JOURNAL OF GENERAL VIROLOGY

RESEARCH ARTICLE

Orba *et al.*, *Journal of General Virology* 2021;102:001518
DOI 10.1099/jgv.0.001518

Diverse mosquito-specific flaviviruses in the Bolivian Amazon basin

Yasuko Orba[1,2,*], Keita Matsuno[2,3], Ryo Nakao[4], Kirill Kryukov[5], Yumi Saito[1], Fumihiko Kawamori[6], Ariel Loza Vega[6], Tokiko Watanabe[7], Tadashi Maemura[8], Michihito Sasaki[1], William W. Hall[2,9,10], Roy A. Hall[11], Juan Antonio Pereira[6], So Nakagawa[12] and Hirofumi Sawa[1,2,10]

Abstract

The genus *Flavivirus* includes a range of mosquito-specific viruses in addition to well-known medically important arboviruses. Isolation and comprehensive genomic analyses of viruses in mosquitoes collected in Bolivia resulted in the identification of three novel flavivirus species. Psorophora flavivirus (PSFV) was isolated from *Psorophora albigenu*. The coding sequence of the PSFV polyprotein shares 60% identity with that of the *Aedes*-associated lineage II insect-specific flavivirus (ISF), Marisma virus. Isolated PSFV replicates in both *Aedes albopictus*- and *Aedes aegypti*-derived cells, but not in mammalian Vero or BHK-21 cell lines. Two other flaviviruses, Ochlerotatus scapularis flavivirus (OSFV) and Mansonia flavivirus (MAFV), which were identified from *Ochlerotatus scapularis* and *Mansonia titillans,* respectively, group with the classical lineage I ISFs. The protein coding sequences of these viruses share only 60 and 40% identity with the most closely related of known lineage I ISFs, including Xishuangbanna aedes flavivirus and Sabethes flavivirus, respectively. Phylogenetic analysis suggests that MAFV is clearly distinct from the groups of the current known *Culicinae*-associated lineage I ISFs. Interestingly, the predicted amino acid sequence of the MAFV capsid protein is approximately two times longer than that of any of the other known flaviviruses. Our results indicate that flaviviruses with distinct features can be found at the edge of the Bolivian Amazon basin at sites that are also home to dense populations of human-biting mosquitoes.

INTRODUCTION

Major *Aedes*-borne flaviviruses, including dengue virus (DENV) 1–4 and Zika virus (ZIKV), circulate in the Bolivian lowland areas but not in the highlands. In lowland tropical regions, these viruses typically circulate during the rainy season from November to April [1, 2]. A human case of Ilheus virus (ILHV) infection was also identified in Magdalena, in the Beni Department in northern Bolivia in 2005 [3]. ILHV is also a mosquito-borne flavivirus that has been detected in and isolated from *Aedes*, *Psorophora*, *Ochlerotatus* and *Culex* species of mosquitoes collected in the Amazon basin regions of both Peru and Brazil [4–6]. Alphaviruses, including Chikungunya virus (CHIKV), Mayaro virus (MAYV) and equine encephalitis viruses (EEVs), have also spread throughout Latin America. CHIKV has been specifically associated with large outbreaks of disease in the tropical areas of Bolivia in 2016 [7].

Received 23 August 2020; Accepted 19 October 2020; Published 08 January 2021
Author affiliations: [1]Division of Molecular Pathobiology, Research Center for Zoonosis Control, Hokkaido University, Sapporo, Japan; [2]International Collaboration Unit, Research Center for Zoonosis Control, Hokkaido University, Sapporo, Japan; [3]Unit of Risk Analysis and Management, Research Center for Zoonosis Control, Hokkaido University, Sapporo, Japan; [4]Laboratory of Parasitology, Faculty of Veterinary Medicine, Hokkaido University, Sapporo, Japan; [5]Department of Genomics and Evolutionary Biology, National Institute of Genetics, Shizuoka, Japan; [6]Faculty of Veterinary Sciences, Gabriel Rene Moreno Autonomous University, Santa Cruz, Bolivia; [7]Department of Molecular Virology, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan; [8]Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin, Madison, Wisconsin, USA; [9]Centre for Research in Infectious Diseases, University College Dublin, Dublin, Ireland; [10]Global Virus Network, Baltimore, Maryland, USA; [11]Australian Infectious Diseases Research Centre, School of Chemistry and Molecular Biosciences, University of Queensland, St Lucia, Queensland, Australia; [12]Department of Molecular Life Science, Tokai University School of Medicine, Kanagawa, Japan.
***Correspondence:** Yasuko Orba, orbay@czc.hokudai.ac.jp
Keywords: Amazon; Bolivia; *Flavivirus*; insect-specific flavivirus; mosquito.
Abbreviations: ISF, Insect-specific flavivirus; MAFV, Mansonia flavivirus; MBFV, mosquito-borne flavivirus; OSFV, Ochlerotatus scapularis flavivirus; PSFV, Psorophora flavivirus.
Repositories: LC567151, LC567152, LC567153.
Two supplementary figures are available with the online version of this article.

Cell fusing agent virus was the first identified insect-specific flavivirus (ISF) and this was originally identified in cultures of *Aedes aegypti* cells [8]. Subsequently, a large number of ISFs have been identified as a result of advances in the metagenomic analysis of field-collected mosquitos. ISFs are believed to replicate exclusively in insect cells and not in vertebrate cells. The lineage I ISFs are phylogenetically distinct from the mosquito- and tick-borne flavivirus pathogens that infect vertebrate hosts; the lineage II ISFs (dual-host affiliated ISFs) constitute clades within the group of pathogenic mosquito-borne flaviviruses (MBFVs) [9]. Replication of lineage II ISFs is also restricted to mosquito cell and these viruses have never been identified in vertebrates. However, it is possible that as yet uncharacterized natural life cycles may exist in which some lineage II ISFs are transmitted from mosquitoes to vertebrate animals with no associated pathogenicity.

In this study, mosquitoes collected in Bolivian lowland areas, including a forested area of the Amazon basin, were investigated in order to identify potential novel flaviviruses and/or alphaviruses.

METHODS

Mosquito collection

Mosquito collections were carried out in a forested area of Trinidad, the capital of Beni Department (14°43′10″S 64°56′45″W) in October 2018 and August 2019, and in Buena Vista, a town in the Santa Cruz Department at the north side of the Amboro National Park (17°22′26′S 63°39′40″W) in October 2018. Mosquitoes were trapped beginning in the afternoon hours until the following morning using Centers for Disease Control and Prevention (CDC) light traps (John W. Hock Co., Gainesville, FL, USA) with CO_2 produced by yeast fermentation, and BG-sentinel traps (Biogents AG, Regensburg, Germany). Hand-nets were used in the morning and early evening collections. Sampling was carried out for one or two nights in each location. Collected mosquitoes were killed by freezing. After species identification based on morphology with reference to identification keys for mosquitoes, 1 to as many as 40 female mosquitoes from each species were pooled [10, 11]. Molecular identification of each species was confirmed by polymerase chain reaction (PCR) amplification and sequencing of the cytochrome oxidase I (COI) gene [12].

