



# Viral discovery and diversity in trypanosomatid protozoa with a focus on relatives of the human parasite *Leishmania*

Danyil Grybchuk<sup>a</sup>, Natalia S. Akopyants<sup>b,1</sup>, Alexei Y. Kostygov<sup>a,c,1</sup>, Aleksandras Konovalovas<sup>d</sup>, Lon-Fye Lye<sup>b</sup>, Deborah E. Dobson<sup>b</sup>, Haroun Zangger<sup>e</sup>, Nicolas Fasel<sup>e</sup>, Anzhelika Butenko<sup>a</sup>, Alexander O. Frolov<sup>c</sup>, Jan Votýpka<sup>f,9</sup>, Claudia M. d'Avila-Levy<sup>h</sup>, Pavel Kulich<sup>i</sup>, Jana Moravcová<sup>j</sup>, Pavel Plevka<sup>j</sup>, Igor B. Rogozin<sup>k</sup>, Saulius Serva<sup>d,l</sup>, Julius Lukeš<sup>g,m</sup>, Stephen M. Beverley<sup>b,2,3</sup>, and Vyacheslav Yurchenko<sup>a,g,n,2,3</sup>

<sup>a</sup>Life Science Research Centre, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic; <sup>b</sup>Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, MO 63110; <sup>c</sup>Zoological Institute of the Russian Academy of Sciences, St. Petersburg, 199034, Russia; <sup>d</sup>Department of Biochemistry and Molecular Biology, Institute of Biosciences, Life Sciences Center, Vilnius University, Vilnius 10257, Lithuania; <sup>e</sup>Department of Biochemistry, University of Lausanne, 1066 Epalinges, Switzerland; <sup>f</sup>Department of Parasitology, Faculty of Science, Charles University, 128 44 Prague, Czech Republic; <sup>g</sup>Biology Centre, Institute of Parasitology, Czech Academy of Sciences, 370 05 České Budějovice, Czech Republic; <sup>h</sup>Coleção de Protozoários, Laboratório de Estudos Integrados em Protozoologia, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, 21040-360 Rio de Janeiro, Brazil; <sup>i</sup>Veterinary Research Institute, 621 00 Brno, Czech Republic; <sup>j</sup>Central European Institute of Technology – Masaryk University, 625 00 Brno, Czech Republic; <sup>k</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894; <sup>l</sup>Department of Chemistry and Bioengineering, Faculty of Fundamental Sciences, Vilnius Gediminas Technical University, Vilnius 10223, Lithuania; <sup>m</sup>University of South Bohemia, Faculty of Sciences, 370 05 České Budějovice, Czech Republic; and <sup>n</sup>Institute of Environmental Technologies, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic

Contributed by Stephen M. Beverley, November 26, 2017 (sent for review October 11, 2017; reviewed by Said A. Ghabrial, Max L. Nibert, and Larry Simpson)

Knowledge of viral diversity is expanding greatly, but many lineages remain underexplored. We surveyed RNA viruses in 52 cultured monoxenous relatives of the human parasite *Leishmania* (*Crithidia* and *Leptomonas*), as well as plant-infecting *Phytomonas*. *Leptomonas pyrrocoris* was a hotbed for viral discovery, carrying a virus (*Leptomonas pyrrocoris ostravirus 1*) with a highly divergent RNA-dependent RNA polymerase missed by conventional BLAST searches, an emergent clade of tombus-like viruses, and an example of viral endogenization. A deep-branching clade of trypanosomatid narnaviruses was found, notable as *Leptomonas seymouri* bearing Narnalike virus 1 (LepseyNLV1) have been reported in cultures recovered from patients with visceral leishmaniasis. A deep-branching trypanosomatid viral lineage showing strong affinities to bunyaviruses was termed “*Leishbunyavirus*” (LBV) and judged sufficiently distinct to warrant assignment within a proposed family termed “*Leishbunyaviridae*.” Numerous relatives of trypanosomatid viruses were found in insect metatranscriptomic surveys, which likely arise from trypanosomatid microbiota. Despite extensive sampling we found no relatives of the totivirus *Leishmanivirus* (LRV1/2), implying that it was acquired at about the same time the *Leishmania* became able to parasitize vertebrates. As viruses were found in over a quarter of isolates tested, many more are likely to be found in the >600 unsurveyed trypanosomatid species. Viral loss was occasionally observed in culture, providing potentially isogenic virus-free lines enabling studies probing the biological role of trypanosomatid viruses. These data shed important insights on the emergence of viruses within an important trypanosomatid clade relevant to human disease.

Trypanosomatidae | Bunyavirales | persistent virus infection | coinfection | coevolution

The ability of viruses to infect virtually any cellular life form on Earth contributes to their immense diversity. While many eukaryotic groups have been probed for the viral presence, the full diversity of viruses remains to be explored (1). Especially promising is the investigation of RNA viruses in simple eukaryotes such as fungi, green algae, diatoms, slime molds, oomycetes, dinoflagellates, apicomplexans, kinetoplastids, diplomonads, and trichomonads (2–4). While originally considered to be little more than evolutionary curiosities, these viruses have started to attract more attention as their important biological roles are now emerging. For example, *Cryphonectria hypovirus 1* plays a key role in limiting pathogenicity to its fungal hosts, with applications

toward biological control (5), and several viruses of the family *Totiviridae* have been implicated in the increased pathogenicity of parasitic protozoa to vertebrate hosts (6, 7).

Most studies reporting unicellular eukaryotic viruses arose from fortuitous discovery of virus-like particles (VLPs) or abundant discrete RNA segments rather than from systematic searches often termed “virus hunting.” Here we present a broad survey of RNA viruses within trypanosomatids, one of the iconic groups of protists. Members of the family Trypanosomatidae exhibit strikingly unusual molecular and biochemical traits (8–12). Several species cause widespread severe illnesses, such as sleeping

## Significance

Largely overlooked, the viruses of protists have started to attract more attention. Several viruses of the family *Totiviridae* are currently implicated in the increased pathogenicity of parasitic protozoa such as *Leishmania* to vertebrate hosts. We conducted a broad survey of RNA viruses within trypanosomatids, one of the iconic groups of protists. These revealed several previously unidentified viral taxa including one designated “*Leishbunyaviridae*” and a highly divergent virus termed “*Leptomonas pyrrocoris ostravirus 1*.” Our studies provide important information on the origins as well as the diversity and distribution of viruses within a group of protists related to the human parasite *Leishmania*.

Author contributions: D.G., N.S.A., A.Y.K., N.F., P.P., S.S., S.M.B., and V.Y. designed research; D.G., N.S.A., A.Y.K., A.K., L.-F.L., D.E.D., H.Z., P.K., and J.M. performed research; A.O.F., J.V., C.M.d.-L., and J.L. contributed new reagents/analytic tools; D.G., N.S.A., A.Y.K., A.B., I.B.R., S.M.B., and V.Y. analyzed data; and D.G., A.Y.K., S.M.B., and V.Y. wrote the paper.

Reviewers: S.A.G., University of Kentucky; M.L.N., Harvard Medical School; and L.S., University of California, Los Angeles.

The authors declare no conflict of interest.

Published under the PNAS license.

Data deposition: The deposition and accession numbers for the sequences reported in this work are summarized in Table S3. Metatranscriptomic contigs assembled from single-read archive depositions, which cannot be deposited in GenBank because they have not been experimentally verified, are listed in Table S4.

<sup>1</sup>N.S.A. and A.Y.K. contributed equally to this work.

<sup>2</sup>S.M.B. and V.Y. contributed equally to this work.

<sup>3</sup>To whom correspondence may be addressed. Email: stephen.beverley@wustl.edu or vyacheslav.yurchenko@osu.cz.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1717806115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1717806115/-DCSupplemental).

sickness, Chagas disease, and kala-azar in humans (13). Monoxenous (with one host) parasites of invertebrates (primarily insects) were ancestors of these dixenous (with two hosts) pathogens and still represent the majority of trypanosomatid lineages (14, 15). Phylogenetic analysis of the Trypanosomatidae has shown convincingly that the transition from a monoxenous to a dixenous state occurred at least three times, giving rise to the genera *Trypanosoma* and *Leishmania* (both parasites of vertebrates), as well as plant-dwelling *Phytomonas* (16).

VLPs were reported from a number of trypanosomatid species including *Endotrypanum schaudinni*, *Leishmania hertigi* [now classified as *Paraleishmania hertigi* (17)], *Phytomonas* spp., *Criethidia pragensis*, *Leptomonas seymouri*, *Angomonas desouzai*, and others (18–23). The molecular era in the research of trypanosomatid viruses began with the pioneering studies of those found in South American *Leishmania* spp. including *Leishmania RNA virus 1* (LRV1) from *Leishmania guyanensis* and *Leishmania braziliensis* (24, 25), and an unrelated RNA virus in *Phytomonas* (21). The biological significance of these lays fallow until the finding that LRV1 was associated with increased disease pathology, parasite numbers, and immune response in animal models (6, 26–29). Subsequent studies provided evidence linking LRV1 to the severity of human leishmaniasis, including acute pathology and drug-treatment failures (30–33), although data relating the viral presence to the chronic mucocutaneous leishmaniasis are mixed (32, 34–36).

Recently, molecular descriptions have been made for the viruses from several additional trypanosomatid species. Among them were a bunyavirus-like virus of *Leptomonas moramango* (37) as well as narnavirus-like viruses of *Leptomonas seymouri* (38) and the dixenous plant pathogen *Phytomonas serpens* (39). Provocatively, *Leptomonas seymouri* has been recovered from cultures from visceral leishmaniasis patients infected with *Leishmania donovani*, and many of such *Leptomonas seymouri* strains bear NLV1 (40). Thus, there appears to be considerable unexplored viral diversity in trypanosomatids, the study of which may contribute to our understanding of the biology of trypanosomatids and their insect and/or plant hosts as well as the origins of viruses in *Leishmania*.

## Results and Discussion

**Screening of Trypanosomatid Isolates.** We surveyed 52 isolates including 44 belonging to the genera *Criethidia* and *Leptomonas* (subfamily Leishmaniinae), as well as eight belonging to *Phytomonas* spp. These originated from diverse insect or plant hosts and geographic regions (Table S1). Total RNA from these isolates was digested with S1 nuclease, removing most cellular RNAs, after which the remaining dsRNA arising from dsRNA viruses or replicative intermediates of ssRNA viruses could then be sensitively detected by gel electrophoresis (see Figs. 1A, 3A, and 4A) (41). From this analysis, 11 Leishmaniinae and three *Phytomonas* spp. exhibited dsRNA bands, while the remainder appeared to lack them (Table 1). Most RNA segments were sequenced, and the sequences of those encoding viral RNA-dependent RNA polymerase (RDRP) were used to assign affiliations to the known viral families.

