Cameroonian blackflies (Diptera: Simuliidae) harbour a plethora of (RNA) viruses

Lander De Coninck[1], Amber Hadermann[2], Ludovica Ingletto[1], Robert Colebunders[2,3], Kongnyu Gamnsi Njamnshi[4], Alfred Kongnyu Njamnshi[4], John L Mokili[5], Joseph Nelson Siewe Fodjo[2#], Jelle Matthijnssens[1#]

1. KU Leuven, Department of Microbiology, Immunology and Transplantation, Rega Institute, Laboratory of Viral Metagenomics, Leuven, Belgium
2. Global Health Institute, University of Antwerp, Antwerp, Belgium.
3. Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom
4. Brain Research Africa Initiative (BRAIN), Yaoundé, Cameroon & Geneva, Switzerland; Neuroscience Lab, Faculty of Medicine and Biomedical Sciences, The University of Yaoundé I, Yaoundé, Cameroon.
5. Viral Information Institute, Department of Biology, San Diego State University San Diego, CA, USA.

[#]Co-last authors

Abstract

Strong epidemiological evidence suggests that onchocerciasis may be associated with epilepsy – hence the name onchocerciasis-associated epilepsy (OAE). However, the pathogenesis of OAE still needs to be elucidated, as recent studies failed to detect *Onchocerca volvulus* in the central nervous system of persons with OAE. Therefore, it was suggested that a potentially neurotropic virus transmitted by blackflies could play a role in triggering OAE. To investigate this hypothesis, adult blackflies were collected in an onchocerciasis-endemic area with a high OAE prevalence in the Ntui Health District, Cameroon. A whole-genome shotgun sequencing approach was used to detect viral sequences in fifty-five pools of ten blackflies. A very high abundance of viral reads was detected across multiple (novel) viral families, including viral families associated with human disease. Although, no genomes closely related to known neurotropic viruses were found in the blackfly virome, the plethora of novel viruses representing novel species, genera and even families, warrant further exploration for their potential to infect vertebrates. These results could serve as a first step for studying the viruses associated with the

hematophagous blackfly, which also could be present in their nematode host *O. volvulus*. Exploring the diversity of viruses in blackflies should be included in the active surveillance of zoonotic diseases.

Introduction

Human onchocerciasis, commonly known as river blindness, is a parasitic disease caused by the filarial worm *Onchocerca volvulus* (Nematoda: Secernentea; Spirurida; Filariidae; Onchocerca). Transmission of the parasite occurs through repeated bites of female blackflies belonging to the *Simulium* species (Brattig, Cheke and Garms 2021). Onchocerciasis is classified as one of the neglected tropical diseases recognized by the World Health Organization (WHO).

Onchocerciasis-endemic regions with high ongoing or past *O. volvulus* transmission are known to have high epilepsy prevalence (Pion and Boussinesq 2012; Levick *et al.* 2017; Colebunders *et al.* 2018b; Siewe Fodjo *et al.* 2018; Boullé *et al.* 2019; Mukendi *et al.* 2019; Raimon *et al.* 2021). While onchocerciasis is commonly thought to affect only the skin and eyes, emerging epidemiological evidence suggests a potential direct or indirect association between *O. volvulus* infection and epilepsy, giving rise to the term "Onchocerciasis-Associated Epilepsy" (OAE), also known as "River Epilepsy" (Chesnais *et al.* 2018, 2020; Colebunders *et al.* 2019, 2021; Hadermann *et al.* 2023). OAE presents with a wide spectrum of epileptic seizures, including generalized tonic-clonic seizures, myoclonic seizures, absence seizures and head nodding seizures (Colebunders *et al.* 2018a). Nodding and Nakalanga syndrome are considered to be phenotypic presentations of OAE (Colebunders *et al.* 2018a; Van Cutsem *et al.* 2023) Nakalanga syndrome is characterized by severe growth retardation, delayed sexual development, mental impairment, facial deformation, kyphoscoliosis, and epileptic seizures (Föger *et al.* 2017; Siewe Fodjo *et al.* 2019).

So far, the pathogenesis of OAE is not yet established. Prior to the introduction of ivermectin mass drug distribution, O. volvulus microfilariae (mf) had been detected in the cerebrospinal fluid (CSF) of individuals with onchocerciasis (Duke, Vincelette and Moore 1976). However, in more recent studies, neither O. volvulus mf nor DNA could be detected in CSF (Winkler *et al.* 2013; Hotterbeekx *et al.* 2020) nor in post-mortem brain tissue of

persons with OAE (Hotterbeekx *et al.* 2019). Therefore, it was suggested that a potentially neurotropic virus transmitted by blackflies, or an endosymbiont of the *O. volvulus* parasite could play a role in the pathogenesis of OAE ((Colebunders *et al.* 2014; Hadermann *et al.* 2023).

Very little is known about the virome of blackflies. Kraberger *et al.* described the ssDNA virome of blackflies in New Zealand and reported high numbers of *Genomoviridae*, *Circoviridae* and *Microviridae* (Kraberger *et al.* 2019). Moreover, invertebrate iridoviruses (family *Iridoviridae*), have been identified in blackfly larvae across the globe (Piégu *et al.* 2013, 2014a, 2014b). In addition, the most studied arbovirus transmitted by blackflies is vesicular stomatitis virus (family *Rhabdoviridae;* genus *Vesiculovirus*), that typically infects cattle, horses, and swine (Howerth, Mead and Stallknecht 2002). Nonetheless, there have been no reports of blackflies transmitting arboviruses to humans to date. However, considering the 1700 *Simulium* species described worldwide, it is plausible that blackflies may transmit unidentified viruses from human and zoonotic origins. The aim of this study was to identify potential neurotropic viruses (in this study defined as viruses related to viruses known to cause infections in the central nervous system, like West Nile virus, Japanese encephalitis virus [*Flaviviridae*] and rabies virus [*Rhabdoviridae*]) transmitted by blackflies, which could play a role in the pathogenesis of OAE.

Materials and Methods

Sample collection

Adult blackflies investigated were collected in the village of Nachtigal (coordinates of the breeding site: N 4°21.146, E 11°37.953; Garmin GPSMAP78). The gathering took place from 7:00 AM to 6:00 PM over three consecutive days in July 2021. Nachtigal is a rural community of the Ntui Health District, an onchocerciasis-endemic area with a high prevalence of OAE (Siewe Fodjo *et al.* 2022), in the forest-savannah transition zone in the Centre Region of Cameroon. This community lives close to the Nachtigal rapids on the Sanaga River, which constitutes one of the main sources of blackfly vectors in Cameroon (Siewe Fodjo *et al.* 2022). The human landing catch method was used to collect the blackflies and mouth aspirators were used to aspirate blackflies into an empty container every hour (Hendy *et al.*

2021). At the end of each catching day, blackflies were stored dry at -20°C in the BRAIN laboratory, Yaoundé, Cameroon until shipping on dry ice a few weeks later. The frozen blackfly samples were then sent to the Laboratory of Viral Metagenomics in Leuven, Belgium, where they were stored at −80°C until the nucleic acids were extracted for sequencing.

Sample processing

A total of fifty-five pools of ten blackflies were analysed. The NetoVIR protocol (Conceição-Neto *et al.* 2015) was adapted for the homogenization of the blackfly pools. Homogenization was carried out in 2mL tubes with 2.8mm zirconium oxide beads at 5000 rpm for 15 seconds. Negative controls were added in the beginning and taken through the whole process (including sequencing) to make sure no contamination happened during the processing of the samples. Samples were centrifuged and filtered through a 0.8 µm filter (Sartorius) to remove cellular debris and enrich the samples for viral particles. This filtrate was treated with a mix of Benzonase (50 U, Novagen) and Micrococcal nuclease (2000 U, New England Biolabs) to digest free-floating nucleic acids. DNA and RNA was simultaneously extracted using the QiaAMP Viral RNA Mini kit (Qiagen) without carrier RNA. A random amplification of both DNA and RNA with the Whole Transcriptome Amplification (WTA2) kit increased the nucleic acid concentration of the samples, and resulting PCR products were further purified and prepared for sequencing with the Nextera XT kit (Illumina). The final sequencing libraries were cleaned up with Agencourt AMPure XP beads (Beckman Coulter, Inc.) using a 0.6 ratio of beads to sample. Paired-end sequencing (2x150bp) was performed on the Novaseq 6000 SP platform at the Nucleomics Core sequencing facility (VIB Leuven).

Raw Reads Processing and Virome Analysis

To process the raw paired-end Illumina reads, an in-house bioinformatics pipeline (ViPER v1.1) (De Coninck 2021) was used with the 'triple assembly' option enabled. Briefly, adapters and low-quality reads were trimmed with Trimmomatic (Bolger, Lohse and Usadel 2014), before a *de novo* assembly with metaSPAdes v3.15.3 (Nurk *et al.* 2017). Resulting fasta files with scaffolds were subsequently concatenated. To remove redundancy in the

data, these scaffolds were clustered on 95% nucleotide identity over 85% coverage of the shortest sequence with the clustering algorithm available from CheckV (Nayfach *et al.* 2020). A length filter of 1000 nucleotides was applied on the sequences before viruses were identified in the data using a combined approach. This approach consisted of: 1) genomad v1.7.0 on default setting with score calibration enabled (Camargo *et al.* 2023); 2) DIAMOND blastx v2.0.11 on the sensitive setting (Buchfink, Xie and Huson 2014) with NCBI's nr database (accessed 23-03-2023); and 3) a Hidden Markov Model (HMM) search of the scaffolds' open reading frames (ORFs), predicted by prodigal (Hyatt *et al.* 2010), against the NeoRdRP HMM dataset (v1.0) (Sakaguchi *et al.* 2022) with HHMER v3.3.2 (Eddy 2011) with a minimum e-value 1e-10. Taxonomy assignment for blastx-identified sequences was decided based on a lowest common ancestor approach from the best twenty-five hits. When possible, the genomad taxonomy output was prioritized for a sequence over the blastx taxonomic assignment. Subsequently, for quality control purposes, CheckV v1.0.1 (Nayfach *et al.* 2020) was employed to assess the genome completeness of the identified viral sequences. Sequences not belonging to the *Riboviria* and less than 50% complete as predicted by CheckV were removed from further analysis. *Riboviria* sequences predicted to be less than 50% complete were kept because CheckV does not handle (segmented) RNA viruses well, which often results in wrong completeness estimations for RNA viruses. Finally, using the sequencing data of the negative controls, sequences that were predicted by the prevalence method of decontam v1.22.0 (Davis *et al.* 2018) to be a contaminant were removed from our dataset.