Detection of flavivirus and alphavirus genes by reverse-transcription PCR (RT-PCR)

Mosquitoes were immersed in minimum essential medium containing 2% foetal bovine serum supplemented with an antibiotic antimycotic solution (penicillin, streptomycin and amphotericin B) and homogenized using a BioMasher (Nippi, Tokyo, Japan). Aliquots of supernatants from the mosquito homogenates (100 µl) were used for RNA extraction using the Direct-Zol kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. The remaining portions of the supernatants were filtered and inoculated onto cells for virus propagation. Pan-flavivirus and pan-alphavirus RT-PCR assays were performed with a PrimeScript One-step RT-PCR kit v.2 (Takara, Shiga, Japan) and degenerate primer sets, including Flavi all S and DEN4 and Flavi all AS2 [13, 14] for identification of flaviviruses, and nsP4-6692F (5′-CAYACRYT RTTYGAYATGTCDGC-3′) and nsP4-7152R (5′-GCRT CDATKATYTTBACYTCCAT-3′) for alphaviruses [15]. The cycling protocol included 30 min of incubation at 50 °C for cDNA synthesis, followed by 2 min of incubation at 94 °C, and 43 cycles each of 94 °C for 30 s, 53 °C (for flaviviruses) or 52 °C (for alphaviruses) for 30 s and 72 °C for 30 s, followed by 72 °C for 5 min. The amplification products were sequenced using a BigDye Terminator v3.0 Cycle Sequencing kit on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

Virus isolation

Viruses were isolated from mosquito homogenates by inoculating C6/36 (*Aedes albopictus*) cells, Vero (African green monkey kidney) cells and BHK-21 (baby hamster kidney) cells in minimum essential medium supplemented with 2% FBS, penicillin, streptomycin, gentamycin and 2 mM L-glutamine. Cultures were maintained in a 5% CO_2 atmosphere at 28 °C for C6/36 cells or at 37 °C for Vero and BHK-21 cells. All virus isolation studies were performed in biosafety level 3 facilities at the Research Center for Zoonosis Control in Sapporo, Japan. Isolation of PSFV by a limiting dilution culture method was confirmed by RT-PCR and RNA sequencing using the supernatants of PSFV-infected cells.

Virus genome sequencing

Whole-genome sequences of flaviviruses were determined by Illumina dye sequencing and rapid amplification of cDNA ends (RACE) analyses. Double-stranded cDNAs were transcribed from RNA extracted from virus isolates or from the mosquito homogenates. These were used for library preparation using Nextera XT DNA Library Prep (Illumina, San Diego, CA, USA), followed by sequencing with a MiSeq Reagent kit v3 (600 cycles) and an Illumina MiSeq System (Illumina). Programs used for the *de novo* assembly of virus genomes included the CLC Genomics Workbench 10 (CLC bio, Qiagen, Hilden, Germany), SPADes [16] and Trinity [17]. Virus-associated contigs longer than 500 nucleotides were identified by BLASTN searches against viral genomes in the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/genome/viruses/).

The 5′- and 3′-sequences at the ends of the flavivirus RNA genomes were determined using RACE analyses; these studies were performed using a 5′-Full RACE Core Set (Takara) or SMARTer RACE 5′/3′ kit (Takara) according the manufacturer's instructions. A poly-A tail was ligated to the isolated RNA using *Escherichia coli* poly (A) polymerase prior to cDNA synthesis in order to amplify 3′-ends using the SMARTer RACE system. Identified terminal sequences were confirmed by Sanger sequencing. Unfortunately, we were unsuccessful in the amplification of the 3′-end of the MAFV or the 3′- or 5′-ends of OSFV untranslated region (UTR) sequences using these RACE systems.

Table 1. Mosquito species analysed in this study

Season	Place	Species	No. of mosquitoes	No. of pools	Alphavirus (+) pool	Flavivirus (+) pool	Identified virus
Oct. 2018	Trinidad	*Psorophora albigenu*	1779	47	0	1*	Psorophora flavivirus (PSFV) bvmq18-51
		Ochlerotatus scapularis	188	6	0	1*	Ochlerotatus scapularis flavivirus (OSFV) bvmq18-25
		Culex tatoi	141	4	0	0	
		Mansonia titillans	25	1	0	1†	Mansonia flavivirus (MAFV) bvmq18-1
		Coquillettidia nigricans	5	1	0	0	
		Others	2	2	0	0	
	Buena Vista /Amboro NP	*Psorophora albigenu*	2	2	0	0	
		Ochlerotatus scapularis	3	1	0	0	
		Culex quinquefasciatus	57	3	0	0	
		Others	2	2	0	0	
Aug. 2019	Trinidad	*Psorophora* sp.	8	3	0	0	
		Ochlerotatus serratus	111	4	0	0	
		Culex quinquefasciatus	24	2	0	0	
		Culex maxi	356	8	0	0	
		Culex spp.	31	5	0	0	
		Mansonia titillans	879	28	0	0	
		Coquillettidia nigricans	3	2	0	0	
		Anopheles oswaldoi	280	7	0	0	
		Uranotaenia sp.	38	2	0	0	
		Others	8	3	0	0	
		Total	3942	133	0	3	

*The viral gene was identified by pan-flavivirus RT-PCR.
†The viral gene was identified by total RNA sequencing.

Analyses of genetics and molecular evolution

Deduced amino acid sequences of the novel flaviviruses were evaluated by BLAST search, and homologies with previously characterized viral structural and non-structural proteins were analysed by the fastp program using GENETYX v.15.

For the phylogenetic analysis, amino acid sequence alignment of flavivirus polyproteins were generated using the L-INS-i program in the MAFFT suite v.7.467 [18]; gaps were removed using the TrimAl program with a gappyout option [19]. The amino acid replacement model of Le_Gascuel_2008 [20] was selected using ProtTest3 [21]. The maximum-likelihood-based phylogenetic tree search was performed using RAxML-NG with 1000 bootstrap replicates [22].

The amino acid sequences of flavivirus capsid proteins were aligned using CLUSTALW [23] and MEGA7 [24]. To predict cleavage sites for each flavivirus protein, SignalP 5.0 and ProP 1.0 were used to identify potential cleavage sites [25, 26], and the TMHMM Server v.2.0 or SOSUI were used to predict transmembrane helices in proteins [27, 28]. The secondary structure of the Mansonia flavivirus (MAFV) capsid protein sequence was predicted using the PSIPRED and DISOPRED server [29–31].

Virus growth in cells

Mosquito *A. albopictus* C6/36 cells, *A. aegypti* CCL-125 cells, *Culex quinquefasciatus* Hsu cells, *Culex tarsalis* Chao Ball cells and mammalian Vero cells, and BHK-21 cells were employed in flavivirus growth assays. CCL-125 cells from the ATCC, Hsu and Chao Ball cells (provided from the University of Queensland, originally from the University of Texas Medical Branch) were maintained in Leibovitz's