**RNA Viruses of *Leptomonas pyrrocoris*.** Three of 18 isolates of *Leptomonas pyrrocoris* (H10, F165, and F19) originating from various locations worldwide (42) exhibited viral dsRNA bands (Table S1). All three bore two common RNAs of 3.5 and 2.2 kb, termed “RNA-T1” and “RNA-T2” (marked by green dots in Fig. 1A), and two (H10 and F19) contained six additional bands termed “RNAs O1–O6” (marked by red dots in Fig. 1A). Sequence analysis of all RNA segments from H10 and F165 suggested the presence of two viruses. The first was distantly related to *Tombusviridae*. It comprised RNAs T1 and T2 and was named “*Leptomonas pyrrocoris* tombus-like virus 1” (hereafter, “LeppyrtLV1”). The second virus comprising RNAs O1–O6 could not be associated with any of known viral groups and was named “*Leptomonas pyrrocoris* ostravirus 1” after the

city of Ostrava, where it was discovered (hereafter “LeppyrtOV1”). PCR tests confirmed the presence/absence of assignments made by S1 nuclease analysis (Table 1).

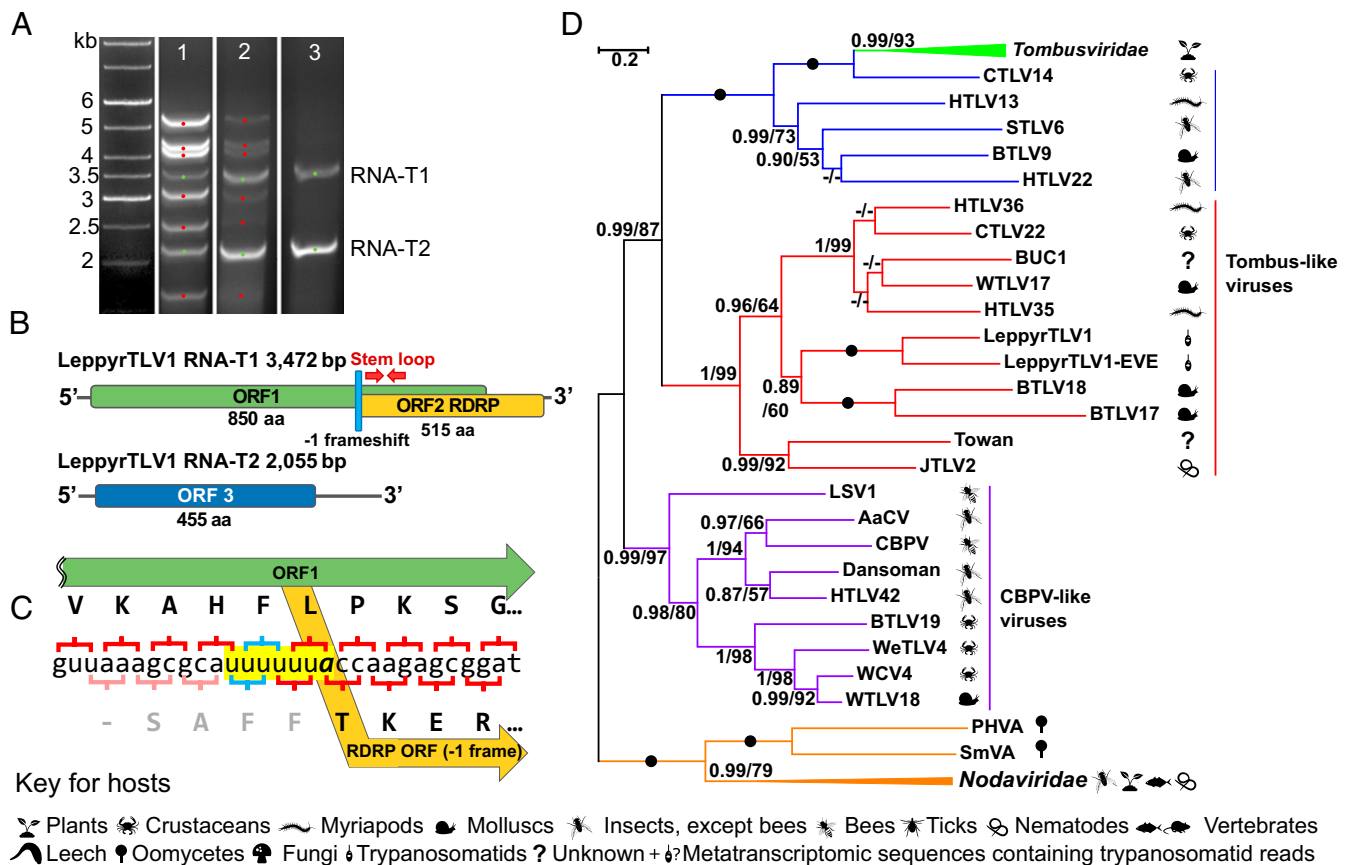
**LeppyrtLV1.** The sequences of segments T1 and T2 in the strains H10 and F165 were highly similar (96.7 and 97.05% nucleotide identity, respectively). RNA-T1 contained two overlapping ORFs with predicted proteins of 850 and 515 aa (Fig. 1B and C). For ORF1, a BLAST search in the National Center for Biotechnology Information (NCBI) nonredundant protein database did not yield any hits. The ORF2 showed a clear homology to viral RDRP (cd01699 in the NCBI Conserved Domain Database, CDD) with closest relationships to positive-strand RNA viruses of the *Tombusviridae*/*Nodaviridae* group (1). The two ORFs showed an overlap of 880 nt (Fig. 1C). A putative slippery sequence, UUUUUUA, was found 6 nt into the overlap, followed by a 127-nt hairpin 6 nt further. Both elements are typical of the –1 ribosomal frameshift of various viruses (43–45). These data suggest that the RDRP of this virus arises through the synthesis of an N-terminal frameshifted protein. While typical *Tombusviridae* encode RDRPs translated as a C-terminal extension of an upstream ORF by stop-codon read-through (46), several examples of –1 ribosomal frameshifting have been reported recently (1, 47, 48). RNA-T2 encoded a single ORF (ORF3) with a predicted protein of 455 aa (Fig. 1B), for which no homologs were identified in BLAST database searches.

Neither RNA T1 nor T2 exhibited conserved terminal sequences, which are also absent in both *Tombusviridae* and *Nodaviridae* (49, 50). Typically tombusviruses are monopartite, and the members of the related family *Nodaviridae* have two segments (49, 51). However, recent studies have shown remarkable variation within both groups (1).

Phylogenetic reconstruction using RDRP sequences placed LeppyrtLV1 within a clade distantly related to *Tombusviridae*, which usually infect plants (Fig. 1D). This clade includes viruses from invertebrates including parasitic nematodes, terrestrial myriapods, bivalves, cephalopods, freshwater crustaceans, and gastropods (Fig. 1D and Table S2) (1). *Pyrrocoris apterus*, the firebug host of *Leptomonas pyrrocoris*, is known to feed on the corpses of invertebrates (52), suggesting this as a possible route of acquisition.

**Endogenous viral element related to LeppyrtLV1.** BLAST searches against the genome assembly of *Leptomonas pyrrocoris* H10 (53) revealed that the ORF H10\_02\_0010 at the rightmost end of the chromosome 2 is homologous to the LeppyrtLV1 RDRP (Fig. S1A). Similar to the RNA-T1 of LeppyrtLV1, an overlapping ORF, H10\_02\_0020, was found immediately upstream. The overlap contained a potential slippery sequence, GGGAAAU, although we did not detect a stem-loop element thereafter (Fig. S1). The ORF H10\_02\_0010 and the LeppyrtLV1 RDRP shared 38% overall amino acid identity, including conservation of key RDRP motifs (Fig. S1D). Whole-transcriptome data for *Leptomonas pyrrocoris* (53) confirmed transcription of both ORFs. No homology was detected between the ORF1 of LeppyrtLV1 and the predicted chromosomal protein H10\_02\_0020. We considered the two ORFs of the chromosome 2 as an endogenous viral element (EVE) related to LeppyrtLV1 and named it “LeppyrtLV-EVE1.”

PCR tests with primers specific to LeppyrtLV-EVE1 RDRP revealed its presence in four additional European isolates (P59, LP, PP1, and PP2), all of whose sequences were identical (Tables S3 and S4). In contrast, this EVE1 region differed by 180 nt substitutions (and 84 indels) from the corresponding part of the LeppyrtLV1 RDRP, while the TLV1 RDRP sequences of strains H10 and F165 differed by only seven nucleotide substitutions. The similarity between EVE1 and TLV1 suggests that a TLV1-like RNA was captured via reverse transcription and integration into the *Leptomonas pyrrocoris* genome. EVEs occur frequently in evolution and are thought to be mediated primarily by reverse transcriptases encoded in host retrotransposons (54–56). Indeed, a number of telomere-associated transposable element (TATE) and spliced leader-associated (SLAC) retroelements have been identified in the *Leptomonas pyrrocoris* genome (53), including one



**Fig. 1.** *Tombus-like virus from *Leptomonas pyrrocoris*.* (A) Agarose gel electrophoresis of S1-digested total RNAs from strains H10 (lane 1), F19 (lane 2), and F165 (lane 3). LeppyRTL1 segments are labeled “RNA-T1” and “RNA-T2” on the right and marked by green dots; LeppyOV1 segments are marked by red dots. The left lane shows a 1-kb DNA ladder. (B) Genome structure of LeppyRTL1. ORFs for different predicted proteins are shown in different colors. A 127-nt stem-loop is found within the predicted N-terminal region of ORF2. (C) Sequence of the ORF1/2 overlap region including a putative slippery sequence (yellow). The RDRP domain is predicted to start from the ACC coding for threonine as previously reported for the UUUUUA slippery sequence (43). (D) Maximum likelihood phylogenetic tree based on RDRP amino acid sequences. Host taxa are shown by symbols defined in the key for hosts. Numbers at the branches indicate Bayesian posterior probability and maximum likelihood bootstrap supports, respectively; those having a Bayesian posterior probability value of 1.0 and maximum likelihood bootstrap support of 100% are marked with black circles. (The scale bar indicates the number of substitutions per site.) The tree was rooted with the sequences of *Nodaviridae*. Abbreviations and GenBank accession numbers are given in Tables S2–S4.

located immediately upstream of the LeppyRTL1-EVE1 (Fig. S14). The high level of sequence divergence with LeppyRTL1 points to a relatively ancient origin of EVE1, perhaps predating the dispersal of *Leptomonas pyrrocoris* across Europe (42).

