comprising short conserved sequence motifs that follow the leader and precede each downstream gene's ORF (48–50). In contrast, the subgenomic mRNAs of RoV do not contain leader sequences (24). Synthesis of subgenomic mRNA templates is believed to be mediated solely by attenuation of minus strand synthesis, involving mRNA transcription on minus-stranded replicative intermediates (nondiscontinuous transcription [NDT]) (26). ToV uses DT to express RNA2 and NDT for RNAs 3 through 5, respectively (51).

To investigate the nature of subgenomic mRNAs in CAVV, total RNA was isolated from CAVV-infected cells and subjected to Northern blot analysis. RNA from noninfected C6/36 cells served as a control. To avoid detection of defective interfering RNAs, cells were infected at a low multiplicity of infection (MOI) with virus obtained from limiting dilution endpoints of early-passage supernatants. Northern blot probes were generated against the most 5' 107 nt and the most 3' 556 nt of the genome (Fig. 4B). Additional probes were generated for the major predicted ORFs (Fig. 4B). Apart from a band corresponding to the genome size, fragments of

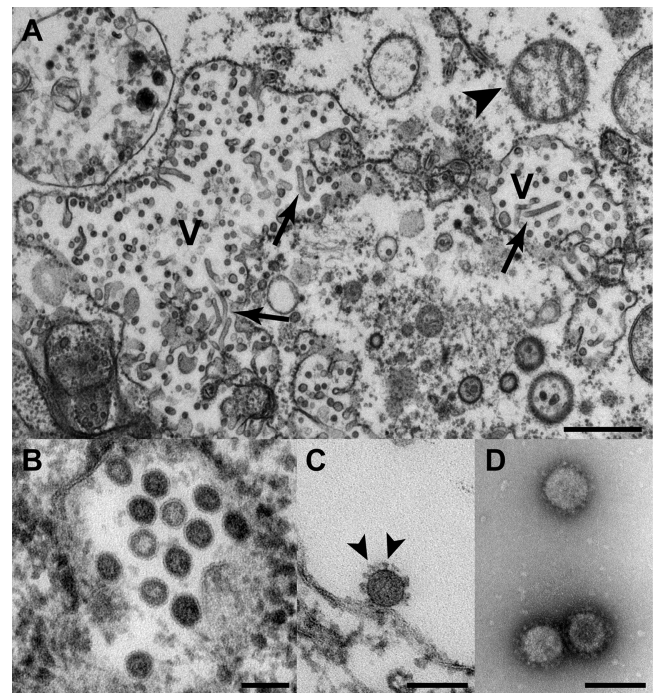


FIG 3 CAVV replication and morphology as observed by transmission electron microscopy. Ultrathin sections of C6/36 cells infected with CAVV at 48 hpi (A to C), showing an overview of the cytoplasm of infected cells (A; bar, 1 μ m; V, vesicle with virus formation; arrowhead, mitochondrion; arrow, tubular structures likely of viral origin) and a higher magnification of vesicles containing spherical, potentially enveloped particles (B; bar, 100 nm) and separation or adsorption of putative virions on cell membranes (C; bar, 100 nm; arrowhead, spikes on virus surface). Negative staining (1% uranyl acetate) of CAVV sedimented by ultracentrifugation through a 36% sucrose cushion (D; bar, 100 nm). It should be noted that better EM results were obtained at 48 hpi than at 24 hpi.

approximately 4.7, 2.7, and 1.8 kb were detected with the 5' probe. These were also represented in a blot with the 3' probe, suggesting that these RNAs are 5' and 3' coterminal with the genome (DT, Fig. 4C). Additional bands of ca. 1.4, 1.2, and 1.0 kb were detected with the 3' probe but not with the 5' probe, compatible with an NDT mechanism.

The ORF2a probe detected the genome and two additional bands, one corresponding to a subgenomic RNA (sgRNA) starting upstream of ORF2a, the other within ORF2a (Fig. 4A). No band corresponding to a separate sgRNA for the predicted ORF2b was observed. A prominent band of 1.8 kb was seen with all probes in and downstream of ORF3a, suggesting an sgRNA starting upstream of ORF3a. A minor band of 1.4 kb was confirmed with the probe placed in ORF3b and all downstream probes. Two additional minor bands of ca. 1.2 and 1.0 kb were seen with the ORF4 probe and the 3'-end probe. According to estimated molecular weights, these corresponded to sgRNAs starting ca. 300 and 500 nt upstream of the initiation codon of ORF4 (₁₉₃₉₉AUG). Functional studies are required to determine these positions more precisely, as size estimates are based on RNA gel electrophoresis only.