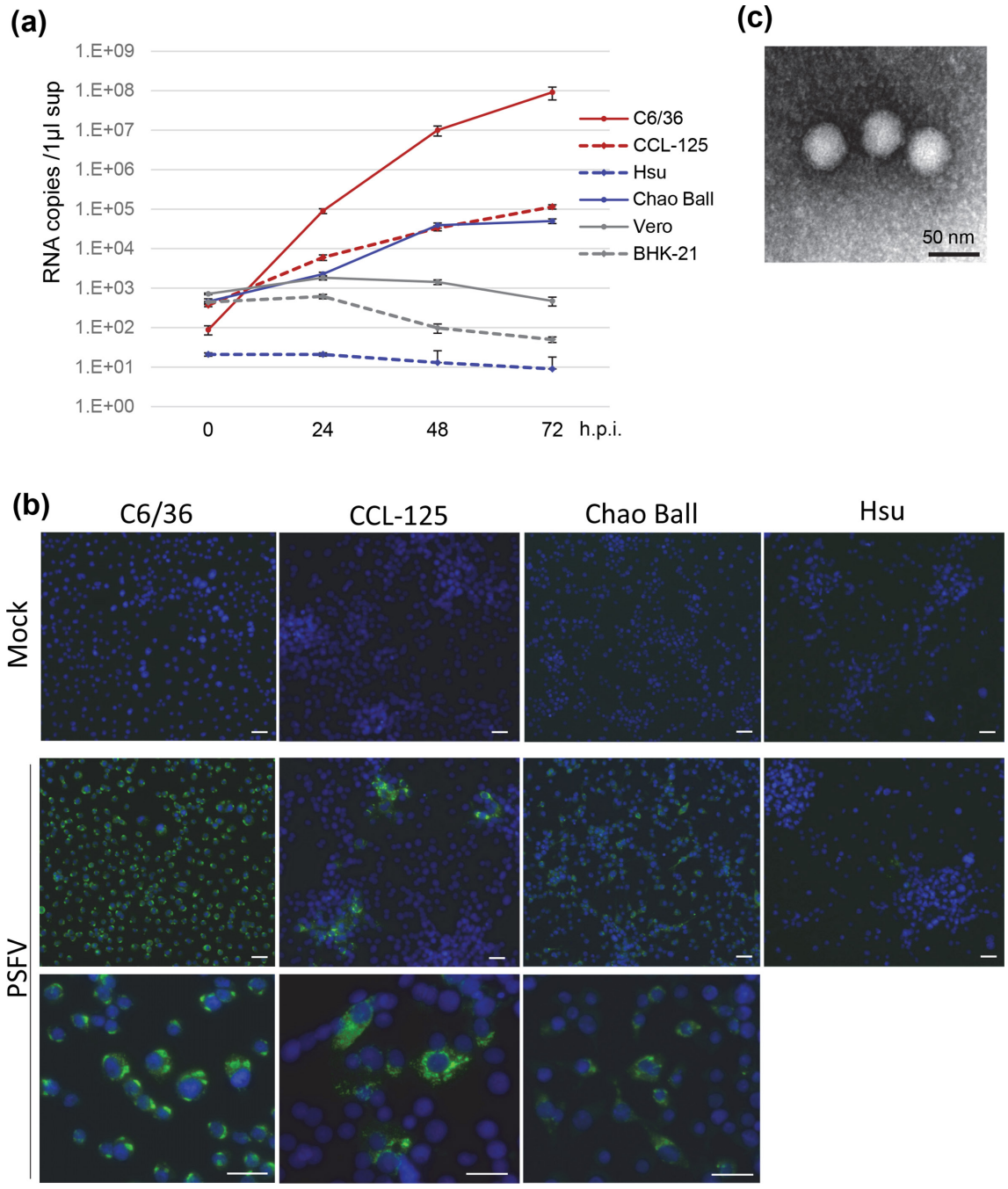

Fig. 1. Characterization of the isolated Psorophora flavivirus (PSFV). (a) Viral RNA copy number of PSFV in supernatants from infected C6/36, CCL-125, Hsu, Chao Ball, Vero or BHK-21 cells at time points indicated as hours post-infection (p.i). Data represent the mean (±SD) of three independent experiments. (b) C6/36, CCL-125, Chao Ball and Hsu cells infected with PSFV or mock for 72 h were stained with anti-NS1 antibody and Alexa-488 conjugated secondary antibody; cell nuclei were counterstained with DAPI. Scale bars, 20 µm. (c) Negatively stained enveloped virus particles ~40 nm in size were detected in the supernatant of PSFV-infected C6/36 cells by electron microscopy; scale bar indicates 50 nm.

L15 medium supplemented with 2% FBS, 10% tryptose phosphate broth, penicillin, streptomycin and 2 mM L-glutamine at 28 °C during virus infection assays. Cells in a 12-well plate were infected with 10^8 copies of PSFV for 1 h at 28 °C or 37 °C; cell supernatants were then removed, and new medium was added after washing. Cell supernatants were collected 0, 24, 48 and 72 h later. The genome copy number of PSFV in the supernatant collected at each time point was quantified by a quantitative RT-PCR assay using PSFV-specific primer sets (PSFV F 5′-GGTTATCACGTG

Table 2. Predicted structure of the viral genomes and encoded proteins

	Genome (nt)	5′-UTR (nt)	3′-UTR (nt)	CDS (aa)	C/ancC (aa)	pre/M (aa)	E (aa)	NS1 (aa)	NS2A (aa)	NS2B (aa)	NS3 (aa)	NS4A (aa)	2 K/NS4B (aa)	NS5 (aa)
PSFV	10831	114	373	3448	105/22	91/74	501	355	227	131	621	128	23/263	807
OSFV	10,449~*	*~33	183~*	3411	121/23	86/62	437	403	195	166	599	142	23/266	888
MAFV	10,982~†	88	325~†	3523	249/20	85/63	426	394	228	148	593	147	23/251	896

*The 5′- and 3′-terminals of UTR sequences have not been determined.
†The 3′-terminal of UTR sequence has not been determined.
ancC, anchor capsid; C, capsid protein; CDS, coding sequence; E, envelope protein; M, membrane protein; NS, non-stractural protein; pre, pre-membrane protein; UTR, untranslated region.

GCAACAGTC and PSFV R 5′-TGCGGTACGCTAAGTC CAGAACG) and the One Step TB Green II kit (Takara) in a StepOnePlus real time PCR system (Applied Biosystems). The copy numbers were estimated by a standard curve method with serially diluted 10^9 copies of RT-PCR product of PSFV genome (5415–9466 nt).

Immunostaining of the PSFV non-structural protein (NS)1 in C6/36, CCL-125 and Hsu cells at 72 h after virus infection was performed with an anti-flavivirus NS1 monoclonal antibody (4G4) [32] (Mozzy Mabs: https://eshop.uniquest.com.au/mozzy-mabs/) and Alexa 488-labelled anti-mouse IgG. Cell nuclei were counterstained with 4′, 6-diamidino-2-phenylindole (DAPI). Fluorescent images were evaluated using the Olympus IX-73 microscope (Olympus, Tokyo, Japan).

Electron microscopy

Culture supernatants of PSFV-infected C6/36 cells were fixed with 0.25% glutaraldehyde and concentrated using an Amicon Ultra 100K filter (Merck Millipore, Darmstadt, Germany). The virus particles were then negatively stained in a 2% phosphotungstic acid solution (pH 5.8) on collodion–carbon-coated copper grids (Nisshin EM Corporation, Tokyo, Japan). Virions were analysed using a H-7650 electron microscope at 80 kV (Hitachi, Kyoto, Japan).

