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

Lander De Coninck[1], Amber Hadermann[2], Ludovica Ingletto[1], Robert Colebunders[2], Kongnyu Gamnsi Njamnshi[3], Alfred Kongnyu Njamnshi[3], John L Mokili[4], 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. Brain Research Africa Initiative (BRAIN), Yaoundé, Cameroon & Geneva, Switzerland; Neuroscience Lab, Faculty of Medicine and Biomedical Sciences, The University of Yaoundé I, Yaoundé, Cameroon.
4. 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 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. A total of fifty-five pools of ten blackflies were analysed (see Supplementary Table S1 for details).

Sample processing

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. 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 (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 (Camargo *et al.* 2023); 2) DIAMOND blastx (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 HH-suite3 (Steinegger *et al.* 2019). 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 (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 (Davis et al. 2018) to be a contaminant were removed from our dataset.

Phylogenetic Analysis of the *Orthornavirae*

In ORFs identified to code for an RNA-dependent RNA polymerase (RdRP) (*vide supra*), we searched for the palmcore sequence using Palmscan2 (https://github.com/rcedgar/palmscan). This palmcore sequence contains the palmprint, three conserved catalytic domains across all known RdRP sequences (domain A, B and C) with two variable regions in between, flanked by ~150 amino acids on both sides (Babaian and Edgar 2022). Furthermore, we added 76,164 palmcore sequences from the RCR90 set of Neri *et al.* (Neri *et al.* 2022). This set contained previously published *Orthornavirae* sequences which were the basis of ICTV's *Orthornavirae* classification (Wolf *et al.* 2018, 2020), as well as new RNA viruses and ten reverse transcriptase (RT) sequences to be used as outgroup for the rooting of the phylogenetic tree. Next, palmcore sequences were aligned a first time with Muscle5 (Edgar 2022) with default settings. After alignment, columns with >90% gaps were removed with trimal (Capella-Gutiérrez, Silla-Martínez and Gabaldón 2009), and gaps were deleted from the sequences to prepare the data for a second alignment. From these trimmed sequences a diversified ensemble was created with Muscle5 (Edgar 2022). This diversified ensemble constitutes 100 multiple sequence alignments (MSAs) of the trimmed 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 -fastest -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.

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.

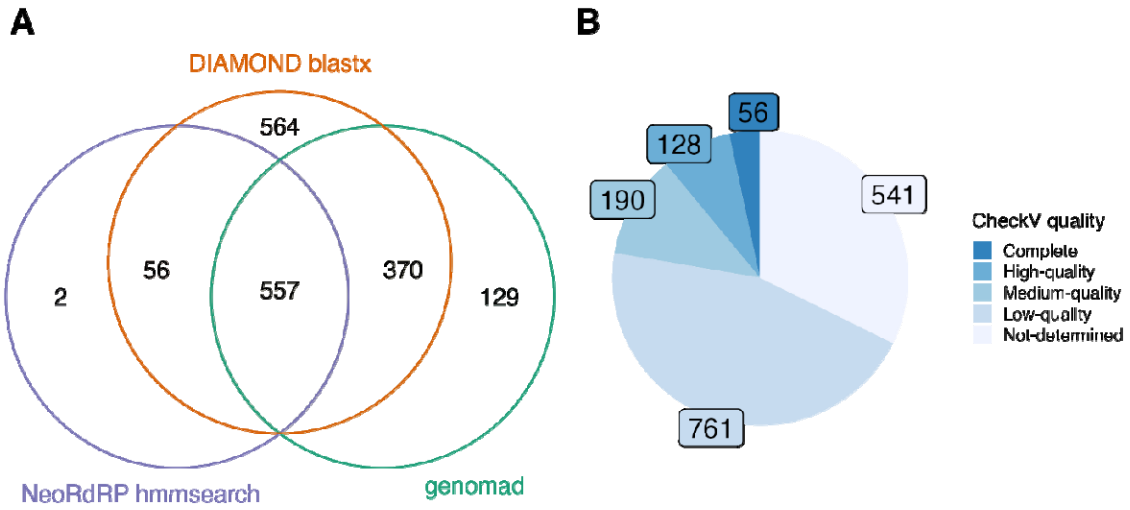
Results

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. 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). 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 (see Supplementary Table S1). After removing contamination predicted with decontam, a total of

175 1,678 scaffolds were identified as viral with a combination of genomad, DIAMOND blastx
176 and an HMM search against the NeoRdRP dataset (see Figure 1A). Of these, 56 were
177 considered to be 100% complete, 128 were of high quality (>90% complete) and 190 of
178 medium quality (>50% complete), the remaining viral sequences were either of low quality
179 or their completeness could not be determined by CheckV (see Figure 1B). However, the
180 low-quality and not-determined groups consisted only of sequences with *Riboviria* or
181 unclassified taxonomic realm assignments. This suggests that these sequences are
182 significantly divergent and might thus be of a higher quality than what is estimated by
183 CheckV (also see Materials). Overall, we concluded that, due to the high convergence of our
184 three identification methods and the completeness estimation of CheckV, we have a robust
185 set of identified viruses.