Phylogenetic Analysis

Based on the taxonomic assignment of the scaffolds (*vide supra*), we downloaded the International Committee on Taxonomy of Viruses (ICTV) representative protein sequences of each family or order present in our dataset (from ICTV's MSL39 v1 VMR) with NCBI Virus. Except for the *Orthomyxoviridae* of which representative PB1 sequences were downloaded from https://github.com/evogytis/orthomyxo-metagenomics/blob/main/data/Fig3/PB1_full.fasta (Dudas and Batson 2023). For the RNA viruses, we first filtered out all non-RdRP proteins and partial RdRPs with a combination of palm_annot (https://github.com/rcedgar/palm_annot; commit 15d9443) and palmscan2

(https://github.com/rcedgar/palmscan) to retrieve the intact palmcore domain (Babaian and Edgar 2022). For the *Genomoviridae* and *Parvoviridae*, we downloaded the Rep and NS1 protein sequences, respectively. For each family or order, a diversified ensemble was created with Muscle5 (Edgar 2022). This diversified ensemble constitutes 100 multiple sequence alignments (MSAs) of the sequences with perturbed alignment parameters and permuted guide trees (see Muscle5 documentation for more information). From these 100 MSAs, the alignment with the highest column confidence, calculated with muscle maxcc, was picked to generate the final phylogenetic tree. Resampling of the diversified ensemble to generate support values for the phylogenetic tree was obtained with muscle resample -conf 0 -gapfract 0.5. Maximum likelihood (ML) phylogenetic trees from the resampled MSAs were inferred with Fasttree (Price, Dehal and Arkin 2010) (-lg -gamma -nosupport). These ML trees served as support for the tree that was generated from the maxcc MSA (also generated with Fasttree and unaltered settings), which was calculated with newick conftree (https://github.com/rcedgar/newick).

Viral host inference

To infer if the *Flaviviridae* and *Rhabdoviridae* viruses in our data could have an invertebrate or vertebrate host, we calculated the relative dinucleotide abundance (RDA) of our sequences and sequences with a known host (see Supplementary Table S2 for accession numbers) with DinuQ (Lytras and Hughes 2020). Next, we performed a principal component analysis in R on all dinucleotide RDA values and plotted the first against the second principal component for all viruses.

Ethical approval

Approval was obtained from the ethics committee of the Cameroon Baptist Convention Health Services (reference number IRB2021-03). We also obtained a research permit from the Ministry of Scientific Research and Innovation (Ref: 000144/MINRESI/B00/C00/C10/C13), as well as an ABS-Nagoya Protocol Prior Informed Consent (Ref: Decision No. 00016/D/MINEPDED/CNA) and ABS Permit No. 00013/MINEPDED/CAN/NP-ABS/ABS-FP) from the Ministry of Environment, Protection of Nature and Sustainable Development.

Results

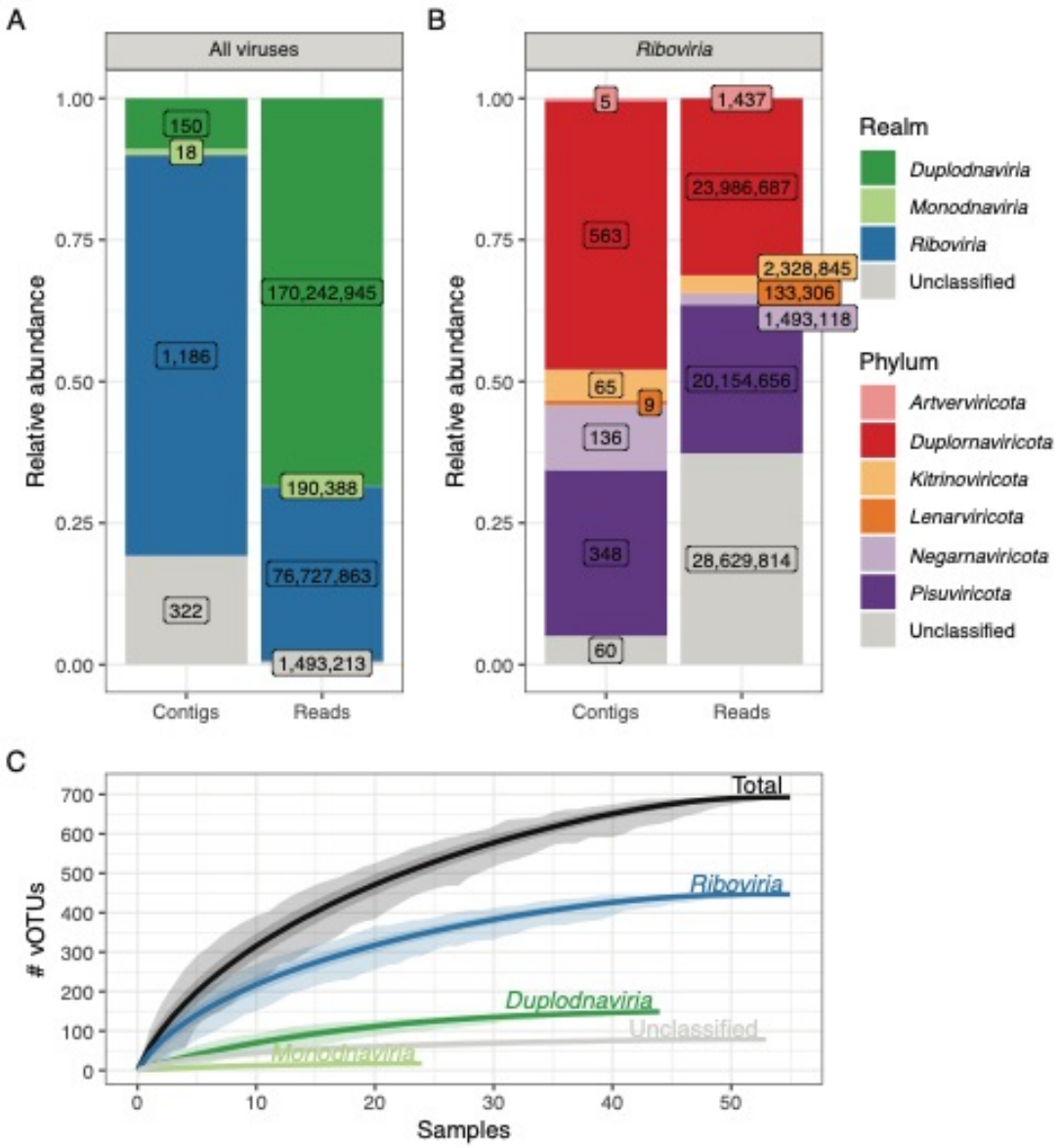
Identification of more than thousand viral contigs

For the fifty-five blackfly pools sequenced, between 6,113,870 and 24,002,232 reads were obtained per pool. In total, around 720 million raw reads were generated from all fifty-five pools. Raw read classification with Kraken2 (Wood, Lu and Langmead 2019) showed that most reads could not be directly classified, followed by on average 30% of the reads assigned to Bacteria (See Supplemental Figure 1B). After trimming, each pool retained between 4,544,534 and 20,790,370 reads, resulting in a 17.5% loss of the initial raw reads (see Supplementary Table S1 and Supplemental Figure 1A). Subsequently, *de novo* assemblies were generated, resulting in a total of 59,645 scaffolds larger than 1000 bp from the 55 pools. The clustering of the assembled scaffolds from all samples based on 95% nucleotide identity over 85% coverage resulted in 45,308 non-redundant scaffolds. After removing contamination predicted with decontam, a total of 1,678 scaffolds were identified as viral with a combination of genomad, DIAMOND blastx and an HMM search against the NeoRdRP dataset (see Supplemental Figure 2A). Of these, 56 were considered to be 100% complete, 128 were of high quality (>90% complete) and 190 of medium quality (>50% complete), the remaining viral sequences were either of low quality or their completeness could not be determined by CheckV (see Supplemental Figure 2B). However, the low-quality and not-determined groups consisted only of sequences with *Riboviria* or unclassified taxonomic realm assignments. This suggests that these sequences are significantly divergent and might thus be of a higher quality than what is estimated by CheckV (also see Materials).

Overall, we concluded that, due to the high convergence of our three identification methods and the completeness estimation of CheckV, we have a robust set of identified viruses.

The blackfly virome is dominated by RNA viruses

When examining the taxonomic distribution of our viral genome sequences, the majority (n=1186) were classified within the *Riboviria* realm, 322 could not be assigned into a specific realm and a small number belonged to the *Monodnavaria* and the *Duplodnaviria* (see Figure 1A). At the read level, a distinct picture arises with approximately two thirds being identified as belonging to the *Duplodnaviria* realm. This is mostly because the average viral genome length of *Duplodnaviria* is much higher than the average genome length for the *Riboviria*, which in our data recruit about one third of the viral reads (see Figure 1A and Supplemental Figure 3). The 322 unclassified sequences recruit only a small fraction (~1.5 million) of the reads. As the diversity seemed the greatest in the *Riboviria* realm, which also contains the vast majority of arboviruses, we further investigated the taxonomic distribution of their scaffolds as well as their reads. Around half of the viral genome sequences belonged to the *Duplornaviricota* phylum (see Figure 1B), while this phylum constituted only one third of the *Riboviria* reads (see Figure 1B). Surprisingly, 81 further unclassified *Riboviria* sequences were responsible for another third of the viral reads in our blackfly samples. These results show that the blackflies contained a high number of completely novel RNA viruses that were present in high abundance. Moreover, we constructed rarefaction curves to assess the depth of sampling required to encompass the full spectrum of viral diversity (Figure 1C). Notably, the curves (total and *Riboviria*) plateaued after approximately fifty pools, suggesting saturation and the absence of novel viral taxa beyond this point. These findings indicate that our sampling efforts comprehensively cover the viral diversity within the blackflies from the onchocerciasis-endemic Nachtigal village (Cameroon).



