JOURNAL OF GENERAL VIROLOGY

RESEARCH ARTICLE
Marklewitz *et al.*, *Journal of General Virology*
DOI 10.1099/jgv.0.001260

MICROBIOLOGY SOCIETY
OPEN MICROBIOLOGY

Diverse novel phleboviruses in sandflies from the Panama Canal area, Central Panama

Marco Marklewitz[1,2,3,†], Larissa C. Dutari[4,5,6,†], Sofia Paraskevopoulou[1], Rachel A. Page[3], Jose R. Loaiza[3,4,6] and Sandra Junglen[1,2,*]

Abstract

The genus *Phlebovirus* (order *Bunyavirales*, family *Phenuiviridae*) comprises 57 viruses that are grouped into nine species-complexes. Sandfly-transmitted phleboviruses are found in Europe, Africa and the Americas and are responsible for febrile illness and infections of the nervous system in humans. The aim of this study was to assess the genetic diversity of sandfly-transmitted phleboviruses in connected and isolated forest habitats throughout the Panama Canal area in Central Panama. In total, we collected 13 807 sandflies comprising eight phlebotomine species. We detected several strains pertaining to five previously unknown viruses showing maximum pairwise identities of 45–78 % to the RNA-dependent RNA polymerase genes of phleboviruses. Entire coding regions were directly sequenced from infected sandflies as virus isolation in cell culture was not successful. The viruses were tentatively named La Gloria virus (LAGV), Mona Grita virus (MOGV), Peña Blanca virus (PEBV), Tico virus (TICV) and Tres Almendras virus (TRAV). Inferred phylogenies and *p*-distance-based analyses revealed that PEBV groups with the Bujaru phlebovirus species-complex, TRAV with the Candiru phlebovirus species-complex and MOGV belongs to the proposed Icoarci phlebovirus species-complex, whereas LAGV and TICV seem to be distant members of the Bujaru phlebovirus species-complex. No specific vector or habitat association was found for any of the five viruses. Relative abundance of sandflies was similar over habitat types. Our study shows that blood-feeding insects originating from remote and biodiverse habitats harbour multiple previously unknown phleboviruses. These viruses should be included in future surveillance studies to assess their geographic distribution and to elucidate if these viruses cause symptoms of disease in animals or humans.

INTRODUCTION

Neotropical sandflies are a diverse group of insects in the order Diptera, family Psychodidae, subfamily Phlebotominae, which has undergone notable taxonomic revision recently [1]. Historically, phlebotomine sandflies have been recognized as vectors of various disease-causing pathogens to humans [2]. In Panama, anthropophilic species of at least six species of phlebotomine sandflies have been implicated as vectors of *Leishmania* parasites, the causing agent of American Cutaneous Leishmaniasis (ACL) [3–5], and sandflies of some of these taxa are also suspected of transmitting phleboviruses to a broad range of animals and humans [6–9]. Sandflies of the six species *Psychodopygus panamensis*, *Nyssomyia trapidoi*, *Lutzomyia gomezi*, *Nyssomyia ylephiletor*, *Lutzomyia*

Received 13 February 2019; Accepted 21 March 2019; Published 03 May 2019
Author affiliations: [1]Institute of Virology, Charité - Universitätsmedizin Berlin, corporate member of Free University Berlin, Humboldt-University Berlin, and Berlin Institute of Health, Berlin, Germany; [2]German Center for Infection Research (DZIF), Berlin, Germany; [3]Smithsonian Tropical Research Institute, Panama City, Republic of Panama; [4]Centro de Biodiversidad y Descubrimiento de Drogas, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT-AIP), Panama City, Republic of Panama; [5]Biotechnology Department, Acharya Nagarjuna University, Guntur, India; [6]Programa Centroamericano de Maestría en Entomología, Vicerrectoría de Investigación y, Postgrado, Universidad de Panamá, Panama City, Republic of Panama.
***Correspondence:** Sandra Junglen, sandra.junglen@charite.de
Keywords: *Bunyavirales*; *Phlebovirus*; species-complex; Neotropics; Panama; Phlebotominae.
Abbreviations: BLAST, Basic Local Alignment Search Tool; CDC, U.S. Center for Disease Control; CF, complement fixation; CHGV, Chagres virus; CL, American Cutaneous Leishmaniasis; DFG, German Science Foundation; GPC, glycoprotein precursor protein; HI, hemagglutination inhibition; ICTV, International Committee on Taxonomy of Viruses; JOAV, Joá virus; LAGV, La Gloria virus; MAFFT, multiple alignment using fast Fourier transform; MEGA, Molecular Evolutionary Genetics Analysis; MOGV, Mona Grita virus; N, nucleocapsid; NGS, next-generation sequencing; NIQV, Nique virus; NPV, Ntepes virus; ORF, open reading frame; PCR, polymerase chain reaction; PEBV, Peña Blanca virus; PTV, Punta Toro virus; RdRp, RNA-dependent RNA polymerase; RT-PCR, reverse transcriptase polymerase chain reaction; RVDB, Reference Viral Database; SENACYT, National Secretariat of Science, Technology and Innovation of Panama; SPP, Priority Program; TICV, Tico virus; TRAV, Tres Almendras virus; URIV, Uriurana virus.
Sequences of LAGV, MOGV, PEBV, TICV and TRAV have been deposited in GenBank under accession number MK524329-MK524350.
†These authors contributed equally to this work

sanguinaria and *Psychodopygus thula* are widespread across forested areas of Panama, feeding opportunistically on animals of various species depending on habitat quality, host availability and biomass [10].

Sandflies almost exclusively transmit viruses of the genus *Phlebovirus* (family *Phenuiviridae,* order *Bunyavirales*). According to the International Committee on Taxonomy of Viruses (ICTV) members of the genus *Phlebovirus* are classified into nine virus species that comprise 57 viruses [11]. In addition, more than 15 putative phlebovirus species have been described that are waiting to be classified by the ICTV. The classified sandfly-transmitted viruses of the New World belong to the species-complexes *Punta Toro*, *Candiru*, *Chilibre* and *Frijoles phlebovirus*. Infections with sandfly-borne phleboviruses can cause unspecific symptoms in humans often misdiagnosed as dengue fever, malaria or influenza [7, 12]. Clinical symptoms are ranging from high fever, severe headache, muscle ache and aseptic meningitis to mild or severe meningoencephalitis [13]. A recent study in Panama has shown that Punta Toro virus (PTV) causes febrile illness in humans with symptoms similar to infections with dengue viruses [12]. Between 1964 and 2009 various members of the PTV species-complex were isolated from phlebotomine sandflies, humans and sentinel hamsters in Panama and Colombia [7]. In addition, Chagres virus (CHGV) has been reported as being responsible for outbreaks of febrile illness in humans [14]. Other phleboviruses detected in Panamanian sandflies in the 1950s are Aguacate, Cacao, Frijoles and Nique viruses [6, 15–17]. More recent contributions to the expansion of the genetic diversity of phleboviruses of the New World have been provided by sequencing of viruses isolated from sandflies (*Phlebotominae* spp.), spiny rats (*Proechimys* sp.) and the Southern two-toed sloth (*Choloepus brasiliensis*) that was collected in Brazil between 1962 and 1985 [18], and 2014–2015 [19]. It is unknown if these viruses are pathogenic for humans. The highest diversity of phleboviruses in Panama was detected in phlebotomine sandflies and human samples from the Central Panama region and the Darien province [7, 20]. Both regions show increasing levels of habitat fragmentation due to land-use change for cattle and crop farming [21, 22]. Despite their impact on human and veterinary health, phleboviruses have been insufficiently studied in Panama [7, 12].