To identify potential fusion sites indicative of DT, one-step RT-PCRs were conducted for each sgRNA detected by Northern blotting. Sense primers were placed at intervals starting from the 5' end of the genome approximately 300 nt into the genomic sequence (see Fig. S2 in the supplemental material). Antisense

TABLE 2 Predicted genome organization of CAVV

ORF	Frame	Genome position (nt)	No. of amino acids	Predicted protein size (kDa)	Predicted protein
1a	+3	363–7862	2,500	288	Polyprotein 1a
1b	+2	7826–15628	2,600	301	ORF1b-encoded part of polyprotein 1ab
2a	+2	15650–18352	900	103	Spike protein
2b	+1	15664–16308	214	24	Nucleocapsid protein
3a	+1	18394–18870	158	18	Membrane protein
3b	+2	18746–19096	116	14	Membrane protein
4	+2	19399–19548	50	6	Unknown <i>Continued on following page</i>

primers were selected according to the projected sizes of the sgRNAs. RT-PCRs targeting the 4.7-kb and 2.7-kb bands resulted in products with sense primers up to at least 126 and 152 nt, respectively, from the 5' end of the genome. The 1.8-kb band was associated with PCR products up to 202 nt from the 5' end of the genome. PCR products of various and unexpected sizes were observed for the smaller potential sgRNAs (i.e., all sgRNAs smaller

than the 1.8-kb band in the Northern blot). Representative PCR products for all putative sgRNAs were cloned and sequenced. PCR products of unexpected sizes from small subgenomic mRNAs yielded sequences indicative of misprimed amplification in various positions of the genome, and it was concluded that these subgenomic mRNAs were generated by NDT, lacking a leader consistent with the Northern blotting results.

