

# SCIENTIFIC REPORTS



OPEN

## Kanyawara Virus: A Novel Rhabdovirus Infecting Newly Discovered Nycteribiid Bat Flies Infesting Previously Unknown Pteropodid Bats in Uganda

Tony L. Goldberg<sup>1,2,3</sup>, Andrew J. Bennett<sup>1</sup>, Robert Kityo<sup>3</sup>, Jens H. Kuhn<sup>4</sup> & Colin A. Chapman<sup>3,5</sup>

Bats are natural reservoir hosts of highly virulent pathogens such as Marburg virus, Nipah virus, and SARS coronavirus. However, little is known about the role of bat ectoparasites in transmitting and maintaining such viruses. The intricate relationship between bats and their ectoparasites suggests that ectoparasites might serve as viral vectors, but evidence to date is scant. Bat flies, in particular, are highly specialized obligate hematophagous ectoparasites that incidentally bite humans. Using next-generation sequencing, we discovered a novel ledantevirus (mononegaviral family *Rhabdoviridae*, genus *Ledantivirus*) in nycteribiid bat flies infesting pteropodid bats in western Uganda. Mitochondrial DNA analyses revealed that both the bat flies and their bat hosts belong to putative new species. The coding-complete genome of the new virus, named Kanyawara virus (KYAV), is only distantly related to that of its closest known relative, Mount Elgon bat virus, and was found at high titers in bat flies but not in blood or on mucosal surfaces of host bats. Viral genome analysis indicates unusually low CpG dinucleotide depletion in KYAV compared to other ledanteviruses and rhabdovirus groups, with KYAV displaying values similar to rhabdoviruses of arthropods. Our findings highlight the possibility of a yet-to-be-discovered diversity of potentially pathogenic viruses in bat ectoparasites.

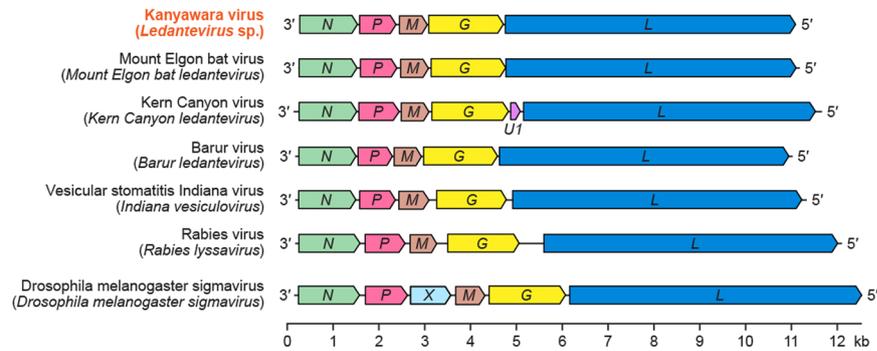
Bats (order Chiroptera) represent the second largest order of mammals after rodents (order Rodentia). Classically, bats are divided into two suborders: megabats (Megachiroptera) and microbats (Microchiroptera). Megabats, also referred to as fruit bats, are assigned to a single family, Pteropidae, whereas microbats are taxonomically more diverse<sup>1</sup>. Both megabats and microbats host numerous, taxonomically diverse viruses. Examples of megabat-borne viruses that are highly virulent for humans are Marburg virus and Nipah virus. Severe acute respiratory syndrome coronavirus is an infamous example of a microbat-transmitted human pathogen<sup>1</sup>. Consequently, characterization of bats, their viromes, and cross-species transmission of bat-borne viruses have become research priorities.

Much less effort has thus far been invested in understanding the role of bat ectoparasites in maintaining viruses in bat populations or potentially transmitting them to humans or mammals of other species. The high degree of specialization and diversity of certain bat ectoparasites suggests that they could, in fact, be reservoirs for certain viruses, maintaining them in their bat hosts. Alternatively, bats could be refractory to infection with ectoparasite viruses, but nevertheless these viruses could be infectious or even pathogenic for other mammals, including humans, and be transmitted through incidental bat ectoparasite bites.

Bat flies are eyeless, wingless, hematophagous dipteran insects (Brachycera: Muscomorpha: Hippoboscoidea) that are obligate bat ectoparasites with off-host breeding life stages. They are assigned to two families, the

<sup>1</sup>Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, Wisconsin, 53706, USA.

<sup>2</sup>Global Health Institute, University of Wisconsin-Madison, Madison, Wisconsin, 53706, USA. <sup>3</sup>Department of Zoology, Makerere University, Kampala, Uganda. <sup>4</sup>Integrated Research Facility at Fort Detrick, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Frederick, Maryland, 21702, USA. <sup>5</sup>Department of Anthropology and School of Environment, McGill University, Montreal, Quebec, H3A 2T7, Canada. Correspondence and requests for materials should be addressed to T.L.G. (email: [tony.goldberg@wisc.edu](mailto:tony.goldberg@wisc.edu))



**Figure 1.** Comparison of the genome organization of Kanyawara virus (KYAV) to that of other members of the mononegaviral family *Rhabdoviridae*. Arrows signify open reading frames. Letters indicate the canonical rhabdovirus nucleoprotein (N), phosphoprotein (P), matrix (M), glycoprotein (G) and polymerase (L) genes and genus-specific open reading frames (U1, X).

monophyletic Nycteribiidae and the probably paraphyletic Streblidae, and they infest bats throughout the Old and New Worlds<sup>2,3</sup>. Bat flies of each family have evolved exquisite morphological and behavioral adaptations to life on bats, reflecting a long history of co-evolution<sup>2</sup>. Bat flies host a diverse community of bacteria, including bartonellae, some of which are zoonotic<sup>4,5</sup>. Bat flies also vector hemsporidian parasites (*Plasmodiidae*: *Polychromophilus melaniferus*) that cause “bat malaria”<sup>6</sup>. On the other hand, evidence for a role of bat flies as reservoirs or vectors of viruses is scant. Only two viruses have been unambiguously identified in bat flies: the putative orthoreovirus Mahlapitsi virus and the putative orthobunyavirus Wolkberg virus, which were both found in the nycteribiid *Eucampsipoda africana* on pteropodid Egyptian rousettes (*Rousettus aegyptiacus*)<sup>7,8</sup>. In addition, rhabdovirus RNA-like sequences were detected in nycteribiids and bats in Spain, but the sequences were too short (108 nt) to unambiguously substantiate virus infection<sup>9</sup>.