RESULTS

Mosquitoes were captured in a forested area of Trinidad in two different seasons, in October 2018 and in August 2019. In October, which is the beginning of rainy season, the main

Table 3. Comparison of the deduced amino acid sequences of structural and non-structural proteins of PSFV with those of other flaviviruses

		PSFV			
		Structural		Non-structural	
		Identity (%)	Similarity (%)	Identity (%)	Similarity (%)
Lineage IIa ISFs	Panmunjeom	47	84	64	90
	Donggang	46	85	64	91
	Marisma	46	85	64	91
	Ilomantsi	45	84	64	91
	Long Pine Key	45	83	63	90
	Lammi	48	85	62	90
	Chaoyang	48	85	62	90
Lineage IIb ISFs	Barkedji	41	81	49	83
	Nhumirim	38	80	49	83
	Nanay	38	78	49	83
Mosquito-borne flaviviruses	West Nile	42	81	50	84
	Zika	37	81	52	84
	Dengue 1	37	77	48	82
	Yellow fever	34	79	48	82

Table 4. Comparison of the deduced amino acid sequences of structural and non-structural proteins of OSFV and MAFV to those of other lineage I ISFs

	OSFV				MAFV			
	Structural		Non-structural		Structural		Non-structural	
	Identity (%)	Similarity (%)	Identity (%)	Similarity (%)	Identity (%)	Similarity (%)	Identity (%)	Similarity (%)
Xishuangbanna	52	84	61	88	24	72	40	76
Menghai	51	84	59	87	24	74	39	75
Cell-fusing agent	36	77	46	80	28	74	40	76
Parramatta River	37	76	45	79	27	74	41	76
Hanko	35	74	44	79	26	73	40	76
Palm Creek	37	79	42	75	24	79	41	77
Culex flavi	36	78	41	75	28	71	41	76
Nienokoue	35	75	41	75	28	71	42	77
Mercadeo	28	74	38	73	29	75	41	75
Culiseta	27	70	40	75	28	75	43	77
Calbertado	27	70	40	73	30	71	39	73
Sabethes flavi	25	68	39	74	29	71	40	75
Anopheles flavi	26	69	40	74	28	70	40	76
MAFV	24	71	40	75	100	100	100	100
PSFV	21	60	29	66	18	58	29	67

mosquito species collected was *Psorophora* (*Ps.*) *albigenu*. A total of 3942 adult female mosquitoes were examined in 133 pools comprising each species for detection and isolation of both flaviviruses and alphaviruses (Table 1).

The pan-flavivirus RT-PCR assay detected two different sequences with some similarity to previously identified flavivirus NS5 genes. These sequences were identified from RNAs extracted from *Ps. albigenu* (1/47 pools) and *Ochlerotatus* (*Och.*) *scapularis* (1/6 pools) collected in October 2018. Notably, no alphavirus genes were detected in any of mosquitoes collected using the pan-alphavirus RT-PCR assay.

A flavivirus was isolated from *Ps. albigenu* homogenate used to inoculate *A. albopictus*-derived C6/36 cells and this was tentatively named Psorophora flavivirus (PSFV). This virus replicated in C6/36, *A. aegypti*-derived CCL-125 and *C. tarsalis*-derived Chao Ball cells without obvious cytopathic effects; in contrast, no replication was detected in *C. quinquefasciatus*-derived Hsu cells or in the mammalian Vero or BHK-21 cells (Fig. 1a). Immunostaining for flavivirus NS1 protein revealed that almost all C6/36 cells were infected with PSFV at 72 h post-inoculation. In contrast, focal infections were observed in CCL-125 cells, and weak positive signals were observed in Chao Ball cells. Very few NS1-positive foci were detected in Hsu cells (Fig. 1b). Typical flavivirus particles that were ~40 nm in size were detected in supernatants of PSFV-infected C6/36 cells by electron microscopy (Fig. 1c).

In efforts to identify the whole-genome sequence of PSFV, RNA was extracted from the supernatants of PSFV-infected cells and analysed by RNA sequencing, followed by 5′- and 3′-RACE assays. We confirmed that the supernatants containing PSFV were not contaminated by other viruses by total RNA sequencing. The complete RNA genome sequence of PSFV (10831 nt) was determined and the viral polyprotein coding sequence (CDS; 3448 aa) and those of specific viral proteins were deduced (Table 2). The polyprotein of flaviviruses is proteolytically cleaved by both viral and host proteases into three structural proteins (capsid, membrane and envelope) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5).

The flavivirus identified in *Och. Scapularis* was named Ochlerotatus scapularis flavivirus (OSFV). This virus was not amplified in any of the cell cultures examined; as such, total RNAs extracted from the OSFV-positive mosquito pool were used for generation of the whole-genome sequence by Illumina sequencing and RACE assays. The complete polyprotein CDS of OSFV was identified (3411 aa), although we have not yet succeeded in identifying the ends of 5′-and 3′-terminal UTR sequences (Table 2).

Finally, RNA sequencing analyses of total RNAs extracted from other representative species of mosquito pools resulted in the identification of another novel flavivirus genome sequence from a pool derived from *Mansonia* (*Ma.*) *titillans*. This flavivirus has tentatively been named Mansonia flavivirus (MAFV). The almost complete genome sequence of MAFV, including the complete polyprotein CDS (3523 aa) and the 5′-UTR, was identified by RACE assays (Table 2). The