186

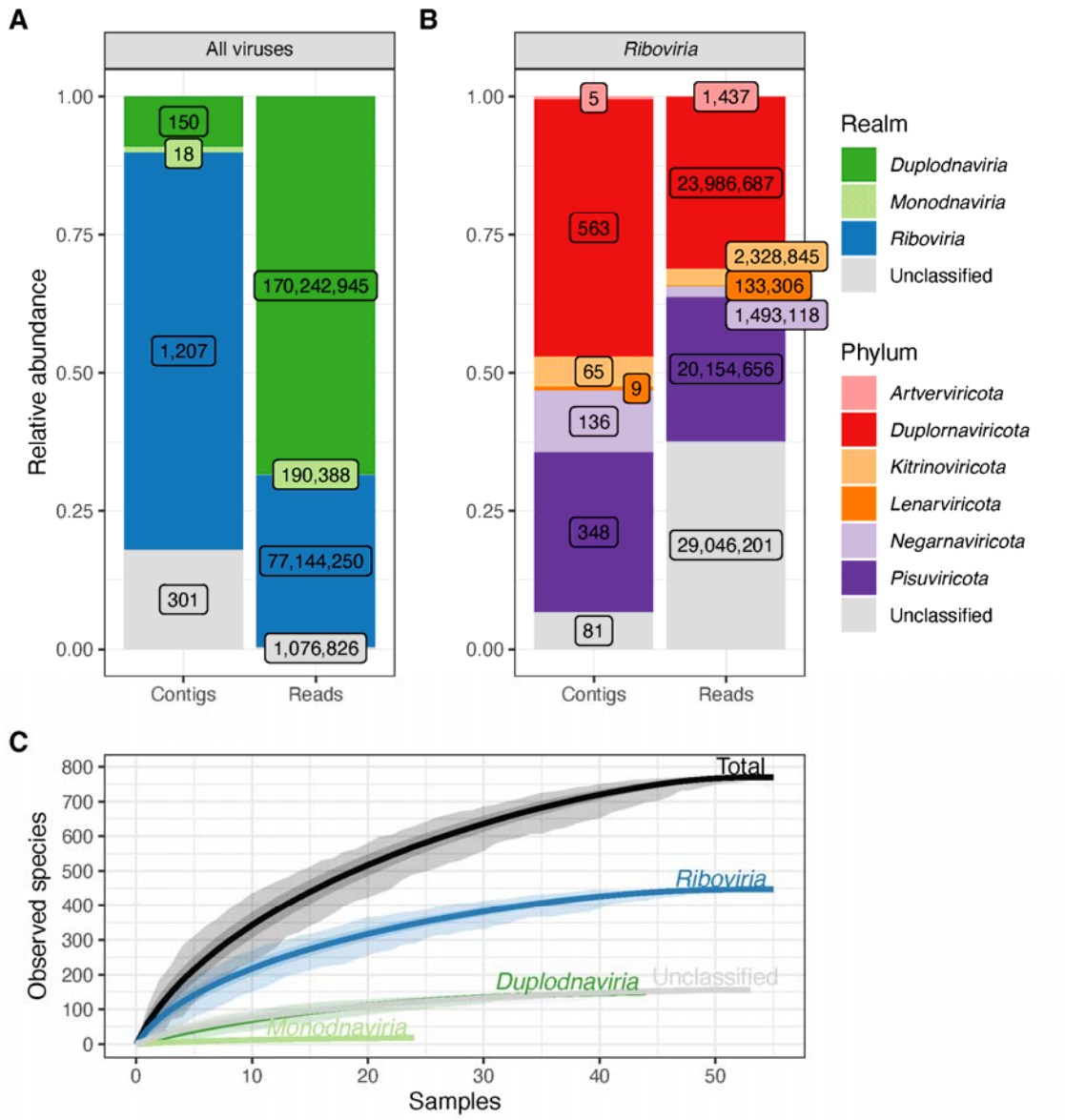


187

188 **Figure 1: A robust set of viruses was identified based on three different identification**
189 **methods and a viral genome completeness estimation.** A) Venn diagram of the identified
190 viral sequences by the various methods. B) Pie chart showing the completeness estimation
191 of CheckV on the identified viral scaffolds.

The blackfly virome is dominated by RNA viruses

When examining the taxonomic distribution of our viral genome sequences, the majority (n=1207) were classified within the *Riboviria* realm, 301 could not be assigned into a specific realm and a small number belonged to the *Monodnavaria* and the *Duplodnaviria* (see Figure 2A). 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 2A). The 301 unclassified sequences recruit only a small fraction (~1 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 2B), while this phylum constituted only one third of the *Riboviria* reads (see Figure 2B). 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 2C). 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).



214

215 **Figure 2: Blackfly viromes are dominated by RNA viruses.** A) Barplots showing the relative
216 abundance of all identified viruses at contig and read level. Labels indicate the absolute
217 numbers. B) Barplots showing the relative abundance of all identified *Riboviria* phyla at
218 contig and read level. Labels indicate the absolute numbers. C) Rarefaction curves for each
219 virus kingdom and the total. Light shaded areas indicate the minimum and maximum
220 observed species across all iterations, the dark shaded areas indicate the 95% confidence
221 around the mean.

222 To visualize the virome similarity of the fifty-five pools, a heatmap with the read count for
223 each viral family on a $\log_2$ scale was constructed (see Figure 3). In total, forty eukaryotic viral
224 families were identified across all samples. In addition, eleven different taxonomic instances
225 of prokaryotic viruses were identified on the family level. Most diversity in the blackflies on
226 the family level is present in the *Riboviria* realm with thirty-five families of eukaryotic RNA
227 viruses detected.

228 The viral families *Caliciviridae, Flaviviridae, Hepeviridae, Orthomyxoviridae, Parvoviridae,*
229 *Phenuiviridae, Picornaviridae, Spinareoviridae, Rhabdoviridae*, and *Togaviridae* include
230 human pathogenic viruses, some of which may also be arboviruses. Notably, known viruses
231 causing neurological complications belong to the families of *Rhabdoviridae*, *Flaviviridae,*
232 *Picornaviridae* as well as *Togaviridae*.

233 Plant and fungal viruses were the most prevalent viruses, with members of *Ghabrivirales*
234 (i.e., *Chrysoviridae*, *Totiviridae*) and *Partitiviridae* being the most abundant. A virus of the
235 family *Flaviviridae* was present in most blackfly pools and in a relatively high abundance.
236 Furthermore, other eukaryotic viral families (main hosts according to ICTV are given in
237 brackets) that were prevalent, were *Curvulaviridae* (fungi), *Dicistroviridae* (invertebrates),
238 *Iflaviridae* (insects), *Genomoviridae* (human, mammals, birds, fungi), *Mymonaviridae* (fungi),
239 *Orthomyxoviridae* (human, birds, pigs), *Parvoviridae* (vertebrates, insects), *Polycipiviridae*
240 (insects), *Rhabdoviridae* (vertebrates, invertebrates, plants), *Totiviridae* (fungi) and
241 *Virgaviridae* (plants). Other viral families of potential clinical interest *Caliciviridae*,
242 *Hepeviridae, Phenuiviridae, Picornaviridae* and *Spinareoviridae* were present in lower
243 abundance and prevalence.

244 Detailed phylogenetic analysis for all above mentioned viral families has been conducted
245 and can be found, together with extra information on closely related viruses, in the
246 Supplemental Information.