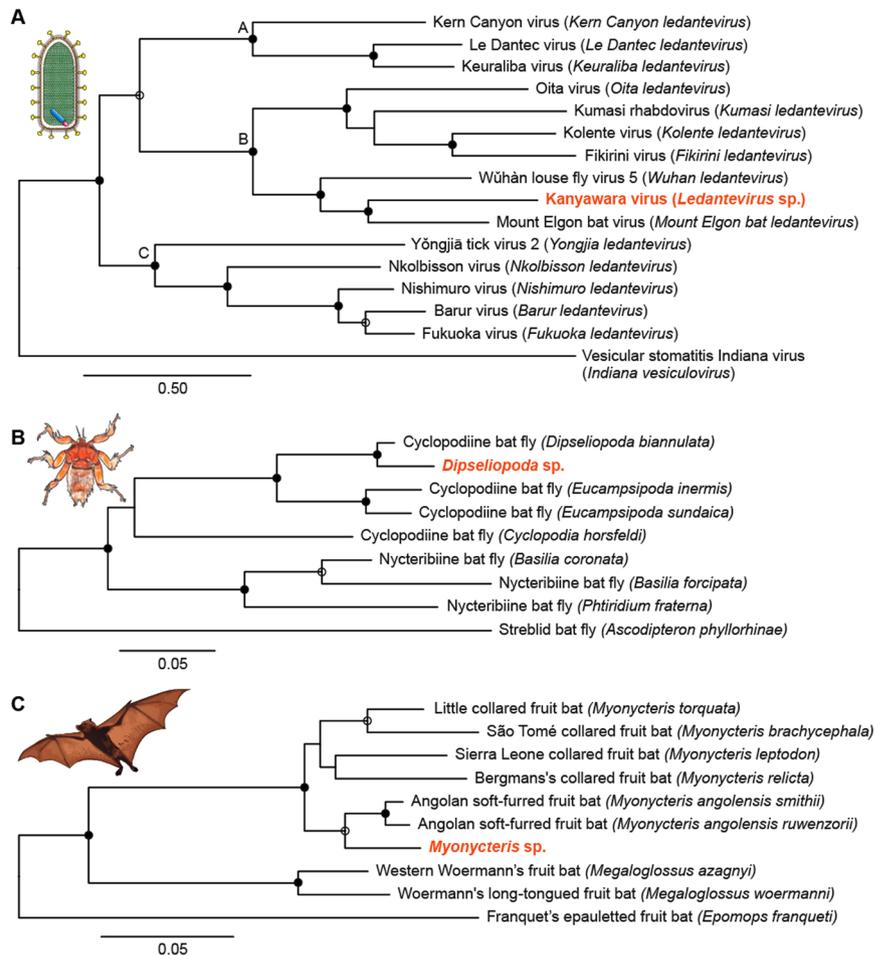
Here, we report the discovery and coding-complete genome sequence of a novel rhabdovirus, Kanyawara virus (KYAV), in a previously unknown nycteribiid bat fly collected from an unclassified megabat in western Uganda. Phylogenetic and genomic analyses of KYAV and its relatives offer new insights into the evolutionary and ecological associations of rhabdoviruses with both bats and arthropods.

## Results

Bat flies were found on six of nine pteropodid bats trapped at the edge of Kibale National Park, western Uganda, in 2010. Next-generation sequencing (NGS) of bat flies yielded  $0.11 \times 10^6$  to  $1.59 \times 10^6$  reads per sample. After quality trimming, rhabdovirus-like sequences were detected in five bat flies, each from a different bat. These sequences mapped with low similarity to conserved regions of rhabdovirus genomes (order *Mononegavirales*, family *Rhabdoviridae*). *De novo* assembly yielded a contiguous sequence of 10,843 nt in one bat fly sample (MPK004), with five open reading frames matching the canonical rhabdovirus genome organization (Fig. 1)<sup>10</sup>. Subsequent analysis of bat fly reads mapped 448 to 206,726 individual reads to this sequence, yielding coding-complete genomes in three other bat fly samples. Rhabdovirus coding genome sequences from bat flies of individual bats were 99.9% and 99.8% similar at the nucleotide and deduced amino acid levels, respectively. Viral read frequencies in the five positive bat fly samples ranged from 8,611 to 262,258 per million, with coverage ranging from 5-fold to 3,632-fold.

Sequencing of sera from the bats on which the bat flies were found yielded  $1.17 \times 10^6$  to  $2.97 \times 10^6$  reads per sample, but no reads mapped to the detected rhabdovirus genome. Application of this method at this sequencing depth is approximately as sensitive as real-time quantitative PCR<sup>11</sup>; therefore, bat sera could confidently be classified as negative for the virus. For further confirmation, however, we also tested all bat sera by PCR, and results were congruent with NGS results (*i.e.*, all bat sera tested negative for the new rhabdovirus). Oral and urogenital swab samples from all bats also tested negative for the new rhabdovirus by PCR.