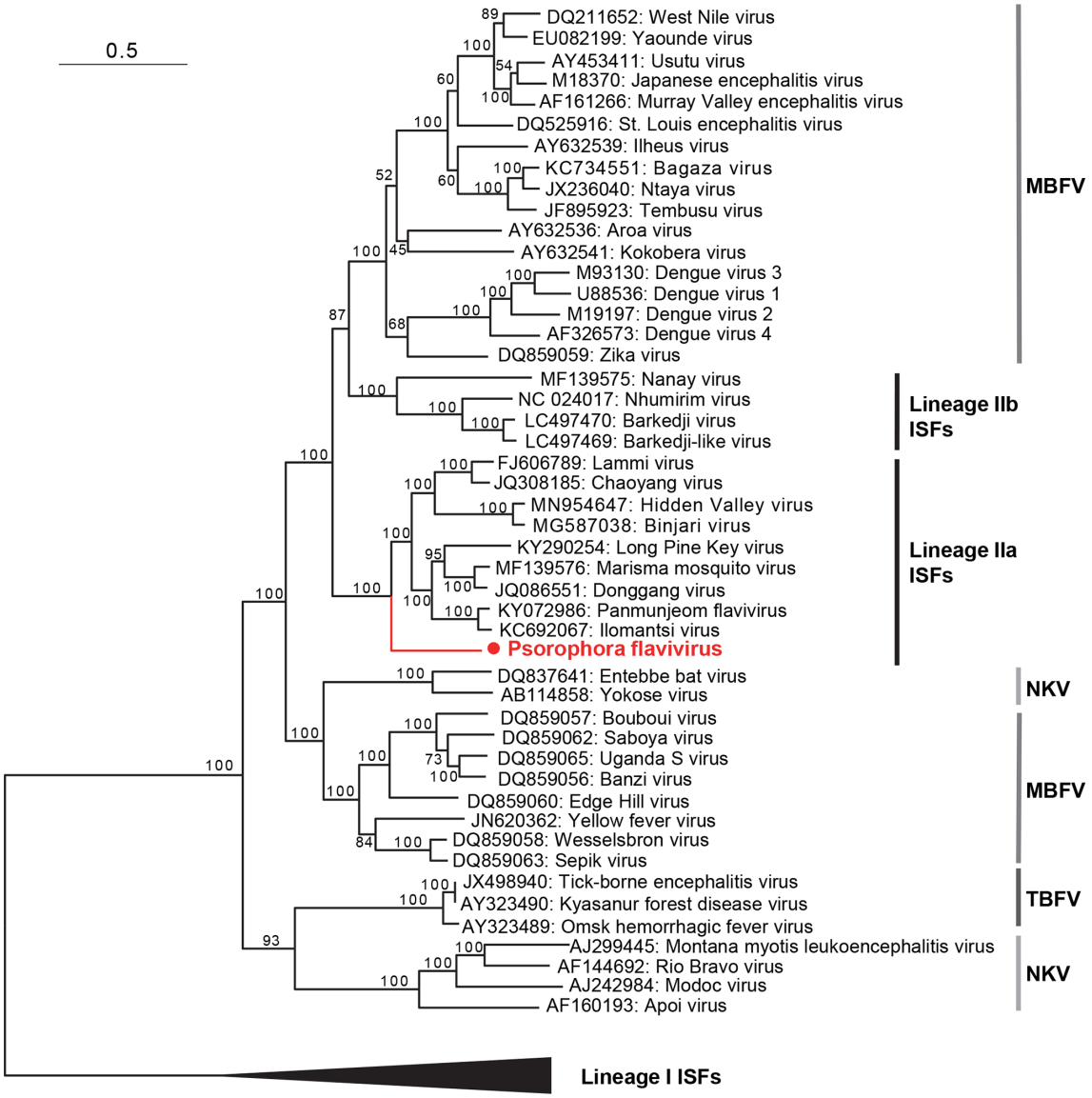

Fig. 2. Molecular phylogenetic analysis of flavivirus polyproteins. The amino acid sequences of flavivirus polyproteins were aligned using the MAFFT program; a maximum-likelihood-based phylogenetic tree was generated using RAxML-NG with 1000 bootstrap replicates; bootstrap values are shown adjacent to the tree branches. The tree is drawn to scale with branch lengths representing the number of substitutions per site. MBFV, mosquito-borne flaviviruses; ISFs, insect-specific flaviviruses; TBFV, tick-borne flaviviruses; NKV, no known vector flaviviruses. Each flavivirus sequence has an appended GenBank accession number.

5′-terminal nucleotide sequence of MAFV, including the UTR and the capsid protein coding region, was confirmed by Sanger sequencing using specific primers. The genome sequences of these three novel flaviviruses have been submitted to the DDBJ/EMBL/GenBank databases and have been assigned accession numbers LC567151 (PSFV), LC567152 (OSFV) and LC567153 (MAFV).

We performed a BLAST analysis, and found that the PSFV CDS shared 60% amino acid sequence identity with the CDS of the lineage II ISF Marisma virus (accession number MF139576). Likewise, the CDSs of OSFV and MAFV shared 60 and 40% identity with previously identified lineage I ISFs, including Xishuangbanna aedes flavivirus (accession number KU201526) and Sabethes flavivirus (accession number MH899446), respectively. When the amino acid homologies of the structural and non-structural proteins were evaluated separately, we found that PSFV structural proteins shared 45–48% identity and non-structural proteins shared 62–64% identity with those of lineage II ISFs. Of note, these flaviviruses are primarily those that were originally isolated from *Aedes* spp. mosquitoes, which are included within the proposed new lineage IIa (Tables 3 and S1 available in the online version of this article) [33–39]. The results of pairwise comparisons and evolutionary distances among OSFV,

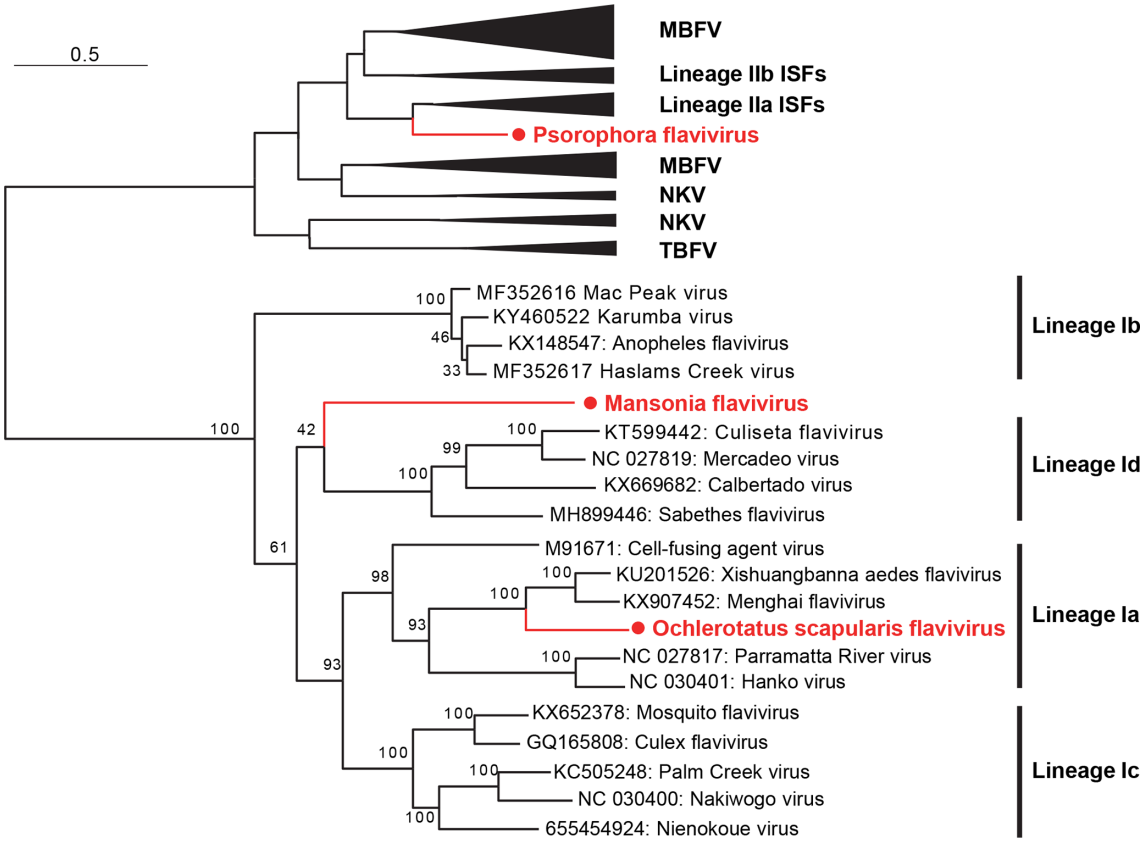

Fig. 3. Molecular phylogenetic analysis of lineage I ISFs polyprotein. The phylogenetic tree focuses on the lineage I ISFs from the phylogenetic analysis shown in Fig. 2. Sublineages Ia, Ib, Ic and Id within lineage I ISFs are indicated by bar lines. The tree is drawn to scale with branch lengths representing the number of substitutions per site. MBFV, mosquito-borne flaviviruses; ISFs, insect-specific flaviviruses; TBFV, tick-borne flaviviruses; NKV, no known vector flaviviruses. Each flavivirus sequence has an appended GenBank accession number.

MAFV and related lineage I ISFs revealed that MAFV is a highly divergent flavivirus with its sequence identities with the closest known relative being 30% with the structural proteins of Calbertado virus and 42% with the non-structural proteins of Nienokoue virus (Tables 4 and S2). The relationships between OSFV and MAFV, which were identified from the same area but from different species of mosquitoes, suggest that they are divergent ISF species in lineage I in that their non-structural proteins share approximately 40% amino acid sequence identity (Table 4).