223

224 **Figure 1: Blackfly viromes are dominated by RNA viruses.** A) Barplots showing the relative
225 abundance of all identified viruses at contig and read level. Labels indicate the absolute
226 numbers. B) Barplots showing the relative abundance of all identified *Riboviria* phyla at
227 contig and read level. Labels indicate the absolute numbers. C) Rarefaction curves for each
228 virus kingdom and the total. Light shaded areas indicate the minimum and maximum
229 observed species across all iterations, the dark shaded areas indicate the 95% confidence
230 around the mean.

231 To visualize the virome similarity of the fifty-five pools, a heatmap with the read count for
232 each viral family on a $\log_2$ scale was constructed (see Figure 2). In total, forty eukaryotic viral
233 families were identified across all samples. In addition, eleven different taxonomic instances
234 of prokaryotic viruses were identified on the family level. Most diversity in the blackflies on
235 the family level is present in the *Riboviria* realm with thirty-five families of eukaryotic RNA
236 viruses detected.

237 The viral families *Flaviviridae, Hepeviridae, Orthomyxoviridae, Parvoviridae, Phenuiviridae,*
238 *Spinareoviridae, Rhabdoviridae*, and *Togaviridae* include human pathogenic viruses, some of
239 which may also be arboviruses. Notably, known viruses causing neurological complications
240 belong to the families of *Rhabdoviridae*, *Flaviviridae, Picornaviridae* as well as *Togaviridae*.

241 Fungal viruses were the most prevalent viruses, with members of *Ghabrivirales* (i.e.,
242 *Chrysoviridae*, *Totiviridae*) and *Partitiviridae* being the most abundant. A virus of the family
243 *Flaviviridae* was present in most blackfly pools and in a relatively high abundance.
244 Furthermore, other eukaryotic viral families (main hosts according to ICTV are given in
245 brackets) that were prevalent, were *Curvulaviridae* (fungi), *Dicistroviridae* (invertebrates),
246 *Iflaviridae* (insects), *Genomoviridae* (human, mammals, birds, fungi), *Mymonaviridae* (fungi),
247 *Orthomyxoviridae* (human, birds, pigs), *Parvoviridae* (vertebrates, insects), *Rhabdoviridae*
248 (vertebrates, invertebrates, plants), *Totiviridae* (fungi) and *Virgaviridae* (plants). From these
249 families we constructed phylogenetic trees, either grouped in their respective order or as a
250 separate family.

Figure 2: Blackflies harbour a large number of viral families, including families of potential clinical interest. Heatmap of all viral families per blackfly pool based on the blastx results (sequence similarity cutoff of at least 70% was used to assign scaffolds to families, otherwise they were assigned to closest related order). The heatmap shows the read count on a log2 scale for eukaryotic virus on family level from all the 55 adult blackfly pools. The number of identified scaffolds is indicated after each viral family name in the heatmap. Hierarchical clustering of the columns is based on the Bray-Curtis distance calculated from the read counts.

RNA virus phylogenetics

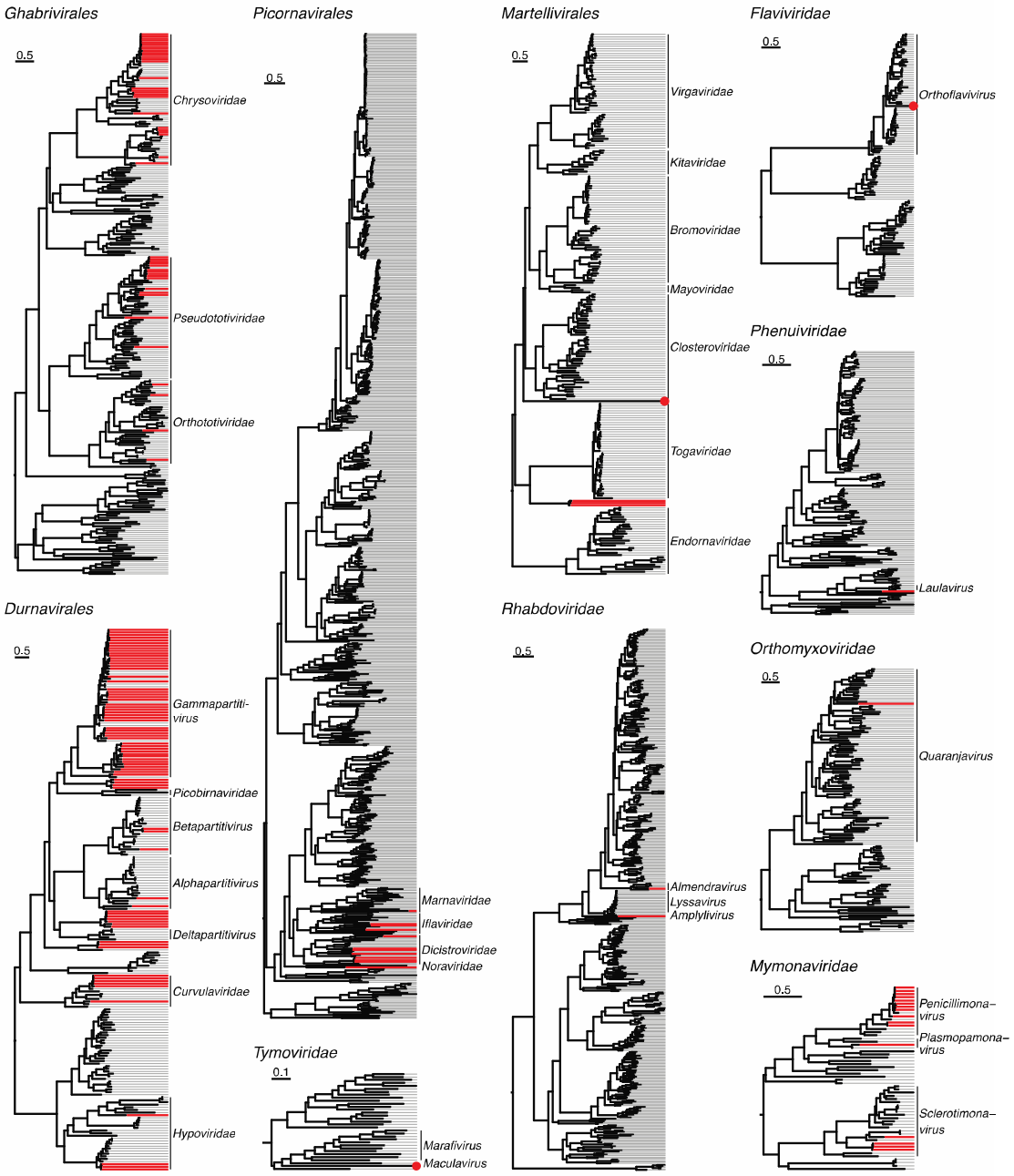
To fully display the diversity of RNA viruses in the blackflies, we made phylogenetic trees of the palmcore sequences of the RdRp protein for the most common and most interesting RNA viruses present. These palmcore sequences might not completely reflect the true phylogeny of the viruses in the tree, but they prove to be a robust way to taxonomically classify sequences as shown by the high monophylicity of established families and genera (see Supplemental Figures 4-13).

The majority of the viruses in the blackflies seem fungi-infecting as they are part of the *Ghabrivirales*, *Durnavirales* (dsRNA) and *Mymonaviridae* (ssRNA-) and cluster together with ICTV exemplar viruses that infect fungi (see Figure 3 and Supplemental Figures 4, 5 and 6, respectively). Within the dsRNA orders, the blackfly virus sequences often form deep branches in the tree suggestive the discovery of new species, genera or maybe even families, thereby expanding the known diversity of the *Partitiviridae* and *Curvulaviridae* in the *Durnavirales*, and the *Chrysoviridae* and *Pseudototiviridae* within the *Ghabrivirales*.

Viruses truly infecting the blackflies are mostly found in the *Picornavirales* order (ssRNA+), within the *Iflaviridae* and *Dicistroviridae* (see Figure 3 and Supplemental Figure 7). Furthermore, within the *Martellivirales*, which contains mostly plant-infecting viruses apart from the *Togaviridae*, we have possibly two new families. Interestingly, the viruses most closely related to the *Togaviridae* are probably segmented viruses as the recovered sequences only contain an RdRP gene. Attempts to retrieve more segments with a contig co-occurrence analysis (Batson *et al.* 2021) were unfortunately unsuccessful due to the low prevalence of these viruses in the blackfly samples. Also blastx searches against NCBI's nr database revealed only sequences with only an RdRP gene and without matching segment in NCBI's nt database coding for a structural protein/proteins.

Furthermore, we have a few singular viruses in the families *Tymoviridae*, *Flaviviridae*, *Phenuiviridae*, and *Orthomyxoviridae* of which the host is most likely the blackfly. In the *Rhabdoviridae*, we find two viruses: one that is clearly a member of the *Almendravirus* genus which contains invertebrate-infecting viruses and one virus that is not a part of a

288 known genus but is part of a large clade that contains viruses infecting vertebrates and
289 invertebrates as well as both (see Supplemental Figure 13).