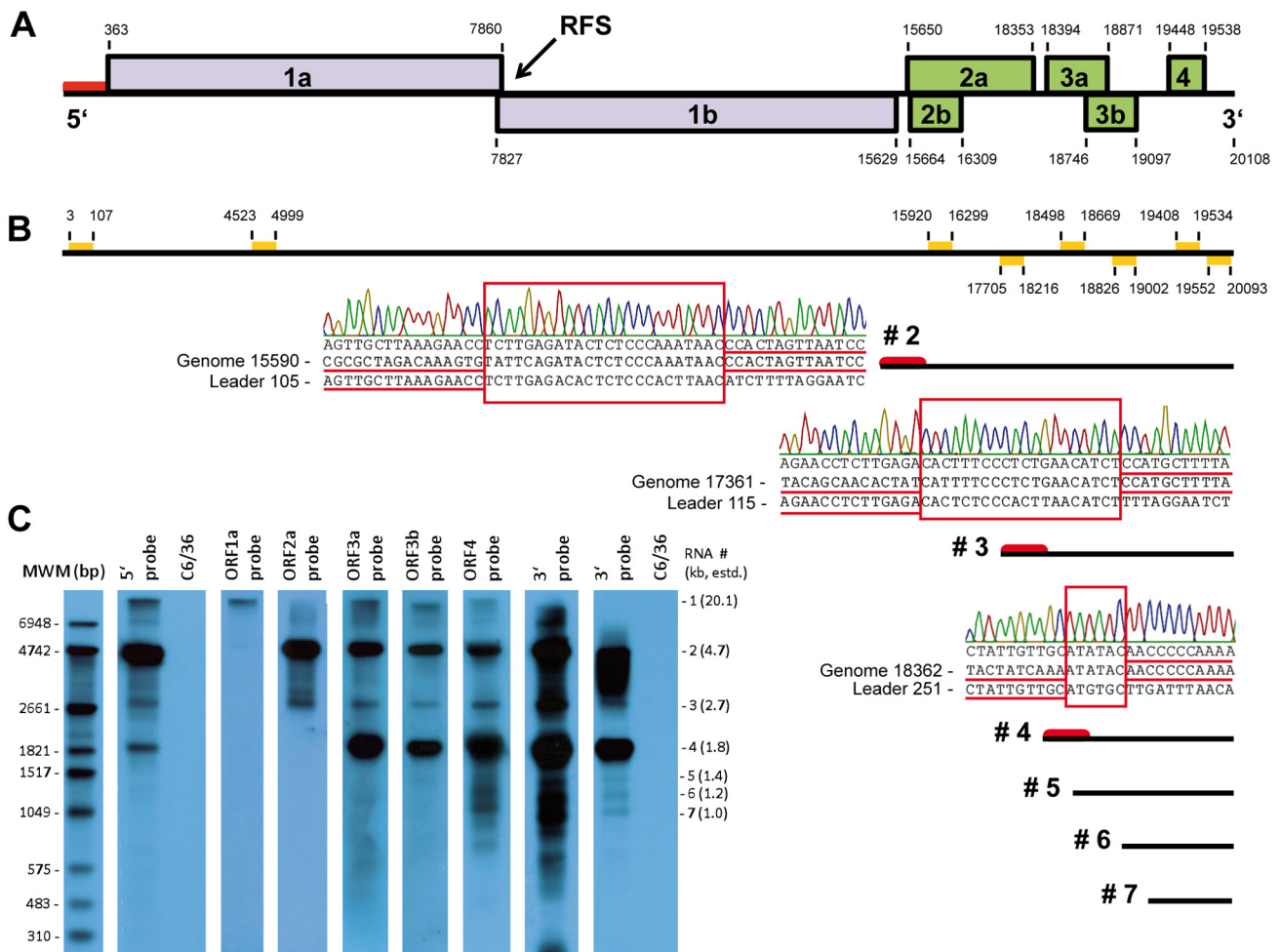


FIG 4 CAVV genome organization and subgenomic mRNA synthesis. (A) Positions and sizes of CAVV ORFs. (B) Positions of probes used for Northern blotting and placement of putative sgRNAs. Putative leader sequences are marked by red bars. Electropherograms shown next to mRNAs 2, 3, and 4 indicate typical leader-body fusion sites identified by RT-PCR. No clear leader-body fusions were identified for RNAs shown without leader symbols (see Fig. S2 in the supplemental material). (C) Detection of CAVV genome and sgRNAs by Northern blot analysis of intracellular viral RNA from infected C6/36 cells. Specific probes for the 5' and 3' prime ends, as well as for each ORF, were employed. The 3'-terminal probe is shown after short and long exposures of the blot. A molecular size marker (MWM) is shown in the left lane. Molecular size indicators on the right summarize estimated sizes of putative subgenomic mRNAs.

The three sgRNAs that were clearly codetected with 5' and 3' probes revealed potential sites of fusions between genome leader and downstream mRNA sequences. Different sites were used for each sgRNA, as shown in Fig. 4B. Nevertheless, all sgRNAs had an A/C-rich region immediately downstream of the fusion region in common. It should be noted that Northern blot detection intensities of the 4.7-kb and 2.7-kb bands corresponded well between the 5' and 3' probes, whereas the detection intensity of the 1.8-kb band was much greater with the 3' probe than with the 5' probe (Fig. 4C). This matched the presence of a much smaller fusion region in the RT-PCR product of the 1.8-kb sgRNA. Given the high intensity of the 1.8-kb band, we suspected that a major fraction of the total amount of this sgRNA would not contain a fused leader element; nonetheless, several variant fusion sites were detected in parallel clones for all subgenomic mRNAs. Further studies are required to characterize in more detail the various CAVV RNA species and their functional relevance in CAVV genome expression.

Initial predictions on structural protein genes. Sequence analyses of proteins predicted to be expressed from the 5 major ORFs in the 3'-proximal region of the CAVV genome revealed little (if any) similarity with other viral (and cellular) proteins, confirming that CAVV diverged profoundly from other nidoviruses and complicating functional assignments of these proteins. As summarized in Table 2, two proteins are expected to be expressed from sgRNA 2. ORF2a is predicted to encode a type I glycoprotein featuring a C-terminal membrane-spanning domain and multiple glycosylation sites. Based on these predictions, the protein likely represents a functional equivalent of the S protein of other nidoviruses, which remains to be confirmed in further studies. The predicted translation start codon of ORF2b is the second AUG on the subgenomic mRNA2, located just downstream of the ORF2a start codon, suggesting that ORF2b may be translated by a leaky scanning mechanism. ORF2b encodes a highly basic protein (pK_a 10.8) with a molecular mass of 24 kDa. Both its size and its charge suggest that this protein may be the viral nucleocapsid protein. Among nidoviruses, this upstream position of the (presumed) N protein gene in the CAVV genome is unusual but has its precedent in members of the family *Roniviridae* (52). ORF3a and -3b encode proteins with predicted molecular masses of 18 and 14 kDa, respectively. Database searches failed to reveal close homologs of these proteins. Protein analysis software predicts the presence of membrane-spanning domains in both proteins (residues 95 to 117 in ORF3a and residues 73 to 95 in ORF3b), suggesting that both proteins are integral membrane proteins. The ORF3a protein likely contains a signal peptidase cleavage site, $_{15}\text{Ala-Met-Ser|Ala-Glu}$, and is predicted to be glycosylated, further supporting a role as a membrane-spanning structural protein of the virus. The specific mechanism used to express the ORF3b gene product is unclear but may involve internal ribosomal entry, as shown previously for several downstream ORFs expressed from coronavirus subgenomic mRNAs (53–56). Further studies are required to establish if ORF3a and ORF3b proteins have functions related to those of the membrane-spanning M and E proteins of other nidoviruses. ORF4 encodes a small protein of 50 amino acid residues with unknown functions.