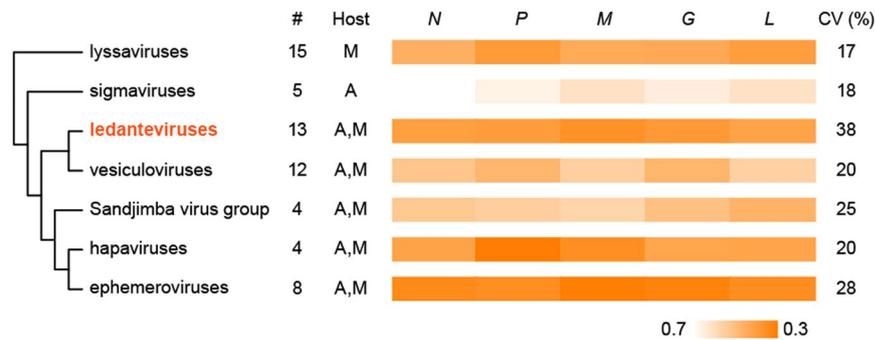
Phylogenetic analysis (Fig. 2A; Supplementary Table S1) indicates the rhabdovirus to be a new member of the recently established genus *Ledantevirus*<sup>12,13</sup>. We named this virus Kanyawara virus (KYAV) after the village closest to the roost from which the bats were sampled. Sequence similarity between KYAV and other ledanteviruses based on concatenated, codon-based alignments of the canonical N, P, G, M, and L genes ranged from 62.4% (Mount Elgon bat virus) to 47.6% (Yongjiā tick virus 2) at the nucleotide level and from 59.3% (Mount Elgon bat virus) to 38.7% (Kern Canyon virus) at the deduced amino acid level, respectively. KYAV fulfills four of the five criteria of the International Committee on Taxonomy of Viruses (ICTV) *Rhabdoviridae* Study Group for classification in the genus *Ledantevirus*: A) the deduced amino acid sequence of the KYAV RNA-dependent RNA polymerase (L) diverges >7% from that of other ledanteviruses (KYAV:Mount Elgon bat virus = 35.2%); B) the deduced amino acid sequence of the KYAV glycoprotein (G) diverges 15% from that of other ledanteviruses (KYAV:Mount Elgon bat virus = 49.0%); C) KYAV has the same genome organization as other ledanteviruses (Fig. 1); and E) KYAV occupies a different ecological niche than other ledanteviruses. Criterion D (“can be distinguished in serological tests”) could not be evaluated due to the absence of a replicating KYAV isolate, but the high divergence of the sequence of KYAV G, the only ledantevirion surface protein, strongly suggests that KYAV is also serologically distinct<sup>14</sup>.



**Figure 2.** Maximum likelihood phylogenetic trees of rhabdoviruses (A), bat flies (B), and pteropodid bats (C). The rhabdovirus phylogeny is based on concatenated codon-based alignments (8,256 positions) of nucleotide sequences of the canonical rhabdovirus nucleoprotein (N), phosphoprotein (P), matrix (M), glycoprotein (G), and RNA-dependent RNA polymerase (L) genes of 15 viruses of the subgroups A–C of the genus *Ledantevirus* with vesicular stomatitis Indiana virus (genus *Vesiculovirus*) as the outgroup. The bat fly phylogeny is based on concatenated codon-based alignments (876 positions) of mitochondrial cytochrome oxidase subunit II (COII) and cytochrome B (CYTB) nucleotide sequences of eight nycteribiids (subfamilies Cyclopediinae and Nycteribiinae) with the streblid *Ascodipteron phyllorhinae* as the outgroup. The bat phylogeny is based on concatenated codon-based alignments (1,820 positions) of mitochondrial cytochrome C subunit I (COI) and CYTB nucleotide sequences of nine pteropodids of the subfamily Epomophorinae, with eight of them belonging to the tribe Myonycterini and Franquet's epauletted fruit bats (tribe Epomophorini) as the outgroup. Circles on nodes indicate statistical confidence based on 1,000 bootstrap replicates of the data (closed circles = 100%; open circles  $\geq 75\%$ ); scale bars indicate nucleotide substitutions per site.

An analysis of the CpG content of the KYAV genome and related rhabdoviruses revealed significant variation (analysis of variance [ANOVA]  $F = 11.443$ ; 6 degrees of freedom;  $P < 0.0001$ ), with low relative CpG depletion in sigmaviruses, vesiculoviruses, and the Sandjimba virus group accounting for this trend (Holm  $T$ -statistic values ranging from 3.77 to 6.86;  $P$  values all  $< 0.01$ ; Supplementary Table S2). Figure 3 shows average CpG depletion by virus group and gene. CpG depletion was least pronounced for the insect-only sigmaviruses<sup>15,16</sup>, but more pronounced in the mammal-specific lyssaviruses<sup>17,18</sup>. These CpG variation patterns were generally consistent across the five canonical rhabdovirus genes *N*, *P*, *G*, *M*, and *L* within each virus group (Fig. 3). Within the genus *Ledantevirus*, KYAV and Oita virus have the lowest CpG depletion values (KYAV: 0.69; Oita virus: 0.72); these values were comparable to values for the insect-specific sigmaviruses (Supplementary Table S2). Variation in CpG frequency also differed significantly among rhabdovirus groups (Levine's  $W$  statistic = 3.29; 6 degrees of freedom;  $P = 0.008$ ). The coefficient of variation in CpG depletion was lowest for sigmaviruses and lyssaviruses and notably higher for the other virus groups (Fig. 3).