290

291 **Figure 3: Phylogenetic trees representing the major RNA virus orders and families**
292 **identified in the blackflies.** Red lines signify novel viruses identified in the blackflies, while
293 grey lines are ICTV references. The scale bars indicate the number of amino acid

substitutions per site. Families and/or genera closest to the novel viruses are shown at the tips of the tree with a bar covering all ICTV exemplar species within that taxonomy (or no bar when there is only one exemplar species).

Novel blackfly rhabdo- and flaviviruses are most likely invertebrate-specific

To investigate whether the *Rhabdoviridae* and *Flaviviridae* member viruses we found in the blackflies could also infect vertebrates, we compared the dinucleotide composition within their genomes with the genomes of viruses with a known host. After coloring the viruses by host, this revealed a clear clustering in vertebrate-infecting vs. insect-specific viruses for the *Flaviviridae* (see Figure 4A), meaning that this family of viruses adapted itself to avoid the vertebrate immune system by using similar dinucleotide abundances in their genome (Lytras and Hughes 2020; Fros *et al.* 2021). For the *Rhabdoviridae* this clustering is less clear, but there seems to be nevertheless a small difference. The flavivirus we found in the blackflies shows a dinucleotide profile suggestive of being an insect-specific virus and also the rhabdoviruses we found in the blackflies seem to be likely infecting invertebrates when combining both the RDA information and the phylogenetic analysis (see Figure 4B and Supplemental Figure 13). Additionally, we added two other virus species recently identified in other studies to the *Rhabdoviridae* dinucleotide analysis: *Onchocerca volvulus* RNA virus 1 (OVRV1) is a rhabdovirus that was found in *Onchocerca volvulus* and elicits an immune response in the host (human as well as cattle and hamsters) (Quek *et al.* 2024), while Mundri virus was detected in a patient with nodding syndrome (Edridge *et al.* 2022). Interestingly, both multiple OVRV1 variants as well as Mundri virus seem to cluster with vertebrate infecting rhabdoviruses, suggesting that these viruses may be capable of infecting vertebrates.

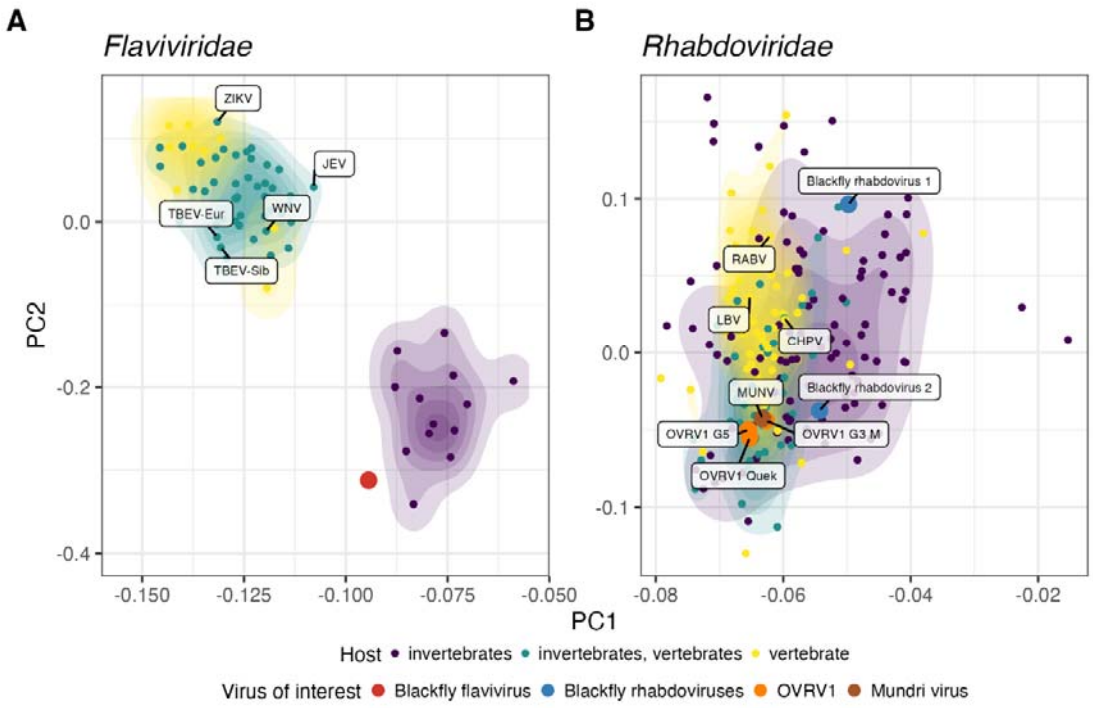


Figure 4: Identified blackfly flavi- and rhabdoviruses are most likely insect-specific. A) PCA plot based on the RDA of selected *Flaviviridae* sequences. The red dot shows the newly identified flavivirus in blackflies. Labels show a selection of neurotropic flaviviruses (TBEV=tick-borne encephalitis virus, JEV=Japanese encephalitis virus, WNV=West Nile virus, ZIKV=Zika virus). B) PCA plot based on the RDA of selected *Rhabdoviridae* sequences. The blue dots show the newly identified blackfly rhabdoviruses, the orange dots show different variants of OVRV1, the brown dot shows Mundri virus (MUNV). Other labelled dots show a selection of neurotropic rhabdoviruses (CHPV=Chandipura virus, LBV=Lagos bat virus, RABV=rabies virus).

Discussion

To our knowledge, this study represents the first comprehensive investigation of DNA and RNA viruses in blackflies utilizing a whole-genome sequencing approach. A very high number of virus genomes, with a dominance of RNA viruses, were identified in blackfly samples. Fifteen good quality (>50% complete) genomes of eukaryotic DNA viruses (eg. *Parvoviridae* and *Genomoviridae*), and 641 RNA viruses (based on the RdRP, of which 466 with a complete palmprint) were detected. A significant proportion of the novel viruses belonged

334 to viral families associated with human disease, eg. *Flaviviridae, Hepeviridae,*
335 *Orthomyxoviridae, Parvoviridae, Phenuiviridae, Picornaviridae, Spinareoviridae,*
336 *Rhabdoviridae*, and *Togaviridae*. However, no viral genomes closely related to known
337 neurotropic viruses were found in the blackfly virome (Supplementary files). Obviously, this
338 does not mean that none of the newly identified viruses could be arboviruses with a
339 neurotropic potential. Additionally, the real host origin for many of the newly discovered
340 viruses remains unknown and further investigations are necessary to determine whether
341 these viruses have the potential to replicate in vertebrate and/or invertebrate hosts, or
342 even if they have a prokaryotic host as could be the case for the *Partitiviridae* and
343 *Picobirnaviridae* (Neri *et al.* 2022; Sadiq, Holmes and Mahar 2024).

344 The only *Flaviviridae* virus in our analysis was observed at a relatively high abundance in
345 most blackfly samples (see Figure 2). Previous phylogenetic analyses and experimental
346 studies (Kuno and Chang 2005; Cook *et al.* 2013) have shown that the phylogenetic tree of
347 the *Flaviviridae* separates into two large clusters. The first cluster contains ISVs which
348 appear to be host-restricted and only replicate in invertebrate cells and primarily infect
349 mosquitoes (Hoshino *et al.* 2009; Guzman *et al.* 2018; Miranda *et al.* 2019; Baidaliuk *et al.*
350 2020). The second cluster comprises human pathogenic arboviruses. These previous studies
351 also indicate that arthropod-borne viruses of vertebrates may have originated from ISVs
352 that evolved from being limited to insects to acquiring the ability to infect vertebrate hosts
353 (Bolling *et al.* 2015). The dinucleotide analysis of the *Flaviviridae* indicated that the flavivirus
354 in our blackfly samples is most likely an insect-specific virus.

355 As blackflies can act as intermediate hosts of pathogens affecting humans and animal health
356 due to their hematophagic habit, several studies (Mead *et al.* 2004; Smith *et al.* 2009;
357 Mesquita *et al.* 2017; Young *et al.* 2021) have focused on the transmission of vesicular
358 stomatitis virus (VSV), an arbovirus belonging to the family *Rhabdoviridae*, genus
359 *Vesiculovirus*, which is known to be transmitted by blackflies. This virus primarily infects
360 livestock, representing an agriculturally relevant pathogen, but zoonotic events have also
361 been reported, highlighting its potential impact on human health (Lichty *et al.* 2004). VSV
362 infections are generally asymptomatic in humans, however, mild flu-like symptoms have

363 been reported in some individuals and a single case of encephalitis in a 3-year-old boy,
364 which was potentially associated with VSV infection, has been reported (Quiroz *et al.* 1988).

365 In 2022, Edridge and colleagues discovered a divergent rhabdovirus in the bloodstream of a
366 15-year-old girl with Nodding syndrome from Mundri West Country in South Sudan (Edridge
367 *et al.* 2022). They named this virus "Mundri virus" (MUNV) and classified it as a novel virus.
368 Subsequent phylogenetic analysis revealed that MUNV belongs to the monophyletic clade of
369 *tibroviruses* within the family *Rhabdoviridae*. The *Tibrovirus* genus is part of a larger
370 phylogenetic group of arthropod-borne rhabdoviruses. Several *tibroviruses* have been
371 isolated from biting midges (*Culicoides* spp.) and cattle, and some instances of human
372 detection have been reported (Cybinski *et al.* 1980; Cybinski and Gard 1986; Gibbs *et al.*
373 1989; Grard *et al.* 2012; Stremlau *et al.* 2015). Human tibrovirus infections may accidentally
374 occur via midge bites (Kuhn *et al.* 2020). Consistent with the findings of Edridge *et al.*, our
375 phylogenetic analysis of *Rhabdoviridae* (see Supplemental Figure 13) showed that MUNV is
376 closely related to the *Tibrovirus* genus, supported by a high bootstrap value. However, the
377 two sequences detected in the blackfly virome were not closely related to the *Tibrovirus* nor
378 *Vesiculovirus* genera. In addition, another novel rhabdovirus, named OVRV1, was discovered
379 by Quek *et al*. in public transcriptomes and viromes of *Onchocerca volvulus* nematodes
380 (Quek *et al.* 2024). Because OVRV1 was shown to elicit an immune response in the
381 vertebrate host of the nematode, it could potentially play a role in the pathogenesis of OAE
382 (Quek *et al.* 2024). We were, however, not able to find this virus back in the virome of the
383 blackflies. Our dinucleotide analysis for the *Rhabdoviridae*, suggested that the viruses we
384 found in the blackflies are invertebrate-specific viruses, in contrast to the previously
385 identified rhabdoviruses MUNV (Edridge *et al.* 2022) and OVRV1 (Quek *et al.* 2024).
386 However, a recent association between MUNV infection and nodding syndrome was
387 investigated in a case-control study in South Sudan. MUNV RNA was only detected in the
388 plasma of one of the 72 persons with Nodding syndrome and in none of the controls. MUNV
389 anti-nucleocapsid IgG antibodies were observed in 28% of Nodding syndrome cases, 22% of
390 household controls, and 16% of community controls without significant differences between
391 cases and both control groups. Therefore, it was concluded that MUNV may not be causally

392 associated with Nodding syndrome (Edridge *et al.* 2022). Similarly, it remains to be shown
393 whether OVRV1 may be associated with OAE.