Phylogenetic analyses reveal a novel cluster of insect nidoviruses. In order to identify the phylogenetic relationship of CAVV with other nidoviruses, we conducted phylogenetic analyses of protein alignments of conserved motifs within 3CL^{pro} (ORF1a),

RdRp, and HEL (ORF1b), as well as the putative structural S protein (Fig. 5). In all phylogenies, CAVV branched from a deep node in the nidovirus tree above the *Roniviridae* family. In the ORF1b and S genes, CAVV had a most recent ancestor in common with the subfamilies *Torovirinae* and *Coronavirinae*. In 3CL^{pro}, it branched from an ancestor it has in common with the subfamily *Torovirinae*. The low bootstrap support at this root point suggested that CAVV might also branch from an ancestor it has in common with *Torovirinae* and *Coronavirinae*. Although formal classification criteria for nidoviruses are not established, this phylogenetic positioning suggested CAVV to be a profoundly separated cluster of nidoviruses that might constitute a new family of nidoviruses with a phylogenetic position between *Roniviridae* and *Coronaviridae* on one side and *Arteriviridae* on the other.

CONCLUSIONS

CAVV is the first mosquito nidovirus and represents the prototype species of a family in the order *Nidovirales* that includes features distinct from those established for the *Arteriviridae*, *Roniviridae*, and *Coronaviridae*. Based on morphology, conserved genome motifs, and phylogenetic relationship, CAVV cannot be assigned to one of the established nidovirus families. Further investigations are required to elucidate further details of the CAVV replication apparatus and structural protein functions. It is unknown whether CAVV infection is restricted to mosquitoes or if transmission to other hosts, potentially vertebrates, occurs. It is interesting that CoV has not been detected in insects, but that the typical reservoir hosts—bats for alpha- and beta-CoV and birds for gamma-CoV—are largely insect feeding. Common ancestors of CAVV and CoV may thus have been insect borne and have diverged after independent host switches to bats and birds. This is in contrast to earlier proposals suggesting acquisition of gamma-CoV by birds from bats via raptors (57) but in agreement with hypotheses that emphasize CoV phylogeny and ecological considerations (58). Moreover, it has been suspected from epidemiological observations that a link between ToV and insects may exist (59). Even though it has not been confirmed that these viruses are carried by insects, the epidemiological implications of insects suggest at least ecologically relevant contact between the virus and its host. An ancestral existence of nidoviruses in arthropods is also supported by the phylogeny of the *Nidovirales* CAVV included; phylogenetically basal RoV is hosted by crustaceans, which, like mosquitoes, belong to the phylum *Arthropoda*. An arthropod host at the root of the *Nidovirales* tree would provide a parsimonious explanation for host associations of several of the known *Nidovirales*.

Our data on CAVV prevalence and divergence demonstrate how a virus may evolve through emergence from a pristine rain-forest habitat into surrounding areas of less host biodiversity due to anthropogenic modification. Critically, while extending out of the primary forest habitat, the virus seemed to narrow its genetic diversity while increasing in prevalence (reverse dilution effect [13, 60, 61]). Further investigations are needed to untangle if the higher prevalence of CAVV in human settlements relates to a higher density of hosts (insects or vertebrates) or virus adaptation. This principal understanding is necessary to develop experimental ecology models of virus-vector dynamics. This study underlines the importance of linking ecosystem biology and virus ecology to unravel the role of ecosystem modifications in the emergence of novel pathogens.