Phylogenetic analysis of mitochondrial DNA sequences from the collected bat flies revealed them to be members of the nycteribiid subfamily Cyclopediinae, representing a putative new species of the genus *Dipseliopoda*. These sequences are approximately as divergent from bat flies of the most closely related cyclopediine bat flies (*D. biannulatus*) as are the cyclopediine bat flies of the species *Eucampsipoda inermis* and *E. sundaica* (Fig. 2B).



**Figure 3.** Relative dinucleotide frequency in seven rhabdovirus groups. Shading indicates relative CpG frequency of the canonical rhabdovirus nucleoprotein (*N*), phosphoprotein (*P*), matrix (*M*), glycoprotein (*G*) and polymerase (*L*) genes, averaged across viruses within each group. Scale (bottom) is inverted, so that darker colors indicate higher relative CpG depletion. Numbers indicate sample sizes of viruses per group, and letters indicate host associations of viruses within groups (A = arthropod; M = mammal). CV indicates the coefficient of variation (percent) of CpG across virus genomes within each group. The cladogram (left) shows non-metric evolutionary relationships among genera, based on<sup>21</sup>. Included in the analysis were all viruses with full *L*, *P*, *G*, *M* and *L* gene sequences available in GenBank and the Virus Pathogen Resource database<sup>58</sup> as of December 12, 2016 (Supplementary Table S2).

Phylogenetic analyses of the sampled bats revealed them to be members of a putative new species, clustering as an outgroup to Angolan soft-furred bats (*Myonycteris angolensis*) and approximately as divergent from those bats as are bats of other species pairs within the genus *Myonycteris* (Fig. 2C).

## Discussion

Viruses of the family *Rhabdoviridae* infect vertebrates, invertebrates, and plants around the world<sup>10, 19</sup>. Their broad host range and wide geographic distribution reflect a deep evolutionary history of lineage-specific adaptation to particular host assemblages and ecologies of transmission<sup>10, 19–21</sup>. Bats are disproportionately represented among mammalian hosts of rhabdoviruses<sup>18, 19</sup>. For example, many viruses of the rhabdoviral genus *Lyssavirus*, including rabies virus, cause bat-borne zoonoses<sup>18, 19</sup>, and bats are the dominant vertebrate hosts for at least two of the three subgroups of the genus *Ledantevirus*<sup>10, 12</sup>. The reasons for this association are not clear but may reflect the unique diversity, biology, or social systems of bats<sup>1, 22, 23</sup>.

Viruses of the family *Rhabdoviridae* also have deep evolutionary relationships with arthropods, as do numerous viruses of other families within the order *Mononegavirales*<sup>24, 25</sup>. These relationships are evident today in the strong ecological associations that many rhabdoviruses maintain with arthropods. Viruses of the genus *Sigmavirus*, for example, are transmitted only vertically among insects<sup>20, 26</sup>, whereas viruses of the genera *Ephemerovirus*, *Tibrovirus*, and *Vesiculovirus* may infect mammals but typically are vectored by biting midges, mosquitoes, sandflies, or ticks<sup>10, 19</sup>. Similarly, plant rhabdoviruses (genera *Cytorhabdovirus* and *Nucleorhabdovirus*) are vectored by aphids, leafhoppers, or plant hoppers, and even fish rhabdoviruses transmitted directly through water may have associations with arthropods<sup>10, 19, 27</sup>. Rhabdovirus genome fragment integration into genomes of arthropods belonging to widely divergent lineages also supports a long history of rhabdovirus-arthropod coevolution<sup>19, 24, 28</sup>.

Despite this family-wide dual adaptation to arthropods and bats, vector-borne transmission of bat-associated rhabdoviruses has proven difficult to confirm. For example, Binger *et al.* searched for the vector of Kumasi rhabdovirus by trapping 1,240 female mosquitoes of six genera close to a large transient breeding colony of African straw-colored fruit bats (*Eidolon helvum*) in Ghana. No infected mosquitoes were identified<sup>29</sup>.

KYAV is a new putative member of the rhabdoviral genus *Ledantevirus*, sorting within subgroup B, which contains bat-associated viruses (Fig. 2A)<sup>12</sup>. The discovery of KYAV in nycteribiid bat flies suggests that KYAV could be a vector-borne virus, with bat flies as vectors. However, we did not find KYAV in the blood or on mucosal surfaces of the bats from which the bat flies were collected. This negative finding may indicate limited or transient viremia in bats, as is characteristic of, for instance, rabies virus<sup>30–32</sup>; however, other rhabdoviruses have been recovered from mucosal surfaces of bats<sup>9, 33</sup>. Alternatively, KYAV may be an insect-specific virus that does not infect bats. The detection of KYAV in 5 out of 6 (83%) bat flies sampled is consistent with this notion because infection rates of arthropod vectors with vector-borne viruses tend to be much lower than this rate, typically below 10%<sup>34</sup>.

The relative CpG dinucleotide frequency in viral genomes varies widely among taxa<sup>35</sup> and within virus groups<sup>36</sup>. CpG depletion has been used as an index of viral host adaptation<sup>16, 37, 38</sup>, although a recent study by Di Giallonardo *et al.* found the measure to be useful only for comparisons of higher taxonomic ranks such as Arthropoda compared to Vertebrata<sup>36</sup>. CpG frequencies in KYAV and related rhabdovirus genomes (Fig. 3) therefore likely reflect a combination of virological factors and host adaptation, with emphasis on the former<sup>36</sup>. In this light, our analyses show that CpG depletion was lowest among the insect-specific sigmaviruses. Genomes of lyssaviruses (including rabies virus), which are transmitted directly between mammals in the absence of arthropod vectors, had higher levels of CpG depletion<sup>17, 18</sup>. Genomes of ephemeroviruses, hapaviruses, and vesiculoviruses, which infect vertebrates

but are vectored by arthropods, had levels of CpG depletion comparable to those of the mammal-specific lyssaviruses, if not somewhat higher. KYAV and Oita virus genomes stand out among ledantavirus genomes by having very high CpG frequencies, similar to insect-specific sigmaviruses. We again caution that dinucleotide composition appears to be shaped more by virus taxon than by host species<sup>36</sup>; however, this metric remains useful for comparing similar viruses that infect very different hosts (e.g., mammals versus arthropods)<sup>16,39,40</sup>.