**LeppyOV1.** The six RNAs O1–O6 of strains H10 and F19 (Fig. 1A) were initially viewed as “satellite” RNAs of LeppyRTL1. However, several observations suggested that they comprise separate virus. First, unlike RTL1 RNAs T1 and T2, the termini of RNAs O1–O6 share common sequences: AAAGAAAAA at the 5’ and ATGAGTTT at the 3’ ends (defined in the presumptive protein-coding strand orientation) (Fig. 2A). Conserved terminal sequences are known to participate in the replication of viruses and often are defining features of viral families (57). Second, in all strains the ratio of RNAs T1 to T2 was relatively constant; the same was true for RNAs O1–O6. However, the overall ratio of both RNA groups was substantially different.

Segments O1–O6 each contained a single ORF, and conventional BLAST searches did not yield any homologs for the corresponding hypothetical proteins. However, search algorithms focused on both structural and sequence homology revealed a putative RDRP motif within the predicted 1,315-aa protein within segment O3 (Fig. 2A), albeit with modest statistical support (NCBI CDD, amino acids 767–870,  $e = 0.89$ ; PHYRE 2, amino acids 684–874, confidence = 56%; HHPRED, amino acids 693–874, confidence = 89.7%) (58, 59). Within this region

we identified conserved viral RDRP motifs responsible for catalytic activity and ribonucleotide selectivity (Fig. 2B) (60–62). Analysis of the base frequencies of codon third positions of the viral ORFs showed significant differences between RTL1 and OV1 and a greater degree between these and the nuclear genome of *Leptomonas pyrrocoris* (Table S5).

Thus, we conclude that RNAs O1–O6 comprise a previously undescribed virus, *Leptomonas pyrrocoris* ostravirus 1 (LeppyOV1). As yet, we have not found a trypanosomatid strain containing this virus alone, which would firmly establish its independence from LeppyRTL1. Further studies are required to address the functional relationships between the six segments of this virus and significance of its co-occurrence with LeppyRTL1.

**A Bunyavirus-Like Genus, “Leishbunyavirus.”** Six isolates showed the presence of dsRNAs related to previously described viruses of *Leptomonas moramango* (37). LepmorLBV1a and b showed features characteristic of many other bunyaviruses, including a trisegmented genome, terminal “panhandle” repeats, and sequence relatedness of the predicted RDRP and nucleocapsid proteins, and were thus assigned as species within the genus *Leishbunyavirus* (LBV) (37). We confirmed the presence of LepmorLBV1a and b in *Leptomonas moramango*, as well as LBV1s in the dixenous phytopathogenic *Phytomonas* sp. TCC231 (PTCCLBV1) and four species of *Crithidia*: *Crithidia otongatchiensis*



**Table 1. Virus-positive trypanosomatid isolates**

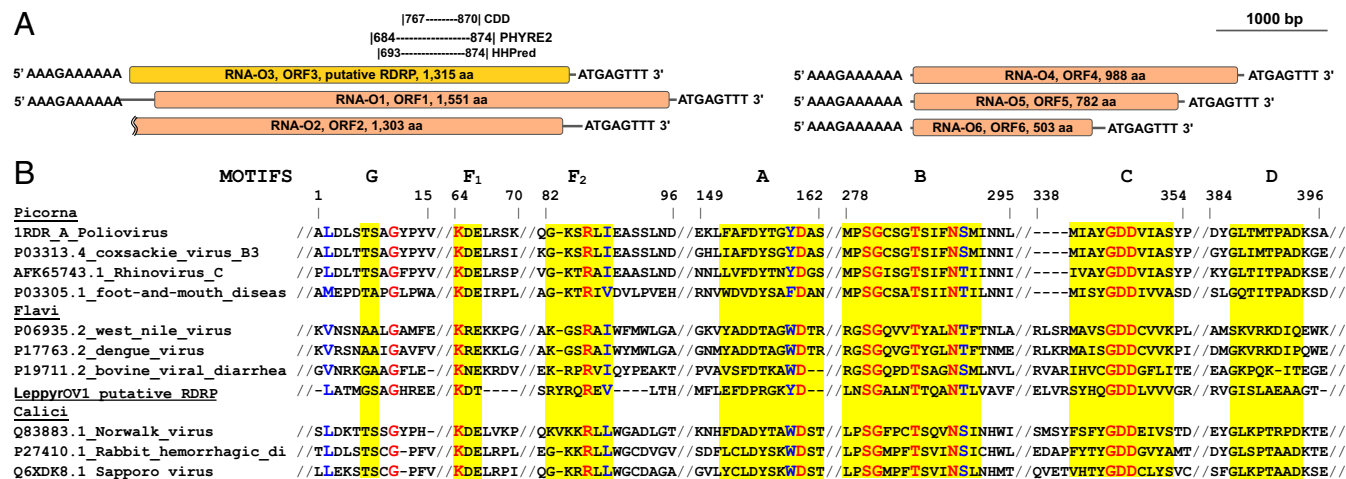
Species	Isolate	Virus name	No. of S1 bands	S1 band sizes, kb
<i>Bunyavirales</i>				
<i>Crithidia</i> sp.	C4	CC4LBV1	ND	ND
<i>Crithidia</i> sp.	G15	CG15LBV1	3	<u>6.0</u> , 1.1, 0.7
<i>Crithidia</i> sp.	ZM	CZMLBV1	3	<u>6.0</u> , 1.9, 0.7
<i>Crithidia abscondita</i>	127AL	CabsLBV1	3	<u>6.0</u> , 1.1, 0.7
<i>Crithidia otongatchiensis</i>	Ecu-08	CotoLBV1	3	<u>6.3</u> , 1.4, <u>0.7</u>
<i>Crithidia pragensis</i>	MCZ-11	CpraLBV1	ND	ND
<i>Leptomonas moramango</i>	MMO-09	LepmorLBV1a	3	<u>6.0</u> , 1.1, <u>0.7</u> (37)
<i>Leptomonas moramango</i>	MMO-09	LepmorLBV1b	3	<u>6.0</u> , 1.1, <u>0.7</u> (37)
<i>Phytomonas</i> sp.	TCC231	PTCCLBV1	2	<u>6.0</u> , 1.0
<i>Narnaviridae</i>				
<i>Phytomonas serpens</i>	9T (UCR)	PserNV1	1	<u>3.8</u> (39)
<i>Phytomonas serpens</i>	30T	PserNV1	1	<u>3.8</u>
<i>Leptomonas seymouri</i>	ATCC	LseyNLV1	2	<u>2.9</u> , <u>1.5</u> (38)
Tombus-like viruses				
<i>Leptomonas pyrrochoris</i>	F165	LeppyrTLV1	2	<u>3.5</u> , <u>2.2</u>
<i>Leptomonas pyrrochoris</i>	F19	LeppyrTLV1	2	<u>3.5</u> , <u>2.2</u>
<i>Leptomonas pyrrochoris</i>	H10	LeppyrTLV1	2	<u>3.5</u> , <u>2.2</u>
Tombus-like nonretroviral EVE				
<i>Leptomonas pyrrochoris</i>	P59	LeppyrTLV-EVE1	N/A	N/A
<i>Leptomonas pyrrochoris</i>	PP1	LeppyrTLV-EVE1	N/A	N/A
<i>Leptomonas pyrrochoris</i>	PP2	LeppyrTLV-EVE1	N/A	N/A
<i>Leptomonas pyrrochoris</i>	LP	LeppyrTLV-EVE1	N/A	N/A
<i>Leptomonas pyrrochoris</i>	H10	LeppyrTLV-EVE1	N/A	N/A
<i>Ostravirus</i>				
<i>Leptomonas pyrrochoris</i>	F19	LeppyrOV1	6	<u>5.2</u> , <u>4.5</u> , <u>4.1</u> , 3.0, 2.5, 1.7
<i>Leptomonas pyrrochoris</i>	H10	LeppyrOV1	6	<u>5.2</u> , <u>4.5</u> , <u>4.1</u> , 3.0, 2.5, 1.7

The sequenced segments are underlined, and the segments containing RDRP are in bold. Accession numbers for viral or related segments sequenced in this work are reported in Tables S3 and S4. N/A, not applicable; ND, not determined.

(CotoLBV1), *Crithidia abscondita* (CabsLBV1), *Crithidia* sp. G15 (CG15LBV1), and *Crithidia* sp. ZM (CZMLBV1) (Fig. 3A and Table 1). PCR tests with primers complementary to the conserved regions of LBV1 RDRPs showed the presence of these viruses in *Crithidia* sp. C4 and *Crithidia pragensis* as well (Table 1). These LBV1-positive strains showed three dsRNAs, except PTCCLBV1 which exhibited only two (Table 1). We sequenced all segments of CotoLBV1 and CabsLBV1 and the

largest segment (completely or partially) of the others (Table 1 and Tables S3 and S4).

**Sequence features and coding potential of LBV1s.** Prototypic bunyaviruses bear three RNA segments, termed “large: (L, 7–12 kb), “medium” (M, 3.2–4.9 kb), and “small” (S, 1–3 kb), encoding RDRP, envelope glycoproteins, and nucleocapsid, respectively (57, 63). The corresponding segments in LBV1s were considerably shorter: 6–6.3 kb (L), 1.0–1.9 kb (M), and 0.7–1.0 kb (S)



**Fig. 2.** *Leptomonas pyrrochoris* ostravirus 1, a unique virus from *Leptomonas pyrrochoris*. (A) Genome structure of LeppyrOV1, showing shared terminal sequences and a single ORF per segment (squiggle marks an incompletely sequenced end). The location of an RDRP domain predicted on RNA-O3 by CDD search, PHYRE2, and HHPred software is shown. (B) Multiple alignments of the LeppyrOV1 putative RDRP with those of *Picornavirales*, *Flavivirales*, and *Calicivirales*. Identical residues are shown in red; similar residues are shown in blue. Amino acid motifs, typically found in viral RDRPs, are highlighted in yellow.

