Title: Viral discovery and diversity in trypanosomatid protozoa with a focus on relatives of the human parasite Leishmania.
Authors: Grybchuk D., Akopyants N.S., Kostygov A.Y., Konovalovas A., Lye L.F., Dobson D.E., Zangger H., Fasel N., Butenko A., Frolov A.O., Votýpka J., d'Avila-Levy C.M., Kulich P., Moravcová J., Plevka P., Rogozin I.B., Serva S., Lukeš J., Beverley S.M., Yurchenko V.

Journal: Proceedings of the National Academy of Sciences of the United States of America

Year: 2018

Volume: 115(3)

Pages: E506-E515

DOI: 10.1073/pnas.1717806115

Running Title: **Viral discovery in trypanosomatids**

Classification: **Biological sciences - Microbiology**

Title: Viral discovery and diversity in trypanosomatids with a focus on relatives of the human parasite *Leishmania*

Authors: Danyil Grybchuk[1], Natalia S. Akopyants[2,$], Alexei Y. Kostygov[1,3,$], Aleksandras Konovalovas[4], Lon-Fye Lye[2], Deborah E. Dobson[2], Haroun Zangger[5], Nicolas Fasel[5], Anzhelika Butenko[1], Alexander O. Frolov[3], Jan Votýpka[6,7], Claudia M. d'Avila-Levy[8], Pavel Kulich[9], Jana Moravcová[10], Pavel Plevka[10], Igor B. Rogozin[11], Saulius Serva[4,12], Julius Lukeš[7,13,14], Stephen M. Beverley[2,*,#] and Vyacheslav Yurchenko[1,7,15,16,*]

[1] Life Science Research Centre, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic; [2] Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, 63110, USA; [3] Zoological Institute of the Russian Academy of Sciences, St. Petersburg, 199034, Russia; [4] Department of Biochemistry and Molecular Biology, Institute of Biosciences, Life Sciences Center, Vilnius University, Vilnius 10257, Lithuania; [5] Department of Biochemistry, University of Lausanne, 1066 Epalinges, Switzerland; [6] Department of Parasitology, Faculty of Science, Charles University, 128 44 Prague, Czech Republic; [7] Biology Centre, Institute of Parasitology, Czech Academy of Sciences, 370 05 České Budejovice (Budweis), Czech Republic; [8] Coleção de Protozoários, Laboratório de Estudos Integrados em Protozoologia, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, 21040-360 Rio de Janeiro, Brazil; [9] Veterinary Research Institute, 621 00 Brno, Czech Republic; [10] Central European Institute of Technology – Masaryk University, 625 00 Brno, Czech Republic; [11] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bldg.38A, room 5N505A, Bethesda, MD 20894, USA; [12] Department of Chemistry and Bioengineering, Faculty of Fundamental Sciences, Vilnius Gediminas Technical University, Vilnius 10223, Lithuania; [13] University of South Bohemia, Faculty of Sciences, 370 05 České Budejovice (Budweis), Czech Republic; [14] Integrated Microbial Biodiversity Program, Canadian Institute for Advanced Research, Toronto, ON M5G 1Z8, Canada; [15] Institute of Environmental Technologies, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic; [16] Department of Pathology, Albert Einstein College of Medicine, Bronx, 10461, USA.

* Corresponding authors: Vyacheslav Yurchenko, Life Science Research Centre, Faculty of Science, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czech Republic, Tel: +420 597 092 326, e-mail vyacheslav.yurchenko@osu.cz; Stephen Beverley, Department of Molecular Microbiology, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis MO 63110 USA; Tel 1-314-705-1622, email stephen.beverley@wustl.edu. These authors contributed equally to this manuscript.

Contributing author: Stephen M. Beverley (NAS 2013)

$ These authors contributed equally to this manuscript.

Keywords

Trypanosomatidae, *Bunyavirales*, persistent virus infection, coinfection, coevolution

Author contributions

Performed research: DG, NSA, AYK, AK, LFL, DED, HZ, PK, JM
Designed research: DG, NSA, AYK, NF, PP, SS, SMB, VY
Analyzed data: DG, NSA, AYK, AB, IBR, SMB, VY
Contributed new reagents or analytic tools: AOF, JV, CAL, JL
Wrote the paper: DG, AYK, SMB, VY

Abstract

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 pyrrhocoris* was a hotbed for viral discovery, carrying a new virus (*Leptomonas pyrrhocoris* 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 narna-like 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 new proposed family termed "Leishbunyaviridae". Numerous relatives of trypanosomatid viruses were found in insect meta-transcriptomic surveys, which likely arise from trypanosomatid microbiota. Despite extensive sampling we found no relatives of the totivirus *Leishmaniavirus* (LRV1/2), implying that it was acquired at about the same time the *Leishmania* became able to parasitize vertebrates. As the new viruses were found in over a quarter of isolates tested, 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.

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 new viral taxa including "Leishbunyaviridae*"* and a highly divergent new virus termed "Leptomonas pyrrhocoris ostravirus 1". Our studies provide important information on the origins as well as diversity and distribution of viruses within a group of protists related to the human parasite *Leishmania*.

\body

Introduction

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 towards 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 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 monoxenous to 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.,

Crithidia 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 *L. guyanensis* and *L. braziliensis* (24, 25), and an unrelated RNA virus in *Phytomonas* (21). The biological significance of these lay 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 *Lep. seymouri* strains bear NLV1 (40). Thus, there appears to be considerable unexplored viral diversity in trypanosomatids, the study of which may contribute to 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 *Crithidia* and *Leptomonas* (subfamily Leishmaniinae), as well as 8 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 (41) (Figs. 1A, 3A, 4A). From this analysis, eleven 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 pyrrhocoris*

Three of 18 isolates of *L. pyrrhocoris* originating from various locations worldwide (42) exhibited viral dsRNA bands (H10, F165, and F19; 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), while two (H10, F19) contained 6 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 new viruses. The first was distantly related to *Tombusviridae*. It comprised RNAs T1 and T2 and was named "Leptomonas pyrrhocoris tombus-like virus 1" (LeppyrTLV1). The second virus comprising RNAs O1-O6 could not be associated to any of known viral groups and was named "Leptomonas pyrrhocoris ostravirus 1" (after the city of Ostrava where it was discovered, hereafter LeppyrOV1). PCR tests confirmed the presence/absence 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% nt identity, respectively). RNA-T1 contained two overlapping ORFs with predicted proteins of 850 and 515 aa (Fig. 1B, C). For ORF1, BLAST search in NCBI non-redundant protein database did not yield any hits. The ORF2 showed a clear homology to viral RDRP (cd01699 in Conserved Domain Database) 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 -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 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, 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, as well as freshwater crustaceans and gastropods (Fig. 1D, Table S2) (1). *Pyrrhocoris apterus*, the firebug host of *L. pyrrhocoris*, is known to feed on the corpses of invertebrates (52), suggesting this as a possible route of acquisition.

Endogenous viral element (EVE) related to *LeppyrTLV1*. BLAST searches against the genome assembly of *L. pyrrhocoris* 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). Similarly to 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 aa identity, including conservation of key RDRP motifs (Fig. S1D). Whole transcriptome data for *L. pyrrhocoris* (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 (Table S3). In contrast, this EVE1 region differed by 180 nt substitutions (and 84 indels) from the

corresponding part of LeppyrTLV1 RDRP, whilethe TLV1 RDRP sequences of strains H10 and F165 differed by only 7 nt substitutions. The similarity between EVE1 and TLV1 suggests that a TLV1-like RNA was captured *via* reverse transcription and integration into the *L. pyrrhocoris* genome. EVEs occur frequently in evolution and are thought to be mediated primarily by reverse transcriptases encoded in host retroposons (54-56). Indeed, a number of TATE and SLACS retroelements have been identified in the *L. pyrrhocoris* genome (53), including one located immediately upstream of the LeppyrTLV-EVE1 (Fig. S1A). The high level of sequence divergence with LeppyrTLV1 points to a relatively ancient origin of EVE1, perhaps predating the dispersal of *L. pyrrhocoris* across Europe (42).

LeppyrOV1. The six RNAs O1-O6 of strains H10 and F19 (Fig. 1A) were initially viewed as "satellite" RNAs of LeppyrTLV1. However, several observations suggested that they comprised separate virus. Firstly, unlike TLV1 RNAs T1 and T2, the termini of RNAs O1-O6 shared common sequences: AAAGAAAAAA 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 replication of viruses and often are defining features of viral families (57). Secondly, in all strains the ratio of RNAs T1 and 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, aa 767-870, e = 0.89; PHYRE 2, aa 684-874, confidence = 56%; HHPRED, aa 693-874, confidence = 89.7%; (58, 59)). Within this region we identified conserved viral RDRP motifs responsible for catalytic activity and ribonucleotide selectivity (60-62) (Fig. 2B). Analysis of the base frequencies of codon third positions of the viral ORFs showed significant differences between TLV1 and OV1, and a greater degree between these and the nuclear genome of *L. pyrrhocoris* (Table S4).

Thus, we conclude that RNAs O1-O6 comprise a novel virus, “Leptomonas pyrrhocoris ostravirus 1” (LeppyrOV1). As yet, we have not found a trypanosomatid strain containing this virus alone, which would firmly establish its independence from LeppyrTLV1. Further studies are required to address the functional relationships between the 6 segments of this virus and significance of its co-occurrence with LeppyrTLV1.

A new bunyavirus-like genus ("Leishbunyavirus" or LBV)

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 the first species within a new genus, “Leishbunyavirus” (LBV) (37). We confirmed the presence of LepmorLBV1a and 1b in *L. moramango,* as well as new LBV1s in the dixenous phytopathogenic *Phytomonas* sp. TCC231 (PTCCLBV1) and four species of *Crithidia: C. otongatchiensis* (CotoLBV1), *Crithidia abscondita* (CabsLBV1), *Crithidia* sp. G15 (CG15LBV1), and *Crithidia* sp. ZM (CZMLBV1) (Fig 3A, 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 *C. pragensis* as well (Table 1). The new LBV1-positive strains showed three dsRNAs, except PTCCLBV1 which exhibited only two (Table 1). We sequenced all segments of CotoLBV1, CabsLBV1, and the largest segment (completely or partially) of the others (Table 1; Table S3).

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; 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 L and M segments, we identified the sequence ACACAAAG at the 5' end (as defined by the viral sense orientation) and the complementary sequence CTTTGTGT at the 3' end. These terminal eight nucleotides are typically found in all viruses belonging to the family *Phenuiviridae* (Table S5).

The ORFs in all completely sequenced L segments encoded putative proteins of ~2,000 aa homologous to bunyaviral RDRP domain (aa 600–1,200; pfam04196, e-values < 10^{-10}, NCBI CDD). The region between aa 85 and 150 showed homology to bunyaviral endonuclease domain, with conserved key residues involved in Mn^{2+} ion coordination and phosphodiester bond cleavage (Fig. S2A). Both the RDRP and endonuclease domains are essential for bunyaviral replication (65).

Various database searches (BLAST/CDD, PHYRE2, 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 2 in CabsLBV1, 1 in CotoLBV1, and 0–1 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 system 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 similarity of those 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 new family "Leishbunyaviridae" within the *Bunyavirales*. RDRP-based phylogenetic trees showed that the LBV1s formed a well-supported clade separate from other major *Bunyavirales* groups (Fig. 3D). The closest family was the *Phenuiviridae*, consistent with the similarities noted earlier in the terminal panhandle elements (Table S5). 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 or greater than other bunyavirus inter-familial divergences, we propose that this clade be recognized as a new family, termed "Leishbunyaviridae".

Identification of new 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 honeybees – Apis bunyavirus 1 (ABV1) and Duke bunyavius (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 data sets. Nonetheless, BLASTN searches of assembled contigs revealed several abundant trypanosomatid

transcripts, such as 18S rRNA or paraflagellar rod proteins (Table S3), 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 (Table S2; Table S3). While the co-occurrence of reads for leishbunyaviruses and trypanosomatids in the meta-transcriptomic 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 LBV1s 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 LBV1s bear hallmarks of infectious bunaviruses, 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 (2.9+1.5 kb and 3.8 kb fragments, respectively; Table 1; Fig. 4A, B), in agreement with previous findings (38-40) that these species bear *L. seymouri* narna-like virus 1 (LepseyNLV1) and *P. 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 a putative sub-terminal hairpin structures (Fig. 4C and S4A). Intriguingly, the very 5' end of the viral RNA showed similarity to the spliced leader (SL) of *P. 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 5 nt and had 3 internal mismatches (Fig. S4B), rendering it unlikely to be functional based on current knowledge of SL function (82). Thus, in the past the *Pser*NV1 may have "snatched" the host's SL, substituting it for the original terminus. In *Lepsey*NLV1 we did not determine the terminal sequences explicitly, however typical narnaviral sub-terminal 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, 4C), 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 with known trypanosomatid sequences, suggesting that it may belong to a yet uncharacterized lineage. Thus and similar to leishbunyaviruses, the insect metatranscriptomic narnavirus may have arisen from its trypanosomatid microbiota.

As it was inferred earlier, *Ourmiavirus* and ourmia-like viruses (family *Ourmiaviridae*) 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, Ourmiaviridae typically contain several segments (87, 88). LepseyNLV1 with its two segments exhibits an independently evolved genome organization being intermediate between those of *Narnaviridae* and *Ourmiaviridae*. 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 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, 5). For leishbunyaviruses, we noticed a gradual decrease of viral dsRNA levels in *C. otongatchiensis* over six months of continuous cultivation, and their disappearance from *C. pragensis* and *L. moramango* after two weeks 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.

There are several non-exclusive mechanisms explaining 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 *L. guyanensis* LRV1 (6, 91). Indeed, 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*, *Bunyavirales*, tombus-like viruses and a previously unknown virus. This 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 retroposons.

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 (92). Narnaviruses, by virtue of lacking either a capsid or an envelope, are only transmitted 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 *L. pyrrhocoris* viruses OV1 and TLV1 have sufficient coding capacity for transmission. Lastly, 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 on the origins and evolution of their viruses (Fig. 5). Firstly, *Leptomonas pyrrhocoris* appears to be a hotbed for viral discovery, with two new viruses (LeppyrTLV1 and LeppyrOV1), and 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 those viruses showing increased likelihood for infectivity (LBV1s, 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 co-evolution. Thus, it is remarkable that for LRV1/2 (93) and, to some extent, for LBVs (Fig. S3), 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 amongst kingdoms.

Notably, in our survey we did not find any LRV-related *Totiviridae* (Table 1, Fig. 5), despite the fact that 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 *L. major* or *L. 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 substituted by other mechanisms contributing to virulence, such as type I interferon induction (29).

Variation in the RNA interference (RNAi) pathway may contribute to the observed patchiness in viral distribution, as this pathway acts as an anti-viral defense mechanism in many species (95). In agreement with this, RNAi pathway (believed to be ancestral to all eukaryotes) is

absent in *Phytomonas* spp., *Leptomonas seymouri* and LRV2-bearing *Leishmania aethiopica* and *L. 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 *L. guyanensis* and *L. braziliensis* have a highly active RNAi pathway (97), and accordingly LRV1 has mastered the ability to co-exist 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 of trypanosomatid viruses in the vertebrate host (6, 27-29, 40, 98). Our studies now suggest that potential role(s) of 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 is 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.

New 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 *L. pyrrhocoris* 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 this suggests that a great number of new 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 (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). 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 single strand (ss) RNA 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 Zymoclean Gel RNA Recovery Kit (Zymo Research).

Viral dsRNA amplification, cloning and sequencing. Gel-extracted dsRNA was polyadenylylated at both 3'-ends using *E. coli* Poly(A) Polymerase (New England Biolabs) and then purified on a PCR-clean up column (ThermoFisher Scientific) according to the manufacturer's protocol. Next, polyadenylylated dsRNA was reverse-transcribed using the Transcriptor First Strand cDNA synthesis kit Roche, and an anchored-oligo (dT) primer QD2-T20 5'-ggcaattaaccctcactatagaattcgttcgatctttttttttttttttttttt-3' (modified from (103)). To prevent renaturation of the complementary RNA strands, DMSO was added to the final concentration of 7.5% and the residual RNA was lysed with 0.1 M NaOH. The complementary 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 (ThermoFisher Scientific) with primer QD2 5'-tcactatagaattcgttcgatc-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 manufacturer's

recommended cycling conditions: 98°C for 10 sec, 55 °C for 30 sec, and 72 °C for 40 sec per kb. Obtained PCR products were cloned into the pTZ57R vector (ThermoFisher Scientific) and sequenced by primer walking. For the analysis described above, we were unable to obtain enough dsRNA from *Crithidia* sp. C4 and *C. pragensis*. In these cases, the partial RNA-dependent RNA polymerase gene (RDRP, ~ 900 bp) was amplified using degenerate primers designed to amplify known LBV1s ((37) and this work; primers LeiBunyaF 5'-ttykcvacnttcaagaaragcac-3' and LeiBunyaR 5'-ccagartcatcwgadgadaccat-3') and the products cloned into the pTZ57R vector and sequenced. To assess the presence of the *Leptomonas pyrrhocoris* RNA virus, total cDNA of all *L. pyrrhocoris* isolates (both positive and negative as judged by gel-based assay) was amplified with primers LpTLV1F 5'-ttactcctataacggggca-3' and LpTLV1R 5'-taaaggagcgaattctgct-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'-cctatgcggatgcactcaa-3' and LpIVR 5'-cttgtgcattttctatccaag-3'. PCR Primers M200 5'-atggctccvvtcaargtwggmat-3' and M201 5'-takccccactcrttrtcrtacca-3' for the glycosomal glyceraldehyde 3-phosphate dehydrogenase (gGAPDH) gene were used as an internal positive control (104). Additional methods for cultivation of trypanosomatids, phylogenetic, genomic and transcriptomic analyses, as well as the negative stain transmission electron microscopy can be found in the Supplementary information.