The aim of this study was to investigate the genetic diversity of phleboviruses in sandflies of forested areas of the Panama Canal area. The lowland tropical rainforest of the Panama Canal area harbours a large diversity of phlebotomine sandfly species of the New World. This region has a long history of arbovirus outbreaks. For example, during the construction of the Panama Canal at the end of the nineteenth century tens of thousands of people died from yellow fever [23].

METHODS

Sandfly collection

Fieldwork was conducted in the Panama Canal area, Central Panama, between 2013 and 2014. Sandflies were collected in landscapes differing in their level of anthropogenic disturbance, (i) in continuous tropical lowland forest at the shore of Gatún Lake (continuous forests), (ii) in small forest fragments (1.5–51 ha) surrounded by agricultural fields (forest fragments) and (iii) in isolated forested islands (5.2–17.5 ha) (forested islands) that are remnants of the original continuous forest and emerged after flooding the Panama Canal in 1914. In each of the three landscape types, five replicate sites were selected resulting in 15 sampling sites in total [24].

Adult sandflies were collected using CDC (U.S. Centers for Disease Control) miniature light traps (John W. Hock Company, Gainesville, FL, USA). The following sampling design was used for each sampling site. In total, nine traps were placed along two parallel transects while two traps were placed in the canopy at approximately 10 m high. Traps operated at night for 12 h. Sandfly specimens were collected after sunrise, identified using female taxonomic keys, and stored in liquid nitrogen until further processing [25, 26].

Phlebotomine sandfly community metrics

Phlebotomine sandfly species diversity was assessed at each landscape type using Shannon and Simpson Alpha diversity indexes [27, 28]. Species relative abundance was measured and compared among landscape types using one-way Kruskal–Wallis nonparametric test as implemented in the PAST3 program (available at https://folk.uio.no/ohammer/past/).

RNA-extraction and PCR screening

Sandflies were pooled according to species, collection site and date by individuals of ten. Specimens containing blood were processed individually. Pools were homogenized in 500 µl L-15 medium (Thermo Fisher Scientific, Waltham, USA) using ceramic beads and a SpeedMill Plus (Analytik Jena, Germany). The suspensions were cleared by centrifugation at 2500 r.p.m. for 10 min. RNA was extracted from 50 µl supernatant using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. cDNA was synthesized using the SuperScript III RT System (Invitrogen, Karlsruhe, Germany) and random hexamer primers according to the instructions from the manufacturing company. Sandfly pools were tested for phleboviruses in a PCR assay using Platinum *Taq* DNA Polymerase (Thermo Fisher Scientific, Waltham, USA) and first round primers F 5'-TCAARAAGAMNCAACATGGTGG, R 5'-TATGCCYTGCATCATYCCWG followed by nested round primers Fn 5'-GGACTTAGAGAGATYTAYGTNTTGG and Rn 5'-ACATGRTGACCYTGRTTCCA. Thermocycler conditions were as follows: 10 cycles of 95 °C for 3 min, 94 °C for 15 s, 55 °C for 20 s and 72 °C for 30 s, then 30 cycles from 95 °C to 15 s, 50 °C to 20 s, 72 °C to 30 s and a final extension step of 72 °C for 5 min. PCR products were visualized on 2 % agarose gels and positive samples were Sanger sequenced (Seqlab, Göttingen, Germany). Obtained sequences were analysed using Geneious R9 [29] and compared to GenBank database (www.ncbi.nlm.nih.gov/genbank/).

Table 1. Sandfly species abundance and diversity across study sites

Abbreviations are as follows: number of specimens (*n*), percent (%), Shannon diversity (*H*), Simpson diversity (λ).

Species	Continous forests				Forest fragments				Forested islands			
	n	%	$-p_i*\ln(p_i)$	p_i^2	*n*	%	$-p_i*\ln(p_i)$	p_i^2	*n*	%	$-p_i*\ln(p_i)$	p_i^2
Psychodopygus panamensis	4010	43.9	0.36	0.19	512	34.36	0.37	0.12	1152	38.51	0.37	0.15
Nyssomyia trapidoi	812	8.89	0.22	0.01	282	18.92	0.32	0.04	213	7.12	0.19	0.01
Lutzomyia gomezi	899	9.84	0.23	0.01	164	11.1	0.24	0.01	69	2.3	0.09	0.00
Bichromomyia olmeca sp.	61	0.66	0.00	0.00	–	–	–	–	14	0.46	0.00	0.00
Psychodopygus thula	292	3.19	0.11	0.00	20	1.34	0.06	0.00	290	9.69	0.23	0.01
Nyssomyia ylephiletor	97	1.06	0.05	0.00	21	1.34	0.06	0.00	46	1.53	0.06	0.00
Pressatia dysponeta	321	3.51	0.12	0.00	7	0.46	0.03	0.00	40	1.34	0.06	0.00
Phlebotominae spp.	2523	27.62	0.36	0.08	469	31.47	0.36	0.10	916	30.7	0.36	0.09
Total	9133	100	H=1.49	λ=0.29	1490	100	H=1.48	λ=0.27	2991	100	H=1.56	λ=0.26

Genome sequencing

RNA derived from phlebovirus-positive sandfly pools was used to sequence full genomes on an Illumina MiSeq next-generation sequencing (NGS) platform. Library preparation and Illumina MiSeq sequencing was carried out by using the SuperScript One-Cycle cDNA Kit, the Nextera XT DNA Library Preparation Kit, and V3 chemistry (2×300 bp) according to the manufacturers' instructions. The individual samples were normalized, pooled and sequenced on an Illumina MiSeq instrument, with a designated yield of ~25 million paired-end reads. After demultiplexing the obtained raw reads were subjected to two filtering steps using Lighter v1.1.0 [30] for correcting sequencing errors and bwa v0.7.17 [31] to subtract host-originated background information using the mitochondrial reference genome of *Phlebotomus papatasi* as no genome of a phlebotomine sandfly species closer related to the species used in this study is available. The remaining reads were *de novo* assembled into contigs with the genome assembler software Spades v3.11.1 [32]. An initial BLASTX search against the Reference Viral Database (RVDB) [33] was performed, followed by a second BLASTX search against two custom databases containing only the S and the M segment sequences of *Phenuiviridae*. Reference mapping of NGS reads to the respective PCR fragment was performed using Geneious mapper [29]. Genome ends were amplified by conventional semi-nested RT-PCR (reverse transcription PCR) using Platinum *Taq* DNA Polymerase (Thermo Fisher Scientific, Waltham, USA), segment-specific primers (Table 1) and a primer containing the terminal conserved nucleotides of phleboviruses fused to an anchor sequence PTN 5'-GACC ATCTAGCGACCTCCACACACAAAG. Products were subsequently Sanger sequenced.

Genome analysis

Full genomes were analysed regarding sizes, putative coding regions and 5'/3' noncoding regions using Geneious R9 [29]. N-glycosylation sites were identified using NetNGlyc v1.0 Server (www.cbs.dtu.dk/services/NetNGlyc/), prediction of transmembrane domains was done using TMHMM Server v2.0 (www.cbs.dtu.dk/services/TMHMM/) and posttranslational GPC processing products were identified using Pfam v31.0 (https://pfam.xfam.org/) as implemented in Geneious [29].

Phylogenetic analysis

Translational alignments of N, Gn, Gc and L genes were performed using the MAFFT E-INS-I algorithm [34]. Modeltest was performed as implemented in MEGA7 to identify the most suitable evolutionary model [35]. Phylogenies were inferred using PhyML with the WAG substitution model as implemented in Geneious R9, with 1000 bootstrap replicates [36].