(Table 1). Within each completely sequenced LBV1 segment we identified a single large ORF (Fig. 3B), in contrast to many bunyaviruses, which can encode multiple ORFs on the M and S segments (57). Bunyaviral RNA segments are typically flanked by panhandle inverted repeats mediating key steps of virus replication, transcription, and translation (64). Although the methods used here did not invariably yield full-length sequences, we were often able to identify panhandles in all fragments. In the L and M segments, we identified the sequence ACACAAAG at the 5' end (as defined by the viral sense orientation) and the complementary sequence CTTGTGT at the 3' end. These terminal eight nucleotides are typically found in all viruses belonging to the family *Phenuiviridae* (Table S6).

The ORFs in all completely sequenced L segments encoded putative proteins of ~2,000 aa homologous to the bunyaviral RDRP domain (amino acids 600–1,200; pfam04196, e-values <10<sup>-10</sup>; NCBI CDD). The region between amino acids 85 and 150 showed homology to the bunyaviral endonuclease domain, with conserved key residues involved in Mn<sup>2+</sup> ion coordination and phosphodiester bond cleavage (Fig. S24). Both the RDRP and endonuclease domains are essential for bunyaviral replication (65).

Various database searches (BLAST/CDD, PHYRE2, and HHpred) with the predicted proteins from the four sequenced M segments returned no hits. However, the predicted M proteins displayed a signal peptide as well as varying numbers of transmembrane domains, ranging from two in CabsLBV1, one in CotoLBV1, and zero or one in LepmorLBV1 (CCTOP, TMPred, and TMHMM algorithms), and N-glycosylation sites in CabsLBV1 and CotoLBV1 (Fig. 3B). These analyses suggest

that LBV1s, much like other bunyaviruses, are able to exploit cellular secretory systems for glycoprotein synthesis and virion assembly (66, 67). Indeed, purified LBV1 virions visualized by negatively stained transmission electron microscopy displayed the typical envelope with surface projections or spikes spread evenly along its surface (Fig. 3C).

The predicted S segment proteins did not yield compelling BLASTP hits. However, PHYRE structural homology searches showed the similarity of those proteins from CabsLBV1 and LepmorLBV1b to the nucleocapsid proteins of *Toscana*- and *Punta Toro* viruses (57.2–89.3% confidence). Alignment of the predicted leishbunyaviral S proteins with the nucleocapsid proteins of other bunyaviruses revealed several universally conserved amino acid motifs (Fig. S2B). Hence, we concluded that LBV1 S segments encode nucleocapsid proteins.

**Phylogenetic analysis of LBV1s suggests classification as a family Leishbunyaviridae within the Bunyvirales.** RDRP-based phylogenetic trees showed that the LBV1s formed a well-supported clade separate from other major Bunyvirales groups (Fig. 3D). The closest family was the *Phenuiviridae*, consistent with the similarities noted earlier in the terminal panhandle elements (Table S6). Many *Phenuiviridae* have been reported from insects and other arthropods, and viruses within the genus *Phlebovirus* are transmitted by the same sand fly species as *Leishmania* (68). However, our data show that LBV1s are far more ancient (Fig. 3D). As the divergence of the LBV1-containing clade from other bunyaviral families is comparable to or greater than other bunyavirus interfamilial divergences, we propose that this clade be recognized as a family, termed *Leishbunyaviridae*.

**A**

kb 1 2 kb 3 4

6 5 6 5

2 1.65 2 1.65

1 0.85 1 0.85

**B**

Leishbunyavirus L segment 6 – 6.3 kb

5' ACACAAAG EN RDRP CTTGTGT 3'

RDRP ~ 2,000 aa

Leishbunyavirus M segment 1.1 – 1.4 kb

5' ACACAAAG CTTGTGT 3'

Putative membrane glycoprotein ~ 320 aa

Leishbunyavirus S segment 0.7 – 1.0 kb

5' ACACAHVG CTTGTGT 3'

Nucleocapsid protein ~ 165 aa

**C**

100 nm

**D**

0.5

0.52/58

0.90/57

1/81

1/99

1/94

0.99/79

1/95

1/92

0.98/59

1/99

0.99/81

1/99

0.99/50

*Phlebovirus*

SFTSV

BHAV

LSV

*Phlebovirus* (uukuniemi group)

*Phasivirus*

*Tenuivirus*

*Goukovirus*

ABV1

CG15LBV1

LmorLBV1b

CC4LBV1

CZMLBV1

CabsLBV1

LmorLBV1a

DuBV

CpraLBV1

HBLV5

CotoLBV1

HBLV6

PTCCLBV1

HHFV

WSV

*Nairoviridae*, *Orthonairovirus*

*Arenaviridae*

*Phasmaviridae*, *Orthophasmavirus*

FERV

JONV

*Hantaviridae*, *Orthohantavirus*

*Fimoviridae*, *Emaravirus*

*Tospoviridae*, *Orthotospovirus*

*Peribunyaviridae*

*Phenuiviridae*

'*Leishbunyaviridae*'

**Fig. 3.** Leishbunyaviruses. (A) Viral dsRNA from *Crithidia otongatchiensis* (lane 1), *Crithidia abscondita* (lane 2), *Crithidia* sp. G15 (lane 3), and *Crithidia* sp. ZM (lane 4). The left lane shows a 1-kb DNA ladder. (B) Genome structure of LBVs. The sizes of segments and their various features (except for terminal complementary sequences) are shown in proportion. EN, endonuclease domain. Orange, teal, and yellow labels in the M segment stand for the signal peptide, glycosylation site(s), and transmembrane domain, respectively. (C) Negative-stain transmission electron micrographs of the virus particle isolated from *Crithidia otongatchiensis*. (Scale bar, 100 nm.) (D) Maximum likelihood phylogenetic tree based on RDRP amino acid sequences. Numbers at the branches indicate Bayesian posterior probability and maximum likelihood bootstrap supports, respectively; those having a Bayesian posterior probability value of 1.0 and maximum likelihood bootstrap support of 100% are marked with black circles. (The scale bar indicates the number of substitutions per site.) The tree was rooted at the midpoint. Abbreviations and GenBank accession numbers are given in Tables S2–S4. The pictograms describing viral hosts are as in Fig. 1.

E510 | www.pnas.org/cgi/doi/10.1073/pnas.1717806115

Grybchuk et al.

**Identification of LBVs within metatranscriptomic viral surveys.** Interestingly, BLAST searches with trypanosomatid LBV1 RDRPs identified several hits in the sequences from metatranscriptomic virus-hunting surveys. These included Huangshi Humpbacked Fly virus (HHFV), Wuhan Spider virus (WSV) (69), Hubei bunya-like virus 5 (HBLV5) from a mix of dipterans, Hubei bunya-like virus 6 (HBLV6) from a horse leech (1), and two from honey bees—*Apis* bunyavirus 1 (ABV1) and Duke bunyavirus (DuBV) (70). On the reconstructed phylogenetic tree all these viruses from metatranscriptomes intermingled with the trypanosomatid LBVs, with high statistical support (Fig. 3D).

Recently, it was proposed that bunyaviruses originated within insects (71, 72), and one explanation for the interdigitation observed here could be multiple transitions of these viruses between arthropods and trypanosomatids. An alternate model is that the insect metatranscriptomic leishbunyaviruses arose not from the insects themselves but from their associated microbiota (73), given that trypanosomatids are well-known parasites of arthropods (14, 74). Thus, we searched the LBV-containing metatranscriptomic sequence read archives (SRAs) for trypanosomatid signatures, a challenging task given the relatively low number of viral reads in these pooled datasets. Nonetheless, BLASTN searches of assembled contigs revealed several abundant trypanosomatid transcripts, such as 18S rRNA or paraflagellar rod proteins (Tables S3 and S4), in the HHFV-, WSV-, HBLV5-, and ABV1-containing SRAs (data for DuBV were not available). Indeed, phylogenetic analysis of these putative transcripts revealed affinities to various trypanosomatids. Based on these data, ABV1 could speculatively be associated with *Lotmaria passim* (subfamily Leishmaniinae), HHFV and HBLV5 with subfamily Strigomonadinae, and HBLV6 and WSV with the genera *Trypanosoma* and *Herpetomonas*, respectively (Tables S2–S4). While the co-occurrence of reads for leishbunyaviruses and trypanosomatids in the metatranscriptomic read sets is not definitive proof that these flagellates actually contained viruses, we consider this a plausible explanation.

These findings provide support for the model postulating a trypanosomatid microbiota origin of LBVs emerging from the metatranscriptome datasets. If borne out, this suggests that, instead of multiple origins from insects, trypanosomatid LBVs may have originated less frequently and perhaps only once. Consistent with the latter, a significant, albeit imperfect, level of phylogenetic congruency can be seen between trypanosomatid LBVs and nuclear genome phylogenies (Fig. S3). However, the possibility of multiple acquisitions of LBVs by trypanosomatids from insects or other trypanosomatids cannot be formally excluded, given that trypanosomatid LBVs bear hallmarks of infectious bunyaviruses and reports of mixed trypanosomatid infections (75–78). Currently we favor a model with a single transition of an ancestral insect virus to a trypanosomatid, but further investigations will be required to rigorously establish this hypothesis.

**Narnaviridae.** In *Leptomonas seymouri* and two isolates of *Phytomonas serpens* we documented the presence of dsRNA (fragments of 2.9 + 1.5 kb and 3.8 kb, respectively) (Fig. 4A and B and Table 1), in agreement with previous findings (38–40) that these species bear *Leptomonas seymouri* narna-like virus 1 (LepseyNLV1) and *Phytomonas serpens narnavirus* 1 (PserNV1). **Sequence features of trypanosomatid narnaviruses.** The genome of PserNV1 was monosegmented (Fig. 4A, lane 2), and its RNA contained a single ORF for RDRP. In contrast, LepseyNLV1 displayed a bipartite organization (Fig. 4B, lanes 1, 4, and 5) with RNA1 encoding RDRP and RNA2 comprising two overlapping ORFs with no homologs identified in database searches. The region of overlap displayed several structural features associated with +1 ribosomal frameshifting, including a hairpin preceded by a slippery sequence (Fig. S4C), suggesting these two ORFs may be expressed as a fusion protein.