ACKNOWLEDGEMENTS. We thank members of our groups, D. Wang and S. Krishnamurthy (WUSM, St. Louis) for discussion, and D. Maslov, L. Simpson, V. Bellofatto for providing *Phytomonas* and *Leptomonas* strains. This work was supported by NIH Grants R01-AI029646 and R56-AI099364 (SMB), the Intramural Research Program of the National Library of Medicine at the NIH (IBR), Grant Agency of Czech Republic awards 17-10656S (VY and JV), 16-18699S (JL and VY), Moravskoslezský kraj research initiative DT01-021358 (VY and AYK), the Russian Foundation for Basic Research (project no. 15-29-02734) and the State assignment for the Zoological Institute AAAA-A17-117030310322-3 (AOF), and the COST action CM1307 (JL). Work in VY lab is financially supported by the Ministry of Education, Youth and Sports of the Czech 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.

Figure Legends.

Figure 1 Tombus-like virus from *Leptomonas pyrrhocoris*. A, Agarose gel electrophoresis of S1-digested total RNAs from strains H10 (lane 1), F19 (lane 2), and F165 (lane 3). LeppyrTLV1 segments are labeled RNA-T1 and RNA-T2 on the right and marked by green dots; LeppyrOV1 segments are marked by red dots. Left lane: 1-kb DNA ladder. B, Genome structure of LeppyrTLV1. 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 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 ML bootstrap supports, respectively; ones having values of 1.0 and 100%, respectively are marked with black circles. The scale bar indicates number of substitutions per site. The tree was rooted with the sequences of *Nodaviridae*. Abbreviations and GenBank accession numbers are given in Tables S2 and S3.

Figure 2 Leptomonas pyrrhocoris ostravirus 1 , a **unique virus from *Leptomonas pyrrhocoris*.** A, Genome structure of LeppyrOV1, showing shared terminal sequences and single ORF per segment (squiggle marks 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 LeppyrOV1 putative RDRP with those of *Picorna- Flavi-* and *Caliciviridae*. Identical residues are shown in red, similar residues - in blue. Amino acid motifs, typically found in viral RDRPs are highlighted in yellow.

Figure 3 Leishbunyaviruses. A, viral dsRNA from *C. otongatchiensis* (lane 1), *C. abscondita* (lane 2), *Crithidia* sp. G15 (lane 3) and *Crithidia* sp. ZM (lane 4). Left lane: 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 *C. otongatchiensis*, scale bar is 100 nm. D, Maximum likelihood phylogenetic tree based on RDRP amino acid sequences. Numbers at the branches indicate Bayesian posterior probability and ML bootstrap supports, respectively; ones having values of 1.0 and 100%, respectively are marked with black circles. The scale bar indicates number of substitutions per site. The tree was rooted at the midpoint. Abbreviations and GenBank accession numbers are given in Tables S2 and S3. The definition of pictograms describing viral hosts is the same as in Fig 1.

Figure 4 Narnaviruses of trypanosomatids. A, Viral dsRNA in two sub-cultures of *Phytomonas serpens* (isolate 9T): 1 – from University of California Riverside and 2 – from Institute of Parasitology in České Budějovice. M – 1-kb DNA ladder. Total RNA (below) was used as a loading control. B, left panel, Viral dsRNA in four sub-cultures of *Leptomonas seymouri* ATCC30220: lanes 1 and 4 – original ATCC culture; lane 2 and 3 – sub-strains 2003WT and 294-1993VB (Rutgers University), lane 5 – culture from Zoological Institute of RAS. *Leishmania guyanensis* (*Lgy*, strain M4147) bearing the 5.3 kb LRV1 served as a positive control. M – 1-kb DNA ladder. Total RNA (below) was used as a loading control for virus-negative sub-strains. 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. *Lepsey*NLV1 and *Pser*NV1 are indicated with trypanosomatid symbol. Numbers at

the branches indicate Bayesian posterior probability and ML bootstrap supports, respectively; ones having values of 1.0 and 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 and S3. The definition of pictograms describing viral hosts is the same as in Fig 1.

Figure 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 whichthe absence or presence of viruses is marked (see graphical legend). Arrows denote hypothetical acquisition of viruses under assumption of single origin in the common ancestor. Maximal bootstrap supports are marked by filled circles, while bootstrap supports over 70% are denoted by open cicles. The scale bar indicates number of substitutions per site.

Table legends

Table 1 Virus-positive trypanosomatid isolates. The sequenced segments are underlined and the segments containing RNA-dependent RNA polymerase are in bold. Accession numbers for viral or related segments sequenced in this work are reported in Table S3.

† sequences published elsewhere. ND, not determined; N/A, not applicable.

References

1. Shi M, *et al.* (2016) Redefining the invertebrate RNA virosphere. *Nature* 540(7634):539-543.
2. Wang AL & Wang CC (1991) Viruses of the protozoa. *Annu Rev Microbiol* 45:251-263.
3. Banik GR, Stark D, Rashid H, & Ellis JT (2014) Recent advances in molecular biology of parasitic viruses. *Infect Disord Drug Targets* 14(3):155-167.
4. Ghabrial SA, Castón JR, Jiang D, Nibert ML, & Suzuki N (2015) 50-plus years of fungal viruses. *Virology* 479-480:356-368.

5. 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(16):7562-7567.
6. Ives A, *et al.* (2011) *Leishmania* RNA virus controls the severity of mucocutaneous leishmaniasis. *Science* 331(6018):775–778.
7. Fichorova RN, *et al.* (2012) Endobiont viruses sensed by the human host - beyond conventional antiparasitic therapy. *PLOS One* 7(11):e48418.
8. 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(5):657-678.
9. Stuart K & Panigrahi AK (2002) RNA editing: complexity and complications. *Mol Microbiol* 45(3):591-596.
10. Záhonová K, Kostygov A, 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(17):2364-2369.
11. Simpson L (1998) A base called J. *Proc Natl Acad Sci U S A* 95(5):2037-2038.
12. Michaeli S (2011) Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol* 6(4):459-474.
13. Nussbaum K, Honek J, Cadmus CM, & Efferth T (2010) Trypanosomatid parasites causing neglected diseases. *Curr Med Chem* 17(15):1594-1617.
14. 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(1):43–52.
15. Flegontov P, *et al.* (2013) *Paratrypanosoma* is a novel early-branching trypanosomatid. *Curr Biol* 23(18):1787–1793.
16. Lukeš J, Skalický T, Týč J, Votýpka J, & Yurchenko V (2014) Evolution of parasitism in kinetoplastid flagellates. *Mol Biochem Parasitol* 195(2):115–122.
17. Kostygov AY & Yurchenko V (2017) Revised classification of the subfamily Leishmaniinae (Trypanosomatidae). *Folia Parasitol* 64:020.
18. Molyneux DH (1974) Virus-like particles in *Leishmania* parasites. *Nature* 249(457):588-589.
19. Croft SL, Chance ML, & Gardener PJ (1980) Ultrastructural and biochemical characterization of stocks of *Endotrypanum*. *Ann Trop Med Parasitol* 74(6):585-589.
20. 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.
21. Marche S, Roth C, Manohar SK, Dollet M, & Baltz T (1993) RNA virus-like particles in pathogenic plant trypanosomatids. *Mol Biochem Parasitol* 57(2):261-267.

22. 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(2):97-112.
23. 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(8):e1005127.
24. Widmer G, Comeau AM, Furlong DB, Wirth DF, & Patterson JL (1989) Characterization of a RNA virus from the parasite *Leishmania*. *Proc Natl Acad Sci U S A* 86(15):5979-5982.
25. Tarr PI, *et al.* (1988) LR1: a candidate RNA virus of *Leishmania*. *Proc Natl Acad Sci U S A* 85(24):9572-9575.
26. 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(3):318-328.
27. 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 U S A* 113(43):11998-12005.
28. Zangger H, *et al.* (2014) *Leishmania aethiopica* field isolates bearing an endosymbiontic dsRNA virus induce pro-inflammatory cytokine response. *PLoS Negl Trop Dis* 8(4):e2836.
29. 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 U S A* 114(19):4987-4992.
30. Hartley MA, Drexler S, Ronet C, Beverley SM, & Fasel N (2014) The immunological, environmental, and phylogenetic perpetrators of metastatic leishmaniasis. *Trends Parasitol*.
31. Ronet C, Beverley SM, & Fasel N (2011) Muco-cutaneous leishmaniasis in the New World: the ultimate subversion. *Virulence* 2(6):547-552.
32. 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(1):112-121.
33. Bourreau 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(1):105-111.
34. Ito MM, *et al.* (2015) Correlation between presence of *Leishmania RNA virus 1* and clinical characteristics of nasal mucosal leishmaniosis. *Braz J Otorhinolaryngol* 81(5):533-540.
35. Cantanhede 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(9):e0004079.

36. Pereira Lde O, *et al.* (2013) Severity of tegumentary leishmaniasis is not exclusively associated with *Leishmania RNA virus 1* infection in Brazil. *Mem Inst Oswaldo Cruz* 108(5):665-667.
37. Akopyants NS, Lye LF, Dobson DE, Lukeš J, & Beverley SM (2016) A novel Bunyavirus-like virus of trypanosomatid protist parasites. *Genome Announc* 4(4): e00715-00716.
38. Lye LF, Akopyants NS, Dobson DE, & Beverley SM (2016) A Narnavirus-like element from the trypanosomatid protozoan parasite *Leptomonas seymouri*. *Genome Announc* 4(4):e00713-00716.
39. Akopyants NS, Lye LF, Dobson DE, Lukeš J, & Beverley SM (2016) A *Narnavirus* in the trypanosomatid protist plant pathogen *Phytomonas serpens*. *Genome Announc* 4(4):e00711-00716.
40. Sukla S, Roy S, Sundar S, & Biswas S (2017) *Leptomonas seymouri narna-like virus 1* and not leishmaniaviruses detected in kala-azar samples from India. *Arch Virol* (in press).
41. Beiting DP, *et al.* (2014) Differential induction of TLR3-dependent innate immune signaling by closely related parasite species. *PLoS One* 9(2):e88398.
42. Votýpka J, *et al.* (2012) Cosmopolitan distribution of a trypanosomatid *Leptomonas pyrrhocoris*. *Protist* 163(4):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(3):447-458.
44. Firth AE, Chung BY, Fleeton MN, & Atkins JF (2008) Discovery of frameshifting in *Alphavirus 6K* resolves a 20-year enigma. *Virol J* 5:108.
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(15):7007-7078.
46. Dreher TW & Miller WA (2006) Translational control in positive strand RNA plant viruses. *Virology* 344(1):185-197.
47. Kim KH & Lommel SA (1998) Sequence element required for efficient -1 ribosomal frameshifting in red clover necrotic mosaic dianthovirus. *Virology* 250(1):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(1-2):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(Pt 8):1855-1866.
50. Na H & White KA (2006) Structure and prevalence of replication silencer-3' terminus RNA interactions in *Tombusviridae*. *Virology* 345(2):305-316.
51. Mandahar CL (1999) *Molecular biology of plant viruses* (Kluwer Academic Publishers, Boston) pp xxix, 281 p.

52. Socha R (1993) *Pyrrhocoris apterus* (Heteroptera) - an experimental model species: a review. *Eur J Entomol* 90:241-286.
53. Flegontov P, *et al.* (2016) Genome of *Leptomonas pyrrhocoris*: 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(4):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(5912):393-396.
57. Viruses ICoTo (2012) *Virus taxonomy : classification and nomenclature of viruses : ninth report of the International Committee on Taxonomy of Viruses* (Academic Press, London ; Waltham; San Diego) p 1327 p.
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(6):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(Web Server issue):W244-248.
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(5):4427-4429.
61. te Velthuis AJW (2014) Common and unique features of viral RNA-dependent polymerases. *Cell Mol Life Sci* 71(22):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 U S A* 107(52):22505-22510.
63. Bishop DH (1986) Ambisense RNA viruses: positive and negative polarities combined in RNA virus genomes. *Microbiol Sci* 3(6):183-187.
64. Elliott RM (1990) Molecular biology of the Bunyaviridae. *J Gen Virol* 71 (Pt 3):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(9):e1001101.
66. Sanchez AJ, Vincent MJ, & Nichol ST (2002) Characterization of the glycoproteins of Crimean-Congo hemorrhagic fever virus. *J Virol* 76(14):7263-7275.
67. Cifuentes-Munoz 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(4):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(8):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.
70. Remnant EJ, *et al.* (2017) A diverse range of novel RNA viruses in geographically distinct honey bee populations. *J Virol* 91(16):e00158-17.
71. Dudas G & Obbard DJ (2015) Are arthropods at the heart of virus evolution? *Elife* 4.
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 U S A* 112(24):7536-7541.
73. Franzosa EA, *et al.* (2015) Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat Rev Microbiol* 13(6):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(8):956-965.
75. Votýpka J, *et al.* (2014) *Kentomonas* gen. n., a new genus of endosymbiont-containing trypanosomatids of Strigomonadinae subfam. n. *Protist* 165(6):825-838.
76. Kostygov AY, Grybchuk-Ieremenko A, Malysheva MN, Frolov AO, & Yurchenko V (2014) Molecular revision of the genus *Wallaceina*. *Protist* 165(5):594-604.
77. Yurchenko V, 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(Pt 4):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(1):243–253.
79. Rodriguez-Cousino N, Solorzano 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(32):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(4):517-525.
81. Günzl A (2010) The pre-mRNA splicing machinery of trypanosomes: complex or simplified? *Eukaryot Cell* 9(8):1159-1170.
82. Mandelboim M, Estrano 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(38):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(1):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(5):e1004362.
85. Rastgou M, *et al.* (2009) Molecular characterization of the plant virus genus *Ourmiavirus* and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. *J Gen Virol* 90(Pt 10):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(5):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(5):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(12):8991-8998.
92. Patterson JL (1993) The current status of *Leishmania RNA virus I*. *Parasitol Today* 9(4):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(12):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(1513):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(2-3):104-115.
97. Lye LF, *et al.* (2010) Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathog* 6(10):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(8):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(4):517-522.

101. Chomczyński P & Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162(1):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(Pt 6):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(2):177–188.

105. Yurchenko V, Lukeš J, Tesařová M, Jirků M, & Maslov DA (2008) Morphological discordance of the new trypanosomatid species phylogenetically associated with the genus *Crithidia*. *Protist* 159(1):99–114.

106. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772-780.

107. Capella-Gutierrez S, Silla-Martinez JM, & Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):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(1):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(3):539–542.

110. Chistyakova LV, Kostygov AY, Kornilova OA, & Yurchenko V (2014) Reisolation and redescription of *Balantidium duodeni* Stein, 1867 (Litostomatea, Trichostomatia). *Parasitol Res* 113(11):4207-4215.

111. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, & Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14(6):587-589.

112. Huang X & Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9(9):868-877.

113. Mullapudi E, *et al.* (2016) Structure and genome release mechanism of the human cardiovirus *Saffold virus 3*. *J Virol* 90(17):7628-7639.

114. Owczarzy R, et al. (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nuc. Acids. Res. 36 (Web Server issue): W163-9.