Virus isolation

Grivet (*Chlorocebus aethiops*) (VeroE6/7) and *Lutzomyia longipalpis* (LL-5) cells were used to perform virus isolation trails from sandfly homogenates as described previously [37].

Table 2. Virus names, designated strains, locality of sandfly specimen collection, sex and species of viruspositive sandflies

Abbreviation of localities are as follows: forest fragment (F), continuous forest (C), forested island (I).

Virus	Strain	Locality	Sex	*Sandfly species*
Tico virus	SP0157/PA/2013	F	f	*Phlebotominae* spp.
Peña Blanca virus	SP1681/PA/2013	C	f	*Phlebotominae* spp.
	SP1683/PA/2013	C	f	*Psychodopygus panamensis*
	SP1684/PA/2013	C	f	*Nyssomyia trapidoi*
La Gloria virus	SP0535/PA/2013	F	f	*Nyssomyia trapidoi*
	SP0538/PA/2013	F	f	*Phlebotominae* spp.
	SP0543/PA/2013	F	m	*Phlebotominae* spp.
	SP0584/PA/2013	F	f	*Psychodopygus panamensis*
	SP1070/PA/2013	C	f	*Nyssomyia trapidoi*
	SP1075/PA/2013	C	f	*Nyssomyia trapidoi*
Mona Grita virus	SP0260/PA/2013	I	f	*Nyssomyia trapidoi*
Tres Almendras virus	SP0412/PA/2013	I	f	*Psychodopygus panamensis*

RESULTS

Sandfly species composition and diversity

In total, 13 692 phlebotomine sandflies were collected in continuous tropical lowland forest (continuous forests), in fragmented forest surrounded by agricultural fields (forest fragments) and in isolated forest on islands in the Panama Canal area, Central Panama (forested islands). Eight different phlebotomine sandfly species were identified, with *Psychodopygus panamensis* as the predominating species (41.6 %) (Table 2). Due to damage of specimens, 28.7 % of the sandflies could not be identified to species level and were summarized as *Phlebotominae* spp..

Great differences were detected in the number of sandflies per landscape type (sandfly abundance). Highest abundance was found in continuous forests (n=9,133, 67 %). Abundance was much lower in forest fragments surrounded by agricultural fields (n=1,490, 15 %) and on forested islands in the Panama Canal (n=2,991, n=18 %). Surprisingly, species richness was similar across the different landscape types, with eight species found in continuous forests and on forested islands in the Panama Canal, and a slightly lower species richness of seven species found in forest fragments surrounded by agricultural fields (Table 2). Except *Bichromomyia olmeca* sp. which was not present in forest fragments, all of the other seven phlebotomine sandfly species were found in all landscape types. Accordingly, no great differences in species diversity entropy (H) and evenness (EH) across landscape types were found, continuous forests H=1.49 and EH=0.51, forest fragments H=1.48 and EH=0.55, and forested islands H=1.56 and EH=0.54. The results from the Kruskal-Wallis H test ($\chi^2(2)=5.527$, p=0.063) suggest that there is no difference between the relative abundance of phlebotomine sandflies among landscape types.

Virus detection and genome analyses

Sandflies were tested for phleboviruses by generic RT-PCR. Sequences with 45–78 % nt identity to the RNA-dependent RNA polymerase (RdRp) gene of phleboviruses were identified in 12 pools (Table 3). The sequences formed five groups with intra-group pairwise nt identities ranging from 96–100 % and inter-group pairwise nt identities ranging from 67–74 %, suggesting the detection of several strains pertaining to five phlebovirus species. Attempts to isolate the viruses in vertebrate (VeroE6/7) and sandfly (LL-5) cells were not successful. Complete coding regions of one strain from each putative species were obtained by NGS and semi-nested RT-PCR from homogenates of infected sandflies. The viruses were tentatively named after the geographic location where the sandflies have been collected, namely La Gloria virus (LAGV), Mona Grita virus (MOGV), Peña Blanca virus (PEBV), Tico virus (TICV) and Tres Almendras virus (TRAV).

All viruses had a tripartite genome organization similar to phleboviruses, comprising a large (L), middle (M) and small (S) segment (Fig. 1). The transduced L segment ORFs of LAGV, MOGV, PEBV, TICV and TRAV showed the highest similarities of 68–91 % to phlebovirus RdRp proteins suggesting the identification of five previously unknown viruses. Bunyaviruses, including phleboviruses, carry the conserved palm motifs Pre-A and A through E on the L segment. These motifs were detected in all five viruses (Fig. 2a). The conserved N-terminal motifs H...D...PD...ExT…K (cation-binding residues and catalytic lysine) of the endonuclease domain of

Table 3. Segment-specific primers used to amplify genome ends

Abbreviations are as follows: Tico virus (TICV), Peña Blanca virus (PEBV), La Gloria virus (LAGV), Mona Grita virus (MOGV) and Tres Almendras virus (TRAV).

Virus	Segment end	Primer sequences	
		First round PCR	Nested round PCR
TICV	M-5'	5'-TGCAGGTTGATCTCAGTTGGG	5'-TTTGAGACCTGGCCGAGTTG
	L-5'	5'-CCACCTTCAGCCCACTATTC	5'-TGGAGGAAGAAATCTGGCAAC
PEBV	M-5'	5'-ACTGATGGCTCGTGAATGGG	5'-ATAACTTGGCTGGCGACTCC
	M-3'	5'-GATTTCGTTCAGTCAGGCGG	5'-CTCATGGATGGGTGGTCCTG
LAGV	S-5'	5'-ACTGAACTGGCAGCTCCTTC	5'-GGTGGCTCTTGTCTTTGTCG
	S-3'	5'-TCACAGGCTTATTCCCACGAG	5'-AGAGCCAGCACAATCATCTTC
	M-5'	5'-TCACCTGATGTCCCGATCTG	5'-TGTCTTAACACTCTGACGGGTG
	L-5'	5'-GGACTGCTGATTAGTGTTGAGC	5'-ACTCAACTCTTAGACTTGATCCC
MOGV	S-5'	5'-GCCTCTAGTCAGAGCGAGAAC	5'-GTCCTCTCTCTTGGACCAGTTG
	M-5'	5'-TTCCACTCTGTTATTTATCCCTTGC	5'-AGTCTGATCCAGAGTAACTGCAC
TRAV	S-5'	5'-CTTGTCTGACATCTTAGTCATCATC	5'-TGGTTTGTTTCCTCGAGTGAG
	S-3'	5'-TGGTGCACTGTGATAGTTGTTG	5'-TTGCGATCAGTAGCACTTCTTG
	M-5'	5'-TCCTGCATCCTGAGAGCTTG	5'-CTTCCCAGTAGTACACTAGCCC
	L-5'	5'-TTGGACCAGACACCTTGTGC	5'-GTGCAGTCAGACGATCTGTTC

phleboviruses were present as well (Fig. 2b). This domain is able to bind and to cleave RNA for transcription processes and is involved in the cap-snatching process typical for negative-sense RNA viral polymerases like in bunya- and orthomyxoviruses [35, 38]. A second bunyavirus typical region, the motif G (RY), was present between aa 668 and aa 675 in all protein sequences of the novel viruses [39]. The conserved arginine is positioned in the polymerase active site to interact with the priming nucleotide triphosphate.