394 Our study is the first to extensively study the DNA and RNA virome of blackflies from an
395 onchocerciasis-endemic area in Africa. Our findings reveal a relative high abundance of the
396 family *Genomoviridae* in the majority of blackfly samples, aligning with the results of
397 Kraberger *et al.*, which demonstrated that the majority of novel ssDNA viral genomes
398 detected in blackflies in New Zealand belonged to the *Genomoviridae* family (Kraberger *et*
399 *al.* 2019). However, in contrast with the blackflies in New Zealand (Kraberger *et al.* 2019),
400 our study identified *Parvoviridae*, in the blackfly virome. Although this viral family exhibited
401 a high abundance, it was only observed in a limited number of blackfly samples.
402 Phylogenetic analysis revealed that one of the *Parvoviridae* genomes detected in this study
403 clustered within the *Densovirinae* subfamily (see Supplemental Figure 15).

404 Most of the detected viruses exhibited low sequence similarities with their closest relatives.
405 Consequently, based on the demarcation criteria established for viral classification, it is
406 reasonable to propose that the majority of sequences identified in our study represent new
407 virus species, genera, and/or families, which warrant further studies. In our project, plant
408 and fungal infecting viruses (mainly members of the *Chrysoviridae* and *Partitiviridae*
409 families) were the most abundant, possibly from the nectar diet of the blackflies. Some
410 members of viral families of clinical interest were detected, albeit at lower abundances and
411 no viral genomes closely related to neurotropic viruses were identified.

412 Our study has some limitations. First of all, it reports the composition of the blackfly virome,
413 including both DNA and RNA viruses but without metadata that would have allowed further
414 associations to be made with *O. volvulus* and potential environmental factors. Further
415 research should compare the virome of blackflies with and without *O. volvulus* infection,
416 and the sex and species of the blackflies should be taken into consideration. It is important
417 to note that only a few blackfly species consume a blood meal (Kraberger *et al.* 2019).
418 Additionally, as less than 1% of blackflies in Cameroon are currently infected with *O.*
419 *volvulus* (Shintouo *et al.* 2020), it is recommended to screen blackflies for *O. volvulus*
420 infection before sequencing data analyses, as previously suggested (Ekangouo *et al.* 2022).

Conclusion

Overall, we provide a glimpse into the viral diversity associated with blackfly vectors in Cameroon. Although no genomes associated with neurotropic viruses were found, a plethora of new virus species, genera, and families were identified. The presence of dominant and under-represented viruses warrants future research to determine the role of these viruses in their hosts and ecosystems. Exploring the diversity of viruses in blackflies should be included in the active surveillance of zoonotic diseases. Our findings constitute a starting point for investigating the viruses associated with the hematophagous blackfly and potentially in their nematode host *O. volvulus*.

Funding

LDC, RC and JNSF were supported by the Research Foundation – Flanders (FWO, grant numbers 11L1323N, G0A0522N and 1296723N, respectively). In addition, RC received funding from the European Research Council (grant 671055) and blackflies were captured in Nachtigal as part of an EDCTP2 project supported by the European Union (grant number TMA2020CDF-3152-SCONE).

Data availability

Sequencing data are available in the SRA database under BioProject accession number PRJNA1088476. Identified viral genomes can be found in following Zenodo repository with a corresponding taxonomy table: https://doi.org/10.5281/zenodo.13737161. The code used for this study is available in a GitHub repository: https://github.com/Matthijnssenslab/2024_XXX_Blackfly_Virome.

Acknowledgements

We thank Jill Swinnen (KU Leuven) and Jara Wagemans (KU Leuven) for their assistance during the processing of the blackfly pools for Illumina sequencing.

445 # References

446 Babaian A, Edgar R. Ribovirus classification by a polymerase barcode sequence. *PeerJ*
447 2022;**10**:e14055.

448 Baidaliuk A, Lequime S, Moltini-Conclois I *et al.* Novel genome sequences of cell-fusing agent
449 virus allow comparison of virus phylogeny with the genetic structure of Aedes aegypti
450 populations. *Virus Evol* 2020;**6**:1.

451 Batson J, Dudas G, Haas-Stapleton E *et al.* Single mosquito metatranscriptomics identifies
452 vectors, emerging pathogens and reservoirs in one assay. *Elife* 2021;**10**, DOI:
453 10.7554/ELIFE.68353.

454 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
455 *Bioinformatics* 2014;**30**:2114–20.

456 Bolling BG, Weaver SC, Tesh RB *et al.* Insect-specific virus discovery: Significance for the
457 arbovirus community. *In Viruses (Vol* 2015;**7**:9.

458 Boullé C, Njamnshi AK, Dema F *et al.* Impact of 19 years of mass drug administration with
459 ivermectin on epilepsy burden in a hyperendemic onchocerciasis area in Cameroon.
460 *Parasit Vectors* 2019;**12**, DOI: 10.1186/S13071-019-3345-7.

461 Brattig NW, Cheke RA, Garms R. Onchocerciasis (river blindness) – more than a century of
462 research and control. *Acta Trop* 2021;**218**:7.

463 Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature*
464 *Methods 2014 12:1* 2014;**12**:59–60.

465 Camargo AP, Roux S, Schulz F *et al.* Identification of mobile genetic elements with geNomad.
466 *Nature Biotechnology 2023* 2023:1–10.

467 Chesnais CB, Bizet C, Campillo JT *et al.* A Second Population-Based Cohort Study in
468 Cameroon Confirms the Temporal Relationship Between Onchocerciasis and Epilepsy.
469 *Open Forum Infect Dis* 2020;**7**:6.

470 Chesnais CB, Nana-Djeunga HC, Njamnshi AK *et al.* The temporal relationship between
471 onchocerciasis and epilepsy: a population-based cohort study. *Lancet Infect Dis*
472 2018;**18**:1278–86.

473 Colebunders R, Abd-Elfarag G, Carter JY *et al.* Clinical characteristics of onchocerciasis-
474 associated epilepsy in villages in Maridi County, Republic of South Sudan. *Seizure*
475 2018a;**62**:108–15.

476 Colebunders R, Hendy A, Nanyunja M *et al.* Nodding syndrome-a new hypothesis and new
477 direction for research. *International Journal of Infectious Diseases* 2014;**27**:74–7.

478 Colebunders R, Njamnshi AK, Menon S *et al.* Onchocerca volvulus and epilepsy: A
479 comprehensive review using the Bradford Hill criteria for causation. *PLoS Negl Trop Dis*
480 2021;**15**:1.

481 Colebunders R, Siewe Fodjo JN, Hopkins A *et al.* From river blindness to river epilepsy:
482 Implications for onchocerciasis elimination programmes. *PLoS Negl Trop Dis* 2019;**13**:7.

483 Colebunders R, Y. Carter J, Olore PC *et al.* High prevalence of onchocerciasis-associated
484 epilepsy in villages in Maridi County, Republic of South Sudan: A community-based
485 survey. *Seizure* 2018b;**63**:93.

486 Conceição-Neto N, Zeller M, Lefrère H *et al.* Modular approach to customise sample
487 preparation procedures for viral metagenomics: A reproducible protocol for virome
488 analysis. *Sci Rep* 2015;**5**, DOI: 10.1038/SREP16532.

489 De Coninck L. ViPER. 2021, DOI: 10.5281/zenodo.5502204.

490 Cook S, Chung BYW, Bass D *et al.* Novel virus discovery and genome reconstruction from
491 field rna samples reveals highly divergent viruses in dipteran hosts. *PLoS One*
492 2013;**8**:11.

493 Van Cutsem G, Siewe Fodjo JN, Dekker MCJ *et al.* Case definitions for onchocerciasis-
494 associated epilepsy and nodding syndrome: A focused review. *Seizure* 2023;**107**:132–5.

495 Cybinski DH, Gard GP. Isolation of a new rhabdovirus in Australia related to Tibrogargan
496 virus. *Aust J Biol Sci* 1986;**39**:225–32.

497 Cybinski DH, St. George TD, Standfast HA *et al.* Isolation of tibrogargan virus, a new
498 Australian rhabdovirus, from Culicoides brevitarsis. *Vet Microbiol* 1980;**5**:301–8.

499 Davis NM, Proctor DiM, Holmes SP *et al.* Simple statistical identification and removal of
500 contaminant sequences in marker-gene and metagenomics data. *Microbiome*
501 2018;**6**:1–14.

502 Dudas G, Batson J. Accumulated metagenomic studies reveal recent migration, whole
503 genome evolution, and undiscovered diversity of orthomyxoviruses. *J Virol* 2023;**97**,
504 DOI: 10.1128/JVI.01056-23/SUPPL_FILE/JVI.01056-23-S0001.PDF.