A phylogenetic tree built using the full-length flavivirus polyprotein CDSs revealed that the separation of PSFV is deepest within the clade of lineage IIa ISFs (Fig. 2). Likewise, MAFV is deeply separate in lineage Id ISFs consisting of Sabethes flavivirus (MH899446), Culiseta flavivirus (KT599442), Mercadeo virus (NC 027819) and Calbertado virus (KX669682) (Fig. 3). In addition, MAFV was frequently located basal to all three *Culicinae*-associated lineages (i.e. Ia, Ic and Id [39]) in the bootstrap trees (data not shown). These results suggest that MAFV may retain ancestral characteristics of the current known *Culicinae*-associated lineage I ISFs. Meanwhile, OSFV clearly belonged to a clade of *Aedes*-associated lineage Ia; these results were consistent with the fact that OSFV was isolated from *Ochlerotatus* mosquitoes, a species that is closely related to mosquitoes of the genus *Aedes* in the tribe *Aedini* [40–42].

Multiple alignments and predictions of individual viral protein sequences revealed that the MAFV CDS is larger than those of any of the other known flaviviruses due to the addition of more than 100 amino acid residues at the N-terminus end of the polyprotein (Fig. 4). A putative cleavage site for the viral protease NS2B/NS3 was predicted at a site 249 aa upstream of the alpha-helices that comprise the anchored capsid region. Typical helical structures and positively charged residues were identified at the C-terminus of the MAFV capsid protein (Fig. 4), although the disordered N-terminal region shares no specific similarity with any other viruses or any known sequences in a BLAST database search. As such, the characteristic capsid protein of MAFV can only be deduced from the genomic sequence.

DISCUSSION

Some arboviruses, including ILHV and MAYV, that have natural enzootic cycles involving forest-dwelling mosquitoes

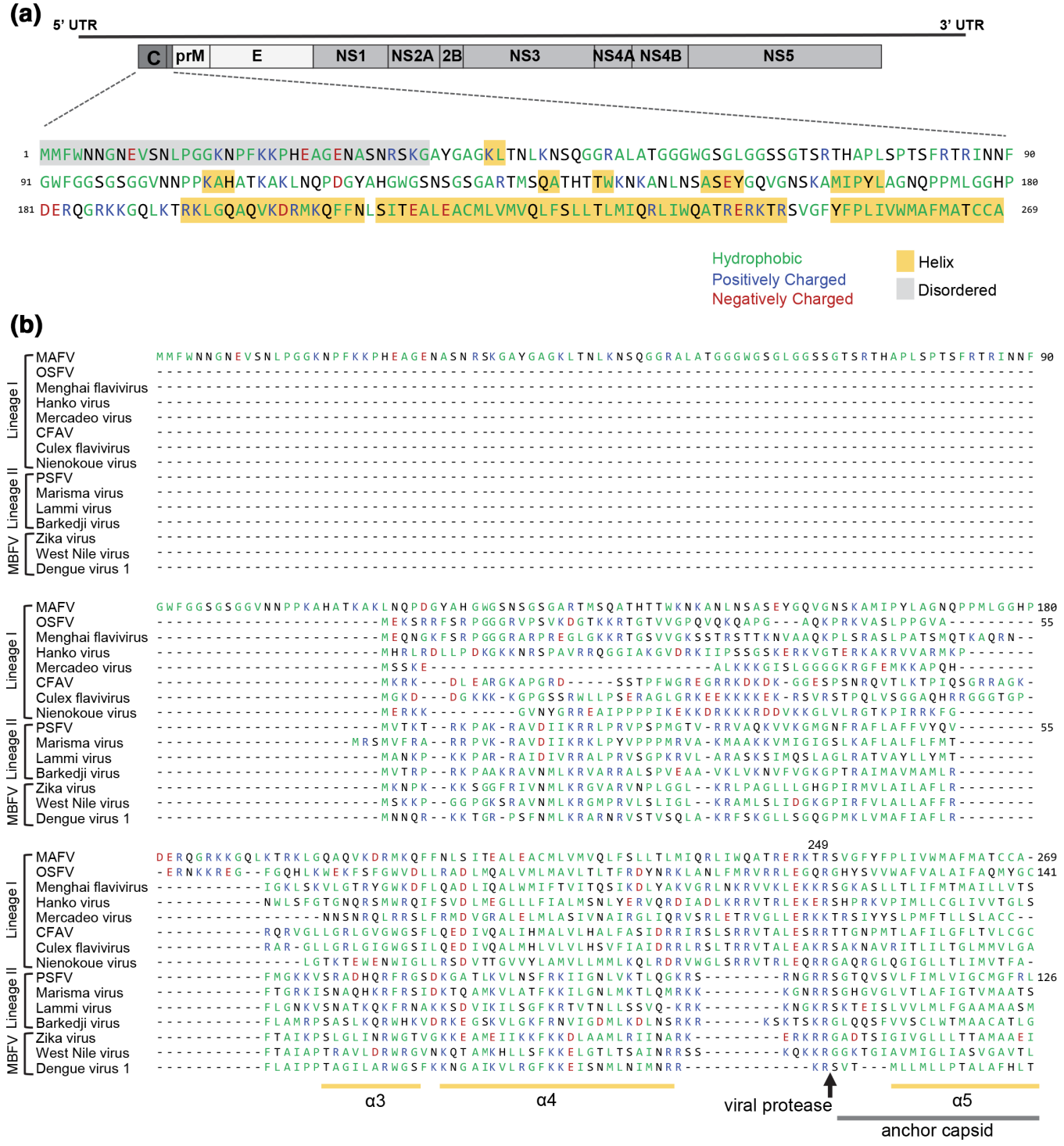

Fig. 4. Sequence of the capsid protein of MAFV. (a) Schematic representation of the flavivirus genome and predicted coding sequences. The deduced amino acid sequence of the MAFV capsid protein is shown with the predicted helical and disordered secondary structure. Amino acid residues forming the helices are highlighted in yellow, and the amino-terminal disordered residues are highlighted in grey. Hydrophobic positively and negatively charged amino acids are indicated in green, blue and red, respectively. (b) Multiple alignment was performed using the deduced amino acid sequences of the capsid proteins of lineage I ISFs, including OSFV and MAFV, lineage II ISFs, and MBFVs. The predicted viral protease NS2B/NS3 cleavage sites are indicated by arrows. The yellow bars represent conserved helices α3, α4 and α5, as reported previously in capsid proteins of ZIKV and DENV. The grey bar represents the membrane anchor regions of the capsid proteins.

and animals in the Amazon basin have the potential to emerge as human pathogens as a result of increased human travel, deforestation and urbanization in tropical areas [43]. Climate change also modifies the distribution of vector mosquitoes and increases the possibility of new arbovirus disease outbreaks in areas with a temperate climate [44]. In late October, the forested area where we collected mosquitoes in Trinidad was inhabited by the mosquito species *Ps. albigenu*, at high density. *Ps. albigenu* is distributed widely in South America, with high population densities in areas of high rainfall [45]. Arboviruses, including ILHV, MAYV, EEVs and DENV, have all been identified in *Psorophora* spp., which are present in numerous South American countries [46–48]. The feeding habitats of the *Psorophora* spp. indicated that they could probably serve as a vector for sylvatic transmission of other arboviruses [49].