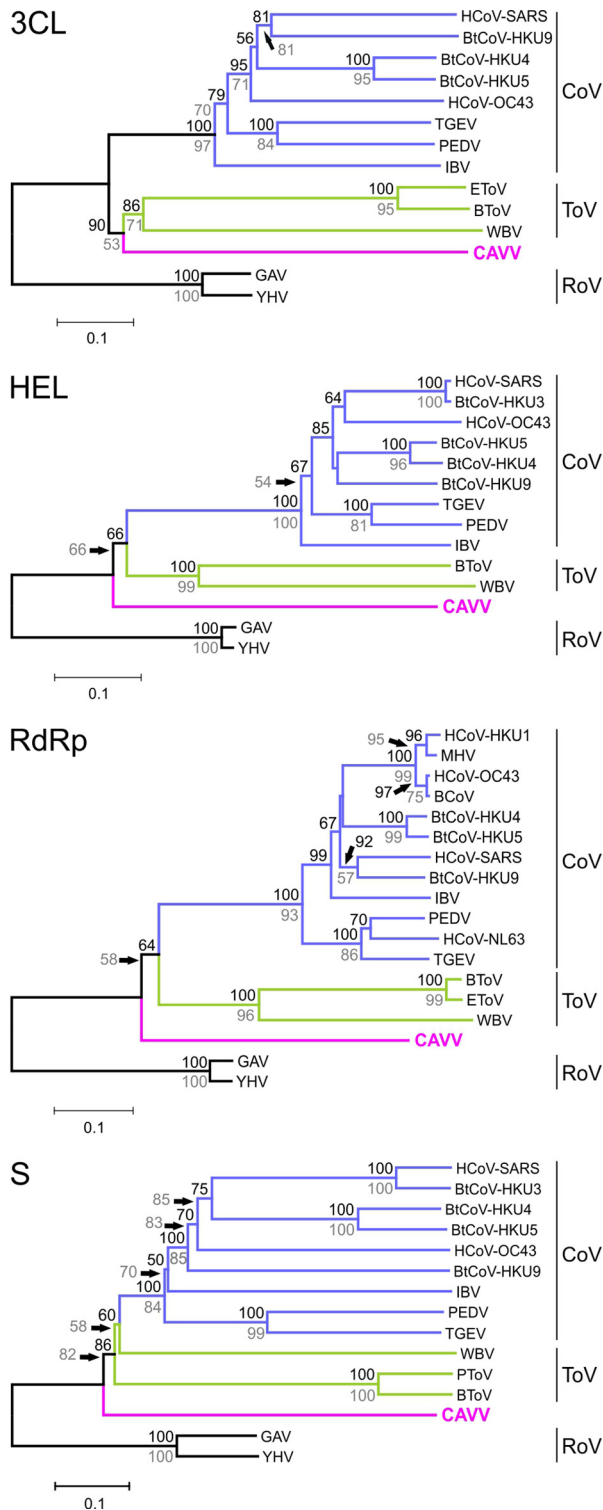


FIG 5 Phylogenetic relationship of CAVV to prototype nidovirus strains for selected genomic regions. Phylogenetic analyses were performed using the NJ algorithm, a BLOSUM62 substitution matrix, and no distance correction. Indels were fully deleted. Significance was tested by bootstrap analysis using 1,000 resampling steps as implemented in MEGA 5.0 (66). Bootstrap values are shown above nodes. Analyses were also performed using the maximum-likelihood algorithm in FastML with the same settings. Results of this analysis are not shown because of congruent topologies. Bootstrap support values (1,000 replicates) from maximum-likelihood analysis are shown in grey below nodes.

MATERIALS AND METHODS

Virus isolation and purification. Virus isolation from 432 pools of 4,839 female mosquito heads was done with *Aedes albopictus* (C6/36) cells as described previously (15). For virus growth kinetics, C6/36 cells were infected with an MOI of 0.1, 0.01, or 0.001 and incubated for 1 h at 28°C. The inoculum was removed, and cells were washed with phosphate-buffered saline (PBS). L-15 medium was added, and cells were incubated for 48 h. Every 3 h, an aliquot of the cell culture supernatant was removed, RNA was extracted using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany), cDNA was synthesized using the SuperScript III RT System (Invitrogen, Karlsruhe, Germany), and CAVV viral genome copy numbers were quantified by real-time RT-PCR (14).

To obtain pure virus stocks, CAVV was titrated on insect cells and cell culture supernatant was harvested at 22 hpi to allow only single-round infections. Numbers of genome copies were measured by real-time RT-PCR, and the infectious supernatant of the highest dilution still showing virus replication was used for retitration of CAVV on insect cells. This procedure was repeated five times.

For purification, CAVV was harvested by freeze-thawing of infected cells. Cell debris was removed by centrifugation at 3,000 rpm for 20 min, followed by ultracentrifugation through a 36% sucrose cushion at 35,000 rpm (SW40 rotor; Beckman) for 2 h at 4°C. The virus pellet was suspended in 150 μ l PBS overnight at 4°C. Further purification was achieved on a continuous gradient of 1 to 2 M sucrose in 0.01 M Tris-HCl–4 mM Na-EDTA at 35,000 rpm (SW40 rotor; Beckman) for 22 h at 4°C. Virus-containing fractions were tested by real-time RT-PCR, and fractions with the highest sequence titers were concentrated through a 36% sucrose cushion at 35,000 rpm (SW40 rotor; Beckman) for 2 h at 4°C. The virus pellet was suspended in 150 μ l PBS buffer overnight at 4°C.

Electron microscopy. For electron microscopy, viral particles were purified through a cushion of 36% sucrose and the pellet was suspended in PBS (15, 16). Viral particles were fixed with 2% paraformaldehyde and analyzed by transmission electron microscopy after negative staining with 1% uranyl acetate (62, 63). For ultrathin sections, infected cells were fixed with 2.5% glutaraldehyde, enclosed in low-melting-point agar, embedded in resin, and evaluated by transmission electron microscopy after ultrathin sectioning.

Genome sequencing. RNA extracted from purified virus preparations was used for unbiased high-throughput sequencing and for conventional sequencing approaches. Genome fragments were generated by adaptor-based random RT-PCR. Following protocols described previously (15, 64), in this specific study, random hexamers linked to a defined primer sequence tail or oligonucleotides that bind to the conserved TRS elements of CoV linked to an oligonucleotide anchor were applied (see Table S1 in the supplemental material).