TABLE 1

Species	Isolate	Virus name	# of S1 bands	S1 band sizes (kb)
Bunyavirales				
Crithidia sp.	C4	CC4LBV1	ND	ND
Crithidia sp.	G15	CG15LBV1	3	**6.0**, 1.1, 0.7
Crithidia sp.	ZM	CZMLBV1	3	**6.0**, 1.9, 0.7
C. abscondita	127AL	CabsLBV1	3	**6.0**, 1.1, 0.7
C. otongatchiensis	Ecu-08	CotoLBV1	3	**6.3**, 1.4, 0,7
C. pragensis	MCZ-11	CpraLBV1	ND	ND
Leptomonas moramango	MMO-09	LepmorLBV1a	3	**6.0**, 1.1, 0.7 †
L. moramango	MMO-09	LepmorLBV1b	3	**6.0**, 1.1, 0.7 †
Phytomonas sp.	TCC231	PTCCLBV1	2	**6.0**, 1.0
Narnaviridae				
P. serpens	9T (UCR)	PserNV1	1	**3.8 †**
P. serpens	30T	PserNV1	1	**3.8**
L. seymouri	ATCC	LseyNLV1	2	**2.9**, 1.5 †
Tombus-like viruses				
L. pyrrhocoris	F165	LeppyrTLV1	2	**3.5**, 2.2
L. pyrrhocoris	F19	LeppyrTLV1	2	**3.5**, 2.2
L. pyrrhocoris	H10	LeppyrTLV1	2	**3.5**, 2.2
Tombus-like non-retroviral endogenous RNA viral element				
L. pyrrhocoris	P59	LeppyrTLV-EVE1	N/A	N/A
L. pyrrhocoris	PP1	LeppyrTLV-EVE1	N/A	N/A
L. pyrrhocoris	PP2	LeppyrTLV-EVE1	N/A	N/A
L. pyrrhocoris	LP	LeppyrTLV-EVE1	N/A	N/A
L. pyrrhocoris	H10	LeppyrTLV-EVE1	N/A	N/A
Ostravirus				
L. pyrrhocoris	F19	LeppyrOV1	6	5.2, 4.5, **4.1**, 3.0, 2.5, 1.7
L. pyrrhocoris	H10	LeppyrOV1	6	5.2, 4.5, **4.1**, 3.0, 2.5, 1.7

Figure 1

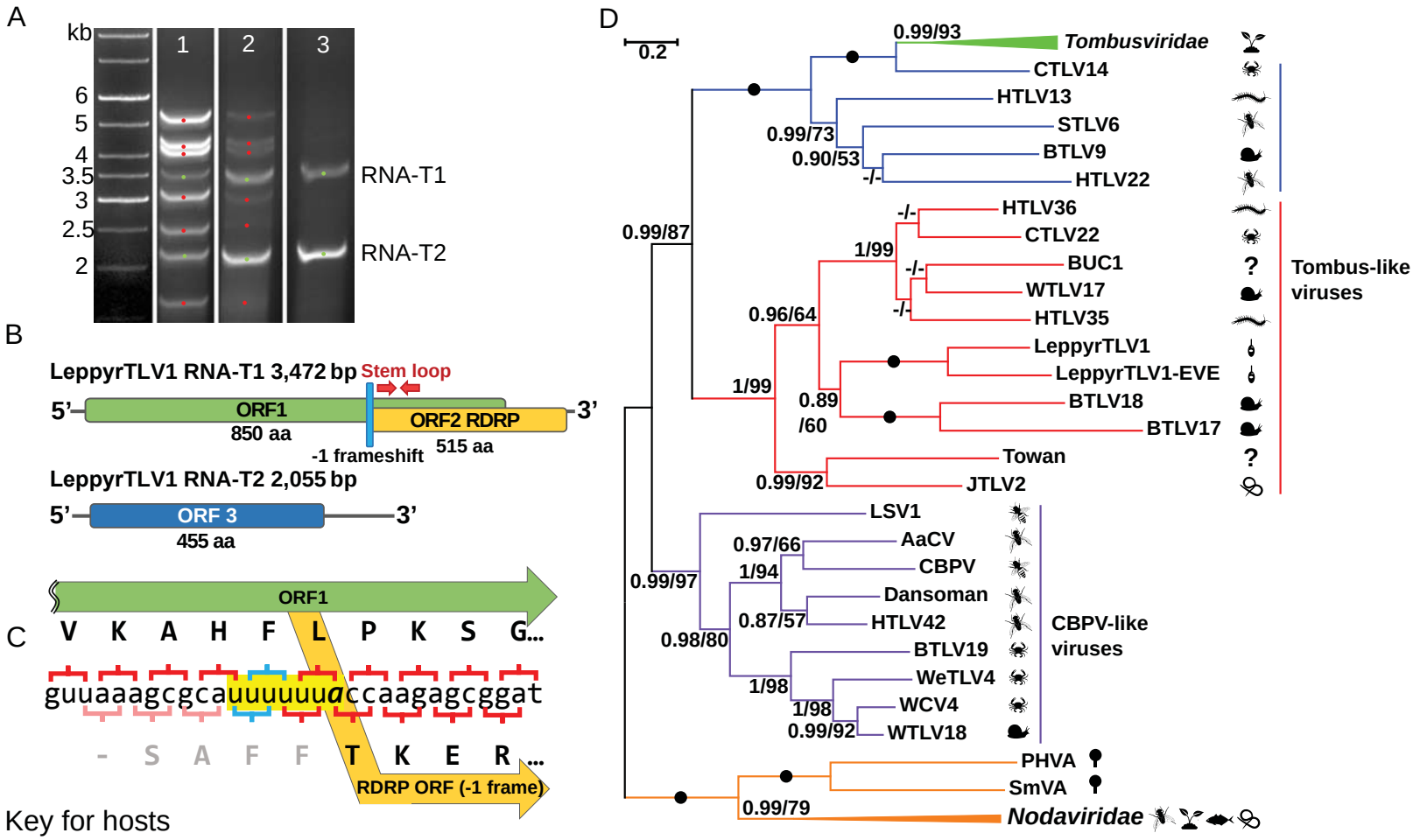
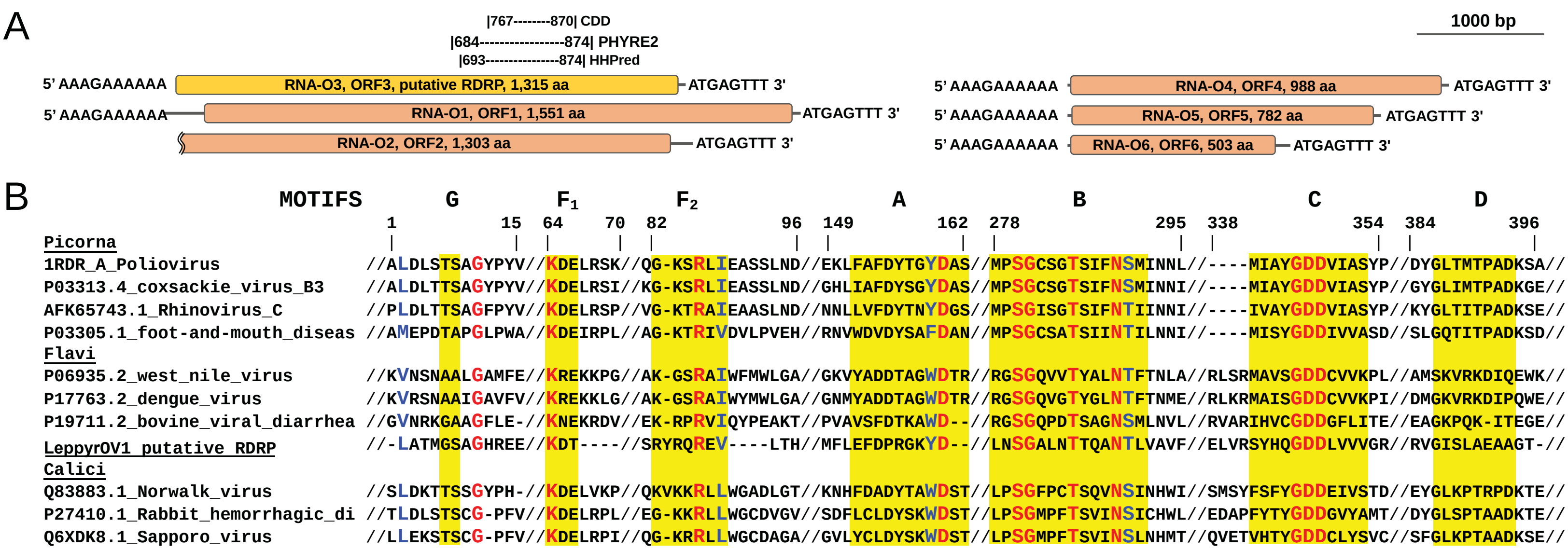

A
kb
6
5
4
3.5
3
2.5
2
1
2
3
RNA-T1
RNA-T2
B
LeppyrTLV1 RNA-T1 3,472 bp
Stem loop
5'
ORF1
850 aa
-1 frameshift
ORF2 RDRP
515 aa
3'
LeppyrTLV1 RNA-T2 2,055 bp
5'
ORF 3
455 aa
3'
C
ORF1
V K A H F L P K S G...
guuaaagcgcauuuuuuuaccaagagcggat
- S A F F T K E R ...
RDRP ORF (-1 frame)
D
0.2
0.99/93
Tombusviridae
CTLV14
HTLV13
0.99/73
STLV6
0.90/53
BTLV9
-/-
HTLV22
-/-
HTLV36
0.99/87
CTLV22
1/99
-/-
BUC1
-/-
WTLV17
-/-
HTLV35
0.96/64
LeppyrTLV1
LeppyrTLV1-EVE
1/99
0.89
/60
BTLV18
BTLV17
Towan
0.99/92
JTLV2
Tombus-like viruses
LSV1
0.97/66
AaCV
1/94
CBPV
0.99/97
Dansoman
0.87/57
HTLV42
0.98/80
BTLV19
WeTLV4
1/98
1/98
WCV4
0.99/92
WTLV18
CBPV-like viruses
PHVA
SmVA
0.99/79
Nodaviridae
Key for hosts
Plants
Crustaceans
Myriapods
Molluscs
Insects, except bees
Bees
Ticks
Nematodes
Vertebrates
Leech
Oomycetes
Fungi
Trypanosomatids
? Unknown
+ ? Metatranscriptomic sequences containing trypanosomatid reads

Figure 2


A
|767--------870| CDD
|684------------------874| PHYRE2
|693----------------874| HHPred
1000 bp
5' AAAGAAAAAA
RNA-O3, ORF3, putative RDRP, 1,315 aa
ATGAGTTT 3'
5' AAAGAAAAAA
RNA-O1, ORF1, 1,551 aa
ATGAGTTT 3'
RNA-O2, ORF2, 1,303 aa
ATGAGTTT 3'
5' AAAGAAAAAA
RNA-O4, ORF4, 988 aa
ATGAGTTT 3'
5' AAAGAAAAAA
RNA-O5, ORF5, 782 aa
ATGAGTTT 3'
5' AAAGAAAAAA
RNA-O6, ORF6, 503 aa
ATGAGTTT 3'
B
MOTIFS G F1 F2 A B C D
1 15 64 70 82 96 149 162 278 295 338 354 384 396
Picorna
1RDR_A_Poliovirus //ALDLSTSAGYPYV//KDELRSK//QG-KSRLIEASSLND//EKLFAFDYTGYDAS//MPSGCSGTSIFNSMINNL//----MIAYGDDVIASYP//DYGLTMTPADKSA//
P03313.4_coxsackie_virus_B3 //ALDLTTSAGYPYV//KDELRSI//KG-KSRLIEASSLND//GHLIAFDYSGYDAS//MPSGCSGTSIFNSMINNI//----MIAYGDDVIASYP//GYGLIMTPADKGE//
AFK65743.1_Rhinovirus_C //PLDLTTSAGFPYV//KDELRSP//VG-KTRAIEAASLND//NNLLVFDYTNYDGS//MPSGISGTSIFNTIINNI//----IVAYGDDVIASYP//KYGLTITPADKSE//
P03305.1_foot-and-mouth_diseas //AMEPDTAPGLPWA//KDEIRPL//AG-KTRIVDVLPVEH//RNVWDVDYSAFDAN//MPSGCSATSIINTILNNI//----MISYGDDIVVASD//SLGQTITPADKSD//
Flavi
P06935.2_west_nile_virus //KVNSNAALGAMFE//KREKKPG//AK-GSRAIWFMWLGA//GKVYADDTAGWDTR//RGSGQVVTYALNTFTNLA//RLSRMAVSGDDCVVKPL//AMSKVRKDIQEWK//
P17763.2_dengue_virus //KVRSNAAIGAVFV//KREKKLG//AK-GSRAIWYMWLGA//GNMYADDTAGWDTR//RGSGQVGTYGLNTFTNME//RLKRMAISGDDCVVKPI//DMGKVRKDIPQWE//
P19711.2_bovine_viral_diarrhea //GVNRKGAAGFLE-//KNEKRDV//EK-RPRVIQYPEAKT//PVAVSFDTKAWD--//RGSGQPDTSAGNSMLNVL//RVARIHVCGDDGFLITE//EAGKPQK-ITEGE//
LeppyrOV1 putative RDRP //-LATMGSAGHREE//KDT----//SRYRQREV----LTH//MFLEFDPRGKYD--//LNSGALNTTQANTLVAVF//ELVRSYHQGDDLVVVGR//RVGISLAEAAGT-//
Calici
Q83883.1_Norwalk_virus //SLDKTTSSGYPH-//KDELVKP//QKVKKRLLWGADLGT//KNHFDADYTAWDST//LPSGFPCTSQVNSINHWI//SMSYFSFYGDDEIVSTD//EYGLKPTRPDKTE//
P27410.1_Rabbit_hemorrhagic_di //TLDLSTSCG-PFV//KDELRPL//EG-KKRLLWGCDVGV//SDFLCLDYSKWDST//LPSGMPFTSVINSICHWL//EDAPFYTYGDDGVYAMT//DYGLSPTAADKTE//
Q6XDK8.1_Sapporo_virus //LLEKSTSCG-PFV//KDELRPI//QG-KRRLLWGCDAGA//GVLYCLDYSKWDST//LPSGMPFTSVINSLNHMT//QVETVHTYGDDCLYSVC//SFGLKPTAADKSE//