The present study has explored the possibility that medically significant arboviruses might be identified in forested areas of the Bolivian Amazon basin region and our overall goal was to collect information that could facilitate preemptive measures against emerging mosquito-borne diseases. Although we identified no mosquito-borne arboviruses, we did instead discover a diverse group of novel flaviviruses from several species of local mosquitoes.

PSFV appears to be an *Aedes*-associated lineage IIa ISF that is closely related to MBFVs. This is consistent with the fact that PSFV was isolated from the genus *Psorophora*, which belongs in the mosquito tribe *Aedini*, as is the mosquito genus, *Aedes*. The results of PSFV growth assays indicated that not only *Aedes*-derived C6/36 and CCL-125 cells, but also *Culex*-derived Chao Ball cells, were susceptible to PSFV infection (Fig. 1b). Further research is required to understand the host species specificity of lineage II ISFs. Although PSFV did not replicate in mammalian BHK-21 or Vero cells, there remains the possibility that PSFV may have one or more other vertebrate hosts that inhabit this region.

Recent reports indicated that ISFs have the potential to inhibit arboviruses via a mechanism known as superinfection exclusion [50]. As such, there are ongoing studies that utilize a lineage II ISF to generate chimeric flaviviruses with MBFVs as candidate vaccine antigens [51]. The differences between lineage II ISFs and mosquito-borne arboviruses involved in host restriction factors are now under investigation to develop strategies to combat arbovirus-associated diseases.

Two novel lineage I (classical) ISFs were also identified in this study. These flaviviruses could not be isolated from the mosquito pools, most likely due to low virus copy numbers, a low susceptibility of C6/36 and Hsu cells, or the presence of other competing mosquito-specific viruses that proliferate at a higher rate. The Illumina sequencing of total RNAs from mosquitoes included an average coverage of OSFV sequence at 27 with that of MAFV at 64. In contrast, the average coverage of other insect-specific RNA viruses such as negevirus and picorna-like virus was more than 10000 (data not shown). However, the complete polyprotein CDS and almost the entire genome sequences of OSFV and MAFV were successfully identified in RNA extracted directly from mosquitoes.

OSFV and especially MAFV display sequences that are evolutionarily diverse are not closely related to those in previously characterized flaviviruses. A phylogenetic analysis suggested that the MAFV may be one of the oldest ISFs that diverged in the lineage of the *Culicinae*-associated lineage I ISFs (Fig. 3). The low sequence identity shared by the MAFV genome with known flaviviruses may be why this virus was not detected by the pan-flavivirus RT-PCR assay.

Investigations on the structure of the flavivirus capsid protein have revealed that both the structure and charge distribution are well conserved among flaviviruses [52–55]. The MAFV capsid proteins contain five α-helices; a host signal peptidase cleaves at the junction between capsid helix α5 and prM at the ER lumen, leaving helix α5 anchored within the ER membrane (anchor C). The viral protease NS2B/NS3 cleaves at the junction between helix α4 with α5 at the protein cytoplasmic face to release the mature capsid protein. The capsid protein generates a dimer at the highly positively charged α4 helices that have been implicated in interactions with viral RNA [56, 57]. As noted above, the deduced amino acid sequence of the MAFV capsid protein is almost two times longer than those of other flaviviruses. The predicted viral protease cleavage site and the α-helix sequence of anchor capsid regions have been identified in the MAFV capsid protein. The highly charged helix α4 domain appears to be conserved among all the lineage II ISFs, including PSFV. The lineage I ISFs, including OSFV and MAFV, also maintain α-helical structures at the C-terminal region of the capsid proteins that include positively charged residues (Fig. 4). The uncharacteristically long N-terminal region of the MAFV capsid protein indicates that it most likely includes one or more regions of disordered structure that have the potential to interact with host and/or viral proteins. Further investigations should reveal novel functions of the MAFV capsid protein that might have been lost during virus evolution.

Funding information
This study was supported by grants for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT) /Japan Society for the Promotion of Science (JSPS) KAKENHI (JP16H05805, JP19H03112 and 20K21298); and grants for Scientific Research on Innovative Areas and International Group from the MEXT/JSPS KAKENHI (JP16H06431, JP16H06429, JP16K21723 and 19H04843); and grants for AMED (JP20WM0225018).

Acknowledgements
We appreciate the cooperation of the Japan International Cooperation Agency (JICA) members, Mariko Tanaka and Madoka Matsuo, with mosquito sampling in Bolivia. Computations in this work were performed in part on the NIG supercomputer at ROIS National Institute of Genetics and SHIROKANE at Human Genome Center (University of Tokyo).