PserNV1 termini were determined by ligating an adapter followed by sequencing across the adapter–virus junction. They revealed features common for *Narnaviridae* (79): short terminal

complementary sequences 5'-ACGC ... GCGT-3' and putative subterminal hairpin structures (Fig. 4C and Fig. S4A). Intriguingly, the very 5' end of the viral RNA showed similarity to the spliced leader (SL) of *Phytomonas serpens* (GenBank X87137). The SL is a 39-nt capped sequence added to the 5' end of every trypanosomatid mRNA by *trans*-splicing (80, 81). However, the PserNV1 SL-related sequence lacked the first five nucleotides and had three internal mismatches (Fig. S4B), rendering it unlikely to be functional based on current knowledge of SL function (82). Thus, in the past the PserNV1 may have “snatched” the host’s SL, substituting it for the original terminus. In LepseyNLV1 we did not determine the terminal sequences explicitly; however, typical narnaviral subterminal hairpins were predicted in the RNA2 assembly (Fig. 4C).

**Phylogeny and evolutionary origins.** RDRP-based phylogenetic reconstruction showed LepseyNLV1 and PserNV1 to be the closest relatives, forming a well-supported clade along with prototypical narnaviruses—*Saccharomyces cerevisiae* 20S and 23S viruses (Fig. 4B and C)—as well as the oomycete-infecting *Phytophthora infestans* RNA virus 4 (PiRV4) (83). Interestingly, we identified a metatranscriptomic virus from the fly *Teleopsis dalmanni* (84), whose transcriptome assembly also contained two contigs (GBBP01074304 and GBBP01074305) corresponding to trypanosomatid 18S rRNA genes. We were not able to closely associate these contigs with known trypanosomatid sequences, suggesting that this virus may belong to a yet-uncharacterized lineage. Thus, similar to leishbunyaviruses, the insect metatranscriptomic *Narnavirus* may have arisen from its trypanosomatid microbiota.

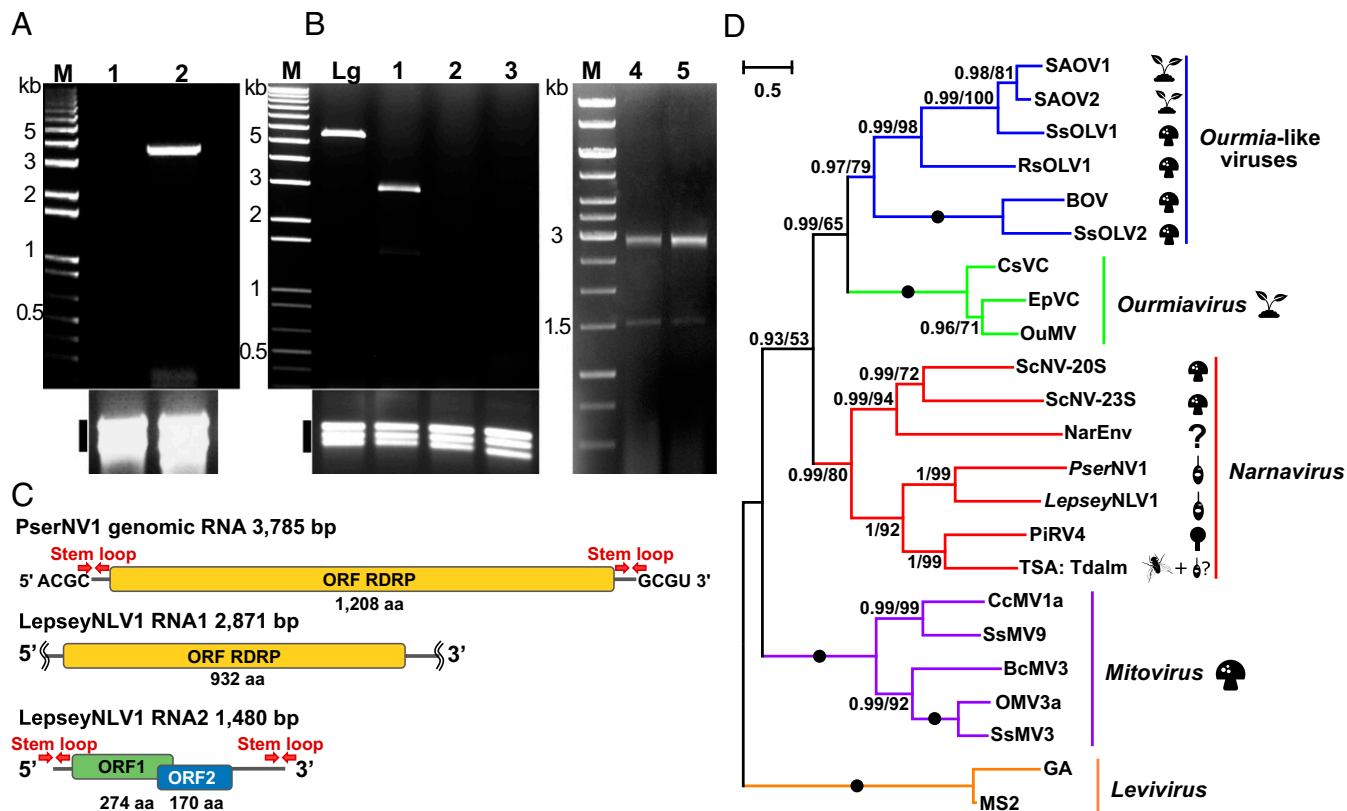
As inferred earlier, *Oourmiavirus* and ourmia-like viruses (family *Oourmiaviridae*) clustered preferentially with *Narnavirus*, while *Mitovirus* (another genus of *Narnaviridae*) was sister to the clade comprising those three groups (85, 86). Previous studies suggested that narnaviruses were ancestral parasites of fungi, which later switched to other organisms (87, 88). Yeasts represent a normal component of insect’s intestine, where they could encounter trypanosomatids (89, 90).

While narnaviruses are typically monosegmented, *Oourmiaviridae* typically contain several segments (87, 88). LepseyNLV1 with its two segments exhibits an independently evolved genome organization being intermediate between those of *Narnaviridae* and *Oourmiaviridae*. While definitive evidence that the two segments represent a single virus is lacking, we consider such an association likely, given that both segments are maintained or lost in parallel as described below.

**Viral Stability.** In the course of our studies we observed that upon in vitro cultivation some viruses could be occasionally lost. PserNV1 was originally found in the 9T strain from the Czech Republic; however, the same strain maintained elsewhere lacked it (Fig. 4A, lanes 1 and 2). Similarly, while LepseyNLV1 occurred in the ATCC30220 isolate, it was absent in the same strain and in a transfectant derivative obtained from another source (Fig. 4B, lanes 1–3). However, it persisted during continuous cultivation (~300 passages) in the Zoological Institute of the Russian Academy of Sciences (Fig. 4B, lanes 4 and 5). For leishbunyaviruses, we noticed a gradual decrease of viral dsRNA levels in *Crithidia otongatchiensis* over 6 mo of continuous cultivation and their disappearance from *Crithidia pragensis* and *Leptomonas moramango* after 2 wk of passaging (although low levels could be detected by RT-qPCR). However, no changes in dsRNA abundance were seen for CabsLBV1, CZMLBV1, CG15LBV1, or PTCCLBV1.

Several nonexclusive mechanisms might explain these observations. Some viruses may be intrinsically unstable or lost because the selective pressures on their trypanosomatid hosts may differ in vitro and in vivo. Alternatively, the culture may be heterogeneous in terms of viral presence, and virus-free cells may outcompete their infected counterparts. Our data collectively suggest that caution is warranted when interpreting viral absence in cultured parasites. Serendipitously, virus-free derivatives may serve as isogenic tools for probing potential roles for viruses in parasite biology, as for *Leishmania guyanensis* LRV1 (6, 91). Indeed,





**Fig. 4.** Narnaviruses of trypanosomatids. (A, Upper) Viral dsRNA in two subcultures of *Phytomonas serpens* (isolate 9T). Lane 1 is from the University of California, Riverside and lane 2 is from the Institute of Parasitology in Budweis, Czech Republic. The lane marked "M" is a 1-kb DNA ladder. (Lower) Total RNA was used as a loading control. (B, Upper) Viral dsRNA in four subcultures of *Leptomonas seymouri* ATCC30220. Lanes 1 and 4 are the original ATCC culture; lanes 2 and 3 are substrains 2003WT and 294-1993VB (Rutgers University), respectively; lane 5 is a culture from Zoological Institute of the Russian Academy of Sciences. *Leishmania guyanensis* (Lgy, strain M4147) bearing the 5.3-kb LRV1 served as a positive control. The lane marked "M" is a 1-kb DNA ladder. (Lower) Total RNA was used as a loading control for virus-negative substrains. (C) Genome structure of LepseyNLV1 and PserNV1. ORFs for different proteins are shown in different colors. The stem-loop structures and terminal complementary sequences are indicated. Squiggles mark incompletely sequenced ends. (D) Maximum likelihood phylogenetic tree of *Narnaviridae* based on RDRP amino acid sequences. LepseyNLV1 and PserNV1 are indicated by a trypanosomatid symbol. Numbers at the branches indicate Bayesian posterior probability and maximum likelihood bootstrap supports, respectively; those having a Bayesian posterior probability value of 1.0 and maximum likelihood bootstrap support value of 100%, respectively are marked with black circles. (The scale bar indicates number of substitutions per site.) The tree was rooted with the sequences of *Leviviridae*. Abbreviations and GenBank accession numbers are given in Tables S2–S4. The pictograms are as in Fig. 1.

the coincidental loss of LepseyNLV1 RNA1 and RNA2 provides some support for a functional association (Fig. 4B).

## Conclusions

Here, we conducted a survey of RNA viruses in two groups of Trypanosomatidae: insect-restricted (monoxenous) relatives of *Leishmania* (*Crithidia* and *Leptomonas*, subfamily Leishmaniinae) and plant-infecting *Phytomonas*. This greatly expanded the known diversity of RNA viruses in these flagellates, showing that trypanosomatids can be infected by various unrelated viruses: *Totiviridae*, *Narnaviridae*, *Bunyvirales*, *tombus-like viruses*, and a previously unknown virus that was termed "Ostravirus" and is currently defined by LeppyrOV1, whose RDRP was so divergent that it escaped generic BLAST searches. We also documented EVE formation in trypanosomatids (LeppyrTLV-EVE1), presumably enabled by the activity of the endogenous retrotransposons.

One interesting question is whether the trypanosomatid viruses can be shed and infect other parasites. Current data suggest that LRV1, like the great majority of other *Totiviridae*, is not shed (92). Narnaviruses, by virtue of lacking either a capsid or an envelope, are transmitted only vertically or during mating (4). However, the presence of an extra segment in LepseyNLV1 (Fig. 4B) might be associated with transmission, as in related ourmiaviruses (85). Similarly, the two *Leptomonas pyrhorcoris* viruses OV1 and TLV1 have sufficient coding capacity for

transmission. Last, for several trypanosomatid LBV1s we visualized the presence of enveloped virions bearing surface proteins (Fig. 3C), the hallmarks of infectious bunyaviruses. These fascinating questions will be addressed in future studies.