The M segments of LAGV, MOGV, PEBV, TICV and TRAV encode one single ORF of 3992–4277 nt in length with the highest similarity (40–86 %) to glycoprotein precursor genes (GPC) of phleboviruses (Fig. 1). The translated GPC is predicted to be posttranslationally cleaved into the glycoproteins Gn and Gc, and the nonstructural protein NSm. Protein sequence searches using the pfam database revealed a similar predicted molecular weight of 59–60 kDa for both Gn and Gc. The predicted molecular weights of the NSm proteins were more variable, ranging from 33 to 44 kDa.

The S segments of the five putative novel viruses LAGV, MOGV, PEBV, TICV and TRAV contained two ORFs (Fig. 1). The ORF closer to the 3′-end showed highest similarities of 64–81 % to the nucleocapsid gene (N) of phleboviruses. The second ORF is encoded in ambisense, which is typical for phleboviruses, and most likely encodes the nonstructural protein NSs.

Phylogenetic analyses and classification

The novel viruses consistently fall into three main clades in the phylogenetic trees of RdRp, N, Gc and Gn proteins. TRAV branches as a sister taxon to Nique virus (NIQV) within the *Candiru phlebovirus* species-complex (Fig. 3). MOGV is part of the *Frijoles phlebovirus* species-complex as it is placed as a sister taxon to Joá virus (JOAV). PEBV is a novel lineage in basal phylogenetic relationship to the *Bujaru phlebovirus* species-complex, which is defined by Bujaru and Munguba viruses. TICV and LAGV share a common ancestor and are placed as two deep rooting lineages in basal position to the *Bujaru phlebovirus* species-complex.

In order to further analyse if the novel viruses may pertain to *Phlebovirus* species-complexes, the range of intra-species pairwise distances (*p*-distance) was analysed based on

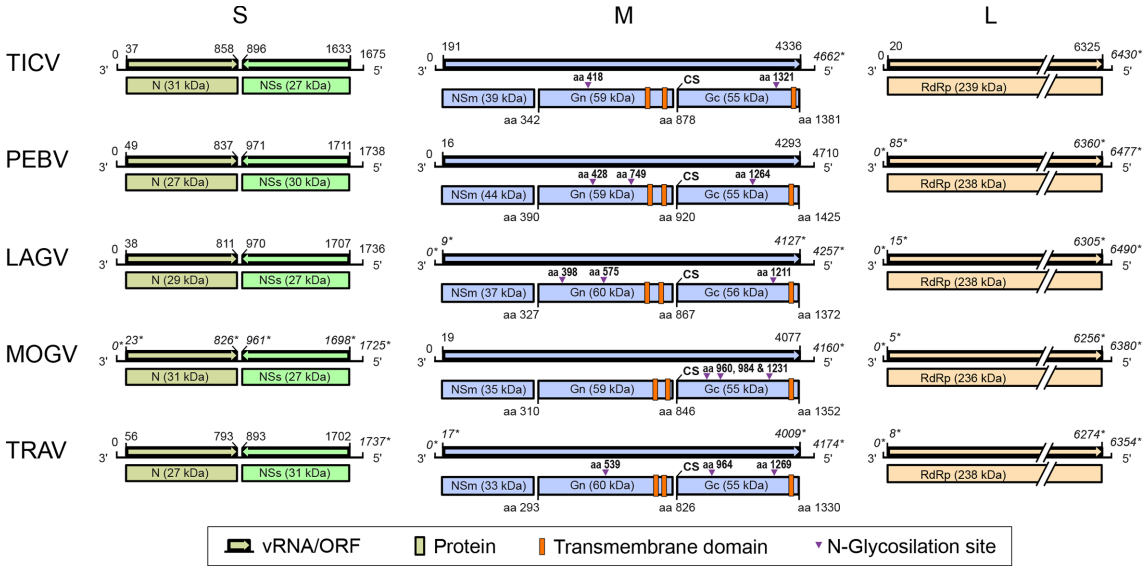

Fig. 1. Schematic genome organization of phleboviruses sequenced in this study. Dark lines represent genome segments, arrows represent coding sequences and boxes represent translated proteins.

Fig. 2. Phleboviral polymerase motifs and endonuclease domain. Sequence alignments of highly conserved motifs within the phleboviral RdRp protein of species-complex prototype viruses and viruses sequenced in this study. Residues conserved throughout all taxa are marked with an asterisk. Highly conserved endonuclease residues are highlighted in red.

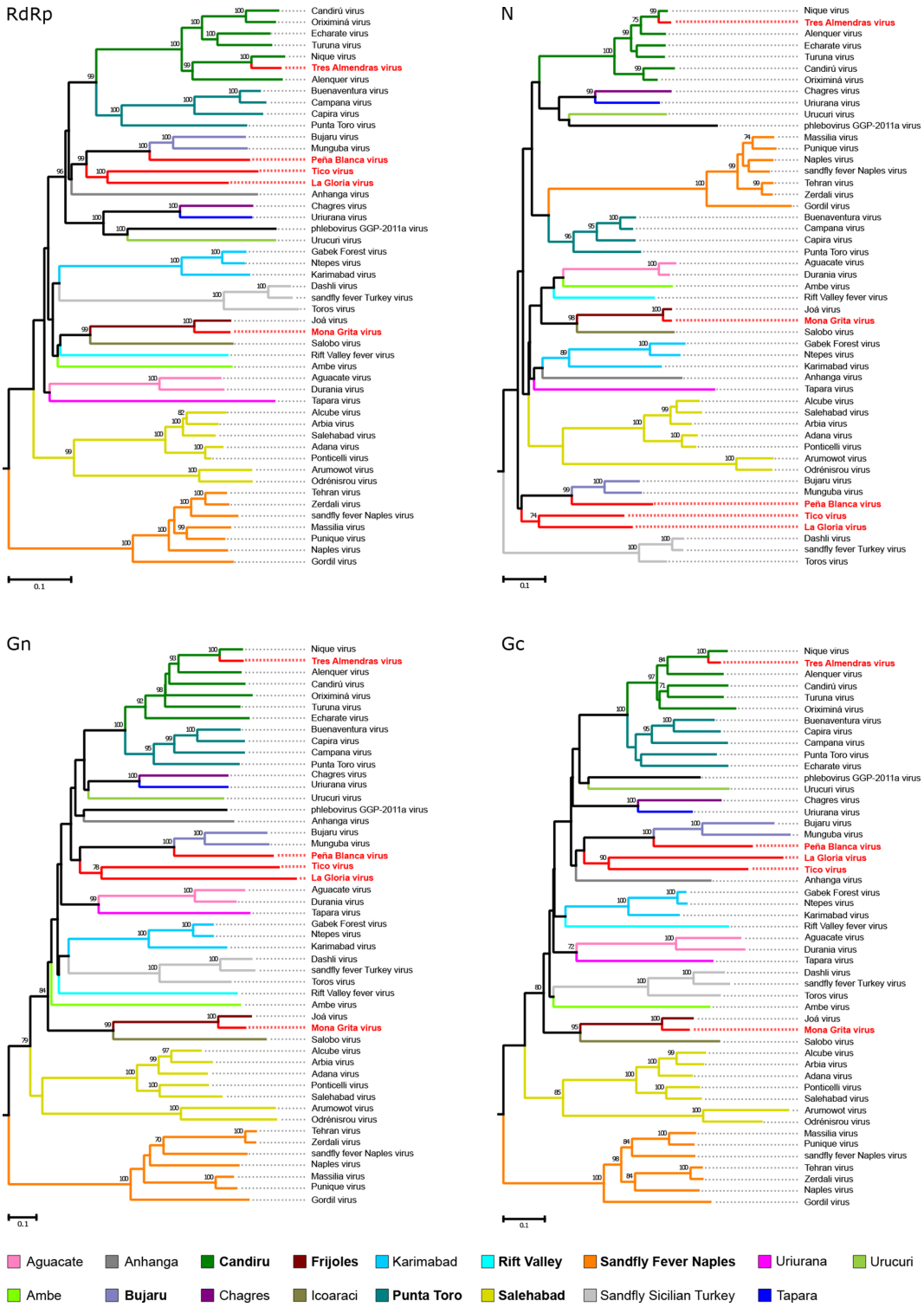

Fig. 3. Phylogenetic analyses of phleboviruses. Maximum-likelihood phylogenies of selected phleboviruses (>5 % aa distance in RdRp proteins) inferred from translated RdRp, N, Gn and Gc genes. Coloured branches represent established (bold type) and suggested species-complexes. Viruses described in this study are shown in red.