The nycteribiid bat flies in which we found KYAV are representatives of a putative new cyclopodiine species within the genus *Dipseliopoda* (Fig. 2B)<sup>2</sup>. This assessment is currently based solely on the phylogeny created here from mitochondrial cytochrome oxidase subunit II and cytochrome C DNA sequence data. Formal classification of these bat flies will have to await morphological characterization and additional genetic analyses.

Likewise, at the time of sampling, we thought based on morphologic characteristics that the collected bats were Angolan soft-furred bats (*Myonycteris angolensis*, also known as *Lissonycteris angolensis*<sup>41</sup>). However, our analysis of mitochondrial DNA sequences placed these bats as outgroup to Angolan soft-furred bats. They appear to be novel members of the Epomophorinae in the genus *Myonycteris* divergent enough to merit consideration as members of a separate species. This assessment is preliminary as it presently relies only on a mitochondrial cytochrome oxidase subunit I and cytochrome C phylogeny (Fig. 2C). Morphological characterization and additional genetic analyses will be required to confirm this taxonomy. Nevertheless, the discovery of a putative new pteropodid bat is surprising given that Kibale National Park is one of the most extensively researched forested areas in Africa<sup>42,43</sup>.

Overall, our data demonstrate that our understanding of the diversity of megabats, their ectoparasites, and their viruses is still fragmented. Viruses of bats are diverse in part because bats themselves are taxonomically diverse<sup>23,44</sup>. Therefore, identifying unknown taxa of megabats would be important for understanding the true diversity of their ectoparasites and associated viruses. Unfortunately, limited biological material and a remote field location in the present case precluded other desirable analyses, such as serologic assessment of bats or other mammals. Future studies using enzyme-linked immunosorbent assay (ELISA) or western blots targeting the major antigenic proteins of KYAV (likely N and/or G) might help elucidate the ecology of this virus in bats and animals of other species. Such studies may also resolve whether the absence of circulating KYAV in the tested bat sera reflects transient viremia, as we speculate, or lack of infection.

Bat flies occasionally bite people<sup>2,45</sup>. Therefore, enigmatic cases of human infection with bat-associated rhabdoviruses may have resulted from incidental bites by bat flies or other bat-associated arthropods. For example, in 1969, Le Dantec virus infected a British dockworker who was bitten by an unidentified insect while unloading peanuts from a ship that had come from Nigeria<sup>29</sup>. Novel, divergent rhabdoviruses have also been found in apparently healthy people in Africa, suggesting unknown pathways of zoonotic transmission<sup>46</sup>. Our identification of an unknown rhabdovirus on unknown bat flies of unknown bats suggests further research on the diversity of these insects and their role as disease vectors might prove fruitful.

## Methods

In February 2010, nine pteropodid bats were mist-netted from an established roost of approximately 12 bats in a peridomestic structure (a storeroom behind a kitchen) at the edge of Kibale National Park, western Uganda<sup>43,47</sup>. Blood, oral swabs, urogenital swabs, and ectoparasites were obtained from bats, and bats were immediately released thereafter. All protocols for animal and sample handling were approved in advance by the Uganda Wildlife Authority, the Uganda National Council for Science and Technology, Makerere University, McGill University, and the University of Wisconsin-Madison, and were performed in accordance with all relevant guidelines and regulations.

Bat flies (the only ectoparasites found on sampled bats) were kept separate by bat and were stored whole at  $-20^{\circ}\text{C}$  in RNAlater solution (Thermo Fisher Scientific, Inc., Waltham, MA, USA). Swab samples were also stored in RNAlater solution at  $-20^{\circ}\text{C}$ . Blood was separated by centrifugation into components for long-term storage at  $-80^{\circ}\text{C}$ . Single bat flies and swab tips were homogenized by bead beating using a portable homogenizer (Terralyzer; Zymo Research Corporation, Irvine, CA, USA). RNA was extracted from all sample types and converted to cDNA in the field using lyophilized reagents (RNA to cDNA EcoDry Premix, TaKaRa Bio USA Inc., Mountain View, CA, USA) and then converted into double-stranded cDNA (NEBNext Second Strand Synthesis Module, New England Biolabs, Ipswich, MA, USA). DNA was stabilized for long-term storage and transport to the USA at ambient temperature (DNastable, Biomatrix, San Diego, CA, USA).

DNA was reconstituted, and libraries were prepared for NGS as previously described<sup>48,49</sup>; this method is approximately as sensitive as real-time quantitative PCR for detecting viruses<sup>41</sup>. Briefly, DNA was purified using Agencourt Ampure XP beads (Beckman Coulter, Brea, CA, USA). Approximately 1 ng of DNA was prepared for sequencing on a MiSeq instrument (Illumina, San Diego, CA, USA) using the Nextera XT DNA sample preparation kit (Illumina). Sequence data were analyzed using CLC Genomics Workbench version 8.5 (CLC bio, Aarhus, Denmark). Low-quality bases were trimmed (phred quality score  $< 30$ ), short reads ( $< 75$  bp) were discarded, and the remaining reads were subjected to *de novo* assembly. Assembled contiguous sequences (contigs) were analyzed for nucleotide- (blastn) and protein-level (blastx) similarity to known viruses in GenBank. All sequences generated in this study were deposited into GenBank. All sequences used for analyses and their accession numbers are listed in Supplementary Table S1.