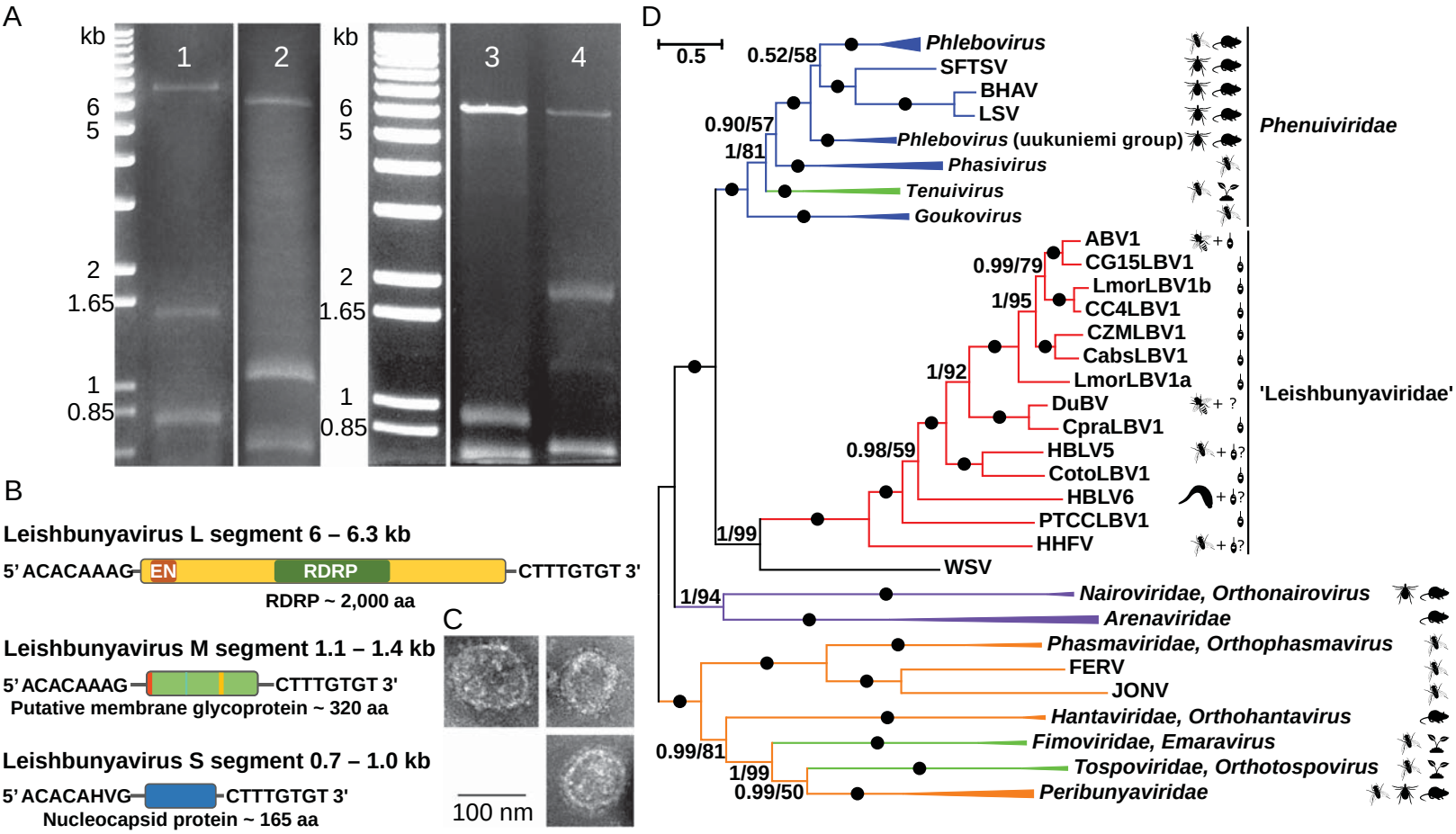
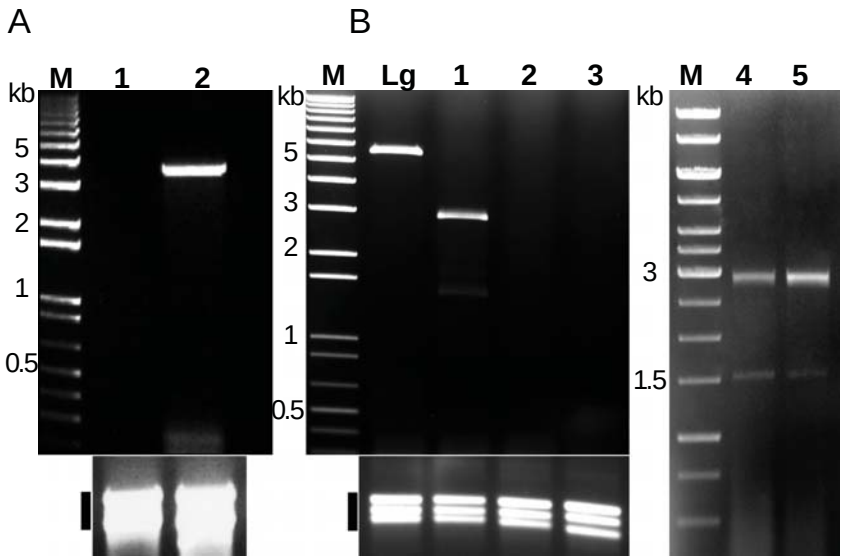
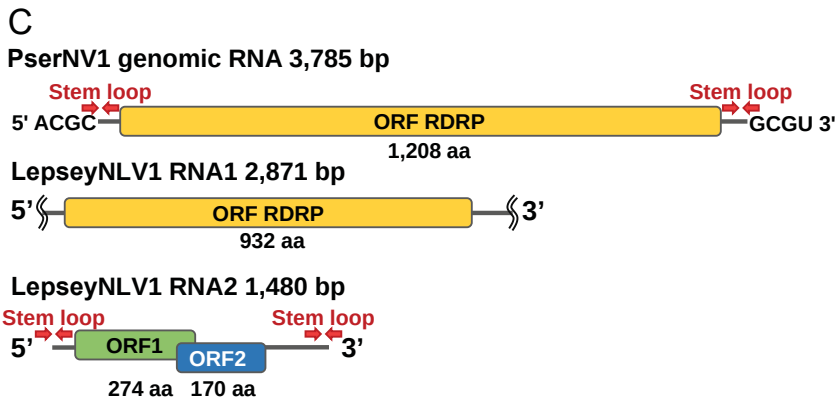

A
kb
1
2
6
5
2
1.65
1
0.85
kb
3
4
6
5
2
1.65
1
0.85
B
Leishbunyavirus L segment 6 – 6.3 kb
5' ACACAAAG
EN
RDRP
CTTTGTGT 3'
RDRP ~ 2,000 aa
Leishbunyavirus M segment 1.1 – 1.4 kb
5' ACACAAAG
CTTTGTGT 3'
Putative membrane glycoprotein ~ 320 aa
Leishbunyavirus S segment 0.7 – 1.0 kb
5' ACACAHVG
CTTTGTGT 3'
Nucleocapsid protein ~ 165 aa
C
100 nm
D
0.5
0.52/58
Phlebovirus
SFTSV
BHAV
LSV
0.90/57
1/81
Phlebovirus (uukuniemi group)
Phasivirus
Tenuivirus
Goukovirus
Phenuiviridae
ABV1
0.99/79
CG15LBV1
LmorLBV1b
1/95
CC4LBV1
CZMLBV1
CabsLBV1
1/92
LmorLBV1a
DuBV
CpraLBV1
0.98/59
HBLV5
CotoLBV1
HBLV6
PTCCLBV1
1/99
HHFV
WSV
'Leishbunyaviridae'
1/94
Nairoviridae, Orthonairovirus
Arenaviridae
Phasmaviridae, Orthophasmavirus
FERV
JONV
Hantaviridae, Orthohantavirus
0.99/81
Fimoviridae, Emaravirus
1/99
Tospoviridae, Orthotospovirus
0.99/50
Peribunyaviridae

Figure 3

Figure 4


A
M 1 2
kb
5
3
2
1
0.5
B
M Lg 1 2 3
kb
5
3
2
1
0.5
M 4 5
kb
3
1.5


C
PserNV1 genomic RNA 3,785 bp
Stem loop
Stem loop
5' ACGC
ORF RDRP
GCGU 3'
1,208 aa
LepseyNLV1 RNA1 2,871 bp
5'
ORF RDRP
3'
932 aa
LepseyNLV1 RNA2 1,480 bp
Stem loop
Stem loop
5'
ORF1
ORF2
3'
274 aa
170 aa

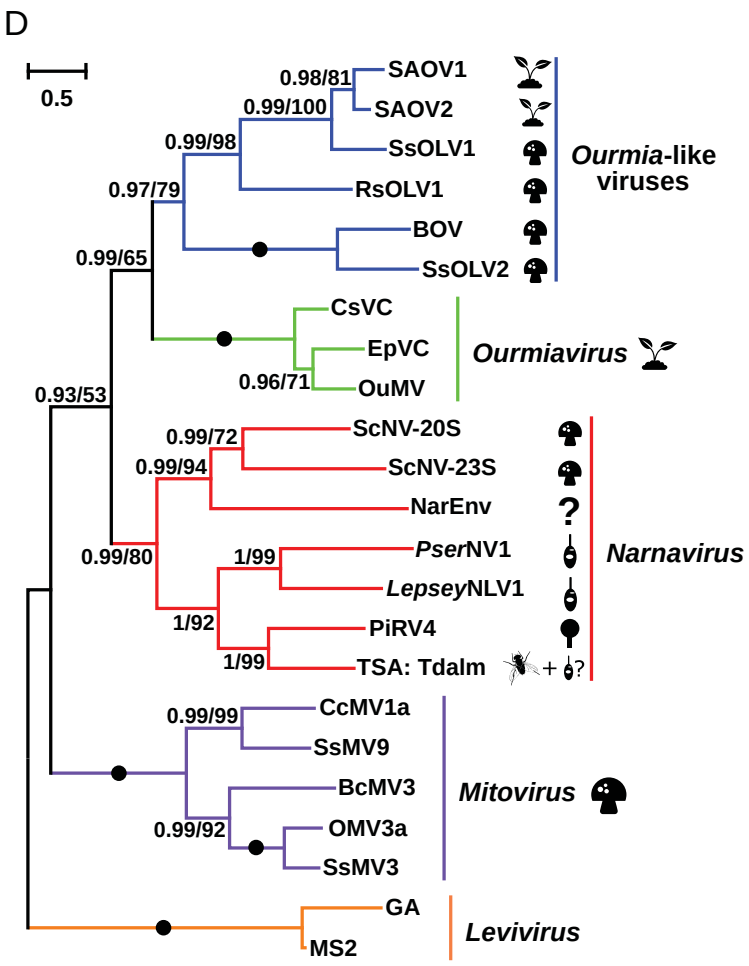
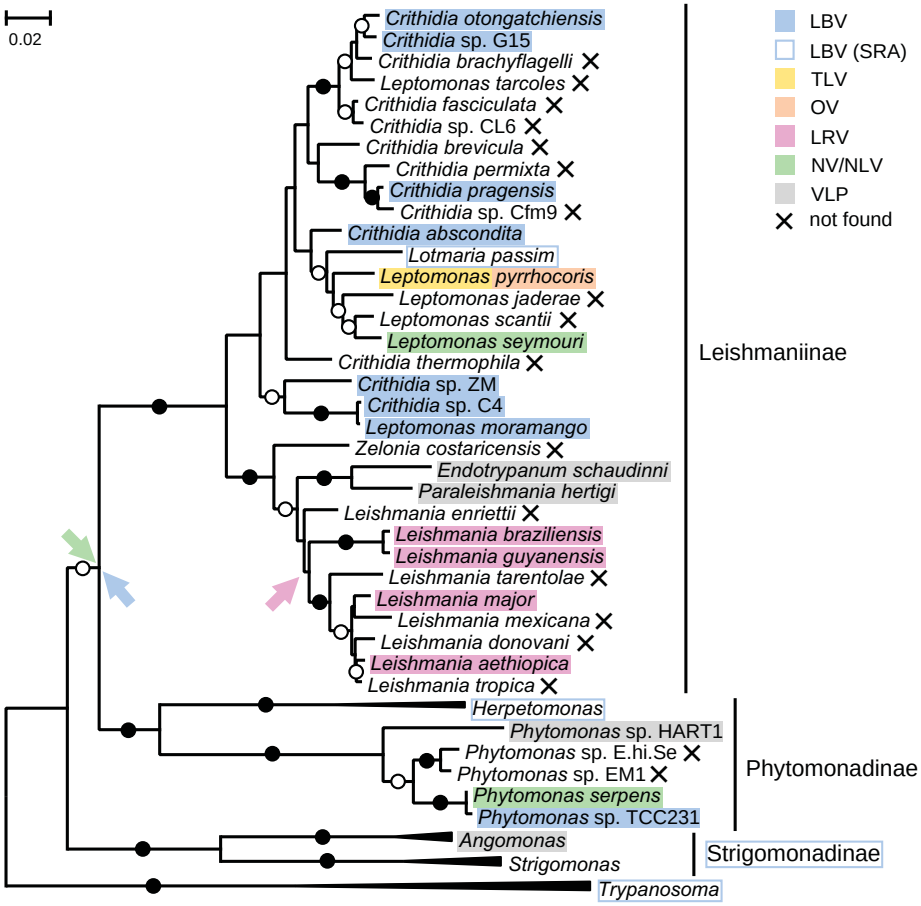

D
0.5
0.98/81
SAOV1
0.99/100
SAOV2
0.99/98
SsOLV1
0.97/79
RsOLV1
BOV
0.99/65
SsOLV2
Ourmia-like viruses
CsVC
EpVC
Ourmiavirus
0.96/71
OuMV
0.93/53
0.99/72
ScNV-20S
0.99/94
ScNV-23S
NarEnv
0.99/80
1/99
PserNV1
LepseyNLV1
1/92
PiRV4
1/99
TSA: Tdalm
+ ?
Narnavirus
0.99/99
CcMV1a
SsMV9
BcMV3
Mitovirus
0.99/92
OMV3a
SsMV3
GA
MS2
Levivirus

Figure 5


0.02
Crithidia otongatchiensis
Crithidia sp. G15
Crithidia brachyflagelli
Leptomonas tarcoles
Crithidia fasciculata
Crithidia sp. CL6
Crithidia brevicula
Crithidia permixta
Crithidia pragensis
Crithidia sp. Cfm9
Crithidia abscondita
Lotmaria passim
Leptomonas pyrrhocoris
Leptomonas jaderae
Leptomonas scantii
Leptomonas seymouri
Crithidia thermophila
Crithidia sp. ZM
Crithidia sp. C4
Leptomonas moramango
Zelonia costaricensis
Endotrypanum schaudinni
Paraleishmania hertigi
Leishmania enriettii
Leishmania braziliensis
Leishmania guyanensis
Leishmania tarentolae
Leishmania major
Leishmania mexicana
Leishmania donovani
Leishmania aethiopica
Leishmania tropica
Herpetomonas
Phytomonas sp. HART1
Phytomonas sp. E.hi.Se
Phytomonas sp. EM1
Phytomonas serpens
Phytomonas sp. TCC231
Angomonas
Strigomonas
Trypanosoma
Leishmaniinae
Phytomonadinae
Strigomonadinae
LBV
LBV (SRA)
TLV
OV
LRV
NV/NLV
VLP
not found

Supporting Information

*Grybchuk et al.,"*Viral discovery and diversity in trypanosomatids with a focus on relatives of the human parasite *Leishmania"*

Supplementary methods

Cultivation of trypanosomatids. Axenic cultures of monoxenous trypanosomatids of the genera *Leptomonas*, *Crithidia*, and *Phytomonas* were obtained from i) Life Science Research Centre, University of Ostrava; ii) Department of Parasitology, Charles University, Prague; iii) Institute of Parasitology, České Budějovice; iv) Fundação Oswaldo Cruz, Rio de Janeiro; and v) Zoological Institute of the Russian Academy of Sciences, St. Petersburg. Trypanosomatids were cultivated in the Brain Heart infusion medium supplemented with 10 μg/ml of hemin, 500 units/ml of penicillin and 0.5 mg/ml of streptomycin as described previously (105) and grown to the late logarithmic stage (10^7 – 10^8 cells/ ml).

Phylogenetic analyses of viruses. RDRP sequences of the new viruses characterized in this work were aligned with the related sequences from GenBank in the online version of MAFFT 7 using E-INS-i method (106). Ambiguous parts of the alignments were removed with the use of TrimAl v. 1.3 (107). Positions with more than 50% gaps were filtered out by Gap Strip/Squeeze V. 2.1.0 (https://www.hiv.lanl.gov/content/sequence/GAPSTREEZE/gap.html). The resulting alignments had the following lengths: 412 aa (the *Narnaviridae* dataset), 273 aa (the *Tombus-/Nodaviridae* dataset), and 1415 aa (the *Bunyaviridae* dataset).

Maximum likelihood phylogenetic inference was performed in IQ-TREE v. 1.4.2 (108) with automatic selection of the best-fit amino acid substitution and site heterogeneity models (4 gamma categories). The best-fitted model parameters defined by Bayesian Information Criterion were LG + F + I + G for *Narnaviridae* and *Bunyaviridae* datasets and Blosum62 + I + G4 *Tombus-/Nodaviridae* dataset. Gaps were treated as missing data. Edge support was estimated with bootstrap test (1,000 "standard" replicates). Bayesian inference was accomplished in MrBayes 3.2.6 (109) with the

analyses ran for 1 million generations (given the observed fast convergence) and trees sampled every 100 generations. The "mixed" amino acid substitution model was used (resulting in 1.0 posterior probability of Blosum62 for all three datasets) with the heterogeneity over sites estimated using G + I model. Amino acid frequencies were fixed to empirical values in *Narnaviridae* and *Bunyaviridae* datasets and estimated from the data matrix in *Tombus-/Nodaviridae* dataset in accordance with the best-fit model defined by IQ-TREE v. 1.4.2 for each dataset. Other parameters were left in their default states.

Phylogeny of trypanosomatid hosts. The core alignments of 18S rRNA and gGAPDH genes were taken from the previous work (76). Ambiguously aligned positions of 18S rRNA gene alignment were removed using Gblocks 0.91b as described previously (110). The concatenated alignments were subjected to maximum likelihood analysis in IQ-TREE (108) with partitioning by gene as well as by codon position for gGAPDH gene. The best partitioned model of nucleotide substitutions (K3Pu + G4, TIM3 + I + G4, TPM2u + G4 for the 1^{st}, 2^{nd} and 3^{rd} codon positions of gGAPDH gene, respectively and TNe + I + G4 for 18S rRNA gene) was selected with the use of ModelFinder (111). The statistical support of branches was estimated using 1,000 replicates of the "standard" bootstrap method.

Genomic and transcriptomic analyses. In order to find trypanosomatid signature sequences in the Sequence Reads Archive (SRA, http://www.ncbi.nlm.nih.gov/sra) and transcriptome shotgun assemblies (TSAs), TBLASTN and BLASTN searches were performed with amino acid (PFR1) and nucleotide (18S rRNA) sequences, respectively. In case of SRA-blast, the retrieved reads were assembled into contigs using CAP3 sequence assembly program with the following parameters: minimal overlap length 20 bp; minimum identity 100% (112). Obtained contigs were extended by successive rounds of BLASTN searches against the original SRA. Final full length contigs as well as hits from TSAs were subjected to blast search (megablast for nucleotide sequences and blastp for translated protein coding sequences) against non-redundant nucleotide collection NCBI database for

identification. Similar approach was used for the search of reads corresponding to nucleocapsid proteins in SRAs, which contained RDRPs closely related to LBVs. Obtained amino acid sequences were aligned with newly identified nucleocapsid proteins of viruses characterized in this work using BLAST pairwise alignment to confirm their identities.