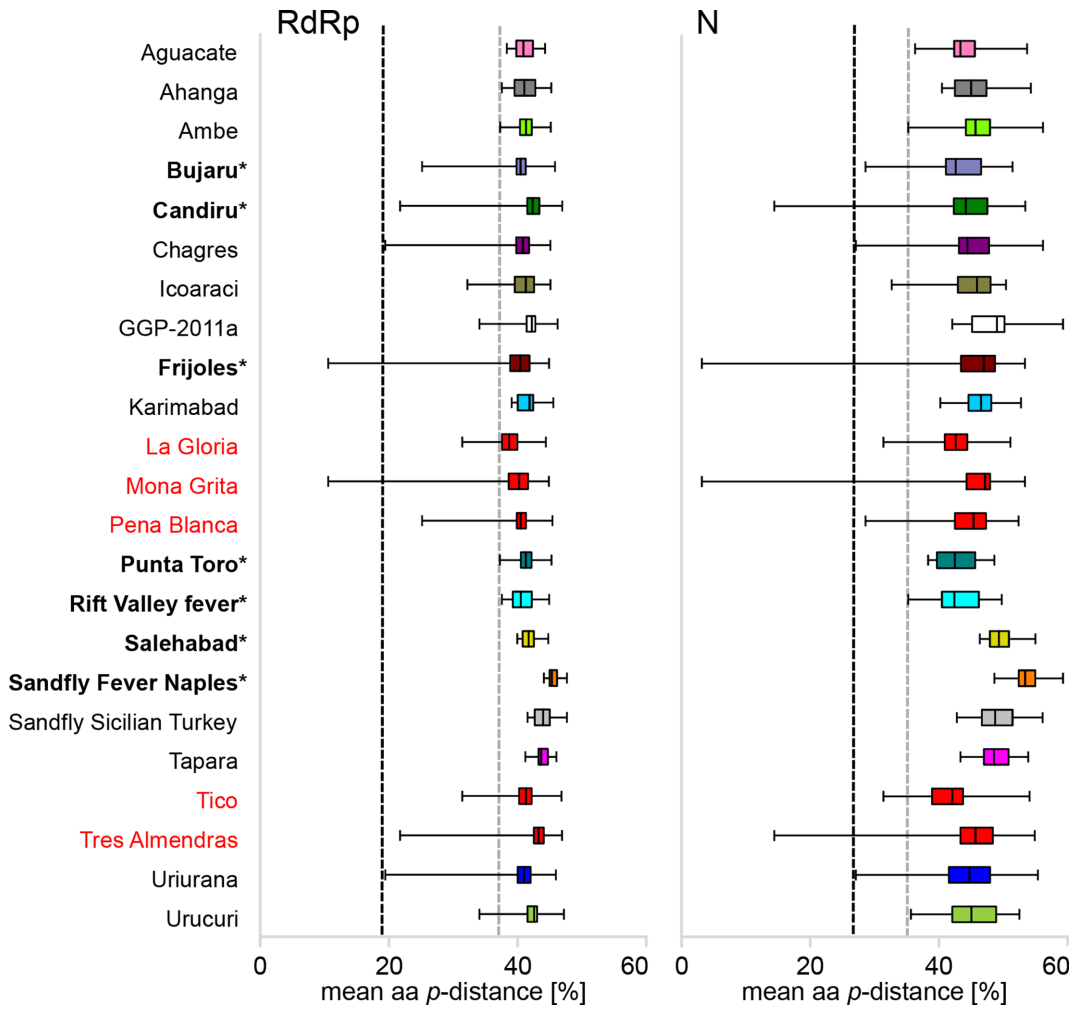

Fig. 4. Mean pairwise distance among established and suggested phlebovirus species-complexes and phleboviruses sequenced in this study. Mean aa *p*-distance comparison of RdRp and N protein among established (bold, asterisk) and suggested *Phlebovirus* species-complexes and viruses sequenced in this study (red). Whiskers represent the data range with the whisker top representing the highest and the whisker bottom the lowest values. Dashed lines represent cut-off value based on projected serological cross reactivity to genetic level using established (grey) and including suggested species-complexes (black).

complete RdRp and N protein sequences among established and suggested species-complexes. The grouping based on genetic distances confirms the phylogeny-based grouping (Fig. 4). The analyses of *p*-distances for established species-complexes revealed 38.4 and 37.6 % as the lowest *p*-distance values among RdRp and N proteins respectively, agreeing with the currently existing genetic cut-off value for members of established species-complexes (shown as a grey dashed line in Fig. 4). Using these cut-off values, three of the novel viruses fall clearly into the diversity of established species-complexes, while two of them show borderline values to be grouped into established species-complexes. MOGV, TRAV and PEBV group with members of the *Frijoles phlebovirus* (*p*-distances of 10.6 and 3.4 % in RdRp and N proteins, respectively), *Candiru phlebovirus* (*p*-distances of 21.7 and 14.5 % in RdRp and N proteins, respectively), and *Bujaru phlebovirus* species-complexes (*p*-distances of 25.2 and 28.6 % in RdRp and N proteins, respectively). LAGV (*p*-distance of 33.7 and 35.2 in RdRp and N proteins, respectively) and TICV (*p*-distance of 36.1 and 36.4 in RdRp and N proteins, respectively) show borderline values to be grouped within the *Bujaru phlebovirus* species-complex.

To further test if LAGV and TICV group with other described phleboviruses, we extended our analysis and included the sequences of more than 15 putative novel phlebovirus species that have not yet been officially classified by the ICTV. These viruses are suggested to form 11 novel phlebovirus species-complexes. Among these, the closely related CHGV and Uriurana virus (URIV) have been suggested to establish distinct virus species-complexes as they do not show serological cross reactivity [14, 18]. Our analyses compare CHGV and URIV to established and suggested species complexes, and reveal *p*-distances of 19.5 and 27.1 % in the RdRp and N