Phylogenetic relationships of relevant trypanosomatid taxa permit a broader view of the origins and evolution of their viruses (Fig. 5). First, *Leptomonas pyrhorcoris* appears to be a hotbed for viral discovery, with two previously unreported viruses (LeppyrTLV1 and LeppyrOV1) and the presence of an EVE. Secondly, narnaviruses, LBV1s, and LRV1/2s appear to be distributed over the trypanosomatid phylogenetic tree in a patchy manner, with many seemingly virus-free lineages interspersed with ones bearing diverse viruses (Fig. 5). This poses a number of challenges. If one postulates the presence of virus in the common ancestor of a particular group (marked by arrows in Fig. 5), viral loss must have occurred independently in a great many subsequent taxa. Alternatively, if one assumes the common ancestor to be virus-free, independent viral acquisitions must have occurred. The chances of this are speculative at best, perhaps being more likely for the viruses showing increased likelihood for infectivity (LBV1s and, conceivably, OV1, TLV1, and NLV1). Superimposed upon or alternative to this is the possibility of viral exchange via infectious shedding during coinfections, as mixed trypanosomatid infections are quite frequent (see above). Importantly, these latter two processes would be expected to further

blur signs of virus–parasite coevolution. Thus, it is remarkable that for LRV1/2 (93) and, to some extent, for LBVs (Fig. S3), the phylogenetic trees for the parasite and their viruses show significant congruency. This suggests that there must be some constraints on the horizontal viral transmission, if present, especially among kingdoms.

Notably, in our survey we did not find any LRV-related *Totiviridae* (Fig. 5 and Table 1), although numerous Leishmaniinae were tested. This suggests that these viruses were acquired upon the involvement of vertebrates into the life cycle of Leishmaniinae. Given the elevated pathogenicity of LRV1-bearing *Leishmania* to the vertebrate host, viral acquisition could be viewed as beneficial for the parasite, if one equates pathogenicity with an increased evolutionary fitness. However, most *Leishmania* except *Viannia* and a handful of *Leishmania major* or *Leishmania aethiopica* isolates lack LRV1/2 (94). This implies that if LRV1/2 presence was indeed beneficial early in evolution, it became less important in modern lineages and/or was replaced by other mechanisms contributing to virulence, such as type I IFN induction (29).

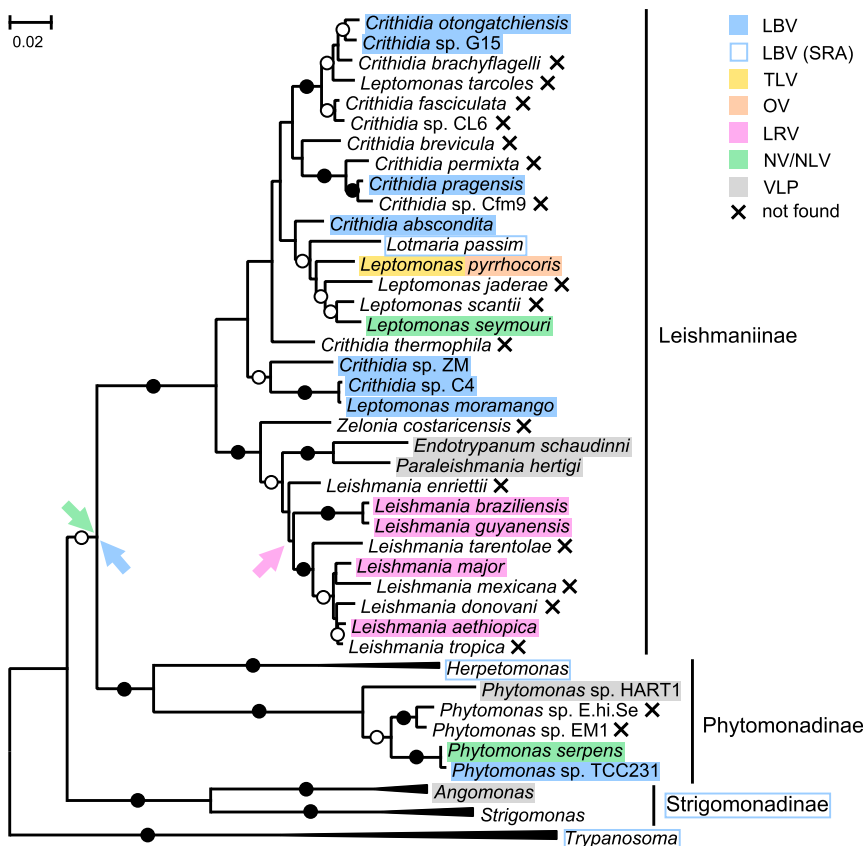
Variation in the RNAi pathway may contribute to the observed patchiness in viral distribution, as this pathway acts as an antiviral defense mechanism in many species (95). In agreement with this, the RNAi pathway (believed to be ancestral to all eukaryotes) is absent in *Phytomonas* spp., *Leptomonas seymouri*, and LRV2-bearing *Leishmania aethiopica* and *Leishmania major* (23, 96, 97). The RNAi pathway may be especially important for narnaviruses, which are presumably defenseless because of the lack of capsids. However, LRV1-containing *Leishmania guyanensis* and *Leishmania braziliensis* have a highly active RNAi pathway (97), and accordingly LRV1 has mastered the ability to coexist in the face of RNAi attack, although under some circumstances RNAi can lead to its elimination (27). In addition, numerous *Crithidia* and *Leptomonas* spp. retain the RNAi pathway (96). It is thus possible that RNAi plays only a weak role in the evolutionary distribution of trypanosomatid viruses.

Several studies have established a role for trypanosomatid viruses in the vertebrate host (6, 27–29, 40, 98). Our studies now suggest that potential role(s) for trypanosomatid viruses in parasite biology within their insect hosts should be considered. While LRV1 and other *Totiviridae* have been implicated in vertebrate pathogenicity (6, 7, 35), there are no direct data concerning the influence of this virus on the relationships between *Leishmania* and sand flies. Given that Toll-like receptors were first discovered in insects (99), and TLR3 specifically was implicated in LRV1 pathogenicity (6), this possibility clearly merits attention. Alternatively, viruses may invade and persist as mere parasitic elements rather than providing any advantage to their trypanosomatid hosts. Resolution of these questions may benefit from the serendipitous identification of virus-free isolates of *Phytomonas serpens* and *Leptomonas seymouri* and their use in studies assessing potential functional roles.

Previously unidentified viruses were found in considerable numbers in the species/isolates tested (Table 1). The actual diversity of trypanosomatids is not known, but at least 600 species have been already described (100). In addition, the example of *Leptomonas pyrrocoris* with its multiple isolates showing variation in viral presence and composition, illustrates another level of diversity. Indeed, as noted in Fig. 5, there are several trypanosomatid lineages for which VLPs have been reported but not studied by modern molecular methods. Furthermore, in various invertebrate metatranscriptomes we found several viruses possibly originating from their trypanosomatid microbiota. Such metatranscriptomes may also provide important new information about the diversity of trypanosomatids themselves. Taken together these findings suggest that a great number of viruses remain to be found in this important group of parasites.

## Methods

**Isolation of Viral RNA and Primary Screening.** Total RNA was isolated from trypanosomatid cultures using the TRI reagent (MRC Inc.) as described previously



**Fig. 5.** Overview of trypanosomatid relationships and viruses. The evolutionary tree shows the maximum likelihood phylogenetic tree of trypanosomatids reconstructed using 18S rRNA and gGAPDH genes, over which the absence or presence of viruses is marked (see graphical legend). Arrows denote hypothetical acquisition of viruses under the assumption of a single origin in the common ancestor. Maximal bootstrap supports are marked by filled circles, and bootstrap supports over 70% are denoted by open circles. (The scale bar indicates the number of substitutions per site.)



(101). For primary screening, 50 µg of total RNA from each sample were treated with RNase-free DNase I (New England Biolabs) and nuclease S1 from *Aspergillus oryzae* (Sigma-Aldrich) (41). The resulting dsRNA was resolved on 0.8% agarose gel and stained with ethidium bromide. For preparative isolation 400 µg of total RNA from virus-positive cultures were digested with DNase I, followed by ssRNA precipitation by LTS solution (2 M LiCl, 150 mM NaCl, 15 mM Tris HCl, pH 8.0) at 4 °C overnight as described previously (102). The ssRNA fraction was removed by centrifugation for 30 min at 20,000 × g at 4 °C, and dsRNA was precipitated by EtOH and visualized as above. Individual dsRNA bands were gel-purified using Zymo-clean Gel RNA Recovery Kit (Zymo Research).

**Viral dsRNA Amplification, Cloning, and Sequencing.** Gel-extracted dsRNA was polyadenylated at both 3' ends using *Escherichia coli* Poly(A) Polymerase (New England Biolabs) and then purified on a PCR clean-up column (Thermo Fisher Scientific) according to the manufacturer's protocol. Next, polyadenylated dsRNA was reverse-transcribed using the Transcriptor First Strand cDNA Synthesis Kit (Roche), and an anchored-oligo (dT) primer QD2-T20 5'-ggcaataacacctactagaaattcggtcgtatcttttttttttttttttttt-3' (modified from ref. 103). To prevent renaturation of the cRNA strands, DMSO was added to a final concentration of 7.5%, and the residual RNA was lysed with 0.1 M NaOH. The cDNA was then reannealed at 65 °C for 90 min followed by gradual cooling to 4 °C (102). The cDNA was purified on PCR clean-up columns and amplified using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) with primer QD2, 5'-tctactatagaattcgttcgac-3', that anneals to the fragment introduced by QD2-T20. The first PCR step that included end repair (72 °C for 5 min) was followed by the manufacturer's recommended cycling conditions: 98 °C for 10 s, 55 °C for 30 s, and 72 °C for 40 s/kb. Obtained PCR products were cloned into the pTZ57R vector (Thermo Fisher Scientific) and sequenced by primer walking. For the analysis described above, we were unable to obtain enough dsRNA from *Crithidia* sp. C4 and *Crithidia pragensis*. In these cases, the partial RDRP gene (~900 bp) was am-

plified using degenerate primers designed to amplify known LBV1s (LeiBunyaF: 5'-ttkvcvacttcaagaaragcac-3' and LeiBunyaR: 5'-ccagartcatcwgadgadaccat-3') (ref. 37 and this work), and the products were cloned into the pTZ57R vector and sequenced. To assess the presence of the *Leptomonas pyrrochoris* RNA virus, total cDNA of all *Leptomonas pyrrochoris* isolates (both positive and negative as judged by gel-based assay) was amplified with primers LpTLV1F 5'-tactctataacggggca-3' and LpTLV1R 5'-taagaggcgaattctgct-3' specific to the RDRP region (~300 bp) of this virus and directly sequenced. Similarly, the occurrence of integrated virus in these isolates was checked by amplification using primers LpIVF 5'-cctatgctcggatgactca-3' and LpIVR 5'-ctgtgactttctatcaag-3'. PCR Primers M200 5'-atggtcctvtaargtggmat-3' and M201 5'-takccccac-trttrtctacca-3' for the glycosomal GAPDH (gGAPDH) gene were used as an internal positive control (104). Additional methods for cultivation of trypanosomatids, phylogenetic, genomic and transcriptomic analyses, and negative-stain transmission electron microscopy can be found in the [Supporting Information](#).