Codon usage analysis in *Leptomonas pyrrhocoris* viruses. Codon frequencies in protein coding regions of the six viral genes were analyzed using the CODONW program (codonw.sourceforge.net).

Negative stain transmission electron microscopy. In brief, gradient-purified virus samples were applied to a carbon-coated copper grid, stained with molybdenum acetate, and examined under Philips 201C transmission electron microscope. The morphologies and sizes of the virus particles were analyzed as described previously (113).

Supplemental Information – Tables.

Table S1. The complete list and properties of the studied trypanosomatid species.

Table S2. Sequences of viral RDRPs with working abbreviations of viral names used in phylogenetic inferences.

Table S3. Accession numbers and/or sequences reported in this work.

Table S4. Codon usage in LeppyrOV1 and LeppyrTLV1 ORFs.

Table S5. Complementary terminal sequences (panhandles) of LBV1s and other Bunyavirales.

Supplemental Information - Figure legends.

Figure S1. LeppyrTLV1 endogenous viral element (EVE1). A, The endogenous viral element is located at the subtelomeric region of the *Leptomonas pyrrhocoris* chromosome (scaffold NW_015438358.1). The EVE1 element is preceded by the reverse transcriptase-coding TATE DNA

transposon indicating the possible mechanism for the LeppyrTLV1 endogenization. B, Comparison of ORF organization of RDRP-coding RNA-T1 of 1 LeppyrTLV1 and its endogenous viral element. C, The overlap region between the ORFs of EVE1 contains a putative slippery sequence (underlined) capable of driving -1 ribosomal frameshift, similarly to LeppyrTLV1. D, Multiple alignments of RDRPs of LeppyrTLV1 and LeppyrTLV1-EVE1 with RNA polymerases of *Picorna- Flavi-* and *Caliciviridae*. Identical residues are indicated in red, similar residues in blue. Each block shows amino acid motifs, typically found in viral RNA polymerases (59).

Figure S2. Leishbunyavirus protein domains. A, Amino acid alignment of the N-terminal endonuclease domain of the L protein of leishbunyaviruses and other bunyaviruses. Functionally important residues are marked with arrowheads. Residues implicated in nucleolytic cleavage in leishbunyaviruses are boxed. B, Amino acid alignment of nucleocapsid proteins of leishbunyaviruses reported in (34) and this work (TrypLBVs) and assembled from metatranscriptomes (MetatranscriptomicLBVs). *Cumuto* and *Gouleako* viruses are shown for comparison. Indels are shown as dashes. Positions with more than 82 % similarity are shaded, 90 % similarity – starred.

Figure S3. Comparison of the phylogenies of leishbunyaviruses and their respective hosts. Thescheme is based on the phylogenetic trees presented in Fig. 3 and 5 with simplifications: 1) nobranch lengths on both trees; 2) only LBVs and LBV-containing trypanosomatids were included;3) all clades with the bootstrap support below 70% were collapsed and reshuffled in order tominimize the number of potential viral transitions between hosts.

Figure S4. Features of genomic RNAs of PserNV1 and LepseyNLV1. A, Comparison of terminal complementary sequences and stem-loop structures with those of narnaviruses from *Saccharomyces cerevisiae*. B, nucleotide alignment of 5' end of PserNV1 and the spliced leader (SL) RNA gene from *Phytomonas serpens*. The boundary with the spliced leader/mini-exon is indicated with the slash

symbol. C, Features of a potential frameshift region between ORF1 and ORF2 of the LepseyNLV1 small segment. The location of a hairpin motif is marked by a bracket, which has a predicted ΔG of -6 kCal (OligoAnalyzer 3.1, (114)). A putative ‘slippery’ pyrimidine rich sequence (CCUCCC) is shown in the box positioned 6 nt upstream.

Table S1. The complete list and properties of the studied trypanosomatid species.

Species	Isolate	Reference to trypanosomatid description	Host	Geographic origin	Presence of viruses
Crithidia abscondita	127AL	Yurchenko et.al., 2009	*Largus* sp.	Ecuador	*Bunyavirales*
Crithidia brachyflagelli	340VL	Jirku et.al., 2012	*Prepos* cf. *accinctus*	Costa Rica	not observed
Crithidia brevicula	101	Kostygov et.al., 2014	*Nabis flavomarginatus*	Russia	not observed
Crithidia brevicula	KV1	Kostygov et.al., 2014	*Gerris lacustris*	Russia	not observed
Crithidia brevicula	F6	Kostygov et.al., 2014	*Nabis flavomarginatus*	Russia	not observed
Crithidia fasciculata	Finn-01.02	Hamilton et.al., 2015	*Drosophila falleni*	USA	not observed
Crithidia fasciculata	COLPROT053	Wallace et. al., 1959	*Phaenicia sericata*	USA	not observed
Crithidia fasciculata	COLPROT048	Wallace, 1943	*Anopheles quadrimaculatus*	France	not observed
Crithidia otongatchiensis	Ecu-08	Yurchenko et.al., 2014	family Syrphidae	Ecuador	*Bunyavirales*
Crithidia permixta	128SI	Yurchenko et.al., 2009	tribe Mirini	Ecuador	not observed
Crithidia pragensis	MCZ-11	Yurchenko et.al., 2014	*Cordilura albipes*	Czech Republic	*Bunyavirales*
Crithidia sp.	G15	Votypka et.al., 2012b	*Rhynocoris rapax*	Ghana	*Bunyavirales*
Crithidia sp.	ZM	Podlipaev, 1987	*Lygocoris lucorum*	Russia	*Bunyavirales*
Crithidia sp.	C4	Podlipaev et.al., 2004	*Limnoporus rufoscutellatus*	Russia	*Bunyavirales*
Crithidia sp.	Cfm9	Merzlyak et.al., 2001	*Nabis flavomarginatus*	Russia	not observed
Crithidia sp.	CL6	Podlipaev et.al., 1999	*Nabis limbatus*	Russia	not observed
Crithidia thermophila	320AR	Jirku et.al., 2012	*Largus maculatus*	Costa Rica	not observed
Crithidia thermophila	COLPROT018	Roitman et.al., 1977	*Cosmoclopius* sp.	Brazil	not observed
Crithidia thermophila	COLPROT054	Roitman et.al., 1977	*Zelus leucogrammus*	Brazil	not observed
Zelonia costaricensis	15EC	Yurchenko et.al., 2006b	*Ricolla simillima*	Costa Rica	not observed
Leptomonas jaderae	34EC	Yurchenko et.al., 2009	*Jadera obscura*	Costa Rica	not observed
Leptomonas moramango	MMO-09	Yurchenko et.al., 2014	*Pachycerina* cf. *vaga*	Madagascar	*Bunyavirales*
Leptomonas pyrrhocoris	F165	Votypka et.al., 2012a	*Pyrrhocoris apterus*	France	LeppyrTLV1
Leptomonas pyrrhocoris	F19	Votypka et.al., 2012a	*Pyrrhocoris apterus*	France	LeppyrTLV1, LeppyrOV1
Leptomonas pyrrhocoris	H10	Votypka et.al., 2012a	*Pyrrhocoris apterus*	Czech Republic	LeppyrTLV1, LeppyrOV1, LeppyrTLV1-EVE
Leptomonas pyrrhocoris	10VL	Votypka et.al., 2012a	*Dysdercus obliquus*	Costa Rica	not observed
Leptomonas pyrrhocoris	121AL	Votypka et.al., 2012a	*Dysdercus obscuratus*	Ecuador	not observed
Leptomonas pyrrhocoris	122AL	Votypka et.al., 2012a	*Dysdercus obscuratus*	Ecuador	not observed
Leptomonas pyrrhocoris	14BT	Votypka et.al., 2012a	*Dysdercus lunulatus*	Costa Rica	not observed
Leptomonas pyrrhocoris	25EC	Votypka et.al., 2012a	*Dysdercus* sp.	Costa Rica	not observed
Leptomonas pyrrhocoris	28EC	Votypka et.al., 2012a	*Dysdercus* sp.	Costa Rica	not observed

Leptomonas pyrrhocoris	324RV	Votypka et.al., 2012a	*Dysdercus obscuratus*	Costa Rica	not observed
Leptomonas pyrrhocoris	329MV	Votypka et.al., 2012a	*Dysdercus mimulis*	Costa Rica	not observed
Leptomonas pyrrhocoris	CH278	Votypka et.al., 2012a	*Dysdercus poecilus*	China	not observed
Leptomonas pyrrhocoris	G58	Votypka et.al., 2012a	*Dysdercus fasciatus*	Ghana	not observed
Leptomonas pyrrhocoris	K06	Votypka et.al., 2012a	*Scantius aegyptius*	Cyprus	not observed
Leptomonas pyrrhocoris	P59	Votypka et.al., 2012a	*Pyrrhocoris marginatus*	Czech Republic	not observed (LeppyrTLV1-EVE)
Leptomonas pyrrhocoris	SERG	Flegontov et.al., 2016	*Dysdercus* sp.	Burkina Faso	not observed
Leptomonas pyrrhocoris	PP1	Frolov et.al., 2014	*Pyrrhocoris apterus*	Russia	not observed (LeppyrTLV1-EVE)
Leptomonas pyrrhocoris	PP2	Frolov et.al., 2014	*Pyrrhocoris apterus*	Russia	not observed (LeppyrTLV1-EVE)
Leptomonas pyrrhocoris	LP	Yurchenko et.al., 2006a	*Pyrrhocoris apterus*	Russia	not observed (LeppyrTLV1-EVE)
Leptomonas scantii	F221	Votypka et.al., 2012a	*Scantius aegyptius*	France	not observed
Leptomonas seymouri	ATCC	Wallace, 1977	*Dysdercus suturellus*	USA	*Narnaviridae*
Leptomonas tarcoles	47VL	Yurchenko et.al., 2008	*Prepops* sp.	Costa Rica	not observed
Phytomonas serpens	9T (CB)	Gibbs, 1957	*Lycopersicon esculentum*	Brazil	not observed
Phytomonas serpens	9T (UCR)	Gibbs, 1957	*Lycopersicon esculentum*	Brazil	*Narnaviridae*
Phytomonas serpens	30T	Gibbs, 1957	*Lycopersicon esculentum*	Brazil	*Narnaviridae*
Phytomonas serpens	1G	Da Silva et.al., 1990	*Phthia picta*	Brazil	not observed
Phytomonas serpens	15T, COLPROT186	Jankevicius, et.al. 1989	*Solanum lycopersicum*	Brazil	not observed
Phytomonas sp.	TCC231	Jankevicius, et.al. 1989	*Lycopersicon esculentum*	Brazil	*Bunyavirales*
Phytomonas sp.	COLPROT079	Kastelein et.al., 1988	*Allamanda cathatica*	Suriname	not observed
Phytomonas sp.	COLPROT080	Conchon et.al., 1989	*Citrus bergamia*	Brazil	not observed

Table S2. Sequences of viral RDRPs with working abbreviations of viral names used in phylogenetic inferences.

Fig. 1 (D) Tombus-like virus from Leptomonas pyrrhocoris

Clade	Short name	Virus name	Acc No	Host
Tombusviridae	JINRV	Japanese iris necrotic ring virus	BAA92792.1	Plants (Iris ensata)
Tombusviridae	TCV	Turnip crinkle virus	AAP78486.1	Plants
Tombusviridae	MNSV	Melon necrotic spot virus	ABC67516.1	Plants
Tombusviridae	CarMV	Carnation mottle virus	CAB38331.1	Plants
Tombusviridae	OMMV	Olive mild mosaic virus	AEC50092.1	Plants (Olea europaea)
Tombusviridae	PoLV	Pothos latent virus	Q9IWA0.1	Plants
Tombusviridae	TBSV	Tomato bushy stunt virus	AAT67237.1	Plants
Tombusviridae	CTLV14	Changjiang tombus-like virus 14	APG76248.1	Crustacea: Procambarus clarkia
Tombus-like viruses	HTLV13	Hubei tombus-like virus 13	APG76577.1	Myriapoda: Scutigeridae
Tombus-like viruses	STLV6	Sanxia tombus-like virus 6	APG76428.1	Insecta: Gerridae sp.
Tombus-like viruses	BTLV9	Beihai tombus-like virus 9	APG76191.1	Mollusca: Octopodidae sp.
Tombus-like viruses	HTLV22	Hubei tombus-like virus 22	APG76327.1	Insecta: Paracercion melanotum (9), Paracercion calamorum (5), Ceriagrion auranticum (10), Brachydiplax chalybea (2), Orthetrum albistylum (1), Pseudothemis zonata (6), Chironomus sp (1)
Tombus-like viruses	HTLV36	Hubei tombus-like virus 36	APG76457.1	Myriapoda: Diplopoda sp. (7), Otostigmus scaber (4), Scolopocryptops sp (3), Otostigmus scaber (1), Myriapoda sp (1)
Tombus-like viruses	CTLV22	Changjiang tombus-like virus 22	APG76278.1	Crustacea: Procambarus clarkia
Tombus-like viruses	BUC1	Brandmavirus UC1	AHA86931.1	unknown
Tombus-like viruses	WTLV17	Wenzhou tombus-like virus 17	APG76615.1	Mollusca: Pomacea canaliculata
Tombus-like viruses	HTLV35	Hubei tombus-like virus 35	APG76480.1	Myriapoda: Diplopoda sp. (7), Otostigmus scaber (4), Scolopocryptops sp (3), Otostigmus scaber (1), Myriapoda sp (1)
Tombus-like viruses	LeppyrTLV1-EVE1	LeppyrTLV1 endogenous virus element		Trypanosomatidae
Tombus-like viruses	LeppyrTLV1	Leptomonas pyrrhocoris tombus-like virus 1		Trypanosomatidae
Tombus-like viruses	BTLV17	Beihai tombus-like virus 17	APG76183.1	Mollusca: Solen strictus
Tombus-like viruses	BTLV18	Beihai tombus-like virus 18	APG76207.1	Mollusca: Octopodidae sp.
Tombus-like viruses	Towan	Towan virus	AOG30801.1	unknown
Tombus-like viruses	JTLV2	Jingmen tombus-like virus 2	APG76305.1	Nematoda: Ascaridia sp.
CBPV-like viruses	LSV1	Lake Sinai virus 1	AEH26193.1	Insecta: Apis mellifera
CBPV-like viruses	AaCV	Anopheline-associated C virus	AGW51750.1	Insecta: Anopheline sp.
CBPV-like viruses	CBPV	Chronic bee paralysis virus	ACO82537.1	Insecta: Apis mellifera
CBPV-like viruses	Dansoman	Dansoman virus	AKH40306.1	Insecta: Drosophila sp.
CBPV-like viruses	HTLV42	Hubei tombus-like virus 42	APG76280.1	Insecta (Diptera): Atherigona orientalis (2), Chrysomya megacephala (6), Lucilia sericata (1), Musca domestica (11), Sarcophaga dux (1), S. peregrine (1), S. sp.(1)
CBPV-like viruses	BTLV19	Beihai tombus-like virus 19	APG76134.1	Crustacea: Amphibalanus rhizophorae
CBPV-like viruses	WeTLV4	Wenling tombus-like virus 4	APG76579.1	Crustacea: Brachyura sp. (2 species, 6), Achelata sp. (3), Penaeoidea sp. (2 species, 3), Ibacus novemdentatus (3), Anomura sp. (3), Penaeidae sp. (2), Charybdis bimaculata (3), Charybdis rufodactylus (3), Latreilliidae sp. (2 species, 3), Ovalipes punctatus (3)
CBPV-like viruses	WCV4	Wenzhou crab virus 4	APG76640.1	Crustacea: Charybdis hellerii (2), Charybdis japonica (9), Charybdis lucifera (1)
CBPV-like viruses	WTLV18	Wenzhou tombus-like virus 18	APG76097.1	Mollusca: Barbatia virescens (12), Sinonovacula constricta (12), Tegillarca granosa (12), Crassostrea ariakensis (12), Mytilus coruscus (12)
CBPV-like viruses	PHVA	Plasmopara halstedii virus A	ADK55578.1	Oomycetes: Plasmopara halstedii
CBPV-like viruses	SmVA	Sclerophthora macrospora virus A	BAC11954.1	Oomycetes: Sclerophthora macrospora
Nodaviridae	FHV	Flock house virus	CAA54399.1	Insecta (Coleoptera): Costelytra zealandica
Nodaviridae	NoV	Nodamura virus	AAF97860.1	Insecta (Diptera): Culex tritaeniorhynchus)
Nodaviridae	PaV	Pariacoto virus	AAF71691	Insecta (Lepidoptera): Spodoptera eridania
Nodaviridae	SJNNV	Striped Jack nervous necrosis virus	BAB64329.1	Chordata: Cocinero sp.