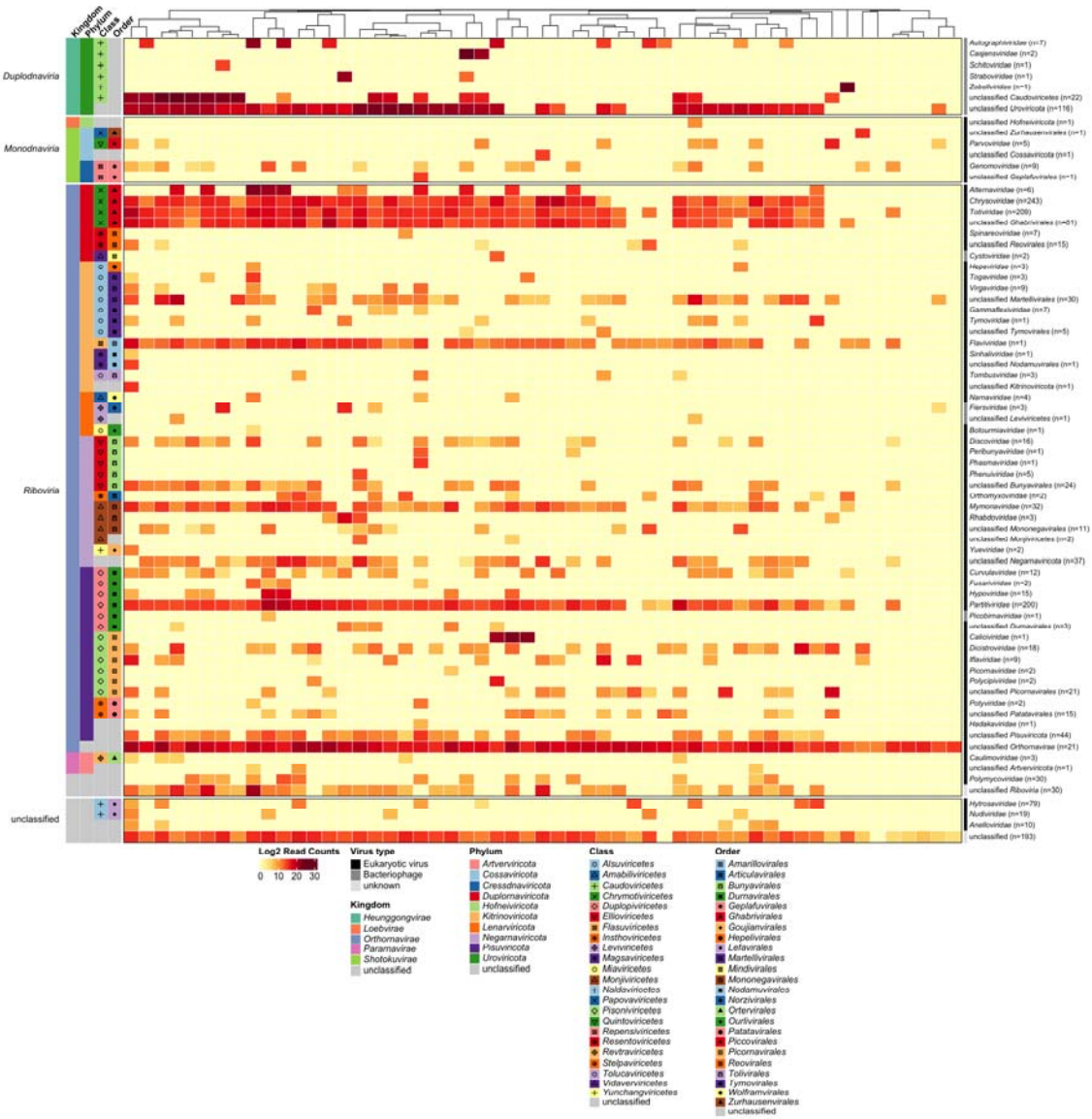


Figure 3: Blackflies harbour a large number of viral families, including families of potential clinical interest. Heatmap of all viral families per blackfly pool. The heatmap shows the read count on a log2 scale for eukaryotic virus on family level from all the 55 adult blackfly pools. The number of identified scaffolds are 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.

Blackfly virome members are found across all *Orthornavirae* phyla, including a potential new phylum of dsRNA viruses

To further explore the *Orthornavirae* diversity in our samples, we generated a ML phylogenetic tree based on the RdRP palmcore of all RdRPs identified in the blackflies, supplemented with previously identified RNA virus sequences (Wolf *et al.* 2018, 2020; Neri *et al.* 2022). In total we could identify 76,097 palmcore sequences with a complete palmprint, of this set 466 were derived from the blackfly data. The cladogram in Figure 4 shows that blackflies host RNA viruses across all *Orthornavirae* phyla (indicated with red dots). Most are situated within the *Negarnaviricota*, the *Duplornaviricota* and the dsRNA *Pisuviricota*. These last two phyla mostly harbour viruses infecting fungi. Interestingly, in our phylogenetic tree the dsRNA *Pisuviricota* were placed in a separate clade from the ssRNA(+) *Pisuviricota*, suggesting that, in contrast to what was previously observed (Wolf *et al.* 2018, 2020; Neri *et al.* 2022), the dsRNA *Pisuviricota* might represent a distinct phylum.

In our search for potential neurotropic viruses, we identified two novel viruses belonging to the *Rhabdoviridae* which are part of the ssRNA(-) *Negarnaviricota*. The most elusive neurotropic member of the *Rhabdoviridae* is rabies virus. Furthermore, also the *Flaviviridae* within the *Kitrinoviricota* harbor known neurotropic viruses (eg. West Nile virus) (Johnson 2016). Within this family we found one novel virus that was consistently present across the blackfly pools (see Figure 3).