Genome characterization and phylogenetic analyses. The nucleotide sequence of the CAVV genome was analyzed for ORFs and translated. Nucleotide and amino acid sequences were compared with other sequences by BLASTn, BLASTx, tBLASTx, and psiBLAST with the GenBank database (<http://www.ncbi.nlm.nih.gov/Genbank>), and protein motifs were identified by web-based comparison to the Pfam database (<http://www.pfam.janelia.org>). Identification of cleavage sites for signal peptides was accomplished by using signalP-NN (<http://www.cbs.dtu.dk/services/SignalP>). Prediction of the hydropathy profile was performed by TMHMM v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>), and N-linked glycosylation sites were identified using the NetNGlyc 1.0 server (<http://www.cbs.dtu.dk/services/NetNGlyc>). RNA folding was modeled by using the Mfold server (<http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi>) (65). For phylogenetic analysis, CAVV amino acid sequences were aligned with representative sequences of other nidoviruses in MEGA v5.0 (66). Alignments were optimized according to published crystal structure predictions. Phylogenetic analysis of amino acid sequences was conducted by the neighbor-joining (NJ) algorithm with the BLOSUM62 substitution matrix for distance correction with 1,000 bootstrap replicates in MEGA v5.0. Maximum-likelihood analyses were done with Fasttree

(67), and tree files were displayed in MEGA versus 5.0. Evolutionary divergence over sequence pairs was estimated in MEGA versus 5.0.

Identification of subgenomic mRNAs. For Northern blotting, the Northern Blot Starter Kit (Roche, Mannheim, Germany) was used. Digoxigenin (DIG)-labeled probes were generated by PCR using the primers shown in Fig. 4B and listed in Table S1 in the supplemental material. Total RNA of isolate CAVV/C79/CI/2004 was extracted with the Qiagen RNeasy Kit (Qiagen, Hilden, Germany) from C6/36 cells at 24 hpi. RNA was separated on a 2% formaldehyde–1.5% agarose gel, blotted onto a nylon membrane (Roche, Mannheim, Germany), and hybridized with the CAVV-specific, DIG-labeled probes. RNAs were analyzed by chemiluminescence using 1:10,000 anti-DIG–alkaline phosphatase Fab fragments and 1:100 CDP-Star reagent (Roche, Mannheim, Germany).

ACKNOWLEDGMENTS

We thank G. Gagné for field assistance and the Ivorian authorities for long-term support, especially the Ministry of the Environment and Forest, as well as the Ministry of Research, the directorship of the Taï National Park, and the Swiss Research Center in Abidjan. We are grateful for logistic support by the Taï chimpanzee project and C. Boesch. We thank A. Kopp, M. Marklewitz, A. Tashmukhamedova, and A. Petrosov for technical assistance.

This project was supported by the Robert Koch Institute and the European Union DG Research through the programs EMPERIE (grant agreement 223498) and EVA (grant agreement 228292), by the Deutsche Forschungsgemeinschaft (grant agreement DR772/3-1), and by the BonFor program of the University of Bonn (grant agreement O-156-0006). Work at the Center for Infection and Immunity at Columbia University was supported by the National Institutes of Health (AI079231, AI079231), the United States Department of Defense, and the United States Agency for International Development (USAID) Emerging Pandemic Threats (EPT) Program, PREDICT project, under the terms of cooperative agreement GHN-A-OO-09-00010-00.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00077-11/-/DCSupplemental>.

Figure S1, PDF file, 0.420 MB.

Figure S2, JPG file, 0.315 MB.

Table S1, PDF file, 0.204 MB.