Fig. 3 (D) Leishbunyaviruses

Clade	Short name	Virus name	Acc No	Host
Phlebovirus	CDUV	Chandiru virus	AEA30057.1	Arthropod vectors (Diptera), vertebrate hosts
Phlebovirus	PTV	Punta Toro phlebovirus	ALL45372.1	Arthropod vectors (Diptera), vertebrate hosts
Phlebovirus	BUJV	Bujaru virus	API68880.1	Arthropod vectors (Diptera), vertebrate hosts
Phlebovirus	Adana	Adana virus	AJK91618.1	Arthropod vectors (Diptera), vertebrate hosts
Phlebovirus	SALV	Salehabad phlebovirus	AGA82741.1	Arthropod vectors (Diptera), vertebrate hosts
Phlebovirus	SFNV	Sandfly fever Naples virus	CAA48478.1	Arthropod vectors (Diptera), vertebrate hosts
Phlebovirus	RVFV	Rift Valley fever virus	ABD51507.1	Arthropod vectors (Diptera), vertebrate hosts
	SFTSV	Severe fever with thrombocytopenia virus	ADZ04470.1	Arthropod vectors (Diptera or ticks), vertebrate hosts
	BHAV	Bhanja virus	AFO66272.1	Arthropod vectors (Diptera or ticks), vertebrate hosts
	LSV	Lone Star virus	AGL50921.1	Arthropod vectors (Diptera or ticks), vertebrate hosts
Phlebovirus (uukuniemi group)	EgAN	EgAN 1825-61 virus	AEL29654.1	Arthropod vectors (ticks), vertebrate hosts
Phlebovirus (uukuniemi group)	UUKV	Uukuniemi phlebovirus	BAA01590.1	Arthropod vectors (ticks), vertebrate hosts
Phlebovirus (uukuniemi group)	Khasan	Khasan virus	AII79370.1	Arthropod vectors (ticks), vertebrate hosts
Phasivirus	Badu	Badu phasivirus	AMA19446.1	Insects (mosquitoes)
Phasivirus	WMPV	Wutai mosquito phasivirus	AJG39270.1	Insects (mosquitoes)
Phasivirus	PCLPV	Phasi Charoen-like phasivirus	AKP18602.1	Insects (mosquitoes)
Phasivirus	WFV	Wuhan fly phasivirus	AJG39259.1	Insects (mosquitoes)
Tenuivirus	RaSV	Ramu stunt virus	ALJ83282.1	Arthropod vectors (Hemiptera), plant hosts
Tenuivirus	RGSV	Rice grassy stunt tenuivirus	BAA89602.1	Arthropod vectors (Hemiptera), plant hosts
Tenuivirus	RiSV	Rice stripe tenuivirus	AFM93792.1	Arthropod vectors (Hemiptera), plant hosts
Goukovirus	Cumuto	Cumuto virus	AHH60917.1	Insects (mosquitoes)
Goukovirus	Gouleako	Gouleako virus	AEJ38175.1	Insects (mosquitoes)
Goukovirus	YIV	Yichang Insect virus	AJG39273.1	Insects (mosquitoes)
Leishbunyaviridae	ABV1	Apis bunyavirus 1	ARO50045.1	Honeybee (Apis mellifera) + Trypanosomatidae (Lotmaria passim)
Leishbunyaviridae	CG15LBV1	Crithidia sp. G15 leishbunyavirus 1		Trypanosomatidae
Leishbunyaviridae	LmorLBV1b	Leptomonas moramango leishbunyavirus 1b		Trypanosomatidae
Leishbunyaviridae	CC4LBV1	Crithidia sp. C4 leishbunyavirus 1		Trypanosomatidae
Leishbunyaviridae	CZMLBV1	Crithidia sp. ZM leishbunyavirus 1		Trypanosomatidae
Leishbunyaviridae	CabsLBV1	Crithidia abscondita leishbunyavirus 1		Trypanosomatidae
Leishbunyaviridae	LmorLBV1a	Leptomonas moramango leishbunyavirus 1a		Trypanosomatidae
Leishbunyaviridae	DuBV	Duke bunyavirus	ARE30258.1	Honeybee (Apis mellifera) + ?
Leishbunyaviridae	CpraLBV1	Crithidia pragensis leishbunyavirus 1		Trypanosomatidae
Leishbunyaviridae	HBLV5	Hubei bunya-like virus 5	APG79301.1	Insects (Diptera) + Trypanosomatidae (Strigomonadinae)
Leishbunyaviridae	CotoLBV1	Crithidia otongatchiensis leishbunyavirus 1		Trypanosomatidae
Leishbunyaviridae	HBLV6	Hubei bunya-like virus 6	APG79326.1	Horse leech (Whitmania pigra) + Trypanosomatidae (Trypanosoma sp)
Leishbunyaviridae	PTCCLBV1	Phytomonas sp. TCC231 leishbunyavirus 1		Trypanosomatidae
Leishbunyaviridae	HHFV	Huangshi Humpbacked Fly virus	AJG39239.1	Humpbacked Fly (Megaselia scalaris) + Trypanosomatidae (Strigomonadinae)
Leishbunyaviridae	WSV	Wuhan Spider virus	AJG39269.1	Arthropods (Araneae) + Diptea, Heteroptera + Trypanosomatidae (Herpetomonas sp)
Nairoviridae, Orthonairovirus	DONV	Dugbe orthonairovirus	AMT75392.1	Arthropod vectors (ticks), vertebrate hosts
Nairoviridae, Orthonairovirus	CCHFONV	Crimean-Congo hemorrhagic fever orthonairovirus	ARB51463.1	Arthropod vectors (ticks), vertebrate hosts
Arenaviridae	LCMAV	Lymphocytic choriomeningitis mammarenavirus	AMR60827.1	Vertebrates
Arenaviridae	MaMAV	Machupo mammarenavirus	AMZ00419.1	Vertebrates

Arenaviridae	GGV	Alethinophid 1 reptarenavirus (Golden Gate virus)	AFP93553.1	Vertebrates
Arenaviridae	Boa	Alethinophid 3 reptarenavirus (Boa arenavirus)	AGH06042.1	Vertebrates
Phasmaviridae, Orthophasmavirus	KPOFV	Kigluaik phantom orthophasmavirus	AIA24559.1	Insects
Phasmaviridae, Orthophasmavirus	WCOPV	Wuchang cockroach orthophasmavirus 1	AJG39258.1	Insects
	FERV	Ferak orthoferavirus	AKN56913.1	Insects
	JONV	Jonchet orthojonvirus	AKN56884.1	Insects
Hantaviridae, Orthohantavirus	HOHV	Hantaan orthohantavirus	APH07644.1	Vertebrates
Hantaviridae, Orthohantavirus	KhaOHV	Khabarovsk orthohantavirus	AIL25337.1	Vertebrates
Fimoviridae, Emaravirus	EMARAV	European mountain ash ringspot-associated virus	AAS73287.2	Arthropod vectors (Eriophyidae), Plant hosts
Fimoviridae, Emaravirus	RREV	Rose rosette emaravirus	ADZ54688.1	Arthropod vectors (Eriophyidae), Plant hosts
Tospoviridae, Orthotospovirus	TSWV	Tomato spotted wilt orthotospovirus	AIY28466.1	Arthropod vectors (Thripidae), Plant hosts
Tospoviridae, Orthotospovirus	IYSV	Iris yellow spot orthotospovirus	ACM89280.1	Arthropod vectors (Thripidae), Plant hosts
Peribunyaviridae	HeHV	Herbert herbevirus	AGX32061.1	Insects
Peribunyaviridae	TaHV	Tai herbevirus	AGX32057.1	Insects
Peribunyaviridae	BUNV	Bunyamwera virus	AKX73309.1	Arthropod vectors (ticks), vertebrate hosts
Peribunyaviridae	OROV	Oropouche orthobunyavirus	ALB07205.1	Arthropod vectors (ticks), vertebrate hosts

Fig. 4 (D) Narnaviruses of trypanosomatids.

Clade	**Short name**	**Virus name**	**Acc No**	**Host**
Ourmia-like viruses	SAOV1	Soybean-associated ourmiavirus 1	ALM62238	Plants
Ourmia-like viruses	SAOV2	Soybean-associated ourmiavirus 2	ALM62250	Plants
Ourmia-like viruses	SsOLV1	Sclerotinia sclerotiorum ourmia-like virus 1	ALD89138	Fungi
Ourmia-like viruses	RsOLV1	Rhizoctonia solani ourmia-like virus 1	ALD89131	Fungi
Ourmia-like viruses	BOV	Botrytis ourmiavirus	CEZ26310	Fungi
Ourmia-like viruses	SsOLV2	Sclerotinia sclerotiorum ourmia-like virus 2	ALD89139	Fungi
Ourmiavirus	CsVC	Cassava virus C	ACI03053	Plants
Ourmiavirus	EpCV	Epirus cherry virus	ACF16357	Plants
Ourmiavirus	OuMV	Ourmia melon virus	ACF16360	Plants
Narnavirus	ScNV-20S	Saccharomyces cerevisiae 20S RNA narnavirus	AAC98925	Fungi
Narnavirus	ScNV-23S	Saccharomyces cerevisiae 23S RNA narnavirus	AAC98708	Fungi
Narnavirus	NarEnv	Narnaviridae environmental sample	AJT39596	unknown
Narnavirus	PserNV1	Phytomoas serpens narnavirus 1		Trypanosomatidae
Narnavirus	LepseyNLV1	Leptomonas seymouri narna-like virus 1		Trypanosomatidae
Narnavirus	PiRV4	Phytophthora infestans RNA virus 4	AEM89291	Oomycete
Narnavirus	TSA: Tdalm	Teleopsis dalmanni transcribed RNA	GBBP01132666.1	Teleopsis dalmanni + Trypanosomatidae (Jaenimonas sp.)
Mitovirus	CcMV1a	Cryphonectria cubensis mitovirus 1a	AAR01970	Fungi (mitochondria)
Mitovirus	SsMV9	Sclerotinia sclerotiorum mitovirus 9	AHF48625	Fungi (mitochondria)
Mitovirus	BcMV3	Botrytis cinerea mitovirus 3	CEZ26302	Fungi (mitochondria)
Mitovirus	OMV3a	Ophiostoma mitovirus 3a	CAA06228	Fungi (mitochondria)
Mitovirus	SsMV3	Sclerotinia sclerotiorum mitovirus 3	AGC24232	Fungi (mitochondria)
Levivirus	GA	Enterobacteria phage GA	CAA27499	Bacteria
Levivirus	MS2	Enterobacteria phage MS2	P00585	Bacteria

Table S3. Accession numbers and/or sequences reported in this work.

Species	Isolate	Viral name	Acc #s
Crithidia otongatchiensis	Ecu-08	*Coto*LBV1	KX373292, KX451144, KX451145, KX683300
Crithidia sp.	G15	CG15LBV1	KX373291
Crithidia sp.	ZM	CZMLBV1	KX373293
Crithidia sp.	C4	CC4LBV1	KY322668
Crithidia abscondita	127AL	CabsLBV1	KX507299, KX507300, KX507301
Crithidia pragensis	MCZ-11	CpraLBV1	KY322669
Leptomonas moramango	MMO-09	LepmorLBV1	KX280012-KX280017
Phytomonas sp.	TCC231	PTCCLBV1	KY322667
Leptomonas pyrrhocoris	F165	LeppyrTLV1	KX373302, KX373303
Leptomonas pyrrhocoris	F19	LeppyrTLV1	MG008317
Leptomonas pyrrhocoris	H10	LeppyrTLV1	KX373300, KX373301
Phytomonas serpens	9T (UCR)	PserNV1	KU882057, KY322666
Phytomonas serpens	30T	PserNV1	-
Leptomonas seymouri	ATCC	LseyNLV1	KU935604, KU935605, KX373304, KX373305
Leptomonas pyrrhocoris	F19	LeppyrOV1	-
Leptomonas pyrrhocoris	H10	LeppyrOV1	KX373294-KX373299
Leptomonas pyrrhocoris	P59	LeppyrTLV1-EVE1	KY364836, KY364842
Leptomonas pyrrhocoris	PP1	LeppyrTLV1-EVE1	KY364837, KY364843
Leptomonas pyrrhocoris	PP2	LeppyrTLV1-EVE1	KY364835, KY364841
Leptomonas pyrrhocoris	LP	LeppyrTLV1-EVE1	KY364839

The following represent metatranscriptomic contigs assembled from SRA depositions. As they have not been experimentally verified they cannot be deposited in GenBank

Contig description			Source of the contig				
Name of contig	Sequence	Identity	SRA accession	Name of virus found in the SRA	SRA name	Biomaterial included in the sample	Reference to the SRA
SRX1712638_HBLV6_PFR	GATTGCGAAGCTGGAGAAGATTGAGGATGAGCTGCGCCGCTCTCAGCTAGACGCGACGGAAATGGCACAGACGCCAGTGATTGTGCTCAAGAACCTCGAG	*Trypanosome* paraflagellar rod protein 2	SRX1712638	Hubei bunya-like virus 6 (HBLV6)	Leech mix Hubei	Whitmania pigra (horse leech)	Shi, 2016
SRX833697_WSV_18S	CCATGCGCAGATCAGACGTAATCTGCCGCAAAAATTTTGCGGTTTCCGCAACATTGGATAACTTGGCGAAACGCCAAGCTAATACATGAACCAACCAGGCGTTCTCCGCCACGGGCGTGCGGGCAACCGTACGTCTAGTGAGACGCCTTGCGAATGAATGACATTAAAACCAATGCCTTCACTGGCAGTAACACCCAGAAGTGTTGACTCAATTCATTCCGTGCGAAAGCCGGATTTCCGGCGTCTTTTTGACGAACAACTGCCCTATCAGCTAGTGATGGCAGTGTAGTGGACTGCCATGGCGTTGACGGGAGCGGGGGATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAATAGCTACCACTTCTACGGAGGGC	*Herpetomonas* sp. 18S ribosomal RNA	SRX833697	Wuhan spader virus (WSV)	Spiders	Neoscona, Parasteatoda, Plexippus, Pirata, Araneae spp.	Li, 2015
SRX1711976_HBLV5_18S	ATCAGCTCGTGATGGCCGTGTAGTGGACTGCCATGGCGTTGACGGGAGCGGGGGATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAATAGCTACCACTTCTACGGAGGGCAGCAGGCGCGCAAATTGCCCAATGTCAAGAAAAAACGATGAGGCAGCGAAAAGAAAT	Strigomonadinae 18S ribosomal RNA	SRX1711976	Hubei bunya-like virus 5 (HBLV5)	Diptera mix Hubei	Drosophila, Episyrphus, Sarcophaga, Muscina, Ptecticus	Shi, 2016
SRX833692_HHFV_18S	CCGGCGTCTTTTGACGAACAACTGCCCTATCAGCTAGTGATGGCCGTGTAGTGGACTGCCATGGCGTTGACGGGAGCGGGGGATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAATAGCTACCACTTCTACGGAGGGCAGCAGGCGCGCAAATTGCCCAATGTCAAGAAAAAACGATGAGGCAGCGAAAAGAAATAGGTTTG	Strigomonadinae 18S ribosomal RNA	SRX833692	Huangshi Humpbacked Fly virus (HHFV)	Insects mix 4 (insect in the mountain)	Psychoda, Velarifictorus, Crocothemis, Phoridae spp., Lampyridae spp., Aphelinus, Hyalopterus, Aulacorthum	Li, 2015