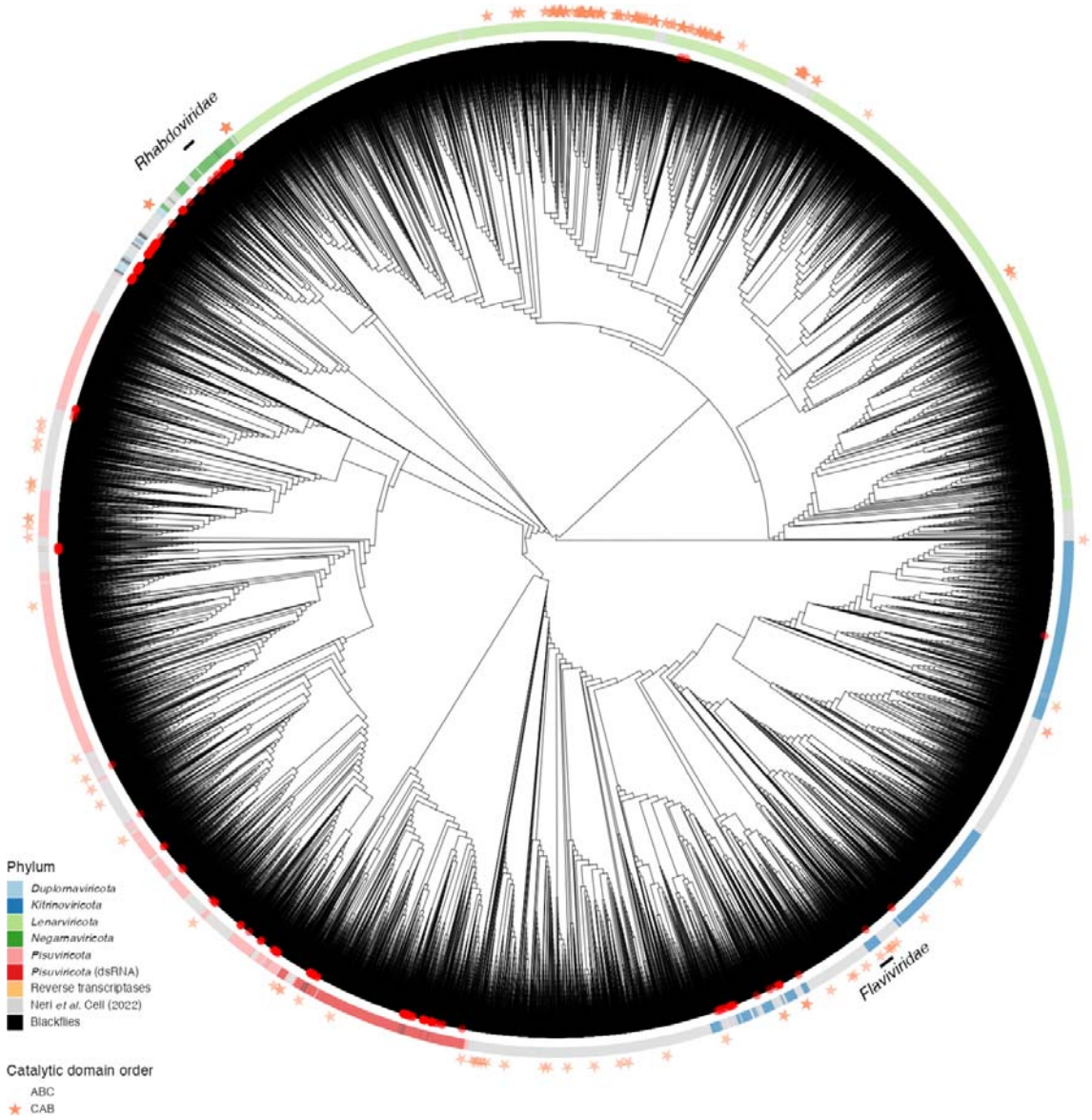


Figure 4: Cameroonian blackflies contain viruses across all phyla of the *Orthornavirae.* Circularized cladogram of *Orthornavirae* viruses based on the phylogeny of the RdRP palmcore domain. The ML tree is rooted on RT sequences as outgroup. Blackfly sequences are indicated with red dots on the leaves of the tree. The ICTV phylum classification of viruses is shown on the outside ring, sequences established in putative new orders/families by Neri *et al.* (2022) are coloured in light grey, while the blackfly sequences are coloured black. Viruses with a permuted catalytic RdRP domain (motif CAB) are reported with an orange star.

Novel blackfly rhabdo- and flaviviruses are most likely insect-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 5A), 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 infecting invertebrates (see Figure 5B). 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.* 2023), 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 also be capable of infecting vertebrates.

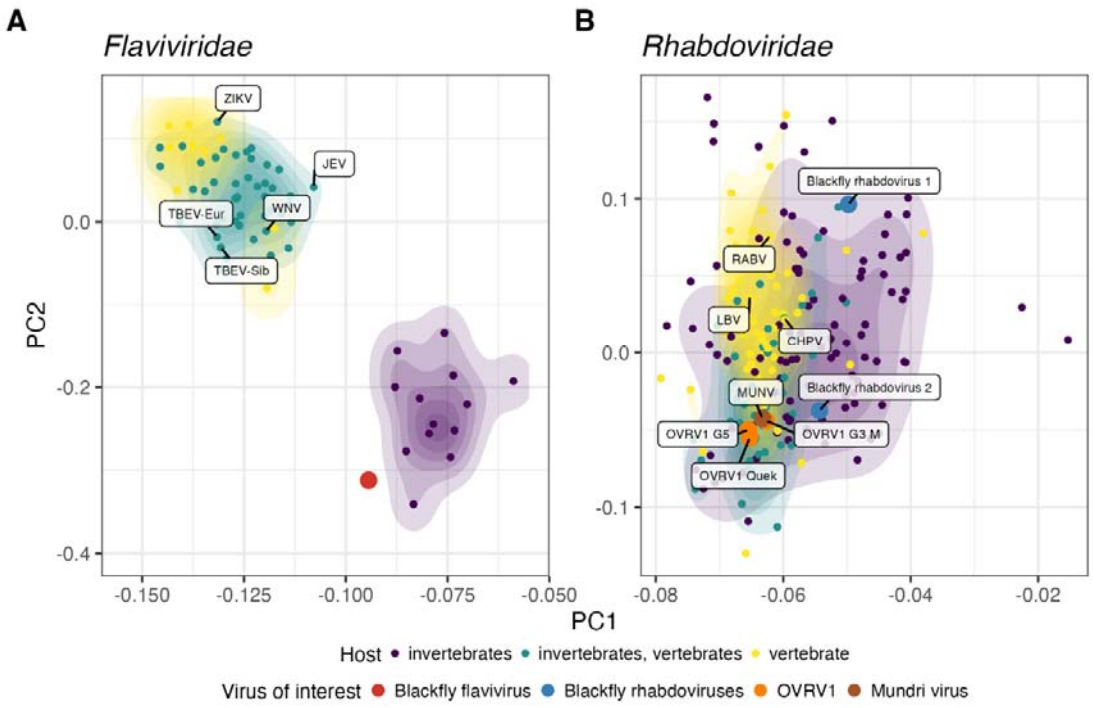


Figure 5: Identified blackfly flavi- and rhabdoviruses are most likely insect-specific. A) PCA plot based on the RDA of selected *Flaviviridae* sequences. The red dot shows the newly identified flavivirus in blackflies. Labels show a selection of neurotropic flaviviruses (TBEV=tick-borne encephalitis 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

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

327 The only *Flaviviridae* virus in our analysis was observed at a relatively high abundance in
328 most blackfly samples (see Figure 3). Previous phylogenetic analyses and experimental
329 studies (Kuno and Chang 2005; Cook *et al.* 2013) have shown that the phylogenetic tree of
330 the *Flaviviridae* separates into two large clusters. The first cluster contains ISVs which
331 appear to be host-restricted and only replicate in invertebrate cells and primarily infect
332 mosquitoes (Hoshino *et al.* 2009; Guzman *et al.* 2018; Miranda *et al.* 2019; Baidaliuk *et al.*
333 2020). The second cluster comprises human pathogenic arboviruses. These previous studies
334 also indicate that arthropod-borne viruses of vertebrates may have originated from ISVs
335 that evolved from being limited to insects to acquiring the ability to infect vertebrate hosts
336 (Bolling *et al.* 2015). The dinucleotide analysis of the *Flaviviridae* indicated that the flavivirus
337 in our blackfly samples is most likely an insect-specific virus. Our phylogenetic analysis
338 revealed that the new blackfly flavivirus forms a deep evolutionary branch within the ISV
339 subcluster, distantly related to Hangzhou flavivirus 3, detected in a fly (see Supplemental
340 Information). These two viruses may form a distinct cluster of ISVs with the fly as the natural
341 host. Given the high abundance of this flavivirus in blackfly pools and the presence of ISVs
342 among them, it is possible that they may become new emerging pathogens.

343 As blackflies can act as intermediate hosts of pathogens affecting humans and animal health
344 due to their hematophagic habit, several studies (Mead *et al.* 2004; Smith *et al.* 2009;
345 Mesquita *et al.* 2017; Young *et al.* 2021) have focused on the transmission of vesicular
346 stomatitis virus (VSV), an arbovirus belonging to the family *Rhabdoviridae*, genus

347 *Vesiculovirus*, which is known to be transmitted by blackflies. This virus primarily infects
348 livestock, representing an agriculturally relevant pathogen, but zoonotic events have also
349 been reported, highlighting its potential impact on human health (Lichty *et al.* 2004). VSV
350 infections are generally asymptomatic in humans, however, mild flu-like symptoms have
351 been reported in some individuals and a single case of encephalitis in a 3-year-old boy,
352 which was potentially associated with VSV infection, has been reported (Quiroz *et al.* 1988).