505 Duke BO, Vincelette J, Moore PJ. Microfilariae in the cerebrospinal fluid, and neurological
506 complications, during treatment of onchocerciasis with diethylcarbamazine.
507 *Tropenmed Parasitol* 1976;**27**:123–32.

508 Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol* 2011;**7**, DOI:
509 10.1371/JOURNAL.PCBI.1002195.

510 Edgar RC. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of
511 sequence homology and phylogeny. *Nature Communications 2022 13:1* 2022;**13**:1–9.

512 Edridge AWD, Abd-Elfarag G, Deijs M *et al.* Divergent Rhabdovirus Discovered in a Patient
513 with New-Onset Nodding Syndrome. *Viruses* 2022;**14**:210.

514 Ekangouo AE, Djeunga HCN, Sempere G *et al.* Bacteriome Diversity of Blackflies' Gut and
515 Association with Onchocerca volvulus, the Causative Agent of Onchocerciasis in Mbam
516 Valley (Center Region, Cameroon). *Pathogens* 2022;**11**, DOI:
517 10.3390/PATHOGENS11010044.

518 Föger K, Gora-Stahlberg G, Sejvar J *et al.* Nakalanga Syndrome: Clinical Characteristics,
519 Potential Causes, and Its Relationship with Recently Described Nodding Syndrome.
520 *PLoS Negl Trop Dis* 2017;**11**, DOI: 10.1371/JOURNAL.PNTD.0005201.

521 Fros JJ, Visser I, Tang B *et al.* The dinucleotide composition of the Zika virus genome is
522 shaped by conflicting evolutionary pressures in mammalian hosts and mosquito
523 vectors. *PLoS Biol* 2021;**19**, DOI: 10.1371/JOURNAL.PBIO.3001201.

524 Gibbs EPJ, Calisher CH, Tesh RB *et al.* Bivens arm virus: a new rhabdovirus isolated from
525 Culicoides insignis in Florida and related to Tibrogargan virus of Australia. *Vet Microbiol*
526 1989;**19**:141–50.

527 Grard G, Fair JN, Lee D *et al.* A novel rhabdovirus associated with acute hemorrhagic fever in
528 central Africa. *PLoS Pathog* 2012;**8**, DOI: 10.1371/JOURNAL.PPAT.1002924.

529 Guzman H, Contreras-Gutierrez MA, Travassos da Rosa APA *et al.* Characterization of Three
530 New Insect-Specific Flaviviruses: Their Relationship to the Mosquito-Borne Flavivirus
531 Pathogens. *Am J Trop Med Hyg* 2018;**98**:410.

532 Hadermann A, Amaral LJ, Van Cutsem G *et al.* Onchocerciasis-associated epilepsy: an update
533 and future perspectives. *Trends Parasitol* 2023;**39**:126–38.

534 Hendy A, Krit M, Pfarr K *et al.* Onchocerca volvulus transmission in the Mbam valley of
535 Cameroon following 16 years of annual community-directed treatment with
536 ivermectin, and the description of a new cytotype of Simulium squamosum. *Parasit*
537 *Vectors* 2021;**14**:563.

538 Hoshino K, Isawa H, Tsuda Y *et al.* Isolation and characterization of a new insect flavivirus
539 from Aedes albopictus and Aedes flavopictus mosquitoes in Japan. *Virology*
540 2009;**391**:1.

541 Hotterbeekx A, Lammens M, Idro R *et al.* Neuroinflammation and not tauopathy is a
542 predominant pathological signature of nodding syndrome. *J Neuropathol Exp Neurol*
543 2019;**78**:1049–58.

544 Hotterbeekx A, Raimon S, Abd-Elfarag G *et al.* Onchocerca volvulus is not detected in the
545 cerebrospinal fluid of persons with onchocerciasis-associated epilepsy. *International*
546 *Journal of Infectious Diseases* 2020;**91**:119–23.

547 Howerth EW, Mead DG, Stallknecht DE. Immunolocalization of vesicular stomatitis virus in
548 black flies (Simulium vittatum). *Ann N Y Acad Sci* 2002;**969**:340.

549 Hyatt D, Chen GL, LoCascio PF *et al.* Prodigal: Prokaryotic gene recognition and translation
550 initiation site identification. *BMC Bioinformatics* 2010;**11**:1–11.

551 Kraberger S, Schmidlin K, Fontenele RS *et al.* Unravelling the single-stranded DNA virome of
552 the New Zealand blackfly. *Viruses* 2019;**11**, DOI: 10.3390/V11060532.

553 Kuhn JH, Pan H, Chiu CY *et al.* Human Tibroviruses: Commensals or Lethal Pathogens?
554 *Viruses* 2020;**12**:3.

555 Kuno G, Chang GJJ. Biological transmission of arboviruses: Reexamination of and new
556 insights into components, mechanisms, and unique traits as well as their evolutionary
557 trends. *In Clinical Microbiology Reviews (Vol* 2005;**18**:4.

558 Levick B, Laudisoit A, Tepage F *et al.* High prevalence of epilepsy in onchocerciasis endemic
559 regions in the Democratic Republic of the Congo. *PLoS Negl Trop Dis*
560 2017;**11**:e0005732.

561 Lichty BD, Power AT, Stojdl DF *et al.* Vesicular stomatitis virus: Re-inventing the bullet.
562 *Trends Mol Med* 2004;**10**:5.

563 Lytras S, Hughes J. Synonymous Dinucleotide Usage: A Codon-Aware Metric for Quantifying
564 Dinucleotide Representation in Viruses. *Viruses 2020, Vol 12, Page 462* 2020;**12**:462.

565 Mead DG, Howerth EW, Murphy MD *et al.* Black fly involvement in the epidemic
566 transmission of Vesicular stomatitis New Jersey virus (Rhabdoviridae: Vesiculovirus).
567 *Vector-Borne and Zoonotic Diseases* 2004;**4**:351–9.

568 Mesquita LP, Diaz MH, Howerth EW *et al.* Pathogenesis of Vesicular Stomatitis New Jersey
569 Virus Infection in Deer Mice (Peromyscus maniculatus) Transmitted by Black Flies (
570 Simulium vittatum). *Vet Pathol* 2017;**54**:1.

571 Miranda J, Mattar S, Gonzalez M *et al.* First report of Culex flavivirus infection from Culex
572 coronator (Diptera: Culicidae), Colombia. *Virol J* 2019;**16**, DOI: 10.1186/S12985-018-
573 1108-2.

574 Mukendi D, Tepage F, Akonda I *et al.* High prevalence of epilepsy in an onchocerciasis
575 endemic health zone in the Democratic Republic of the Congo, despite 14 years of
576 community-directed treatment with ivermectin: A mixed-method assessment. *Int J*
577 *Infect Dis* 2019;**79**:187–94.

578 Nayfach S, Camargo AP, Schulz F *et al.* CheckV assesses the quality and completeness of
579 metagenome-assembled viral genomes. *Nature Biotechnology 2020 39:5* 2020;**39**:578–
580 85.

581 Neri U, Wolf YI, Roux S *et al.* Expansion of the global RNA virome reveals diverse clades of
582 bacteriophages. *Cell* 2022;**185**:4023-4037.e18.

583 Nurk S, Meleshko D, Korobeynikov A *et al.* MetaSPAdes: A new versatile metagenomic
584 assembler. *Genome Res* 2017;**27**:824–34.

585 Piégu B, Guizard S, Spears T *et al.* Complete genome sequence of invertebrate iridescent
586 virus 22 isolated from a blackfly larva. *Journal of General Virology* 2013;**94**:9.

587 Piégu B, Guizard S, Spears T *et al.* Complete genome sequence of invertebrate iridovirus IIV-
588 25 isolated from a blackfly larva. *Arch Virol* 2014a;**159**:5.

589 Piégu B, Guizard S, Yeping T *et al.* Complete genome sequence of invertebrate iridovirus
590 IIV22A, a variant of IIV22, isolated originally from a blackfly larva. *Stand Genomic Sci*
591 2014b;**9**:3.

592 Pion SDS, Boussinesq M. Significant Association between Epilepsy and Presence of
593 Onchocercal Nodules: Case-Control Study in Cameroon. *Am J Trop Med Hyg*
594 2012;**86**:557.

595 Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for
596 Large Alignments. *PLoS One* 2010;**5**, DOI: 10.1371/JOURNAL.PONE.0009490.

597 Quek S, Hadermann A, Wu Y *et al.* Diverse RNA viruses of parasitic nematodes can elicit
598 antibody responses in vertebrate hosts. *Nature Microbiology 2024* 2024:1–18.

599 Quiroz E, Moreno N, Peralta PH *et al.* A human case of encephalitis associated with vesicular
600 stomatitis virus (Indiana serotype) infection. *Am J Trop Med Hyg* 1988;**39**:3.

601 Raimon S, Dusabimana A, Abd-Elfarag G *et al.* High Prevalence of Epilepsy in an
602 Onchocerciasis-Endemic Area in Mvolo County, South Sudan: A Door-To-Door Survey.
603 *Pathogens* 2021;**10**, DOI: 10.3390/PATHOGENS10050599.

604 Sadiq S, Holmes EC, Mahar JE. Genomic and phylogenetic features of the Picobirnaviridae
605 suggest microbial rather than animal hosts. *bioRxiv* 2024:2024.02.04.578841.

606 Sakaguchi S, Urayama SI, Takaki Y *et al.* NeoRdRp: A Comprehensive Dataset for Identifying
607 RNA-dependent RNA Polymerases of Various RNA Viruses from Metatranscriptomic
608 Data. *Microbes Environ* 2022;**37**, DOI: 10.1264/JSME2.ME22001.

609 Shintouo CM, Nguve JE, Asa FB *et al.* Entomological assessment of onchocerca species
610 transmission by black flies in selected communities in the west region of cameroon.
611 *Pathogens* 2020;**9**:9.