REFERENCES

- Boesch C, Boesch-Achermann H. 2000. The chimpanzees of the Tai forest: behavioural ecology and evolution. Oxford University Press, Oxford, United Kingdom.
- Myers N. 1990. The biodiversity challenge: expanded hot-spots analysis. *Environmentalist* 10:243–256.
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853–858.
- Chapman CA, Balcomb SR, Gillespie TR, Skorupa JP, Struhsaker TT. 2000. Long-term effects of logging on African primate communities: a 28-year comparison from Kibale National Park, Uganda. *Conserv. Biol.* 14:207–217.
- Hall JS, Harris DJ, Medjibe V, Ashton PMS. 2003. The effects of selective logging on forest structure and tree species composition in a Central African forest: implications for management of conservation areas. *Forest Ecol. Manag.* 183:249–264.
- Van Gernerden BS, Oliff H, Parren MPE, Bongers F. 2003. The pristine rain forest? Remnants of historical human impacts on current tree species composition and diversity. *J. Biogeogr.* 30:1381–1390.
- Doherty PF, Grubb TC. 2002. Survivorship of permanent-resident birds in a fragmented forested landscape. *Ecology* 83:844–857.
- Gascon C, Williamson GB, da Fonseca GA. 2000. Ecology. Receding forest edges and vanishing reserves. *Science* 288:1356–1358.
- Meyer CFJ, Fründ J, Lizano WP, Kalko EKV. 2008. Ecological correlates of vulnerability to fragmentation in neotropical bats. *J. Appl. Ecol.* 45:381–391.
- Morgan K, et al. 2009. Molecular phylogenetics and biogeography of the Neocellia Series of *Anopheles* mosquitoes in the Oriental Region. *Mol. Phylogenet. Evol.* 52:588–601.
- Thomas CD. 2000. Dispersal and extinction in fragmented landscapes. *Proc. Biol. Sci.* 267:139–145.
- Wade TG, Riitters KH, Wickham JD, Jones KB. 2003. Distribution and causes of global forest fragmentation. *Conserv. Ecol.* 7:7.
- Keesing F, et al. 2010. Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* 468:647–652.
- Junglen S, et al. 2009. Examining landscape factors influencing relative distribution of mosquito genera and frequency of virus infection. *Ecohealth* 6:239–249.
- Junglen S, et al. 2009. A new flavivirus and a new vector: characterization of a novel flavivirus isolated from *Uranotaenia* mosquitoes from a tropical rain forest. *J. Virol.* 83:4462–4468.
- Quan PL, et al. 2010. Moussa virus: a new member of the Rhabdoviridae family isolated from *Culex decens* mosquitoes in Cote d'Ivoire. *Virus Res.* 147:17–24.
- Siddell SG, Ziebuhr J, Snijder EJ. 2005. Coronaviruses, toroviruses and arteriviruses, 10th ed. Hodder Arnold, London, United Kingdom.
- Cowley JA, Dimmock CM, Spann KM, Walker PJ. 2000. Gill-associated virus of *Penaeus monodon* prawns: an invertebrate virus with ORF1a and ORF1b genes related to arteri- and coronaviruses. *J. Gen. Virol.* 81:1473–1484.
- Schütze H, et al. 2006. Characterization of *White bream virus* reveals a novel genetic cluster of nidoviruses. *J. Virol.* 80:11598–11609.
- Beaudette FR, Hudson CB. 1937. Cultivation of the virus of infectious bronchitis. *J. Am. Vet. Med. Assoc.* 90:51–58.
- Poon LL, et al. 2005. Identification of a novel coronavirus in bats. *J. Virol.* 79:2001–2009.
- Tang XC, et al. 2006. Prevalence and genetic diversity of coronaviruses in bats from China. *J. Virol.* 80:7481–7490.
- Woo PC, Lau SK, Yuen KY. 2006. Infectious diseases emerging from Chinese wet-markets: zoonotic origins of severe respiratory viral infections. *Curr. Opin. Infect. Dis.* 19:401–407.
- Cowley JA, Dimmock CM, Walker PJ. 2002. Gill-associated nidovirus of *Penaeus monodon* prawns transcribes 3'-coterminal subgenomic mRNAs that do not possess 5'-leader sequences. *J. Gen. Virol.* 83:927–935.
- Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. 2006. Nidovirales: evolving the largest RNA virus genome. *Virus Res.* 117:17–37.
- Pasternak AO, Spaan WJ, Snijder EJ. 2006. Nidovirus transcription: how to make sense...? *J. Gen. Virol.* 87:1403–1421.
- Sittidilokratna N, Dangtip S, Cowley JA, Walker PJ. 2008. RNA transcription analysis and completion of the genome sequence of yellow head nidovirus. *Virus Res.* 136:157–165.
- Gosert R, Kanjanahaluethai A, Egger D, Bienz K, Baker SC. 2002. RNA replication of mouse hepatitis virus takes place at double-membrane vesicles. *J. Virol.* 76:3697–3708.
- Knoops K, et al. 2008. SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol.* 6:e226.
- Schlegel A, Giddings TH, Jr, Ladinsky MS, Kirkegaard K. 1996. Cellular origin and ultrastructure of membranes induced during poliovirus infection. *J. Virol.* 70:6576–6588.
- Ziebuhr J, Snijder EJ, Gorbalenya AE. 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* 81:853–879.
- Brierley I, Dos Ramos FJ. 2006. Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res.* 119:29–42.
- Namy O, Moran SJ, Stuart DI, Gilbert RJ, Brierley I. 2006. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* 441:244–247.
- Cheng A, et al. 2005. Expression, purification, and characterization of SARS coronavirus RNA polymerase. *Virology* 335:165–176.
- Bhardwaj K, Guarino L, Kao CC. 2004. The severe acute respiratory syndrome coronavirus Nsp15 protein is an endoribonuclease that prefers manganese as a cofactor. *J. Virol.* 78:12218–12224.
- Ivanov KA, et al. 2004. Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc. Natl. Acad. Sci. U. S. A.* 101:12694–12699.
- Minskaia E, et al. 2006. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 103:5108–5113.
- Snijder EJ, et al. 2003. Unique and conserved features of genome and

- proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331:991–1004.
39. von Grothuss M, Wyrwicz LS, Rychlewski L. 2003. mRNA cap-1 methyltransferase in the SARS genome. *Cell* 113:701–702.
 40. Putics A, Filipowicz W, Hall J, Gorbalenya AE, Ziebuhr J. 2005. ADP-ribose-1"-monophosphatase: a conserved coronavirus enzyme that is dispensable for viral replication in tissue culture. *J. Virol.* 79:12721–12731.
 41. Saikatendu KS, et al. 2005. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1"-phosphate dephosphorylation by a conserved domain of nsP3. *Structure* 13:1665–1675.
 42. Brierley I. 1995. Ribosomal frameshifting viral RNAs. *J. Gen. Virol.* 76(Pt. 8):1885–1892.
 43. Brierley I, Jenner AJ, Inglis SC. 1992. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* 227:463–479.
 44. Jacks T, Madhani HD, Masiarz FR, Varmus HE. 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* 55:447–458.
 45. Kim KH, Lommel SA. 1994. Identification and analysis of the site of –1 ribosomal frameshifting in red clover necrotic mosaic virus. *Virology* 200: 574–582.
 46. Kim KH, Lommel SA. 1998. Sequence element required for efficient –1 ribosomal frameshifting in red clover necrotic mosaic dianthovirus. *Virology* 250:50–59.
 47. Sawicki SG, Sawicki DL. 1995. Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. *Adv. Exp. Med. Biol.* 380:499–506.
 48. van der Most RG, Luytjes W, Rutjes S, Spaan WJ. 1995. Translation but not the encoded sequence is essential for the efficient propagation of the defective interfering RNAs of the coronavirus mouse hepatitis virus. *J. Virol.* 69:3744–3751.
 49. Lai MM, Cavanagh D. 1997. The molecular biology of coronaviruses. *Adv. Virus Res.* 48:1–100.
 50. Snijder EJ, Meulenberg JJ. 1998. The molecular biology of arteriviruses. *J. Gen. Virol.* 79(Pt. 5):961–979.
 51. van Vliet AL, Smits SL, Rottier PJ, de Groot RJ. 2002. Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. *EMBO J.* 21:6571–6580.
 52. Cowley JA, Cadogan LC, Spann KM, Sittidilokratna N, Walker PJ. 2004. The gene encoding the nucleocapsid protein of Gill-associated nidovirus of *Penaeus monodon* prawns is located upstream of the glycoprotein gene. *J. Virol.* 78:8935–8941.
 53. Jendrach M, Thiel V, Siddell S. 1999. Characterization of an internal ribosome entry site within mRNA 5' of murine hepatitis virus. *Arch. Virol.* 144:921–933.
 54. Liu DX, Inglis SC. 1992. Internal entry of ribosomes on a tricistronic mRNA encoded by infectious bronchitis virus. *J. Virol.* 66:6143–6154.
 55. O'Connor JB, Brian DA. 2000. Downstream ribosomal entry for translation of coronavirus TGEV gene 3b. *Virology* 269:172–182.
 56. Thiel V, Siddell SG. 1994. Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5'. *J. Gen. Virol.* 75(Pt. 11):3041–3046.
 57. Vijaykrishna D, et al. 2007. Evolutionary insights into the ecology of coronaviruses. *J. Virol.* 81:4012–4020.
 58. Woo PC, Lau SK, Huang Y, Yuen KY. 2009. Coronavirus diversity, phylogeny and interspecies jumping. *Exp. Biol. Med.* (Maywood) 234: 1117–1127.
 59. Pignatelli J, et al. 2009. Molecular characterization of a new PTOV strain. Evolutionary implications. *Virus Res.* 143:33–43.
 60. Johnson PT, Thielges DW. 2010. Diversity, decoys and the dilution effect: how ecological communities affect disease risk. *J. Exp. Biol.* 213: 961–970.
 61. Ostfeld RS. 2009. Biodiversity loss and the rise of zoonotic pathogens. *Clin. Microbiol. Infect.* 15(Suppl. 1):40–43.
 62. Biel SS, Gelderblom HR. 1999. Electron microscopy of viruses. Oxford University Press, Oxford, United Kingdom.
 63. Hayat MA. 2000. Principles and techniques of electron microscopy: biological applications, 4th ed. Cambridge University Press, Cambridge, United Kingdom.
 64. Stang A, Korn K, Wildner O, Uberla K. 2005. Characterization of virus isolates by particle-associated nucleic acid PCR. *J. Clin. Microbiol.* 43: 716–720.
 65. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406–3415.
 66. Tamura K, et al. 4 May 2011, posting date. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 10.1093/molbev/msr121.
 67. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490.