353 In 2022, Edridge and colleagues discovered a divergent rhabdovirus in the bloodstream of a
354 15-year-old girl with Nodding syndrome from Mundri West Country in South Sudan (Edridge
355 *et al.* 2022). They named this virus "Mundri virus" (MUNV) and classified it as a novel virus.
356 Subsequent phylogenetic analysis revealed that MUNV belongs to the monophyletic clade of
357 *tibroviruses* within the family *Rhabdoviridae*. The *Tibrovirus* genus is part of a larger
358 phylogenetic group of arthropod-borne rhabdoviruses. Several *tibroviruses* have been
359 isolated from biting midges (*Culicoides* spp.) and cattle, and some instances of human
360 detection have been reported (Cybinski *et al.* 1980; Cybinski and Gard 1986; Gibbs *et al.*
361 1989; Grard *et al.* 2012; Stremlau *et al.* 2015). Human tibrovirus infections may accidentally
362 occur via midge bites (Kuhn *et al.* 2020). Consistent with the findings of Edridge *et al.*, our
363 phylogenetic analysis of *Rhabdoviridae* (see Supplemental Information) showed that MUNV
364 is closely related to the Tibrovirus genus, supported by a high bootstrap value. However, the
365 two sequences detected in the blackfly virome were not closely related to the *Tibrovirus* nor
366 *Vesiculovirus* genera. In addition, another novel rhabdovirus, named OVRV1, was discovered
367 by Quek *et al*. in public transcriptomes and viromes of *Onchocerca volvulus* nematodes
368 (Quek *et al.* 2023). Because OVRV1 was shown to elicit an immune response in the
369 vertebrate host of the nematode, it could potentially play a role in the pathogenesis of OAE
370 (Quek *et al.* 2023). We were, however, not able to find this virus back in the virome of the
371 blackflies. Our dinucleotide analysis for the *Rhabdoviridae*, suggested that the viruses we
372 found in the blackflies are invertebrate-specific viruses, in contrast to the previously
373 identified rhabdoviruses MUNV (Edridge *et al.* 2022) and OVRV1 (Quek *et al.* 2023).
374 However, a recent association between MUNV infection and nodding syndrome was
375 investigated in a case-control study in South Sudan. MUNV RNA was only detected in the
376 plasma of one of the 72 persons with Nodding syndrome and in none of the controls. MUNV

377 anti-nucleocapsid IgG antibodies were observed in 28% of Nodding syndrome cases, 22% of
378 household controls, and 16% of community controls without significant differences between
379 cases and both control groups. Therefore, it was concluded that MUNV may not be causally
380 associated with Nodding syndrome (Edridge *et al.* 2022). Similarly, it remains to be shown
381 whether OVRV1 may be associated with OAE.

382 Our study is the first to extensively study the DNA and RNA virome of blackflies from an
383 onchocerciasis-endemic area in Africa. Our findings reveal a relative high abundance of the
384 family *Genomoviridae* in the majority of blackfly samples, aligning with the results of
385 Kraberger *et al.*, which demonstrated that the majority of novel ssDNA viral genomes
386 detected in blackflies in New Zealand belonged to the *Genomoviridae* family (Kraberger *et*
387 *al.* 2019). However, in contrast with the blackflies in New Zealand (Kraberger *et al.* 2019),
388 our study identified *Parvoviridae*, in the blackfly virome. Although this viral family exhibited
389 a high abundance, it was only observed in a limited number of blackfly samples.
390 Phylogenetic analysis revealed that one of the *Parvoviridae* genomes detected in this study
391 clustered within the recently defined *Hamaparvovirinae* subfamily, containing many viruses
392 identified in vertebrate hosts, making it an interesting candidate to further study for its
393 ability to infect vertebrate cells.

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

402 Our study has some limitations. First of all, it reports the composition of the blackfly virome,
403 including both DNA and RNA viruses but without metadata that would have allowed further
404 associations to be made with *O. volvulus* and potential environmental factors. Further
405 research should compare the virome of blackflies with and without *O. volvulus* infection,

and the sex and species of the blackflies should be taken into consideration. It is important to note that only a few blackfly species consume a blood meal (Kraberger *et al.* 2019). Additionally, as less than 1% of blackflies in Cameroon are currently infected with *O. volvulus* (Shintouo et al. 2020), it is recommended to screen blackflies for *O. volvulus* 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. Viral genomes are submitted to GenBank under accession numbers XXX-XXX. 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.

References

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

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

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

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

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

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

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

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

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972.

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

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

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

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

467 Colebunders R, Njamnshi AK, Menon S *et al.* Onchocerca volvulus and epilepsy: A
468 comprehensive review using the Bradford Hill criteria for causation. *PLoS Negl Trop Dis*
469 2021;**15**:1.

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

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

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

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

479 Cook S, Chung BYW, Bass D *et al.* Novel virus discovery and genome reconstruction from
480 field rna samples reveals highly divergent viruses in dipteran hosts. *PLoS One*
481 2013;**8**:11.

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

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

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

488 Davis NM, Proctor DiM, Holmes SP *et al.* Simple statistical identification and removal of
489 contaminant sequences in marker-gene and metagenomics data. *Microbiome*
490 2018;**6**:1–14.

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

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

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

498 Ekangouo AE, Djeunga HCN, Sempere G *et al.* Bacteriome Diversity of Blackflies' Gut and
499 Association with Onchocerca volvulus, the Causative Agent of Onchocerciasis in Mbam
500 Valley (Center Region, Cameroon). *Pathogens* 2022;**11**, DOI:
501 10.3390/PATHOGENS11010044.

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

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

508 Gibbs EPJ, Calisher CH, Tesh RB *et al.* Bivens arm virus: a new rhabdovirus isolated from
509 Culicoides insignis in Florida and related to Tibrogargan virus of Australia. *Vet Microbiol*
510 1989;**19**:141–50.

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

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

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

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

522 Hoshino K, Isawa H, Tsuda Y *et al.* Isolation and characterization of a new insect flavivirus
523 from Aedes albopictus and Aedes flavopictus mosquitoes in Japan. *Virology*
524 2009;**391**:1.

525 Hotterbeekx A, Lammens M, Idro R *et al.* Neuroinflammation and not tauopathy is a
526 predominant pathological signature of nodding syndrome. *J Neuropathol Exp Neurol*
527 2019;**78**:1049–58.

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

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

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

535 Johnson BW. Neurotropic flaviviruses. *Neurotropic Viral Infections: Volume 1: Neurotropic*
536 *RNA Viruses* 2016:229–58.

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

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

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

544 Levick B, Laudisoit A, Tepage F *et al.* High prevalence of epilepsy in onchocerciasis endemic
545 regions in the Democratic Republic of the Congo. *PLoS Negl Trop Dis*
546 2017;**11**:e0005732.

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

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

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

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

557 Miranda J, Mattar S, Gonzalez M *et al.* First report of Culex flavivirus infection from Culex
558 coronator (Diptera: Culicidae), Colombia. *Virol J* 2019;**16**, DOI: 10.1186/S12985-018-
559 1108-2.

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

564 Nayfach S, Camargo AP, Schulz F *et al.* CheckV assesses the quality and completeness of
565 metagenome-assembled viral genomes. *Nature Biotechnology 2020 39:5* 2020;**39**:578–
566 85.

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

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

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

573 Piégu B, Guizard S, Spears T *et al.* Complete genome sequence of invertebrate iridovirus IIV-
574 25 isolated from a blackfly larva. *Arch Virol* 2014a;**159**:5.