**ACKNOWLEDGMENTS.** We thank members of our groups and D. Wang and S. Krishnamurthy (Washington University School of Medicine in St. Louis) for discussion and D. Maslov, L. Simpson, and V. Bellofatto for providing *Phytomonas* and *Leptomonas* strains. This work was supported by NIH Grants R01-AI029646 and R56-AI099364 (to S.M.B.), the Intramural Research Program of the National Library of Medicine at the NIH (I.B.R.), Grant Agency of Czech Republic Awards 17-10656S (to V.Y. and J.V.) and 16-18699S (to J.L. and V.Y.), Moravskoslezský kraj Research Initiative DT01-021358 (to V.Y. and A.Y.K.), the Russian Foundation for Basic Research Project no. 15-29-02734 and the State Assignment for the Zoological Institute AAAA-A17-117030310322-3 (to A.O.F.), and the European Cooperation in Science and Technology (COST) action CM1307 (to J.L.). Work in the V.Y. laboratory is financially supported by the Ministry of Education, Youth and Sports of the Czech Republic in the National Feasibility Program I, project LO1208 TEWEP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

- Shi M, et al. (2016) Redefining the invertebrate RNA virosphere. *Nature* 540:539–543.
- Wang AL, Wang CC (1991) Viruses of the protozoa. *Annu Rev Microbiol* 45:251–263.
- Banik GR, Stark D, Rashid H, Ellis JT (2014) Recent advances in molecular biology of parasitic viruses. *Infect Disord Drug Targets* 14:155–167.
- Ghabrial SA, Castón JR, Jiang D, Nibert ML, Suzuki N (2015) 50-plus years of fungal viruses. *Virology* 479:480:356–368.
- Chen B, Geletka LM, Nuss DL (2000) Using chimeric hypoviruses to fine-tune the interaction between a pathogenic fungus and its plant host. *J Virol* 74:7562–7567.
- Ives A, et al. (2011) *Leishmania* RNA virus controls the severity of mucocutaneous leishmaniasis. *Science* 331:775–778.
- Fichorova RN, et al. (2012) Endobiont viruses sensed by the human host—Beyond conventional antiparasitic therapy. *PLoS One* 7:e48418.
- Opperdoes FR, Butenko A, Flegontov P, Yurchenko V, Lukeš J (2016) Comparative metabolism of free-living *Bodo saltans* and parasitic trypanosomatids. *J Eukaryot Microbiol* 63:657–678.
- Stuart K, Panigrahi AK (2002) RNA editing: Complexity and complications. *Mol Microbiol* 45:591–596.
- Záhonová K, Kostygov AY, Ševčíková T, Yurchenko V, Eliáš M (2016) An unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. *Curr Biol* 26:2364–2369.
- Simpson L (1998) A base called J. *Proc Natl Acad Sci USA* 95:2037–2038.
- Michaeli S (2011) *Trans*-splicing in trypanosomes: Machinery and its impact on the parasite transcriptome. *Future Microbiol* 6:459–474.
- Nussbaum K, Honek J, Cadmus CM, Efferth T (2010) Trypanosomatid parasites causing neglected diseases. *Curr Med Chem* 17:1594–1617.
- Maslov DA, Votýpka J, Yurchenko V, Lukeš J (2013) Diversity and phylogeny of insect trypanosomatids: All that is hidden shall be revealed. *Trends Parasitol* 29:43–52.
- Flegontov P, et al. (2013) *Paratrypanosoma* is a novel early-branching trypanosomatid. *Curr Biol* 23:1787–1793.
- Lukeš J, Skalický T, Týč J, Votýpka J, Yurchenko V (2014) Evolution of parasitism in kinetoplastid flagellates. *Mol Biochem Parasitol* 195:115–122.
- Kostygov AY, Yurchenko V (2017) Revised classification of the subfamily Leishmaniinae (Trypanosomatidae). *Folia Parasitol* 64:020.
- Molyneux DH (1974) Virus-like particles in *Leishmania* parasites. *Nature* 249:588–589.
- Croft SL, Chance ML, Gardener PJ (1980) Ultrastructural and biochemical characterization of stocks of *Endotrypanum*. *Ann Trop Med Parasitol* 74:585–589.
- Soares MJ, Motta MC, de Souza W (1989) Bacterium-like endosymbiont and virus-like particles in the trypanosomatid *Crithidia desouzai*. *Microbios Lett* 41:137–141.
- Marche S, Roth C, Manohar SK, Dollet M, Baltz T (1993) RNA virus-like particles in pathogenic plant trypanosomatids. *Mol Biochem Parasitol* 57:261–267.
- Yurchenko V, et al. (2014) Ultrastructure and molecular phylogeny of four new species of monoxenous trypanosomatids from flies (Diptera: Brachycera) with redefinition of the genus *Wallaceina*. *Folia Parasitol* 61:97–112.
- Kraeva N, et al. (2015) *Leptomonas seymouri*: Adaptations to the dixenous life cycle analyzed by genome sequencing, transcriptome profiling and co-infection with *Leishmania donovani*. *PLoS Pathog* 11:e1005127.
- Widmer G, Comeau AM, Furlong DB, Wirth DF, Patterson JL (1989) Characterization of a RNA virus from the parasite *Leishmania*. *Proc Natl Acad Sci USA* 86:5979–5982.
- Tarr PI, et al. (1988) LR1: A candidate RNA virus of *Leishmania*. *Proc Natl Acad Sci USA* 85:9572–9575.
- Eren RO, et al. (2016) Mammalian innate immune response to a *Leishmania*-resident RNA virus increases macrophage survival to promote parasite persistence. *Cell Host Microbe* 20:318–328.
- Brettmann EA, et al. (2016) Tilting the balance between RNA interference and replication eradicates *Leishmania RNA virus 1* and mitigates the inflammatory response. *Proc Natl Acad Sci USA* 113:11998–12005.
- Zangger H, et al. (2014) *Leishmania aethiopica* field isolates bearing an endosymbiotic dsRNA virus induce pro-inflammatory cytokine response. *PLoS Negl Trop Dis* 8:e2836.
- Rossi M, et al. (2017) Type I interferons induced by endogenous or exogenous viral infections promote metastasis and relapse of leishmaniasis. *Proc Natl Acad Sci USA* 114:4987–4992.
- Hartley MA, Drexler S, Ronet C, Beverley SM, Fasel N (2014) The immunological, environmental, and phylogenetic perpetrators of metastatic leishmaniasis. *Trends Parasitol* 30:412–422.
- Ronet C, Beverley SM, Fasel N (2011) Mucocutaneous leishmaniasis in the New World: The ultimate subversion. *Virulence* 2:547–552.
- Adaui V, et al. (2016) Association of the endobiont double-stranded RNA virus LRV1 with treatment failure for human leishmaniasis caused by *Leishmania braziliensis* in Peru and Bolivia. *J Infect Dis* 213:112–121.
- Boureaud E, et al. (2016) Presence of *Leishmania RNA Virus 1* in *Leishmania guyanensis* increases the risk of first-line treatment failure and symptomatic relapse. *J Infect Dis* 213:105–111.
- Ito MM, et al. (2015) Correlation between presence of *Leishmania RNA virus 1* and clinical characteristics of nasal mucosal leishmaniasis. *Rev Bras Otorrinolaringol (Engl Ed)* 81:533–540.
- Cantanhêde LM, et al. (2015) Further evidence of an association between the presence of *Leishmania RNA Virus 1* and the mucosal manifestations in tegumentary leishmaniasis patients. *PLoS Negl Trop Dis* 9:e0004079.
- Pereira LdeO, et al. (2013) Severity of tegumentary leishmaniasis is not exclusively associated with *Leishmania RNA virus 1* infection in Brazil. *Mem Inst Oswaldo Cruz* 108:665–667.
- Akopyants NS, Lye LF, Dobson DE, Lukeš J, Beverley SM (2016) A novel Bunyavirus-like virus of trypanosomatid protist parasites. *Genome Announc* 4:e00715–e00716.
- Lye LF, Akopyants NS, Dobson DE, Beverley SM (2016) A Narnavirus-like element from the trypanosomatid protozoan parasite *Leptomonas seymouri*. *Genome Announc* 4:e00713–e00716.
- Akopyants NS, Lye LF, Dobson DE, Lukeš J, Beverley SM (2016) A Narnavirus in the trypanosomatid protist plant pathogen *Phytomonas serpens*. *Genome Announc* 4:e00711–e00716.
- Sukla S, Roy S, Sundar S, Biswas S (2017) *Leptomonas seymouri* narna-like virus 1 and not leishmaniviruses detected in kala-azar samples from India. *Arch Virol* 162:3827–3835.
- Beiting DP, et al. (2014) Differential induction of TLR3-dependent innate immune signaling by closely related parasite species. *PLoS One* 9:e88398.