Genetic similarity between Kanyawara virus (KYAV) and its relatives was assessed using pairwise sequence comparisons in the computer program MEGA7<sup>50</sup>. Maximum likelihood phylogenetic analyses were conducted on codon-based alignments of concatenated virus genes, with poorly aligned regions removed. Alignments were created using the MAFFT algorithm<sup>51</sup> implemented in Translator X<sup>52</sup>, with the Gblocks algorithm<sup>53</sup> applied to remove poorly aligned regions. Trees were constructed using the maximum likelihood method implemented in PhyML<sup>54</sup>, with best-fit models of molecular evolution estimated from the data using jModeltest<sup>55</sup>. Trees were displayed using FigTree<sup>56</sup>. The same phylogenetic methods were applied to mitochondrial gene sequences of bats

(COI and CYTB) and bat flies (COII and CYTB nucleotide sequences) extracted from deep sequence data to investigate the taxonomy of sampled bats and bat flies.

Based on coding-complete virus genome sequences detected in the course of bioinformatics analysis (see Results), PCR primers KYAV-10368 (5'-GCGAACCCGACGATCATAGT-3') and KYAV-10695 (5'-GCTGTGCATTCCAGTCTCCT-3') were designed to amplify a 327-bp region of the KYAV RNA-dependent RNA polymerase (*L*) gene. PCR conditions were optimized using bat fly cDNA samples known to be positive and negative for KYAV infection by NGS. The optimized PCR was used to test cDNA extracted from swab samples. PCRs were conducted using the HiFi kit (Kappa Biosystems, Wilmington, MA, USA), with 40 cycles at 94 °C for 15 sec, 56 °C for 15 sec, and 72 °C for 15 sec. Amplicons were visualized on 2% agarose electrophoretic gels stained with ethidium bromide.

The relative frequency of CpG dinucleotide pairs was calculated for each of the canonical rhabdovirus genes (*N*, *P*, *G*, *M*, and *L*) from KYAV and related rhabdoviruses using the R Biostrings package<sup>57</sup>. Rhabdovirus sequences included in this analysis were selected objectively, using all coding-complete genomes from major rhabdovirus clades (genera) with at least four members available in GenBank and the Virus Pathogen Resource database<sup>58</sup> as of December 12, 2016. Differences in mean CpG depletion among groups were evaluated for statistical significance using ANOVA, with pairwise differences between groups examined using the Holm *post-hoc* method, which adjustments for multiple testing<sup>59</sup>. Variances in CpG depletion among groups were compared using Levine's test<sup>60</sup>. Statistical analyses were performed in the computer package R<sup>61</sup>.

**Data Availability.** All data generated during the current study are available in GenBank (accession numbers KY385385-385392) or are included in this published article and its Supplementary Information files.