575 Piégu B, Guizard S, Yeping T *et al.* Complete genome sequence of invertebrate iridovirus
576 IIV22A, a variant of IIV22, isolated originally from a blackfly larva. *Stand Genomic Sci*
577 2014b;**9**:3.

578 Pion SDS, Boussinesq M. Significant Association between Epilepsy and Presence of
579 Onchocercal Nodules: Case-Control Study in Cameroon. *Am J Trop Med Hyg*
580 2012;**86**:557.

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

583 Quek S, Hadermann A, Wu Y *et al.* The RNA virome of human and animal parasitic
584 nematodes. *Res Sq* 2023, DOI: 10.21203/RS.3.RS-3265990/V1.

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

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

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

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

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

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

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

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

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

610 Steinegger M, Meier M, Mirdita M *et al.* HH-suite3 for fast remote homology detection and
611 deep protein annotation. *BMC Bioinformatics* 2019;**20**:1–15.

612 Stremlau MH, Andersen KG, Folarin OA *et al.* Discovery of novel rhabdoviruses in the blood
613 of healthy individuals from West Africa. *PLoS Negl Trop Dis* 2015;**9**, DOI:
614 10.1371/JOURNAL.PNTD.0003631.

615 Vinkeles Melchers NVS, Mollenkopf S, Colebunders R *et al.* Burden of onchocerciasis-
616 associated epilepsy: First estimates and research priorities. *Infect Dis Poverty* 2018;**7**,
617 DOI: 10.1186/S40249-018-0481-9.

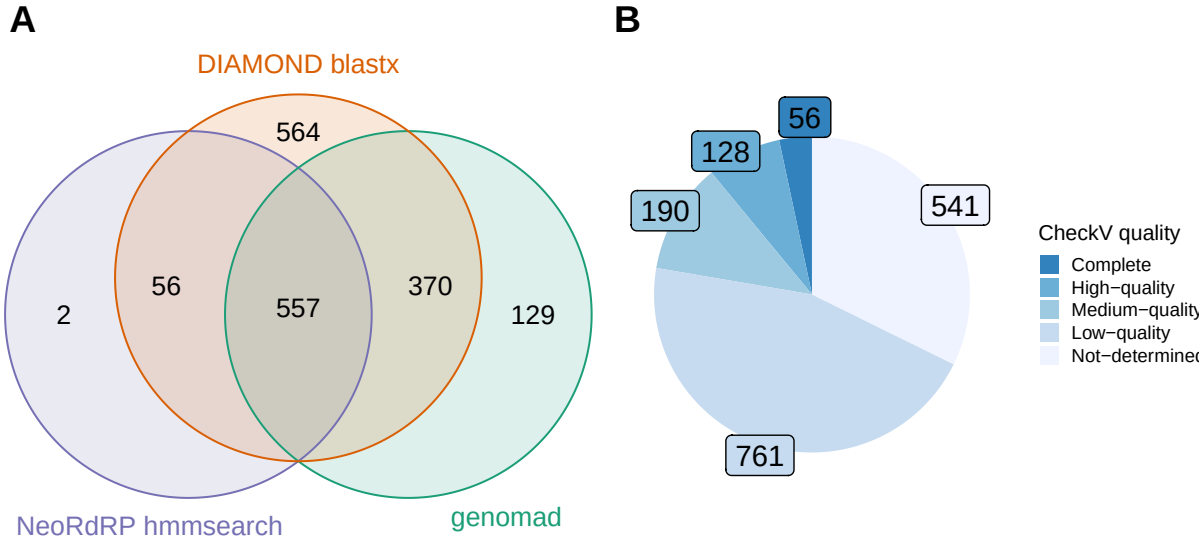
618 Winkler AS, Friedrich K, Velicheti S *et al.* MRI findings in people with epilepsy and nodding
619 syndrome in an area endemic for onchocerciasis: An observational study. *Afr Health Sci*
620 2013;**13**:529–40.

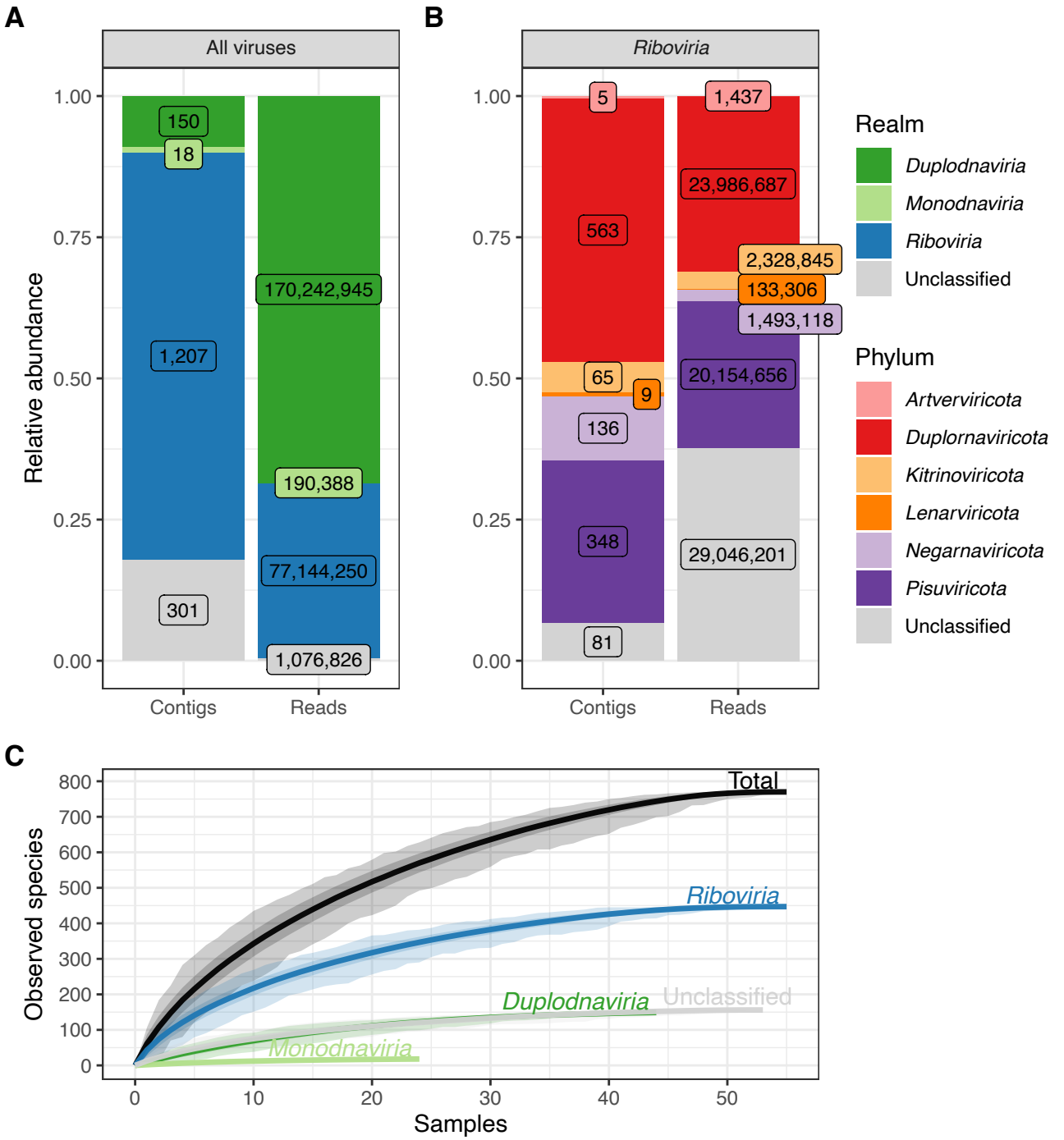
621 Wolf YI, Kazlauskas D, Iranzo J *et al.* Origins and evolution of the global RNA virome. *mBio*
622 2018;**9**, DOI: 10.1128/MBIO.02329-18/ASSET/42CAF1D8-B7AA-433C-9128-
623 816182956BD1/ASSETS/GRAPHIC/MBO0061842030010.JPEG.