612 Siewe Fodjo JN, Mandro M, Mukendi D *et al.* Onchocerciasis-associated epilepsy in the
613 Democratic Republic of Congo: Clinical description and relationship with microfilarial
614 density. *PLoS Negl Trop Dis* 2019;**13**, DOI: 10.1371/JOURNAL.PNTD.0007300.

615 Siewe Fodjo JN, Ngarka L, Njamnshi WY *et al.* Onchocerciasis in the Ntui Health District of
616 Cameroon: epidemiological, entomological and parasitological findings in relation to
617 elimination prospects. *Parasit Vectors* 2022;**15**, DOI: 10.1186/S13071-022-05585-0.

618 Siewe Fodjo JN, Tatah G, Tabah EN *et al.* Epidemiology of onchocerciasis-associated epilepsy
619 in the Mbam and Sanaga river valleys of Cameroon: Impact of more than 13 years of
620 ivermectin. *Infect Dis Poverty* 2018;**7**:1.

621 Smith PF, Howerth EW, Carter D *et al.* Mechanical transmission of vesicular stomatitis New
622 Jersey virus by Simulium vittatum (Diptera: Simuliidae) to domestic swine (Sus scrofa). *J*
623 *Med Entomol* 2009;**46**:6.

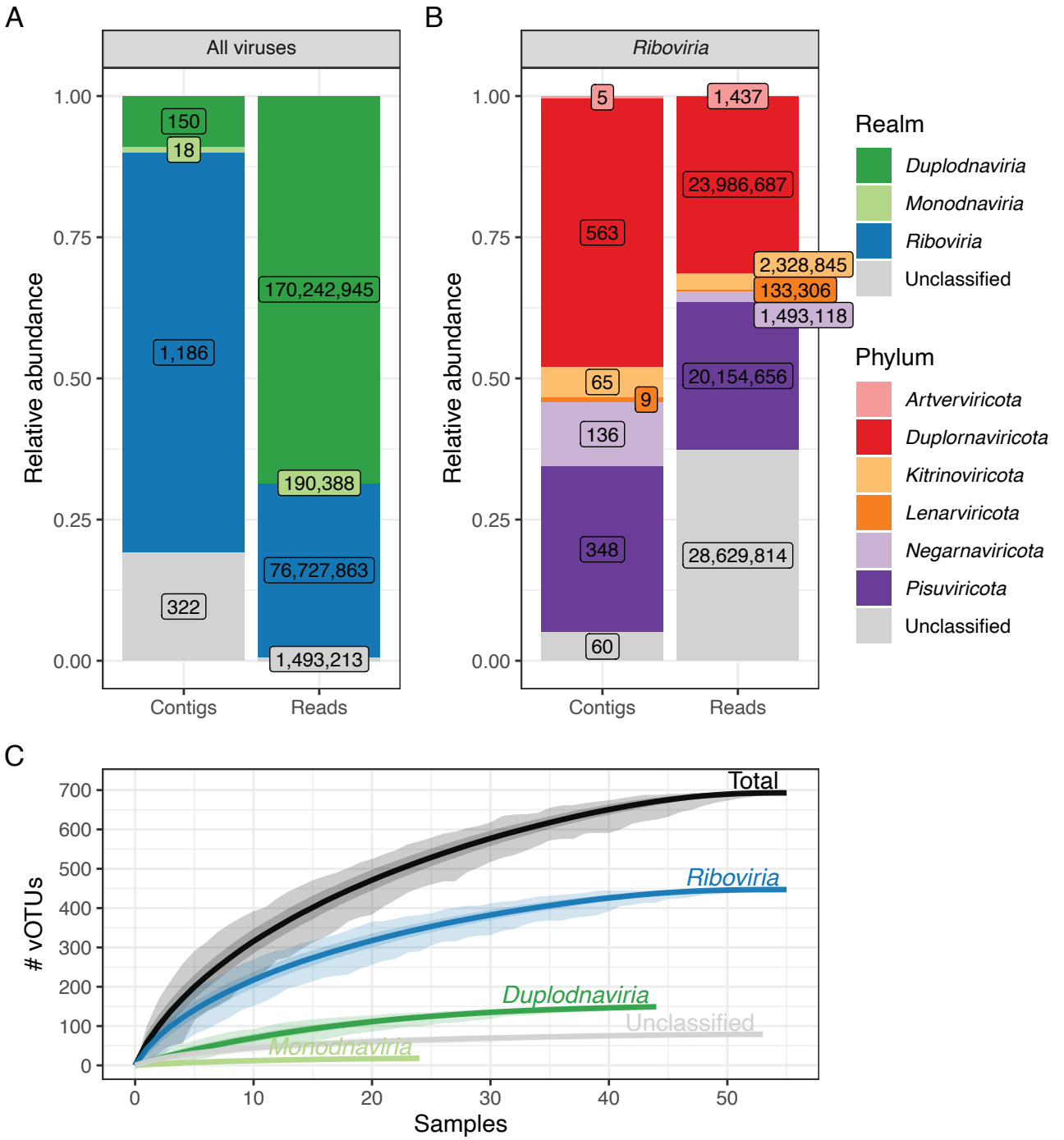
624 Stremlau MH, Andersen KG, Folarin OA *et al.* Discovery of novel rhabdoviruses in the blood
625 of healthy individuals from West Africa. *PLoS Negl Trop Dis* 2015;**9**, DOI:
626 10.1371/JOURNAL.PNTD.0003631.

627 Winkler AS, Friedrich K, Velicheti S *et al.* MRI findings in people with epilepsy and nodding
628 syndrome in an area endemic for onchocerciasis: An observational study. *Afr Health Sci*
629 2013;**13**:529–40.

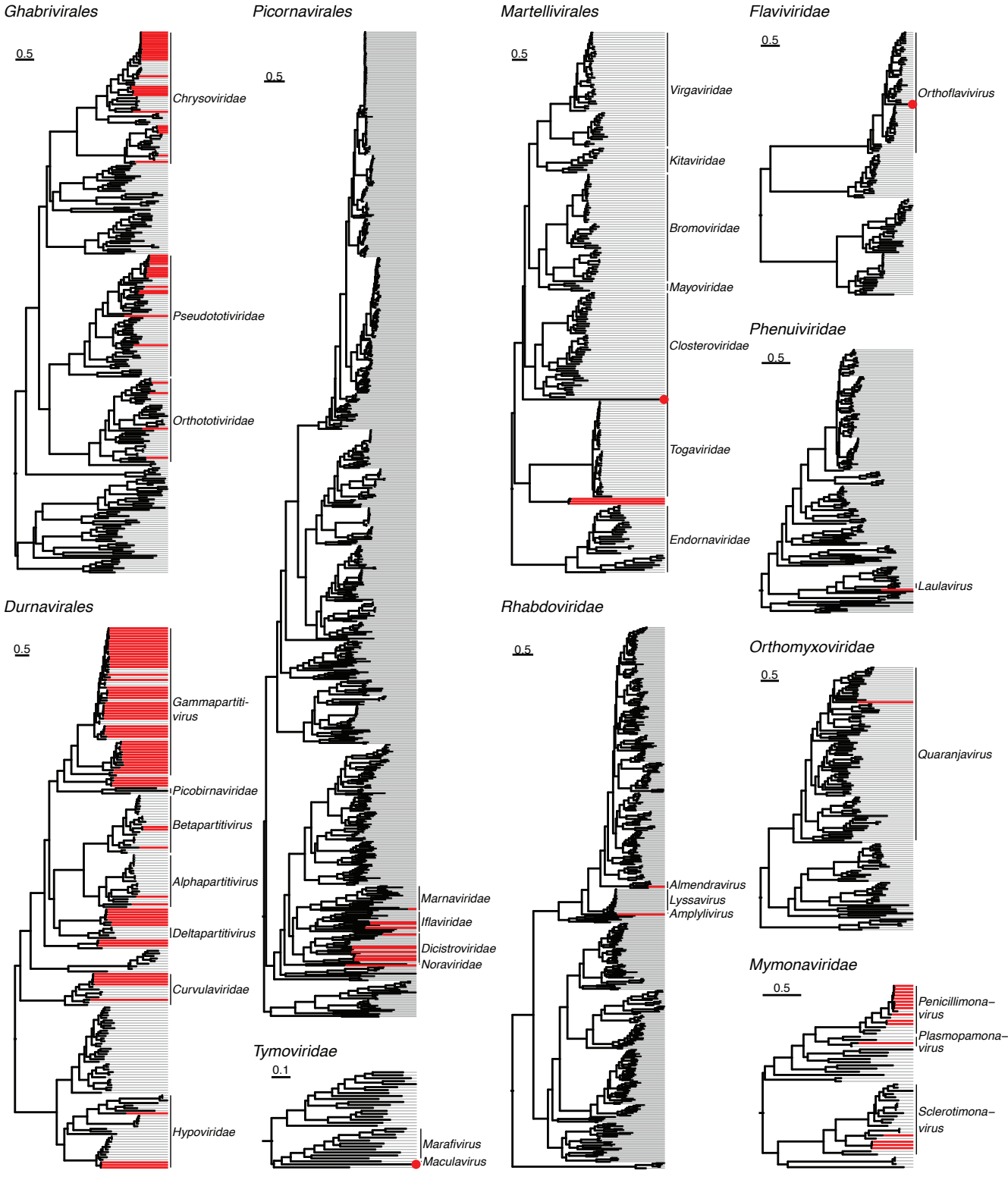
630 Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*
631 2019;**20**:1–13.