## References

- Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V. & Schountz, T. Bats: important reservoir hosts of emerging viruses. *Clin Microbiol Rev* **19**, 531–545, doi:10.1128/CMR.00017-06 (2006).
- Dick, C. W. & Patterson, B. D. Bat flies: obligate ectoparasites of bats. In *Micromammals and Macroparasites—From Evolutionary Ecology to Management* (eds S. Morand, B. R. Krasnov, & R. Poulin) 179–194 (Springer, 2006).
- Petersen, F. T., Meier, R., Kutty, S. N. & Wiegmann, B. M. The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers. *Mol Phylogenet Evol* **45**, 111–122, doi:10.1016/j.ympev.2007.04.023 (2007).
- Morse, S. F. *et al.* Global distribution and genetic diversity of *Bartonella* in bat flies (Hippoboscoidea, Streblidae, Nycteribiidae). *Infect Genet Evol* **12**, 1717–1723, doi:10.1016/j.meegid.2012.06.009 (2012).
- Wilkinson, D. A. *et al.* The bacteriome of bat flies (Nycteribiidae) from the Malagasy Region: a community shaped by host ecology, bacterial transmission mode, and host-vector specificity. *Appl Environ Microbiol* **82**, 1778–1788, doi:10.1128/AEM.03505-15 (2016).
- Obame-Nkoghe, J. *et al.* Bat flies (Diptera: Nycteribiidae and Streblidae) infesting cave-dwelling bats in Gabon: diversity, dynamics and potential role in *Polychromophilus melanipherus* transmission. *Parasit Vectors* **9**, 333, doi:10.1186/s13071-016-1625-z (2016).
- Jansen van Vuren, P. *et al.* Isolation of a novel fusogenic orthoreovirus from *Eucampsipoda africana* bat flies in South Africa. *Viruses* **8**, 65, doi:10.3390/v8030065 (2016).
- Jansen van Vuren, P. *et al.* Isolation of a novel orthobunyavirus from bat flies (*Eucampsipoda africana*). *J Gen Virol*. doi:10.1099/jgv.0.000753 (2017).
- Aznar-Lopez, C. *et al.* Detection of rhabdovirus viral RNA in oropharyngeal swabs and ectoparasites of Spanish bats. *J Gen Virol* **94**, 69–75, doi:10.1099/vir.0.046490-0 (2013).
- Dietzgen, R. G. *et al.* Rhabdoviridae. In *Virus Taxonomy—Ninth Report of the International Committee on Taxonomy of Viruses* (eds Andrew M. Q. King, Michael J. Adams, Eric B. Carstens, & Elliot J. Lefkowitz) 686–713 (Elsevier/Academic Press, 2011).
- Kurth, K., Sibley, S. & Goldberg, T. L. Metagenomic assessment of adventitious viruses in commercial bovine sera. *Biologicals* in press (2017).
- Blasdell, K. R. *et al.* *Ledantevirus*: a proposed new genus in the *Rhabdoviridae* has a strong ecological association with bats. *Am J Trop Med Hyg* **92**, 405–410, doi:10.4269/ajtmh.14-0606 (2015).
- Amarasinghe, G. K. *et al.* Taxonomy of the order *Mononegavirales*: update 2017. *Arch Virol*, Apr 7. doi:10.1007/s00705-017-3311-7 (2017).
- Walker, P. J. *et al.* One new genus (*Ledantevirus*) including 14 new species in the family Rhabdoviridae. International Committee on Taxonomy of Viruses (ICTV) TaxoProp 2016.006a-dM. <https://talk.ictvonline.org/ICTV/proposals/2016.006a-dM.A.v2.Ledantevirus.pdf> (2016).
- Chantawannakul, P. & Cutler, R. W. Convergent host-parasite codon usage between honeybee and bee associated viral genomes. *J Invertebr Pathol* **98**, 206–210, doi:10.1016/j.jip.2008.02.016 (2008).
- Upadhyay, M., Sharma, N. & Vivekanandan, P. Systematic CpT (ApG) depletion and CpG excess are unique genomic signatures of large DNA viruses infecting invertebrates. *PLoS One* **9**, e111793, doi:10.1371/journal.pone.0111793 (2014).
- Shackelton, L. A., Parrish, C. R. & Holmes, E. C. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J Mol Evol* **62**, 551–563, doi:10.1007/s00239-005-0221-1 (2006).
- Lobo, F. P. *et al.* Virus-host coevolution: common patterns of nucleotide motif usage in *Flaviviridae* and their hosts. *PLoS One* **4**, e6282, doi:10.1371/journal.pone.0006282 (2009).
- Dietzgen, R. G., Kondo, H., Goodin, M. M., Kurath, G. & Vasilakis, N. The family *Rhabdoviridae*: mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. *Virus Res* **227**, 158–170, doi:10.1016/j.virusres.2016.10.010 (2017).
- Longdon, B. *et al.* The evolution, diversity, and host associations of rhabdoviruses. *Virus Evol* **1**, vev014, doi:10.1093/ve/vev014 (2015).
- Walker, P. J. *et al.* Evolution of genome size and complexity in the *Rhabdoviridae*. *PLoS Pathog* **11**, e1004664, doi:10.1371/journal.ppat.1004664 (2015).
- Dobson, A. P. Virology. What links bats to emerging infectious diseases? *Science* **310**, 628–629, doi:10.1126/science.1120872 (2005).
- Luis, A. D. *et al.* A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc Biol Sci* **280**, 20122753, doi:10.1098/rspb.2012.2753 (2013).
- Li, C.-X. *et al.* Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife* **4**, e05378, doi:10.7554/eLife.05378 (2015).
- Shi, M. *et al.* Redefining the invertebrate RNA virosphere. *Nature* **540**, 539–543, doi:10.1038/nature20167 (2016).
- Longdon, B. *et al.* Vertically transmitted rhabdoviruses are found across three insect families and have dynamic interactions with their hosts. *Proc Biol Sci* **284**, doi:10.1098/rspb.2016.2381 (2017).