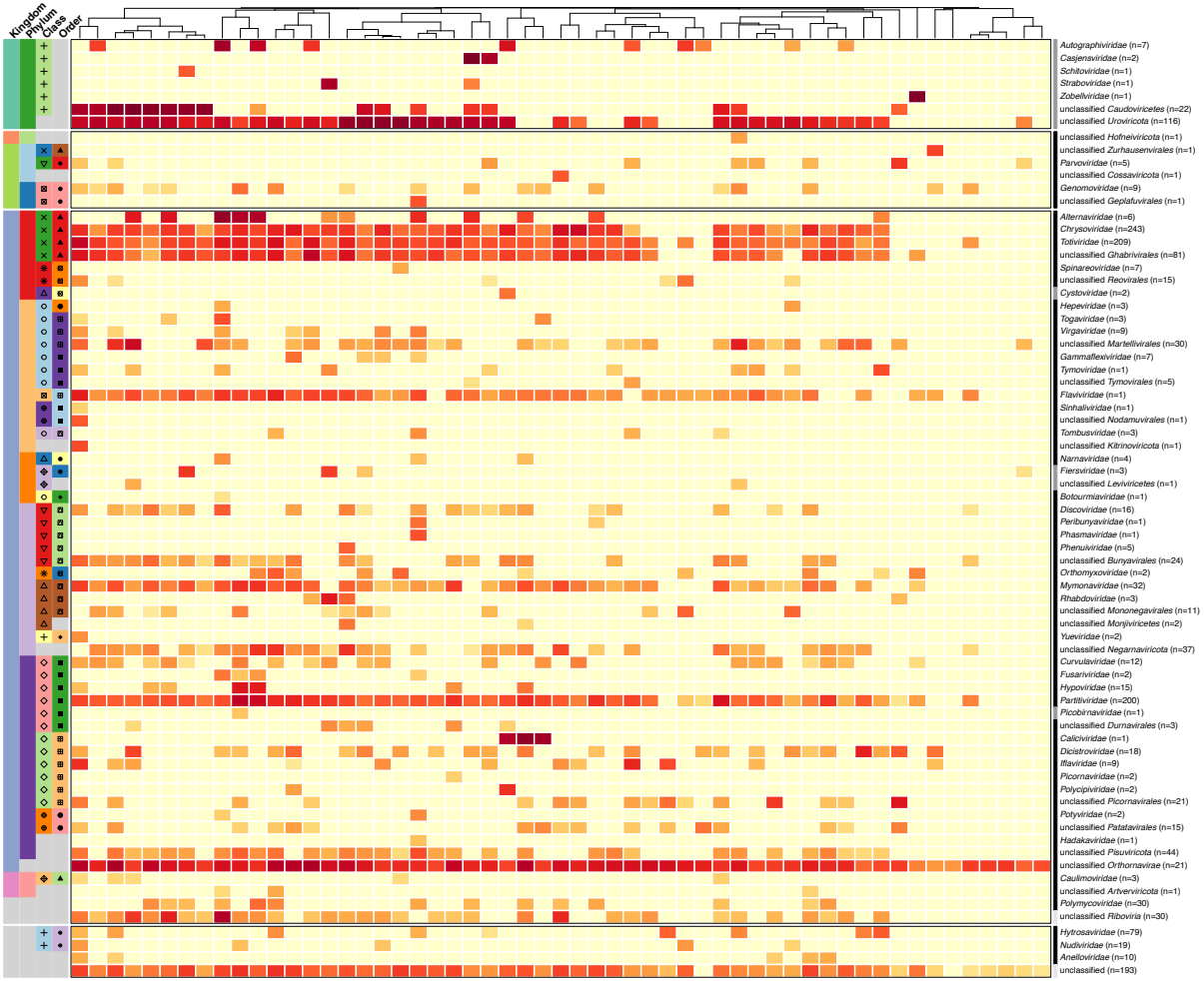
624 Wolf YI, Silas S, Wang Y *et al.* Doubling of the known set of RNA viruses by metagenomic
625 analysis of an aquatic virome. *Nature Microbiology 2020 5:10* 2020;**5**:1262–70.