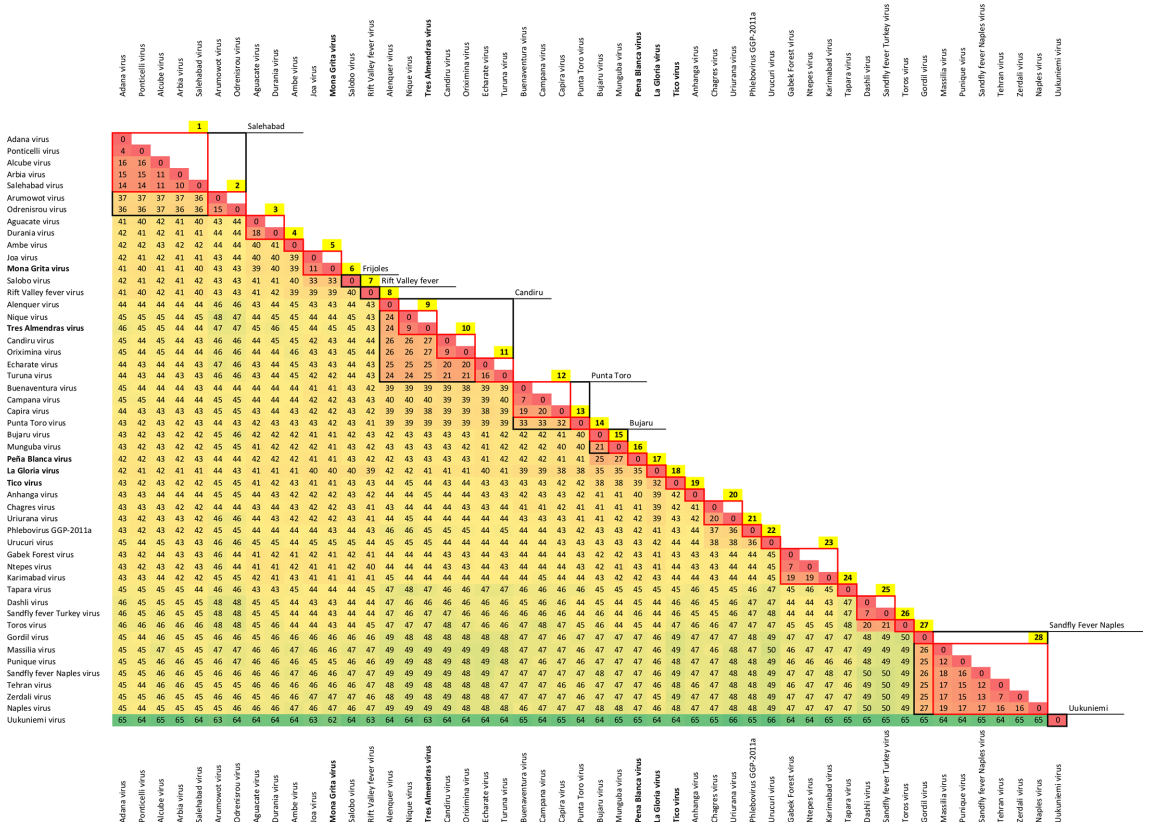

Fig. 5. Distance matrix of phleboviruses. Distance matrix of the RdRp protein of selected phleboviruses (<95 % aa identity, RdRp) and viruses sequenced in this study (bold). ICTV classified species-complexes are illustrated by black frames. Shades of green to yellow to red are shown in accordance to the percentage of aa distance (green=66 %; yellow=31 %; dark red=4 % aa distance). Putative species-complexes applying ≤20 % aa distance as the cut-off value are numbered by yellow boxes and framed in red.

proteins, respectively. These values represent the lowest values among all established and suggested species complexes and may define a novel cut-off for species-complex demarcation (shown as black, dashed line in Fig. 4).

In a next step, we sought to identify genetic-based species demarcation criteria. As the intra-species genetic diversity of phleboviruses is unknown, we used *Rift Valley fever virus* (RVFV) as a surrogate since it represents the only established phlebovirus species in a phylogenetic solitary position, from which numerous strains have been sequenced. We compared the pairwise distances among all available RVFV strains in their RdRp genes. The analysis of 154 RVFV strains reveals maximum pairwise nucleotide and aa intraspecies genetic distances of 5 and 2 %, respectively. Next, all available Diptera-associated phlebovirus protein sequences, which were ≤5 % different from the closest related sequence on aa level, were selected as they are likely to present distinct species. Comparing the pairwise distances among the selected sequences reveals that all known Diptera-associated phleboviruses form 48 distinct species when using a conservative cut-off value of 5 % aa distance as species demarcation criterion. These species may be grouped into 28 species-complexes when applying a cut-off value of 20 % as species-complex demarcation criterion (Fig. 5).

Virus-vector associations

Virus infection rates were very low and prevented comprehensive analyses of virus prevalence patterns in the different types of landscape. In general, virus abundance was mainly linked to a geographic location. For example, two of the five viruses have been exclusively detected in sandflies collected on forested islands, MOGV in *Nyssomyia trapidoi* and TRAV in *Psychodopygus panamensis,* although both vector species were found in high numbers in all landscape types (Table 3). Furthermore, TICV was detected in one pool of *Phlebotominae* spp. collected in a forest fragment, and PEBV was found in three pools of sandflies collected in continuous forest containing specimens of *Psychodopygus panamensis, Nyssomyia trapidoi* and *Phlebotominae* spp.. LAGV was detected in six pools also consisting of *Psychodopygus panamensis, Nyssomyia trapidoi* and *Phlebotominae* spp. originating from continuous forest and forest fragments. LAGV was also detected in a pool of male *Phlebotominae* spp. sandflies suggesting a vertical transmission. The results from the correspondence analysis indicate that phlebotomine sandfly

species composition and habitat type are not significantly associated ($\chi^2(2)=5.527$, p=0.063, df=16).

DISCUSSION

Insects originating from biodiverse tropical forests have been shown to be infected with genetically highly diverse viruses. While the majority of viruses is found in mosquitoes [17, 37, 40, 41] other blood-feeding arthropods such as sandflies are largely unexplored for viral infections. Here, we collected sandflies in remote forest areas of the Panama Canal area and screened them for infections with phleboviruses. We detected and genetically characterized five previously unknown phleboviruses found in *Psychodopygus panamensis* and *Nyssomyia trapidoi* species. Although a comparatively large number of sandflies was analysed (*n*=13 807), which have been collected within 1 year, we did not detect any of the sandfly-associated viruses known to occur in Panama. The latter viruses are mainly found in urban areas and their prevalence may be low in forested areas with comparatively little landscape modification.

Continuous forests showed the highest abundance of sandflies overall, which may be due to the extensive presence of leaf-litter with decomposing organic matter for larval feeding, plus abundant burrows and crevices that serve as optimal breeding and resting sites for adult phlebotomine sandflies in more humid and undisrupted forest environments [25]. We did not find differences in phlebotomine sandfly species richness among landscape types, suggesting that phlebotomine sandflies are probably more resilient to forest fragmentation than other blood-feeding dipterans [42]. However, this could also be an artefact of our sampling strategy as CDC light traps tend to over select for certain species. Hence, community metrics might be better estimated using a combination of sampling strategies instead of using a single trapping method. *Bichromomyia olmeca* sp. was the only taxon exclusively found in continuous forest. The absence of *Bichromomyia olmeca* sp. from other landscape types is in line with findings by Fairchild and Theodor [43]. Our results, in terms of phlebotomine sandfly species richness and overall species abundance, are very similar to those reported by Azpurua and colleagues who also investigated sandfly community structure and pathogen transmission in the Panama Canal area using CDC light traps [5]. However, like Azpurua *et al.*, we only surveyed for phlebotomine sandflies at night, therefore diurnal species were not part of our analysis. Further studies will have to corroborate whether diurnal species fit the community outcomes presented here.

PEBV, TRAV and LAGV were detected in *Psychodopygus panamensis,* which is a known vector of NIQV in Panama [17]. NIQV belongs to the *Candiru phlebovirus* species-complex and is the closest relative to TRAV. Some members of the *Candiru phlebovirus* species-complex have been shown to infect humans [17]. Whether TRAV can also infect humans and cause symptoms of disease needs further investigation. MOGV groups as a sister taxon to JOAV within the *Frijoles phlebovirus* species-complex and was detected in *Nyssomyia trapidoi*. JOAV was isolated from an unidentified species of phlebotomine sandflies in Brazil [18]. The vector of TICV could not be not identified.