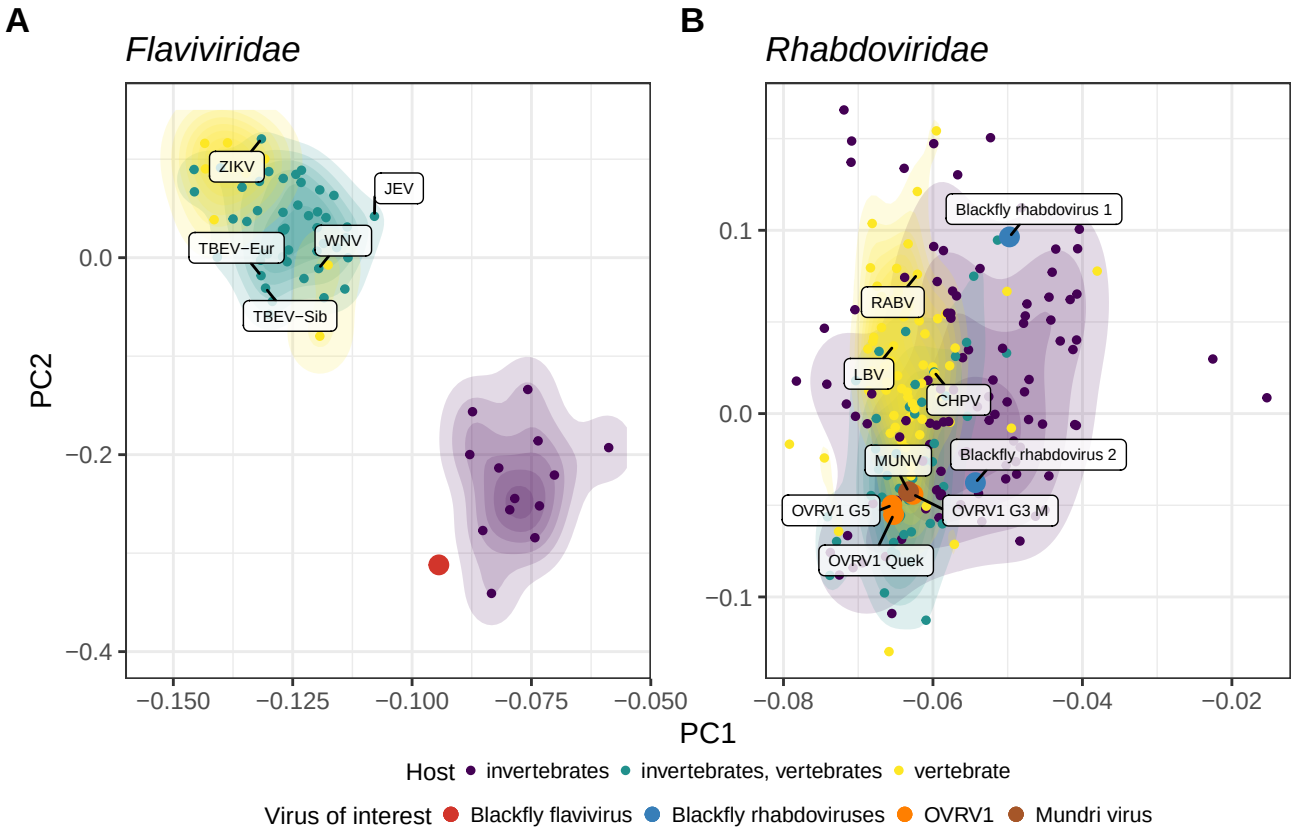
632 Young KI, Valdez F, Vaquera C *et al.* Surveillance along the Rio Grande during the 2020
633 Vesicular Stomatitis Outbreak Reveals Spatio-Temporal Dynamics of and Viral RNA
634 Detection in Black Flies. *Pathogens* 2021;**10**:10.

635


A
All viruses
Relative abundance
1.00
0.75
0.50
0.25
0.00
150
18
1,186
322
170,242,945
190,388
76,727,863
1,493,213
Contigs
Reads
B
Riboviria
5
563
65
9
136
348
60
1,437
23,986,687
2,328,845
133,306
1,493,118
20,154,656
28,629,814
Realm
Duplodnaviria
Monodnaviria
Riboviria
Unclassified
Phylum
Artverviricota
Duplornaviricota
Kitrinoviricota
Lenarviricota
Negarnaviricota
Pisuviricota
Unclassified
C
vOTUs
700
600
500
400
300
200
100
0
Total
Riboviria
Duplodnaviria
Unclassified
Monodnaviria
0
10
20
30
40
50
Samples

Kingdom
Phylum
Class
Order
Duplodnaviria
Monodnaviria
Riboviria
unclassified
Autographiviridae (n=7)
Casjensviridae (n=1)
Schitoviridae (n=1)
Straboviridae (n=1)
unclassified Caudoviricetes (n=24)
unclassified Uroviricota (n=116)
unclassified Hofneiviricota (n=1)
unclassified Zurhausenvirales (n=1)
Parvoviridae (n=2)
unclassified Piccovirales (n=3)
unclassified Cossaviricota (n=1)
Genomoviridae (n=8)
unclassified Geplafuvirales (n=2)
Alternaviridae (n=5)
Chrysoviridae (n=75)
Totiviridae (n=108)
unclassified Ghabrivirales (n=351)
Spinareoviridae (n=1)
unclassified Reovirales (n=21)
Cystoviridae (n=1)
unclassified Mindivirales (n=1)
Hepeviridae (n=1)
unclassified Hepelivirales (n=2)
Togaviridae (n=1)
Virgaviridae (n=4)
unclassified Martellivirales (n=37)
Gammaflexiviridae (n=4)
Tymoviridae (n=1)
unclassified Tymovirales (n=8)
Flaviviridae (n=1)
unclassified Nodamuvirales (n=2)
Tombusviridae (n=2)
unclassified Tolivirales (n=1)
unclassified Kitrinoviricota (n=1)
Narnaviridae (n=1)
unclassified Wolframvirales (n=3)
unclassified Norzivirales (n=3)
unclassified Leviviricetes (n=1)
Botourmiaviridae (n=1)
Discoviridae (n=10)
Peribunyaviridae (n=1)
unclassified Bunyavirales (n=36)
unclassified Articulavirales (n=2)
Mymonaviridae (n=22)
Rhabdoviridae (n=1)
unclassified Mononegavirales (n=23)
unclassified Monjiviricetes (n=2)
unclassified Goujianvirales (n=2)
unclassified Negarnaviricota (n=37)
Curvulaviridae (n=5)
Fusariviridae (n=2)
Hypoviridae (n=7)
Partitiviridae (n=138)
unclassified Durnavirales (n=81)
Dicistroviridae (n=8)
Iflaviridae (n=1)
unclassified Picornavirales (n=44)
unclassified Patatavirales (n=17)
unclassified Pisuviricota (n=45)
unclassified Orthornavirae (n=12)
Caulimoviridae (n=2)
unclassified Ortervirales (n=1)
unclassified Artverviricota (n=1)
Polymycoviridae (n=16)
unclassified Riboviria (n=32)
Hytrosaviridae (n=3)
unclassified Lefavirales (n=95)
Anelloviridae (n=9)
unclassified (n=215)
Log2 Read counts
0 10 20 30
Virus type
Eukaryotic virus
Bacteriophage
unknown
Kingdom
Heunggongvirae
Loebvirae
Orthornavirae
Pararnavirae
Shotokuvirae
unclassified
Phylum
Artverviricota
Cossaviricota
Cossaviricota
Cressdnaviricota
Cressdnaviricota
Duplornaviricota
Hofneiviricota
Kitrinoviricota
Lenarviricota
Negarnaviricota
Pisuviricota
Uroviricota
unclassified
Class
Alsuviricetes
Amabiliviricetes
Caudoviricetes
Chrymotiviricetes
Duplopiviricetes
Ellioviricetes
Flasuviricetes
Insthoviricetes
Leviviricetes
Magsaviricetes
Miaviricetes
Monjiviricetes
Naldaviricetes
Papovaviricetes
Pisoniviricetes
Quintoviricetes
Repensiviricetes
Resentoviricetes
Revtraviricetes
Stelpaviricetes
Tolucaviricetes
Vidaverviricetes
Yunchangviricetes
unclassified
Order
Amarillovirales
Articulavirales
Bunyavirales
Durnavirales
Geplafuvirales
Ghabrivirales
Goujianvirales
Hepelivirales
Lefavirales
Martellivirales
Mindivirales
Mononegavirales
Nodamuvirales
Norzivirales
Ortervirales
Ourlivirales
Patatavirales
Piccovirales
Picornavirales
Reovirales
Tolivirales
Tymovirales
Wolframvirales
Zurhausenvirales
unclassified


Ghabrivirales
0.5
Chrysoviridae
Pseudototiviridae
Orthototiviridae
Picornavirales
0.5
Marnaviridae
Iflaviridae
Dicistroviridae
Noraviridae
Tymoviridae
0.1
Marafivirus
Maculavirus
Martellivirales
0.5
Virgaviridae
Kitaviridae
Bromoviridae
Mayoviridae
Closteroviridae
Togaviridae
Endornaviridae
Flaviviridae
0.5
Orthoflavivirus
Phenuiviridae
0.5
Laulavirus
Durnavirales
0.5
Gammapartiti-
virus
Picobirnaviridae
Betapartitivirus
Alphapartitivirus
Deltapartitivirus
Curvulaviridae
Hypoviridae
Rhabdoviridae
0.5
Almendravirus
Lyssavirus
Amplylivirus
Orthomyxoviridae
0.5
Quaranjavirus
Mymonaviridae
0.5
Penicillimona–
virus
Plasmopamona–
virus
Sclerotimona–
virus


A
Flaviviridae
ZIKV
JEV
TBEV−Eur
WNV
TBEV−Sib
0.0
−0.2
−0.4
PC2
−0.150
−0.125
−0.100
−0.075
−0.050
B
Rhabdoviridae
Blackfly rhabdovirus 1
RABV
LBV
CHPV
MUNV
Blackfly rhabdovirus 2
OVRV1 G5
OVRV1 G3 M
OVRV1 Quek
0.1
0.0
−0.1
−0.08
−0.06
−0.04
−0.02
PC1
Host
invertebrates
invertebrates, vertebrates
vertebrate
Virus of interest
Blackfly flavivirus
Blackfly rhabdoviruses
OVRV1
Mundri virus