SRX2422212_ABV1_colonyA_18S	TGACAGTAAAACCAATGCCTTCACTGGCAGTAACACCCAGACGTGTTGACTCAATTCATTCCGTGCGAAAGCCGGCTTGTTCCGGCGTCTTTTGACGAACAACTGCCCTATCAGCTGGTGATGGCCGTGTAGTGGACTGCCATGGCGTTGACGGGAGCGGGGGATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAATAGCTACCACTTCTACGGAGGGCAGCAGGCGCGCAAATTGCCCAATGTCAAAACAAAACGATGAGGCAGCGAAAAGAAATAGAGTTGTCAGTCCATTTGGATTGTCATTTCAATGAGGGATATTTAAACCCATCGAAAATCTAGTAACAATTGGAGGACAAGTCTGGTGCCAGCACCCGCGGTAATTCCAGCTCCAAAAGCGTATAT	Leishmaniinae 18S ribosomal RNA	SRX2422225	Apis bunyavirus 1	RNA-seq of Apis mellifera: South Africa colony 11	Apis mellifera	Remnant, 2017 in press
SRX833692_HHFV_nuc	GGAGGACGATCTGCAGATTCAATATGTCTGCCTTAGAGGTTGAAGAGATTTGTTTTAACCTAGCCTATGTGGCTGTGACGCCTTGGGACACACGTGTGGCCTTAGAGAGAGATGGAAGGTCCGAGGTTGCGAAGTTAATAGCCATCGTTGTGGGCTGCAGGGGTACCAACCTCGCCAAGGTCGTTAAAAGGAGTAATAATCCTGAAAAGGCCCTCGAATACGTGAATAGCATCCGCTCTTTCATGGCCAACCAGACTGTCTCCTTGGGNNCTGTCTCCTTGGGACATATTGCCGCCGCTTACCCCGAGGTAGTCTACGATGCAAGAATTGCGTCCAAGAGTGCTATCTCTGTCAACACCCTCCAGTTCCTGAAATTCGCAGGTCTGACCAAGACGGTGTGGCTGAAGGCTAATCACGACTTCTTCGAACTCACAGGCCAAAGTTCAGAGAAGTTTGATGAGCTGTCCGATGTGATATGGGCAGACGATAGCTTCGCTGGCGTGGTAGGTCACAGG	LBV nucleocapsid	SRX833692	Huangshi Humpbacked Fly virus (HHFV)	Insects mix 4 (insect in the mountain)	Psychoda, Velarifictorus, Crocothemis, Phoridae spp., Lampyridae spp., Aphelinus, Hyalopterus, Aulacorthum	Li, 2015
SRX833697_WSV_nuc	GGGGGATTGGAGAATCAGGCAAAACTGGTTGCGATGATCGTCGGGTGCAGGGGAACCAATCTTGATAAGATCCAAGGTTCTTCGAGCAATCCGGAGATGGCACGTCGCCTTGTTGGCATTGCCCGTAGCCTGTCAGCAGACCTGGGCGTCTCAGTGGCTCACATTGCTTCTGCTTATCCAGAGGTGCTTTATGATGCCAGGAAGAGGTGCGGCAAGTCAGATAAGTGTGAACACTATTTCTTCCTTGAAGGTGACGGCCTCACCAAGGATGCTTGGCTGAGGGCAAACAGGGAGTTCTGCCAGCTGGTGGGGCTTGACTACCAGAAGTTTCAACGCATTTCTGAAATGATCTGGTCTGATAGCAGCTCATCGCCC	LBV nucleocapsid	SRX833697	Wuhan spader virus (WSV)	Spiders	Neoscona, Parasteatoda, Plexippus, Pirata, Araneae spp.	Li, 2015
SRX1711976_HBLV5_nuc_contig6	ATGTCTGAGCCAGCTGGAGATAGCCACATGATGCTTGAGCTTGACGATATAGTTGAAAGCCTTAAGTATCAGGGGTTGAGTCCATGGGAAACGAGAGAGAGGATGTGCAGAGAAGGTCATGAGAAGATGGGGAGGCTTATTGGCTATGTAGTAGGATTGAGGGGAACGAACCTTGCTAAAATGGCTTATAAGAGTGTGGATCGGGAGAAGGCTGCAGTGATCGTGGAGAGAATCAATAGGTATATGGTAACTCAGCCTGTATCTATTGGACATATCGCTGCAGCTTTTCCTGATTGTGTTTACGATGCAAGACTTAGAACTGAAAACAGAAGATTTGTGAATACGCTACAGTTCTTAAAGCACAATGGTCTGCGGAAAGAGGTTTGGTTTCAAGCTAATTTAGATTTCTTACAAACCACGAGGCAAGATGCTAGCAAGTTTGTGGATCTGGCAGATATTATTTGGAGAGATGATAGTTTTGCTGGGATTGTGGGAGATAGGATGCCTTGTTAGTCAGTTGCTTAGAAACCTATGGAATTCAAGTTATGGCGC	LBV nucleocapsid	SRX1711973	Hubei bunya-like virus 5 (HBLV5)	Diptera mix Hubei	Drosophila, Episyrphus, Sarcophaga, Muscina, Ptecticus	Shi, 2016
SRX1711976_HBLV5_nuc_contig15	CAATTTTTGATTTAAACGGCAGGTTGGCACTATCCCTCACCGGCCAAATATTAAGCTGGACCAGCTGGGAGCCTATCACCAACAATTTTGGAGAAAGAATCATCATCCCAAATTGCATCTGAAACGGCAAGAAACCGGTCCGGACTTAGATGCACTAAACGACAAAAATCCAAGTTCGCCTGTAGCCACACGTCTTTGGTAAGTCCATTAAATTTCAAGAACTGAAGTGTGTTCACTGAGATTAAAGAACTGCATTCAACCCGCACATCGTACATCACCTCGGGGAAAGCTGAAGCAATGTGTGCTGCAGAGACTCCATAAAGAGCCACCATTTCTCTTACAACAGCAATGAGCTCTGCAGCCTTCTGTTTGTCCTTCGAGCGACCAGCAATCTTATCTAGCCTTGTTCCTCGAGCTGCGACCAAGATAGCAACAATTTTTGGCTTTCTCCAAATGTCCTGCTTGGATAATCTTCATCCTAGTGGACCATGGTGATACCCCCCTCGTATTGGATTGCATCAAGTAGCGCAATCTCTTTCGTGTCTAGGGAGGAGCTCATTTTTCTCTCTATAGGGAAAATAGTGCTTTCGTCTGG	LBV nucleocapsid	SRX1711974	Hubei bunya-like virus 5 (HBLV5)	Diptera mix Hubei	Drosophila, Episyrphus, Sarcophaga, Muscina, Ptecticus	Shi, 2016
SRX1711976_HBLV5_nuc_contig18	GCCTTTCTTACTCTTCCTTCGTCGATGCAGTCGGTGGCCTCGAGGCTCTGGCTGAAGCTCTCAGCTATAGGGGTAGAAGTCCCTACCAGACTCGAGAGCTGCTGATATCCGAAGGACTAGAAGGACCGGCCAGAATTCTTTCTTTCACCGTAGGTTGTCGGGGCACCAATCTTATGAAGCTCCGAGAAAAGAGTGAGAAGTCAGAAGTACTCGATAGAGTGCTTGCTGCTTTGTCCTTGGTGACCAGTGCTCACAAAGAGTCCTTAGGTCATCTGGCTGCTGCTTTCCCTGAGAACTTGTATGATGGTAGAGTGGTGAGCAATTCGCCCTTCTCCGTCAACACTATCCAGGCTCTGAAGTTCAATGGCTTGGAGAAGAGTGACTGGGTAGCAGCCAACATCAAAATGCTGGAGTCAGTGAAGGGTGATGTTGCGAAGTTCACAGCCATCGCTGATATTGTTTGGGGAGACGACAAGTTTTCAGGAGTGATGGGAGATAGGAAGCCCTCCCCTGTGTAATTAGCGACTGAATCTACCAATTAGAGGGAGTGTTAGTTGTAGATGTAGT	LBV nucleocapsid	SRX1711975	Hubei bunya-like virus 5 (HBLV5)	Diptera mix Hubei	Drosophila, Episyrphus, Sarcophaga, Muscina, Ptecticus	Shi, 2016
SRX1711976_HBLV5_nuc_contig8	CTCCAACCACTCCTGCGAAGCGGTCGTCCTCCCAGATCAAGTCCGCAATGGACTCAAATCTGGGGATCTCCCCCGCCCTCTTCACGAGGGTCAACATCTCACGATTGGCTTTGAGCCAATCATCCTTCTCAAGCCCAGCGTGCTTGAGGAAGGCGAGCGTGTTGACACTAACGGGGGAGCCGCCCGCAACTCTCCCATCATAGACCTCCTCCGGGAAGGCCGCCGCGAGGTGCCCTAAAGACACCCCGTACGTCGCCACAAGAGCCCTGGCATTAGCTGCCAGGGCCCGGGCGACCTCAGGACCTTTACTCCGG	LBV nucleocapsid	SRX1711976	Hubei bunya-like virus 5 (HBLV5)	Diptera mix Hubei	Drosophila, Episyrphus, Sarcophaga, Muscina, Ptecticus	Shi, 2016

SRX1712638_HBLV6_nuc	CCCCCACCTAGGATGAATTTGAGCACTATTTCCTCCGCTGACGTTACAGCATTCTCAAATGAAGAGATGACTATTGCCCTAAATCAAATCACTACAGATATCCATTATGAAGGTAAGTCTCCTGAGCAAACTAGGATAGACATAGAGAATGCTCACTTAGGTAATGCTGCTAGAATCATAGCTTACACTGTGGGGACTGCTGGAACCAATTGGGAAAGCAATGCACATAAGTTCTCTAATAGTCAAACATTAGTGAACTTCATACGACTTCATACTGCAATATGTACTAGTTTGCGTAGACAAGTCTCATTGGCACACATAGCAGCAGCATACCCAGCAGAGTTGTATGATGGTAGAAAGGCATCTCGCAGTCAAATCTCTGTTGATACGATCCCTAGTCTGAAGCACAAAGGGCTAACAAAAGAACGGTGGCTAGAAGCAACTGAGAATTTTATCCGATTAACTTCAAGCACCCCTGCCACAGCATTATTCAATTTTTACAAAATAGCTGATATGATCTGGAATAATGCACCTGACATAATGGGTGATAGGGATCCTGAAAATCCA	LBV nucleocapsid	SRX1712638	Hubei bunya-like virus 6 (HBLV6)	Leech mix Hubei	Whitmania pigra (horse leech)	Shi, 2016

The table includes viral sequences generated in current study as well as partial sequences of trypanosomarid-specific and LBV-like nucleocapsid protein genes assembled from viral metatranscriptomes.

Table S4. Codon usage in LeppyrOV1 and LeppyrTLV1 ORFs.

LeppyrOV1	**T3**	**C3**	**A3**	**G3**	**GC**
ORF1	0.40	0.30	0.18	0.31	0.51
ORF2	0.35	0.32	0.19	0.33	0.53
ORF3	0.39	0.28	0.20	0.34	0.52
ORF4	0.33	0.34	0.17	0.34	0.55
ORF5	0.39	0.27	0.15	0.33	0.57
ORF6	0.45	0.27	0.16	0.32	0.48
average	0.38	0.30	0.18	0.33	0.53
LeppyrTLV1	**T3**	**C3**	**A3**	**G3**	**GC**
ORF1	0.34	0.26	0.33	0.32	0.45
ORF2	0.41	0.29	0.28	0.23	0.47
ORF3	0.34	0.24	0.20	0.22	0.47
average	0.36	0.26	0.27	0.26	0.46
Leppyr GENOME	**T3**	**C3**	**A3**	**G3**	**GC**
	0.14	0.38	0.1	0.39	0.77

A3, C3, G3, T3 are frequencies of the corresponding nucleotides in the 3rd position; GC is the overall GC content

Table S5. Complementary terminal sequences (panhandles) of LBV1s and other *Bunyavirales*.

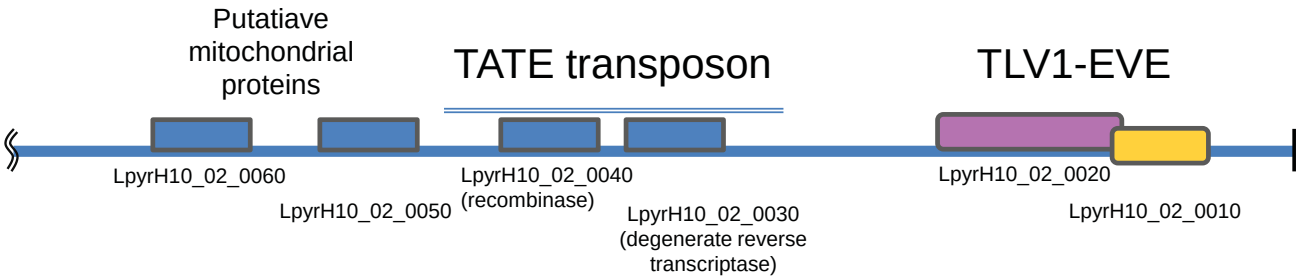
LBV1s		
L segment	5'	3'
CG15LBV1	ACACAAAGAGAA	TGTTCTTTGTGT
LepmorLBV1b	ACACAAAGACAA	TATTCTTTGTGT
LepmorLBV1a	ACACAAAGATAA	ND
CZMLBV1	ACACAAAGAGAA	TGTTCTTTGTGT
CotoLBV1	ACACAAAGACGA	TATTCTTTGTGT
PTCCLBV1	ACACAAAGAAGA	TATTCTTTGTGT
consensus	ACACAAAGANAA	TRTTCTTTGTGT
M segment	5'	3'
CabsLBV1-M	ACACAAAGAGAA	TGTTCTTTGTGT
CotoLBV1-M	ACTCAAAGACGA	ND
LepmorLBV1a-M	ACACAAAGATAA	TATTCTTTGTG-
LepmorLBV1b-M	ACACAAAGACAA	ND
consensus	ACWCAAAGABRA	TRTTCTTTGTGT
S segment	5'	3'
CabsLBV1-S	ACACACGGAAAA	TATTCTTTGTGT
CotoLBV1-S	CACAACGACGA	ND
LepmorLBV1a-S	ACACATAGACAA	ND
LepmorLBV1b-S	ACACACAGATAA	ND
consensus	ACACAHVGAHRA	TRTTCTTTGTGT
Phenuiviridae		
Gouleako virus	5'	3'
L segment	ACACAAAGACAC	TGGACTTTGTGT
M segment	ACACAGTGACCC	GGGACTTTGTGT
S segment	ACACAGTGACCT	GGGACTTTGTGT
Rift Valley fever virus	5'	3'
L segment	ACACAAAGGCGC	CGGTCTTTGTGT
M segment	ACACAAAGACGG	CGGTCTTTGTGT
S segment	ACACAAAGACCC	GGAGCTTTGTGT
Rice grassy stunt virus	5'	3'
L segment	ACACAAAGTCCT	CAGACTTTGTGT
M segment	ACACAAAGTCCT	CAGACTTTGTGT
S segment	ACACAAAGTCCT	CAGACTTTGTGT

Other *Bunyavirales*		
Hantaan Orthohantavirus	5'	3'
L segment	TAGTAGTAGACT	AGCATACTACTA
M segment	TAGTAGTAGACT	AGCATACTACTA
S segment	TAGTAGTAGACT	AGCATACTACTA
Crimean-Congo hemorrhagic fever Orthonairovirus		
	5'	3'
L segment	TCTCAAAGATAT	ATTTCTTTGAGA
M segment	TCTCAAAGAAAT	ATTTCTTTGAGA
S segment	TCTCAAAGAAAC	ATTTCTTTGAGA

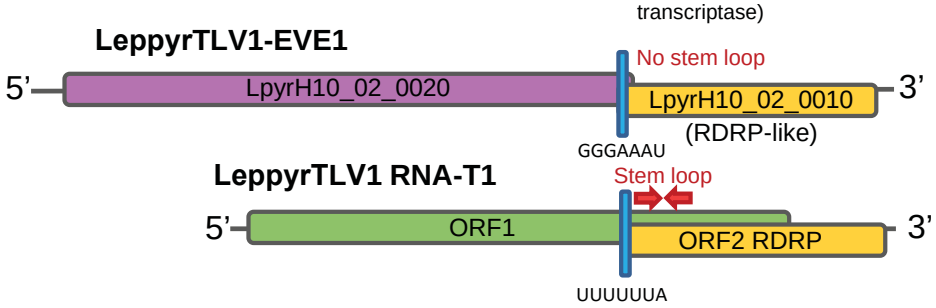
Alignments of 5’ and 3’ terminal sequences L, M and S segments of LBV1s and selected other *Bunyavirales* are shown. Nucleotides involved in bulge formation are underlined. ND – not determined.