27. Faisal, M. & Winters, A. D. Detection of viral hemorrhagic septicemia virus (VHSV) from *Diporeia* spp. (Pontoporeiidae, Amphipoda) in the Laurentian Great Lakes, USA. *Parasit Vectors* **4**, 2, doi:10.1186/1756-3305-4-2 (2011).
28. Fort, P. *et al.* Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. *Mol Biol Evol* **29**, 381–390, doi:10.1093/molbev/msr226 (2012).
29. Binger, T. *et al.* A novel rhabdovirus isolated from the straw-colored fruit bat *Eidolon helvum*, with signs of antibodies in swine and humans. *J Virol* **89**, 4588–4597, doi:10.1128/JVI.02932-14 (2015).
30. Sims, R. A., Allen, R. & Sulkin, S. E. Studies on the pathogenesis of rabies in insectivorous bats. III. Influence of the gravid state. *J Infect Dis* **112**, 17–27 (1963).
31. Burne, J. C. Viræmia in rabies. *Lancet* **295**, 195–196 (1970).
32. Reaves, E. J. *et al.* Susceptibility and lack of evidence for a viremic state of rabies in the night owl monkey *Aotus nancymaae*. *Viol J* **9**, 95, doi:10.1186/1743-422X-9-95 (2012).
33. Vázquez-Morón, S. *et al.* Asymptomatic rhabdovirus infection in meridional serotine bats (*Eptesicus isabellinus*) from Spain. *Dev Biol (Basel)* **131**, (311–316 (2008).
34. Bustamante, D. M. & Lord, C. C. Sources of error in the estimation of mosquito infection rates used to assess risk of arbovirus transmission. *Am J Trop Med Hyg* **82**, 1172–1184, doi:10.4269/ajtmh.2010.09-0323 (2010).
35. Glass, J. L. *et al.* CG dinucleotide clustering is a species-specific property of the genome. *Nucleic Acids Res* **35**, 6798–6807, doi:10.1093/nar/gkm489 (2007).
36. Di Giallonardo, F., Schlub, T. E., Shi, M. & Holmes, E. C. Dinucleotide composition in animal RNA viruses is shaped more by virus family than host species. *J Virol*. doi:10.1128/JVI.02381-16 (2017).
37. Kapoor, A., Simmonds, P., Lipkin, W. I., Zaidi, S. & Delwart, E. Use of nucleotide composition analysis to infer hosts for three novel picorna-like viruses. *J Virol* **84**, 10322–10328, doi:10.1128/JVI.00601-10 (2010).
38. Cheng, X. *et al.* CpG usage in RNA viruses: data and hypotheses. *PLoS One* **8**, e74109, doi:10.1371/journal.pone.0074109 (2013).
39. Upadhyay, M. *et al.* CpG dinucleotide frequencies reveal the role of host methylation capabilities in parvovirus evolution. *J Virol* **87**, 13816–13824, doi:10.1128/JVI.02515-13 (2013).
40. Blitvich, B. J. & Firth, A. E. Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses* **7**, 1927–1959, doi:10.3390/v7041927 (2015).
41. Nesi, N. *et al.* Molecular systematics and phylogeography of the tribe Myonycterini (Mammalia, Pteropodidae) inferred from mitochondrial and nuclear markers. *Mol Phylogenet Evol* **66**, 126–137, doi:10.1016/j.ympev.2012.09.028 (2013).
42. Chapman, C. A., Struhsaker, T. T. & Lambert, J. E. Thirty years of research in Kibale National Park, Uganda, reveals a complex picture for conservation. *Int J Primatol* **26**, 539–555, doi:10.1007/s10764-005-4365-z (2005).
43. Plumtre, A. J. *et al.* The biodiversity of the Albertine Rift. *Biol Cons* **134**, 178–194, doi:10.1016/j.biocon.2006.08.021 (2007).
44. Quan, P.-L. *et al.* Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proc Natl Acad Sci USA* **110**, 8194–8199, doi:10.1073/pnas.1303037110 (2013).
45. Wenzel, R. L. & Tipton, V. J. In *Ectoparasites of Panama* (eds Wenzel R. L. & V. J. Tipton) 677–723 (Field Museum of Natural History, 1966).
46. Stremlau, M. H. *et al.* Discovery of novel rhabdoviruses in the blood of healthy individuals from West Africa. *PLoS Negl Trop Dis* **9**, e0003631, doi:10.1371/journal.pntd.0003631 (2015).
47. Kityo, R. & Kerbis, J. C. Observations on the distribution and ecology of bats in Uganda. *J E Afr Nat Hist* **85**, 49–63 (1996).
48. Kurth, K., Sibley, S. & Goldberg, T. L. Metagenomic assessment of adventitious viruses in commercial bovine sera. *Biologicals*, in press (2017).
49. Sibley, S. D. *et al.* Novel reovirus associated with epidemic mortality in wild largemouth bass (*Micropterus salmoides*). *J Gen Virol* **97**, 2482–2487, doi:10.1099/jgv.0.000568 (2016).
50. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**, 1870–1874, doi:10.1093/molbev/msw054 (2016).
51. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066 (2002).
52. Abascal, F., Zardoya, R. & Telford, M. J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res* **38**, W7–13, doi:10.1093/nar/gkq291 (2010).
53. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564–577, doi:10.1080/10635150701472164 (2007).
54. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321, doi:10.1093/sysbio/syq010 (2010).
55. Santorum, J. M., Darriba, D., Taboada, G. L. & Posada, D. *jmodeltest.org*: selection of nucleotide substitution models on the cloud. *Bioinformatics* **30**, 1310–1311, doi:10.1093/bioinformatics/btu032 (2014).
56. Rambaut, A. FigTree, v1.4.3. <http://tree.bio.ed.ac.uk/software/figtree/> (2016).
57. Pagès, H., Aboyou, P., Gentleman, R. & DebRoy, S. Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2.42.1. <http://bioconductor.org/packages/Biostrings/> (2016).
58. Pickett, B. E. *et al.* ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res* **40**, D593–598, doi:10.1093/nar/gkr859 (2012).
59. Aickin, M. & Gensler, H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *Am J Public Health* **86**, 726–728 (1996).
60. Levene, H. Robust tests for equality of variance. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (eds I. Olkin *et al.*) 278–292 (Stanford University Press, 1960).
61. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org/> (2013).

## Acknowledgements

We thank the Uganda Wildlife Authority and Uganda National Council of Science and Technology for permission to conduct this research. We also thank May Nakityo and Rhoda Nankabirwa for assistance with fieldwork and Makerere University Biological Field Station for logistic support. We are grateful to Laura Bollinger and Jiro Wada (National Institutes of Health, National Institute of Allergy and Infectious Diseases (NIH/NIAID) Integrated Research Facility at Fort Detrick, Frederick, MD, USA) for editing the manuscript and improving figures, respectively, and to two anonymous reviewers for providing valuable suggestions. This work was funded in part by United States National Institutes of Health grant TW009237 as part of the joint NIH-NSF Ecology of Infectious Disease program, the Natural Sciences and Engineering Research Council of Canada, the National Geographic Society, and by the University of Wisconsin-Madison John D. MacArthur Fellowship Program. This work was also supported in part through Battelle Memorial Institute's prime contract with the NIAID under Contract No. HHSN2722007000161. A subcontractor to Battelle Memorial Institute who performed this work is: J.H.K., an employee of Tunnell Government Services, Inc.

### Author Contributions

R.K. collected samples, T.L.G. performed experiments, T.L.G. and A.J.B. analyzed the data; T.L.G., R.K., A.J.B., J.H.K., and C.A.C. wrote the manuscript. All authors reviewed and improved the final manuscript.

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-05236-w](https://doi.org/10.1038/s41598-017-05236-w)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017