Author contributions
Investigation: Y. O., Y. S., K. M., R. N., H. S., S. N., K. K., T. W., T. M., W. W. H. Supervision: Y. O., S. N., J. A. P., H. S. Resources: Y. O., K. M., R. N., F. K., A. L., T. W., T. M., R. A. H., J. A. P., M. S., K. K., S. N. Project Administration: J. A. P., H. S. Writing – original draft preparation: Y. O., S. N., K. M., R. A. H., W. W. H. Writing – review and editing: all.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Saba Villarroel PM, Nurtop E, Pastorino B, Roca Y, Drexler JF *et al*. Zika virus epidemiology in Bolivia: a seroprevalence study in volunteer blood donors. *PLoS Negl Trop Dis* 2018;12:e0006239.
2. Roca Y, Baronti C, Revollo RJ, Cook S, Loayza R *et al*. Molecular epidemiological analysis of dengue fever in Bolivia from 1998 to 2008. *Vector Borne Zoonotic Dis* 2009;9:337–344.
3. Venegas EA, Aguilar PV, Cruz C, Guevara C, Kochel TJ *et al*. Ilheus virus infection in human, Bolivia. *Emerg Infect Dis* 2012;18:516–518.
4. Laemmert HW, Hughes TP. The virus of Ilhéus encephalitis; isolation, serological specificity and transmission. *J Immunol* 1947;55:61–67.
5. Turell MJ, O'Guinn ML, Jones JW, Sardelis MR, Dohm DJ *et al*. Isolation of viruses from mosquitoes (Diptera: Culicidae) collected in the Amazon Basin region of Peru. *J Med Entomol* 2005;42:891–898.
6. Vieira CJdaSP, Andrade CD, Kubiszeski JR, Silva DJFda, Barreto ES, Vieira C, Silva D *et al*. Detection of Ilheus virus in mosquitoes from Southeast Amazon, Brazil. *Trans R Soc Trop Med Hyg* 2019;113:424–427.
7. Organization PAH. Geographic spred of Chkungunya in the Americas 2018.
8. Stollar V, Thomas VL. An agent in the *Aedes aegypti* cell line (Peleg) which causes fusion of *Aedes albopictus* cells. *Virology* 1975;64:367–377.
9. Patterson EI, Villinger J, Muthoni JN, Dobel-Ober L, Hughes GL. Exploiting insect-specific viruses as a novel strategy to control vector-borne disease. *Curr Opin Insect Sci* 2020;39:50–56.
10. (WRBU) TWRBU. Mosquito identification resources. http://wrbu.si.edu/aors/aors_Keys.html 2019.
11. Segura MNO. Atlas de Culicídeos na Amazônia brasileira: características específicas de insetos hematófagos da família Culicidae: Instituto evandro chagas 2007.
12. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 1994;3:294–299.
13. Patel P, Landt O, Kaiser M, Faye O, Koppe T *et al*. Development of one-step quantitative reverse transcription PCR for the rapid detection of flaviviruses. *Virol J* 2013;10:58.
14. Orba Y, Hang'ombe BM, Mweene AS, Wada Y, Anindita PD *et al*. First isolation of West Nile virus in Zambia from mosquitoes. *Transbound Emerg Dis* 2018;65:933–938.
15. Torii S, Orba Y, Hang'ombe BM, Mweene AS, Wada Y *et al*. Discovery of Mwinilunga alphavirus: a novel alphavirus in Culex mosquitoes in Zambia. *Virus Res* 2018;250:31–36.
16. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
17. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA *et al*. Full-Length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;29:644–652.
18. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
19. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–1973.
20. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol* 2008;25:1307–1320.
21. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 2011;27:1164–1165.
22. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;35:4453–4455.
23. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA *et al*. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007;23:2947–2948.
24. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.
25. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O *et al*. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 2019;37:420–423.
26. Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. *Protein Eng Des Sel* 2004;17:107–112.
27. Möller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 2001;17:646–653.
28. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 1998;14:378–379.
29. Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res* 2019;47:W402–W407.
30. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 1999;292:195–202.
31. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* 2015;31:857–863.
32. Clark DC, Lobigs M, Lee E, Howard MJ, Clark K *et al*. In situ reactions of monoclonal antibodies with a viable mutant of Murray Valley encephalitis virus reveal an absence of dimeric NS1 protein. *J Gen Virol* 2007;88:1175–1183.
33. Takhampunya R, Kim HC, Tippayachai B, Lee DK, Lee WJ *et al*. Distribution and mosquito hosts of Chaoyang virus, a newly reported flavivirus from the Republic of Korea, 2008–2011. *J Med Entomol* 2014;51:464–474.
34. Huhtamo E, Putkuri N, Kurkela S, Manni T, Vaheri A *et al*. Characterization of a novel flavivirus from mosquitoes in northern Europe that is related to mosquito-borne flaviviruses of the tropics. *J Virol* 2009;83:9532–9540.
35. Vázquez A, Sánchez-Seco MP, Palacios G, Molero F, Reyes N *et al*. Novel flaviviruses detected in different species of mosquitoes in Spain. *Vector Borne Zoonotic Dis* 2012;12:223–229.
36. Huhtamo E, Cook S, Moureau G, Uzcátegui NY, Sironen T *et al*. Novel flaviviruses from mosquitoes: Mosquito-specific evolutionary lineages within the phylogenetic group of mosquito-borne flaviviruses. *Virology* 2014;464-465:320–329.
37. Korkusol A, Takhampunya R, Hang J, Jarman RG, Tippayachai B *et al*. A novel flavivirus detected in two *Aedes* spp. collected near the demilitarized zone of the Republic of Korea. *J Gen Virol* 2017;98:1122–1131.
38. Harrison JJ, Hobson-Peters J, Colmant AMG, Koh J, Newton ND *et al*. Antigenic characterization of new lineage II insect-specific flaviviruses in Australian mosquitoes and identification of host restriction factors. *mSphere* 2020;5.
39. Colmant AMG, Hobson-Peters J, Bielefeldt-Ohmann H, van den Hurk AF, Hall-Mendelin S. A new clade of insect-specific flaviviruses from Australian. *mSphere* 2017;2.
40. Zhang X, Guo X, Fan H, Zhao Q, Zuo S *et al*. Complete genome sequence of Menghai flavivirus, a novel insect-specific flavivirus from China. *Arch Virol* 2017;162:1435–1439.
41. Fan H, Zhao Q, Guo X, Sun Q, Zuo S *et al*. Complete genome sequence of Xishuangbanna flavivirus, a novel mosquito-specific flavivirus from China. *Arch Virol* 2016;161:1723–1727.
42. McLean BJ, Hobson-Peters J, Webb CE, Watterson D, Prow NA *et al*. A novel insect-specific flavivirus replicates only in Aedes-derived cells and persists at high prevalence in wild Aedes vigilax populations in Sydney, Australia. *Virology* 2015;486:272–283.

43. Weaver SC, Reisen WK. Present and future arboviral threats. *Antiviral Res* 2010;85:328–345.
44. Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. *PLoS Negl Trop Dis* 2019;13:e0007213.
45. Alencar J, Pacheco JB, Dos Santos Silva J, Silva SOF, Guimarães Anthony Érico. Influence of climatic factors on *Psorophora* (*Janthinosoma*) *Albigenu* in Pantanal landscape, Mato Grosso state, Brazil. *J Am Mosq Control Assoc* 2018;34:177–181.
46. Turell MJ, Jones JW, Sardelis MR, Dohm DJ, Coleman RE *et al*. Vector competence of Peruvian mosquitoes (Diptera: Culicidae) for epizootic and enzootic strains of Venezuelan equine encephalomyelitis virus. *J Med Entomol* 2000;37:835–839.
47. Groot H, Morales A, Vidales H. Virus isolations from forest mosquitoes in San Vicente de Chucuri, Colombia. *Am J Trop Med Hyg* 1961;10:397–402.
48. Serra OP, Cardoso BF, Ribeiro AL, Santos FA, Slhessarenko RD. Mayaro virus and dengue virus 1 and 4 natural infection in culicids from Cuiabá, state of Mato Grosso, Brazil. *Mem Inst Oswaldo Cruz* 2016;111:20–29.
49. Mucci LF, Júnior RP, de Paula MB, Scandar SA, Pacchioni ML *et al*. Feeding habits of mosquitoes (Diptera: Culicidae) in an area of sylvatic transmission of yellow fever in the state of São Paulo, Brazil. *J Venom Anim Toxins Incl Trop Dis* 2015;21:6.
50. Öhlund P, Lundén H, Blomström AL. Insect-specific virus evolution and potential effects on vector competence. *Virus Genes* 2019;55:127–137.
51. Hobson-Peters J, Harrison JJ, Watterson D, Hazlewood JE, Vet LJ *et al*. A recombinant platform for flavivirus vaccines and diagnostics using chimeras of a new insect-specific virus. *Sci Transl Med* 2019;11:eaax7888.
52. Tan TY, Fibriansah G, Kostyuchenko VA, Ng T-S, Lim X-X *et al*. Capsid protein structure in Zika virus reveals the flavivirus assembly process. *Nat Commun* 2020;11:895.
53. Dokland T, Walsh M, Mackenzie JM, Khromykh AA, KH E. West Nile virus core protein; tetramer structure and ribbon formation. *Structure* 2004;12:1157–1163.
54. Oliveira ERA, Mohana-Borges R, de Alencastro RB, Horta BAC. The flavivirus capsid protein: structure, function and perspectives towards drug design. *Virus Res* 2017;227:115–123.
55. Sotcheff S, Routh A. Understanding flavivirus capsid protein functions: the tip of the iceberg. *Pathogens* 2020;9:42.
56. Jones CT, Ma L, Burgner JW, Groesch TD, Post CB *et al*. Flavivirus capsid is a dimeric alpha-helical protein. *J Virol* 2003;77:7143–7149.
57. Ma L, Jones CT, Groesch TD, Kuhn RJ, Post CB. Solution structure of dengue virus capsid protein reveals another fold. *Proc Natl Acad Sci U S A* 2004;101:3414–3419.