Figure S1

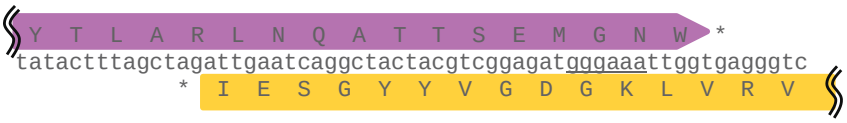
A



B



C



D

MOTIFS		F_1		F_2		A	
Picorna							
1RDR_A_Poliovirus	(155) VTYVKDELRS	(164)	(169) EQGKSRLIEASSLND	(183)	(228) --KLFAFDYTGYDASLSP	(243)	
P03313.4_coxsackie_virus_B3	(156) VTYVKDELRS	(165)	(170) AKGKSRLIEASSLND	(184)	(230) --HLIAFDYSGYDASLSP	(244)	
AFK65743.1_Rhinovirus_C	(139) ITYLKDELRS	(148)	(153) KVGKTRAIEAASLND	(167)	(213) --NLLVFDYTNYDGSLNP	(227)	
P03305.1_foot-and-mouth_diseas	(160) QTFLKDEIRP	(169)	(174) RAGKTRIVDVLPVEH	(188)	(233) YRNVWDVDYSAFDANHCS	(250)	
Flavi							
P06935.2_west_nile_virus	(455) NMMGKREKKP	(464)	(469) KAKGSRAIWFMWLGA	(483)	(529) GGKVYADDTAGWDTRITK	(546)	
P17763.2_dengue_virus	(452) NMMGKREKKL	(461)	(466) KAKGSRAIWYMWLGA	(480)	(526) GGNMYADDTAGWDTRITE	(543)	
P19711.2_bovine_viral_diarrhea	(259) TAIPKNEKRD	(268)	(280) VEKRPRVIQYPEAKT	(294)	(338) EPVAVSFDTKAWDTQVTS	(355)	
Calici							
Q83883.1_Norwalk_virus	(143) TAALKDELVK	(152)	(158) QKVKKRLLWGADLGT	(172)	(216) YKNHFDADYTAWDSTQNR	(233)	
P27410.1_Rabbit_hemorrhagic_di	(142) ACGLKDELRP	(151)	(156) KEGKKRLLWGCDVGV	(170)	(216) ASDFLCLDYSKWDSTMSP	(233)	
Q6XDK8.1_Sapporo_virus	(315) KLALKDELRP	(324)	(329) AQGKRRLLWGCDAGA	(343)	(387) GGVLYCLDYSKWDSTQHP	(404)	
LeppyrTLV1-EVE1	(158) -SFIKDEGYD	(166)	(167) EYKMPRAINSYSDST	(181)	(221) DSKVLYTDFSHFESHHRG	(238)	
LeppyrTLV1	(157) DSFIKDESYD	(166)	(167) ELKAPRSIMSYPDDV	(181)	(221) GGKVGSTDFSSYECHHSG	(238)	

MOTIFS		B		C		D	
Picorna							
1RDR_A_Poliovirus	(286) MPSGCSGTSIFNSMINNLII	(305)	(322) KMIAYGDDVIASYP	(335)	(348) KDYGLTM---T	(355)	
P03313.4_coxsackie_virus_B3	(288) MPSGCSGTSIFNSMINNIII	(307)	(324) RMIAYGDDVIASYP	(337)	(350) KGYGLIM---T	(357)	
AFK65743.1_Rhinovirus_C	(268) MPSGISGTSIFNTIINNIII	(287)	(304) KIVAYGDDVIASYP	(317)	(330) VKYGLTI---T	(337)	
P03305.1_foot-and-mouth_diseas	(296) MPSGCSATSIINTILNNIYV	(315)	(332) TMISYGDDIVVASD	(345)	(358) KSLGQTI---T	(365)	
Flavi							
P06935.2_west_nile_virus	(602) RGSGQVVTYALNTFTNLAVQ	(621)	(662) RMAVSGDDCVVKP-	(674)	(687) NAMSKVRKDIQ	(697)	
P17763.2_dengue_virus	(598) RGSGQVGTYGLNTFTNMEAQ	(617)	(656) RMAISGDDCVVKP-	(668)	(681) NDMGKVRKDIP	(691)	
P19711.2_bovine_viral_diarrhea	(403) RGSGQPDTSAGNSMLNVLTM	(422)	(442) RIHVCGDDGFLITE	(455)	(471) HEAGKPQK-IT	(480)	
Calici							
Q83883.1_Norwalk_virus	(279) LPSGFPCTSQVNSINHWIIT	(298)	(318) YFSFYGDDEIVSTD	(331)	(344) KEYGLKP---T	(351)	
P27410.1_Rabbit_hemorrhagic_di	(279) LPSGMPFTSVINSICHWLLW	(298)	(321) PFYTYGDDGVYAMT	(334)	(349) RDYGLSP---T	(356)	
Q6XDK8.1_Sapporo_virus	(450) LPSGMPFTSVINSLNHMTYF	(469)	(493) TVHTYGDDCLYSVC	(506)	(521) TSFGLKP---T	(528)	
LeppyrTLV1-EVE1	(284) LMSGALWTSFQNSFLNFFVM	(303)	(323) KMFIEGDDGIMKHF	(336)	(344) KRLGLCLK-IN	(353)	
LeppyrTLV1	(288) LMSGALWTSFQNSFLNLMMI	(307)	(325) DTFIEGDDGIFRAF	(338)	(346) ARLGIKLK-IS	(355)	

Figure S2

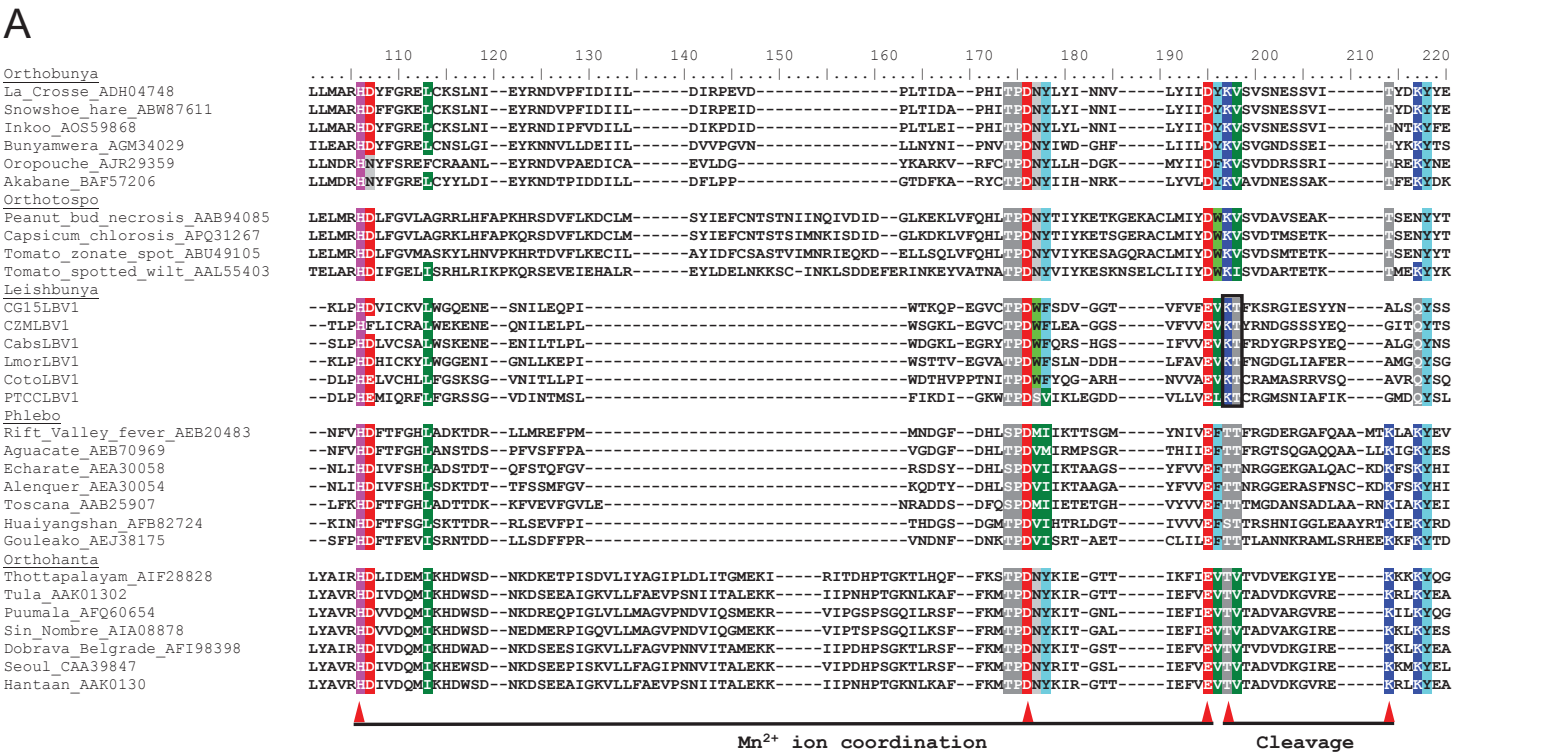

A
Orthobunya
La_Crosse_ADH04748
Snowshoe_hare_ABW87611
Inkoo_AOS59868
Bunyamwera_AGM34029
Oropouche_AJR29359
Akabane_BAF57206
Orthotospo
Peanut_bud_necrosis_AAB94085
Capsicum_chlorosis_APQ31267
Tomato_zonate_spot_ABU49105
Tomato_spotted_wilt_AAL55403
Leishbunya
CG15LBV1
CZMLBV1
CabsLBV1
LmorLBV1
CotoLBV1
PTCCLBV1
Phlebo
Rift_Valley_fever_AEB20483
Aguacate_AEB70969
Echarate_AEA30058
Alenquer_AEA30054
Toscana_AAB25907
Huaiyangshan_AFB82724
Gouleako_AEJ38175
Orthohanta
Thottapalayam_AIF28828
Tula_AAK01302
Puumala_AFQ60654
Sin_Nombre_AIA08878
Dobrava_Belgrade_AFI98398
Seoul_CAA39847
Hantaan_AAK0130
Mn2+ ion coordination
Cleavage
B
TrypLBVs
CabsLBV1
LepmorLBV1A
LepmorLBV1B
CotLBV1
MetatranscriptomicLBVs
WSV
HHFV
HubeiBLV5_15
HubeiBLV5_6
HubeiBLV5_11_1_14
HubeiBLV5_18
HubeiBLV6
Goukovirus(outgroup)
Cumuto_virus
Gouleako_virus

Figure S3

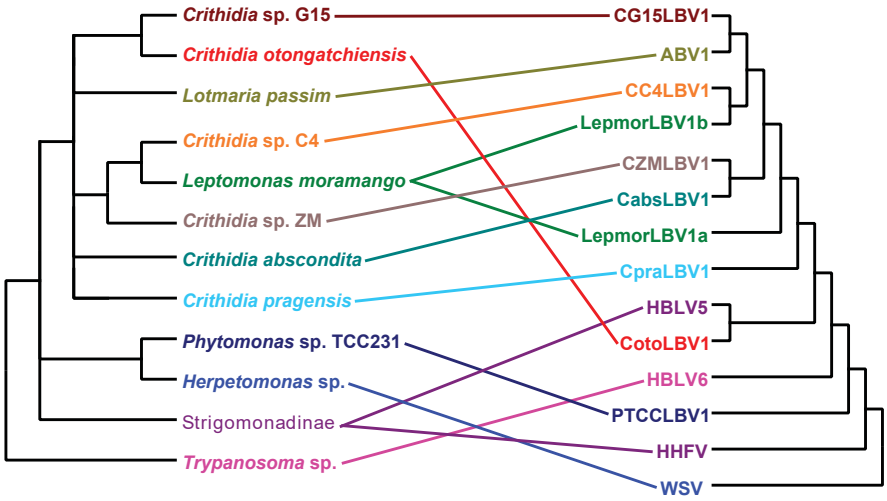
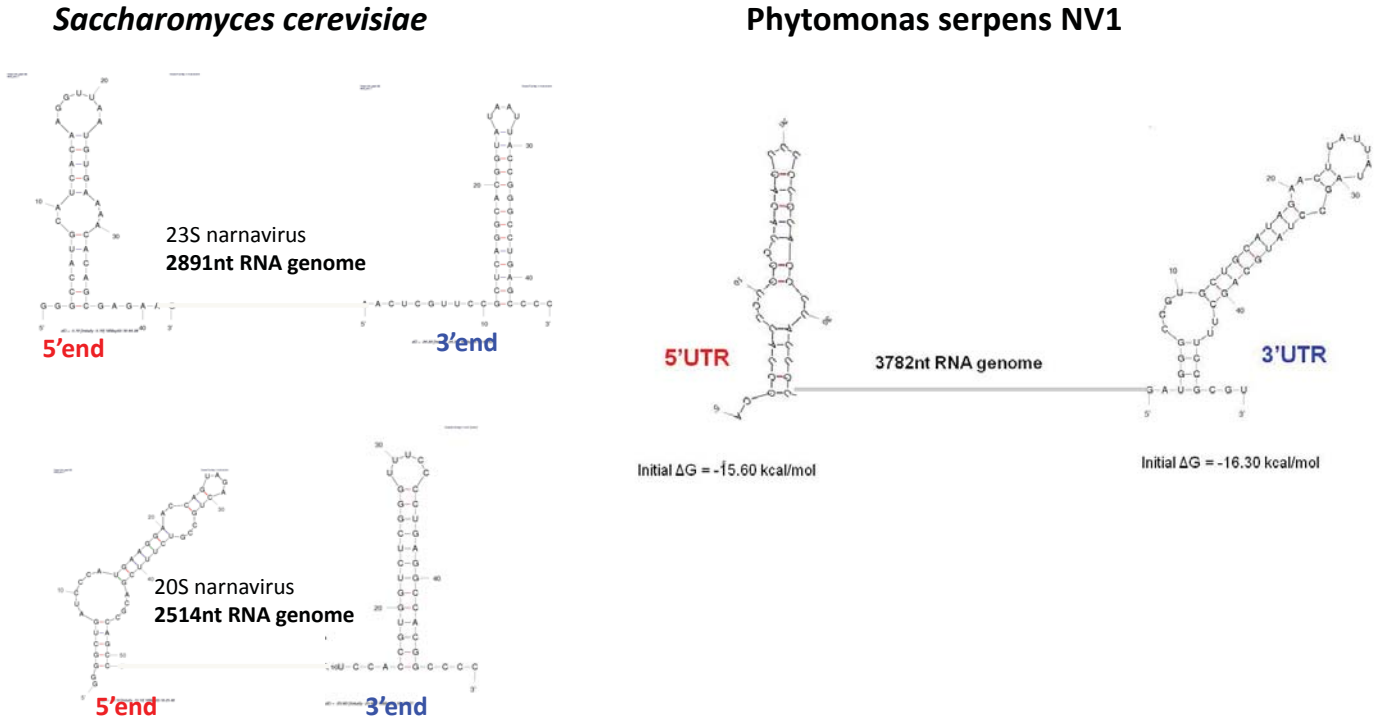

Crithidia sp. G15
Crithidia otongatchiensis
Lotmaria passim
Crithidia sp. C4
Leptomonas moramango
Crithidia sp. ZM
Crithidia abscondita
Crithidia pragensis
Phytomonas sp. TCC231
Herpetomonas sp.
Strigomonadinae
Trypanosoma sp.
CG15LBV1
ABV1
CC4LBV1
LepmorLBV1b
CZMLBV1
CabsLBV1
LepmorLBV1a
CpraLBV1
HBLV5
CotoLBV1
HBLV6
PTCCLBV1
HHFV
WSV

Figure S4

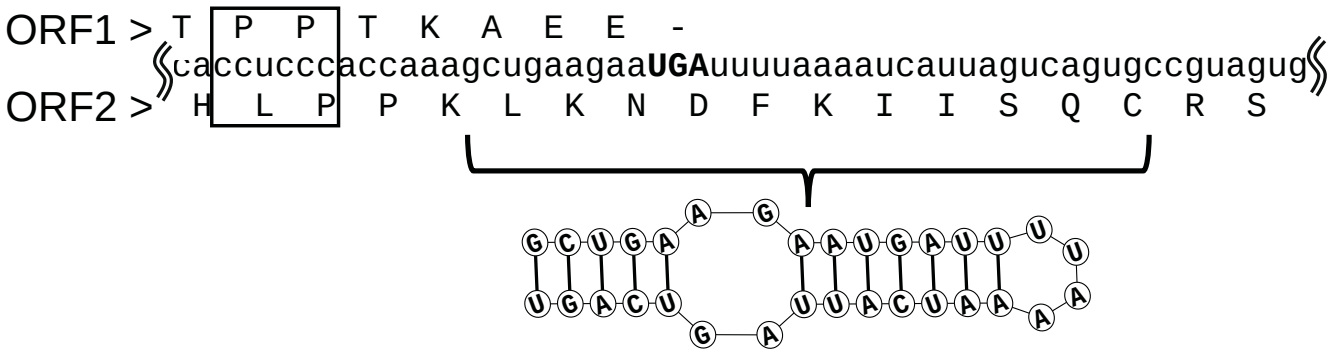
A



B

```
SL  AACTAACGCTATTCTAGATACAGTTTCTGTACTTTATTG/GTATGAGAAACTTCCAGAA
          |||||||||| | ||||||||||||||| ||||||   |  |
PserNV1 ACGCTATTCTGGCTACAGTTTCTGTACCTTATTG/TAAGAACTTCAAAGAGCTC
```

C


ORF1 > T P P T K A E E -
caccucccaccaaagcugaagaaUGAuuuuaaaaucauuagucagugccguagug
ORF2 > H L P P K L K N D F K I I S Q C R S