Due to the huge expansion of sequence information from phleboviruses, which were not isolated in cell culture, there is an urgent need to establish genetic-based classification criteria. Until recently, the classification of phleboviruses into species-complexes was only based on the presence or absence of serological cross-reactivity. Distinct species were differentiated by fourfold difference in neutralization tests, which requires the availability of reference viruses and antisera, as well as high laboratory safety standards. To circumvent this, haemagglutination inhibition (HI) and complement fixation (CI) tests have been widely used to test for serological cross-reactivity, leading to the establishment of species-complexes as these tests cannot differentiate between serologically closely related viruses. In 2015, the ICTV has started an extensive reorganization of the bunyavirus taxonomy, including the genus *Phlebovirus*. The genomes of nearly all phlebovirus isolates that had only been characterized by serology were sequenced and classified into distinct viruses (designated member 'virus' by the ICTV), which were assigned to species-complexes (designated 'species' by the ICTV). This classification largely reflects the previous serology-based classification [11]. Although, the genus *Phlebovirus* currently contains nine species-complexes, which comprise 57 recognized viruses, no genetic-based species demarcation criteria have been defined yet. In addition, the grouping of viruses into species-complexes seems to be a rather rough classification, as for example the *Salehabad phlebovirus* species-complex is extremely diverse and includes viruses with distances of up to 37 % within their RdRp proteins. Like in orthonairo- and orthohantaviruses, the members of the genus *Phlebovirus* are not classified in species but assigned to species-complexes. Our analyses of the intraspecies genetic diversity of 154 RVFV strains yielded a maximum distance of 2 % distance in RdRp protein sequences. Thus, a species demarcation criterion of 5 % aa distance between RdRp protein sequences would represent a rather conservative cut-off value. Such a criterion would be in line with the recently discovered novel sandfly-borne Ntepes virus (NPV), that groups with other members of the so far unclassified '*Karimabad phlebovirus*' species-complex [44]. NPV is at least 7 % distant in the RdRp protein from other phleboviruses (Fig. 5) and did not react with related viruses in neutralization assays suggesting that it defines a distinct species [45]. Furthermore, distinct species in a related bunyavirus genus, the genus *Orthobunyavirus*, shall be differentiated by at least 4 % distance in their RdRp proteins [38]. Taken together, these data suggest that cut-off values of 4–5 % aa distance within RdRp proteins are appropriate in a biological sense as species demarcation criteria.

Our phylogenetic analyses differed from the one presented by Nunes-Neto and colleagues [18]. In their analyses, Uriurana virus groups together with Tapara virus and the authors propose to classify both viruses into a single species-complex. In contrast, our phylogenetic analysis places Uriurana virus as a sister taxon to Chagres virus (Fig. 3), a topology which

has also been found by others [19]. Chagres virus has been isolated from a human in 1960 from Panama [14] and Uriurana virus was isolated in 1985 from Phlebotominae sp. collected in Brazil [18]. Despite their long history, both viruses are not officially classified by the ICTV. According to their genetic distance of 20 % in their RdRp protein sequences (Fig. 5), Uriurana virus and Chagres virus seem to be two distinct species of a single species-complex, tentatively named '*Chagres phlebovirus*' species-complex.

The detection and genomic characterization of five novel viruses identified in Panamanian phlebotomine sandflies species show that the taxonomic diversity of phleboviruses is larger than what has been reported before in the Panama Canal area. In this study, *Psychodopygus panamensis* and *Nyssomyia trapidoi* were the most prevalent species showing the highest overall proportions regardless of landscape type. The fact that both species have been previously incriminated as vectors of human pathogens in Panama, combined with their considerable abundance, is very suggestive of their involvement in the transmission of LAGV, MOGV, PEBV, TICV and TRAV. However, future studies will have to investigate if these sandfly species are competent vectors, as well as if these viruses infect mammals. Our findings underline the importance to extend surveillance to blood-feeding arthropods other than mosquitoes, including hard ticks, biting-midges and sandflies.

Funding information
This research was funded by the German Science Foundation (DFG) and is part of the DFG Priority Program SPP 1596 Ecology and Species Barriers in Emerging Infectious Diseases (grant JU 2857/3-1 and JU 2857/3-2 to S.J.). Larissa C. Dutari was supported by the National Secretariat of Science, Technology and Innovation of Panama (SENACYT), through the international internship scholarship with code APY-GC-2014-031 (Convocations for the Generation of Scientific and Technological Capabilities 2014 - 2015).

Acknowledgements
We thank Gloria Marklewitz and Bastian Sauer for their support in sandfly collection. For granting us permission to perform field-work on their terrains we are grateful to the private landowners. We also thank Stefan Brändel for his help with the research permit applications. For excellent technical assistance we would like to acknowledge Pascal Trippner and Tobias Bleicker. Finally, we thank the Ministry of Environment of Panama (MiAmbiente) for granting the permits for collection and export of biological samples.