42. Votýpka J, et al. (2012) Cosmopolitan distribution of a trypanosomatid *Leptomonas pyrrocoris*. *Protist* 163:616–631.
43. Jacks T, Madhani HD, Masiarz FR, Varmus HE (1988) Signals for ribosomal frameshifting in the *Rous sarcoma virus* gag-pol region. *Cell* 55:447–458.
44. Firth AE, Chung BY, Fleton MN, Atkins JF (2008) Discovery of frameshifting in *Alphavirus 6K* resolves a 20-year enigma. *Virology* 375:108.
45. Atkins JF, Loughran G, Bhatt PR, Firth AE, Baranov PV (2016) Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res* 44:7007–7078.
46. Dreher TW, Miller WA (2006) Translational control in positive strand RNA plant viruses. *Virology* 344:185–197.
47. Kim KH, Lommel SA (1998) Sequence element required for efficient -1 ribosomal frameshifting in red clover necrotic mosaic dianthovirus. *Virology* 250:50–59.
48. Olivier V, et al. (2008) Molecular characterisation and phylogenetic analysis of *Chronic bee paralysis virus*, a honey bee virus. *Virus Res* 132:59–68.
49. Johnson KN, Johnson KL, Dasgupta R, Gratsch T, Ball LA (2001) Comparisons among the larger genome segments of six nodaviruses and their encoded RNA replicases. *J Gen Virol* 82:1855–1866.
50. Na H, White KA (2006) Structure and prevalence of replication silencer-3' terminus RNA interactions in *Tombusviridae*. *Virology* 345:305–316.
51. Schneemann A, Ball LA, Delsert C, Johnson JE, Nishizawa T (2005) Family *Nodaviridae*. *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, eds Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (Elsevier, New York), pp 865–872.
52. Socha R (1993) *Pyrrocoris apterus* (Heteroptera) - an experimental model species: A review. *Eur J Entomol* 90:241–286.
53. Flegontov P, et al. (2016) Genome of *Leptomonas pyrrocoris*: A high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci Rep* 6:23704.
54. Feschotte C, Gilbert C (2012) Endogenous viruses: Insights into viral evolution and impact on host biology. *Nat Rev Genet* 13:283–296.
55. Bruenn JA, Warner BE, Yerramsetty P (2015) Widespread mitovirus sequences in plant genomes. *PeerJ* 3:e876.
56. Geuking MB, et al. (2009) Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. *Science* 323:393–396.
57. Viruses ICoTo (2012) *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses* (Academic, London), p 1327.
58. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845–858.
59. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–W248.
60. Routhier E, Bruenn JA (1998) Functions of conserved motifs in the RNA-dependent RNA polymerase of a yeast double-stranded RNA virus. *J Virol* 72:4427–4429.
61. te Velthuis AJW (2014) Common and unique features of viral RNA-dependent polymerases. *Cell Mol Life Sci* 71:4403–4420.
62. Gong P, Peersen OB (2010) Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc Natl Acad Sci USA* 107:22505–22510.
63. Bishop DH (1986) Ambisense RNA viruses: Positive and negative polarities combined in RNA virus genomes. *Microbiol Sci* 3:183–187.
64. Elliott RM (1990) Molecular biology of the Bunyaviridae. *J Gen Virol* 71:501–522.
65. Reguera J, Weber F, Cusack S (2010) *Bunyaviridae* RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. *PLoS Pathog* 6:e1001101.
66. Sanchez AJ, Vincent MJ, Nichol ST (2002) Characterization of the glycoproteins of Crimean-Congo hemorrhagic fever virus. *J Virol* 76:7263–7275.
67. Cifuentes-Muñoz N, Salazar-Quiroz N, Tischler ND (2014) Hantavirus Gn and Gc envelope glycoproteins: Key structural units for virus cell entry and virus assembly. *Viruses* 6:1801–1822.
68. Moriconi M, et al. (2017) Phlebotomine sand fly-borne pathogens in the Mediterranean Basin: Human leishmaniasis and phlebovirus infections. *PLoS Negl Trop Dis* 11:e0005660.
69. Li CX, et al. (2015) Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* 4:e05378.
70. Remnant EJ, et al. (2017) A diverse range of novel RNA viruses in geographically distinct honey bee populations. *J Virol* 91:e00158–17.
71. Dudas G, Obbard DJ (2015) Are arthropods at the heart of virus evolution? *eLife* 4:e06837.
72. Marklewitz M, Zirkel F, Kurth A, Drosten C, Junglen S (2015) Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. *Proc Natl Acad Sci USA* 112:7536–7541.
73. Franzosa EA, et al. (2015) Sequencing and beyond: Integrating molecular 'omics' for microbial community profiling. *Nat Rev Microbiol* 13:360–372.
74. d'Avila-Levy CM, et al. (2015) Exploring the environmental diversity of kinetoplastid flagellates in the high-throughput DNA sequencing era. *Mem Inst Oswaldo Cruz* 110:956–965.
75. Votýpka J, et al. (2014) *Kentomonas* gen. n., a new genus of endosymbiont-containing trypanosomatids of Strigomonadinae subfam. n. *Protist* 165:825–838.
76. Kostygov AY, Grybchuk-Ieremenko A, Malysheva MN, Frolov AO, Yurchenko V (2014) Molecular revision of the genus *Wallaceina*. *Protist* 165:594–604.
77. Yurchenko VY, Lukeš J, Jirků M, Maslov DA (2009) Selective recovery of the cultivation-prone components from mixed trypanosomatid infections: A case of several novel species isolated from Neotropical Heteroptera. *Int J Syst Evol Microbiol* 59:893–909.
78. Votýpka J, et al. (2010) Probing into the diversity of trypanosomatid flagellates parasitizing insect hosts in South-West China reveals both endemism and global dispersal. *Mol Phylogenet Evol* 54:243–253.
79. Rodríguez-Cousiño N, Solórzano A, Fujimura T, Esteban R (1998) Yeast positive-stranded virus-like RNA replicons. 20 S and 23 S RNA terminal nucleotide sequences and 3' end secondary structures resemble those of RNA coliphages. *J Biol Chem* 273:20363–20371.
80. Murphy WJ, Watkins KP, Agabian N (1986) Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: Evidence for trans splicing. *Cell* 47:517–525.
81. Günzl A (2010) The pre-mRNA splicing machinery of trypanosomes: Complex or simplified? *Eukaryot Cell* 9:1159–1170.
82. Mandelboim M, Estraño CL, Tschudi C, Ullu E, Michaeli S (2002) On the role of exon and intron sequences in trans-splicing utilization and cap 4 modification of the trypanosomatid *Leptomonas collosoma* SL RNA. *J Biol Chem* 277:35210–35218.
83. Cai G, Myers K, Fry WE, Hillman BI (2012) A member of the virus family *Narnaviridae* from the plant pathogenic oomycete *Phytophthora infestans*. *Arch Virol* 157:165–169.
84. Reinhardt JA, et al. (2014) Meiotic drive impacts expression and evolution of x-linked genes in stalk-eyed flies. *PLoS Genet* 10:e1004362.
85. Rastgou M, et al. (2009) Molecular characterization of the plant virus genus *Ourniavirinae* and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. *J Gen Virol* 90:2525–2535.
86. Hillman BI, Cai G (2013) The family *narnaviridae*: Simplest of RNA viruses. *Adv Virus Res* 86:149–176.
87. Koonin EV, Dolja VV, Krupovic M (2015) Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* 479–480:2–25.
88. Dolja VV, Krupovic M (2013) Accelerating expansion of the viral universe. *Curr Opin Virol* 3:542–545.
89. Araujo JP, Hughes DP (2016) Diversity of entomopathogenic fungi: Which groups conquered the insect body? *Adv Genet* 94:1–39.
90. Mitchell PL (2004) Heteroptera as vectors of plant pathogens. *Neotrop Entomol* 33:519–545.
91. Ro YT, Scheffter SM, Patterson JL (1997) Hygromycin B resistance mediates elimination of *Leishmania* virus from persistently infected parasites. *J Virol* 71:8991–8998.
92. Patterson JL (1993) The current status of *Leishmania RNA virus 1*. *Parasitol Today* 9:135–136.
93. Widmer G, Dooley S (1995) Phylogenetic analysis of *Leishmania RNA virus* and *Leishmania* suggests ancient virus-parasite association. *Nucleic Acids Res* 23:2300–2304.
94. Hartley MA, Ronet C, Zangger H, Beverley SM, Fasel N (2012) *Leishmania* RNA virus: When the host pays the toll. *Front Cell Infect Microbiol* 2:99.
95. Obbard DJ, Gordon KH, Buck AH, Jiggins FM (2009) The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc Lond B Biol Sci* 364:99–115.
96. Matveyev AV, et al. (2017) The evolutionary loss of RNAi key determinants in kinetoplastids as a multiple sporadic phenomenon. *J Mol Evol* 84:104–115.
97. Lye LF, et al. (2010) Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathog* 6:e1001161.
98. Ghosh S, Banerjee P, Sarkar A, Datta S, Chatterjee M (2012) Coinfection of *Leptomonas seymouri* and *Leishmania donovani* in Indian leishmaniasis. *J Clin Microbiol* 50:2774–2778.
99. Krautz R, Arefin B, Theopold U (2014) Damage signals in the insect immune response. *Front Plant Sci* 5:342.
100. Podlipaev SA (2000) Insect trypanosomatids: The need to know more. *Mem Inst Oswaldo Cruz* 95:517–522.
101. Chomczyński P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156–159.
102. Potgieter AC, et al. (2009) Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes. *J Gen Virol* 90:1423–1432.
103. Liu H, et al. (2012) Evolutionary genomics of mycovirus-related dsRNA viruses reveals cross-family horizontal gene transfer and evolution of diverse viral lineages. *BMC Evol Biol* 12:91.
104. Maslov DA, Yurchenko VY, Jirků M, Lukeš J (2010) Two new species of trypanosomatid parasites isolated from Heteroptera in Costa Rica. *J Eukaryot Microbiol* 57:177–188.
105. Yurchenko V, Lukeš J, Tesarová M, Jirků M, Maslov DA (2008) Morphological discordance of the new trypanosomatid species phylogenetically associated with the genus *Crithidia*. *Protist* 159:99–114.
106. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30:772–780.
107. Capella-Gutiérrez S, Silla-Martinez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
108. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274.
109. Ronquist F, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542.
110. Chistyakova LV, Kostygov AY, Kornilova OA, Yurchenko V (2014) Reisolation and redescription of *Balantidium duodenis* Stein, 1867 (Litostomatea, Trichostomatia). *Parasitol Res* 113:4207–4215.
111. Kalyanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589.
112. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9:868–877.
113. Mullapudi E, et al. (2016) Structure and genome release mechanism of the human cardiomyocyte *Saffold virus 3*. *J Virol* 90:7628–7639.
114. Owczarzyk R, et al. (2008) IDT SciTools: A suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res* 36:W163–W169.