Author contributions
S. J. conceived and designed the study. M. M. and R. A. P. coordinated and M. M. performed field work. L. C. D. identified the sandfly species. M. M. and L. C. D. performed the experiments. M. M. analysed the data. M. M. and L. C. D. wrote the first draft of the manuscript. S. P. performed the bioinformatic analyses. All authors played a vital part in the preparation and revision of the final version of the manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Shimabukuro PHF, de Andrade AJ, Galati EAB. Checklist of American sand flies (Diptera, Psychodidae, Phlebotominae): genera, species, and their distribution. *Zookeys* 2017;660:67–106.
2. Ready PD. Biology of phlebotomine sand flies as vectors of disease agents. *Annu Rev Entomol* 2013;58:227–250.
3. Telford SR, Herrer A, Christensen HA. Enzootic cutaneous leishmaniasis in eastern Panama. *Annals of Tropical Medicine & Parasitology* 2016;66:173–179.
4. Christensen HA, Herrer A. Development of a Panamanian strain of *Leishmania mexicana* in co-indigenous *Lutzomyia sanguinaria* and *Lu. gomezi* (Diptera: Psychodidae). *J Med Entomol* 1980;17:188–189.
5. Azpurua J, De La Cruz D, Valderama A, Windsor D. *Lutzomyia* sand fly diversity and rates of infection by *Wolbachia* and an exotic *Leishmania* species on Barro Colorado Island, Panama. *PLoS Negl Trop Dis* 2010;4:e627.
6. Palacios G, da Rosa AT, Savji N, Sze W, Wick I *et al*. Aguacate virus, a new antigenic complex of the genus *Phlebovirus* (family *Bunyaviridae*). *J Gen Virol* 2011;92:1445–1453.
7. Palacios G, Wiley MR, Travassos da Rosa APA, Guzman H, Quiroz E, Ladner JT, Tesh RB *et al*. Characterization of the Punta Toro species complex (genus *Phlebovirus*, family *Bunyaviridae*). *J Gen Virol* 2015;96:2079–2085.
8. Tesh RB, Chaniotis BN, Peralta PH, Johnson KM. Ecology of viruses isolated from Panamanian phlebotomine sandflies. *Am J Trop Med Hyg* 1974;23:258–269.
9. Tesh RB, Boshell J, Young DG, Morales A, Ferra de Carrasquilla C, De Rodriguez C, Travassos da rosa APA *et al*. Characterization of five new phleboviruses recently isolated from sand flies in tropical America. *Am J Trop Med Hyg* 1989;40:529–533.
10. Dutari LC, Loaiza JR. American cutaneous leishmaniasis in Panama: a historical review of entomological studies on anthropophilic *Lutzomyia* sand fly species. *Parasit Vectors* 2014;7:218.
11. Maes P, Adkins S, Alkhovsky SV, Avšič-Županc T, Ballinger MJ *et al*. Taxonomy of the order *Bunyavirales*: second update 2018. *Arch Virol* 2019;164:927–941.
12. Gundacker ND, Carrera J-P, Castillo M, Díaz Y, Valenzuela J, Valenzuela J, Pascale JM, Tamhane A *et al*. Clinical manifestations of Punta Toro virus species complex infections, Panama, 2009. *Emerg Infect Dis* 2017;23:872–874.
13. Baldelli F, Ciufolini MG, Francisci D, Marchi A, Venturi G *et al*. Unusual presentation of life-threatening Toscana virus meningoencephalitis. *Clin Infect Dis* 2004;38:515–520.
14. Peralta PH, Shelokov A, Brody JA. Chagres virus: a new human isolate from Panama. *Am J Trop Med Hyg* 1965;14:146–151.
15. Lambert AJ, Lanciotti RS. Consensus amplification and novel multiplex sequencing method for S segment species identification of 47 viruses of the *Orthobunyavirus*, *Phlebovirus*, and *Nairovirus* genera of the family *Bunyaviridae*. *J Clin Microbiol* 2009;47:2398–2404.
16. Xu F, Liu D, Nunes MRT, DA Rosa APAT, Tesh RB, DA ROSA APAT, *et al*. Antigenic and genetic relationships among Rift Valley fever virus and other selected members of the genus *Phlebovirus* (*Bunyaviridae*). *Am J Trop Med Hyg* 2007;76:1194–1200.
17. Palacios G, Tesh R, Travassos da Rosa A, Savji N, Sze W *et al*. Characterization of the Candiru antigenic complex (*Bunyaviridae:* phlebovirus), a highly diverse and reassorting group of viruses affecting humans in tropical America. *J Virol* 2011;85:3811–3820.
18. Nunes-Neto JP, Souza WMde, Acrani GO, Romeiro MF, Fumagalli M *et al*. Characterization of the Bujaru, frijoles and Tapara antigenic complexes into the sandfly fever group and two unclassified phleboviruses from Brazil. *J Gen Virol* 2017;98:585–594.
19. de Carvalho MS, de Lara Pinto AZ, Pinheiro A, Rodrigues JSV, Melo FL *et al*. Viola phlebovirus is a novel Phlebotomus fever serogroup member identified in *Lutzomyia* (*Lutzomyia*) *longipalpis* from Brazilian Pantanal. *Parasit Vectors* 2018;11:405.
20. Xu F, Chen H, Travassos da Rosa APA, Tesh RB, Xiao S-Y. Phylogenetic relationships among sandfly fever group viruses (*Phlebovirus: Bunyaviridae*) based on the small genome segment. *J Gen Virol* 2007;88:2312–2319.
21. Rompré G, Robinson WD, Desrochers A. Causes of habitat loss in a Neotropical landscape: the Panama canal corridor. *Landsc Urban Plan* 2008;87:129–139.

22. Condit R, Robinson WD, Ibáñez R, Aguilar S, Sanjur A *et al*. The status of the Panama canal watershed and its biodiversity at the beginning of the 21st century: long-term ecological studies reveal a diverse flora and fauna near the Panama canal, harbored within a corridor of forest stretching from the Caribbean to the Pacific, but deforestation, land degradation, erosion, and overhunting remain threats. *Bioscience* 2001;51:389–398.
23. Panama Canal Authority. 2010. A history of the Panama canal: French and American construction efforts. http://www.pancanal.com/eng/history/history/index.html [accessed 12 February 2019].
24. Schmid J, Rasche A, Eibner G, Jeworowski L, Page RA *et al*. Ecological drivers of *Hepacivirus* infection in a neotropical rodent inhabiting landscapes with various degrees of human environmental change. *Oecologia* 2018;188:289–302.
25. Young DG, Duncan MA. *Guide to the identification and geographic distribution of Lutzomyia sand Flies in Mexico, the West Indies, Central and South America*. Gainesville, Florida, USA: Associated Publishers; 1994.
26. Chaniotis BN. Use of external characters for rapid identification of phlebotomine sandflies in vector studies. *J Med Entomol* 1974;11:501–501.
27. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948;27:623–656.
28. Simpson EH. Measurement of diversity. *Nature* 1949;163:688.
29. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M *et al*. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012;28:1647–1649.
30. Song L, Florea L, Langmead B. Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biol* 2014;15:509.
31. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–595.
32. 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.
33. Goodacre N, Aljanahi A, Nandakumar S, Mikailov M, Khan AS *et al*. A reference viral database (RVDB) to enhance bioinformatics analysis of high-throughput sequencing for novel virus detection. *mSphere* 2018;3:00069–18.
34. Katoh K, Misawa K, Kuma K-ichi, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30:3059–3066.
35. Reguera J, Malet H, Weber F, Cusack S. Structural basis for encapsidation of genomic RNA by *La Crosse* Orthobunyavirus nucleoprotein. *Proceedings of the National Academy of Sciences* 2013;110:7246–7251.
36. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;52:696–704.
37. Marklewitz M, Zirkel F, Kurth A, Drosten C, Junglen S. Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. *Proc Natl Acad Sci U S A* 2015;112:7536–7541.
38. Reguera J, Gerlach P, Rosenthal M, Gaudon S, Coscia F *et al*. Comparative structural and functional analysis of bunyavirus and arenavirus Cap-Snatching endonucleases. *PLoS Pathog* 2016;12:e1005636.
39. Gerlach P, Malet H, Cusack S, Reguera J. Structural insights into bunyavirus replication and its regulation by the vRNA promoter. *Cell* 2015;161:1267–1279.
40. Hermanns K, Zirkel F, Kopp A, Marklewitz M, Rwego IB *et al*. Discovery of a novel alphavirus related to Eilat virus. *J Gen Virol* 2017;98:43–49.
41. Carrera J-P, Guzman H, Beltrán D, Díaz Y, López-Vergès S *et al*. Mercadeo virus: a novel mosquito-specific flavivirus from Panama. *Am J Trop Med Hyg* 2015;93:1014–1019.
42. Loaiza JR, Dutari LC, Rovira JR, Sanjur OI, Laporta GZ *et al*. Disturbance and mosquito diversity in the lowland tropical rainforest of central Panama. *Sci Rep* 2017;7:7248.
43. Fairchild GB, Theodor O. On *Lutzomyia flaviscutellata* (Mangabeira) and *L. Olmeca* (Vargas and Diaz-Najera) (Diptera: Psychodidae). *J Med Entomol* 1971;8:153–159.
44. Tchouassi DP, Marklewitz M, Chepkorir E, Zirkel F, Agha SB *et al*. Sand fly-associated phlebovirus with evidence of neutralizing antibodies in humans, Kenya. *Emerg Infect Dis* 2019;25:681–690.
45. Lambert AJ, Adkins S, Alkhovsky SV, Beer M, Blair CD *et al*. Thirty-eight new species within the genus *Orthobunyavirus*. *Pending Proposal awaiting ICTV ratification*. https://talk.ictvonline.org/files/proposals/animal_dsrna_and_ssrna-_viruses/m/animal_rna_minus_ec_approved/7656 2019 [accessed 2 December 2019].