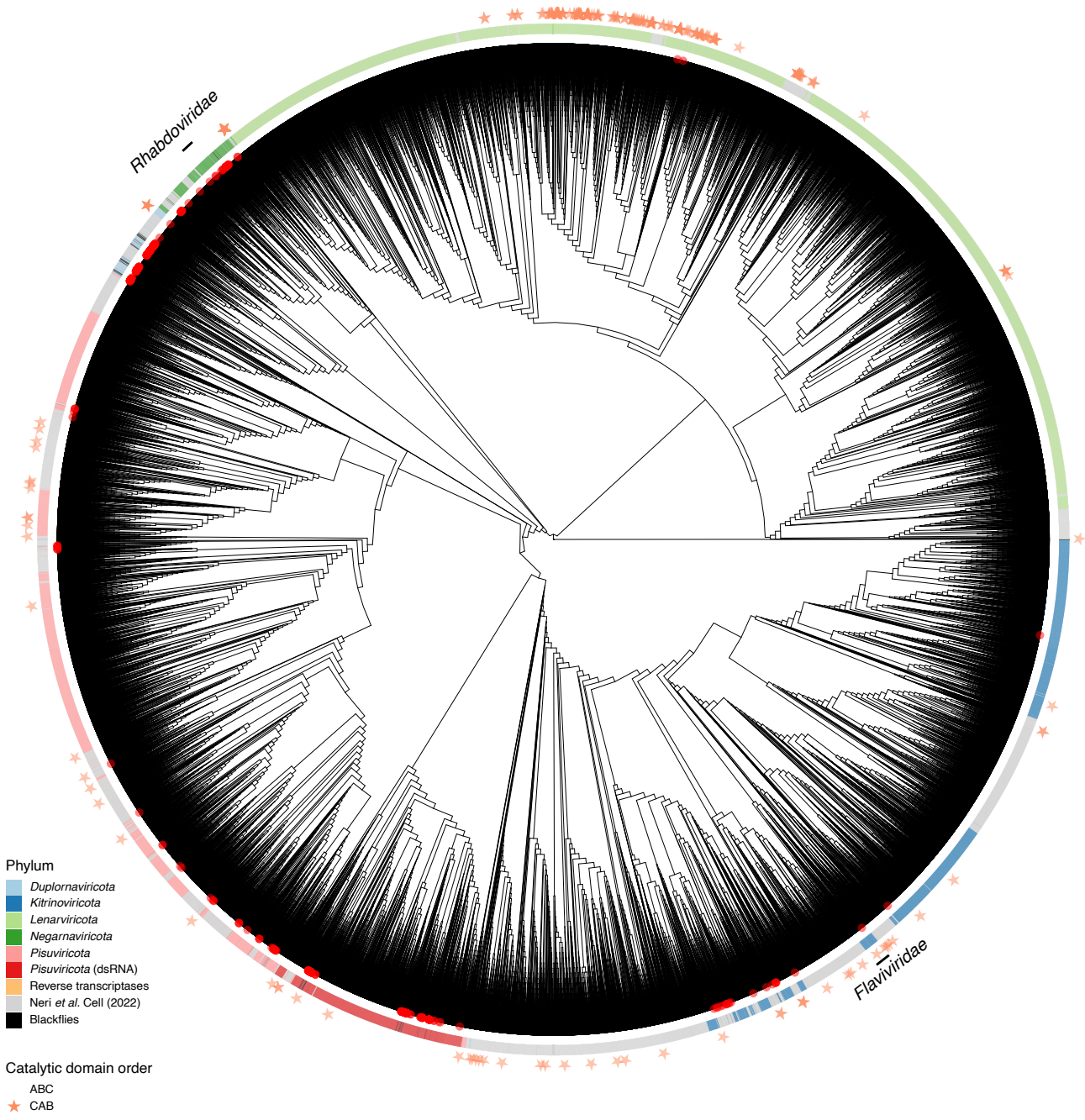
626 Young KI, Valdez F, Vaquera C *et al.* Surveillance along the Rio Grande during the 2020
627 Vesicular Stomatitis Outbreak Reveals Spatio-Temporal Dynamics of and Viral RNA
628 Detection in Black Flies. *Pathogens* 2021;**10**:10.

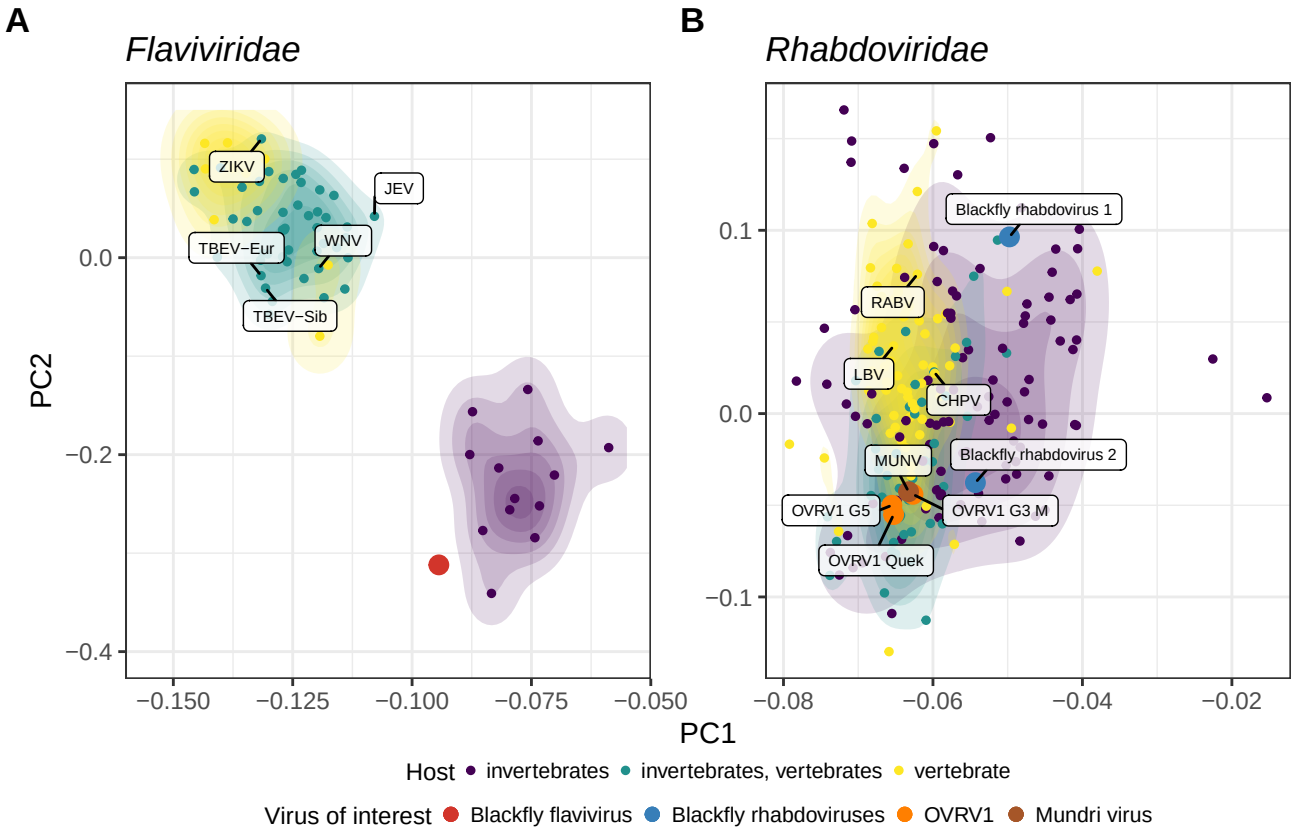
629


A
DIAMOND blastx
564
56
557
370
2
129
NeoRdRP hmmsearch
genomad
B
56
128
190
541
761
CheckV quality
Complete
High−quality
Medium−quality
Low−quality
Not−determined


A
All viruses
Relative abundance
1.00
0.75
0.50
0.25
0.00
150
18
1,207
301
170,242,945
190,388
77,144,250
1,076,826
Contigs
Reads
B
Riboviria
5
563
65
9
136
348
81
1,437
23,986,687
2,328,845
133,306
1,493,118
20,154,656
29,046,201
Realm
Duplodnaviria
Monodnaviria
Riboviria
Unclassified
Phylum
Artverviricota
Duplornaviricota
Kitrinoviricota
Lenarviricota
Negarnaviricota
Pisuviricota
Unclassified
C
Observed species
800
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=2)
Schitoviridae (n=1)
Straboviridae (n=1)
Zobellviridae (n=1)
unclassified Caudoviricetes (n=22)
unclassified Uroviricota (n=116)
unclassified Hofneiviricota (n=1)
unclassified Zurhausenvirales (n=1)
Parvoviridae (n=5)
unclassified Cossaviricota (n=1)
Genomoviridae (n=9)
unclassified Geplafuvirales (n=1)
Alternaviridae (n=6)
Chrysoviridae (n=243)
Totiviridae (n=209)
unclassified Ghabrivirales (n=81)
Spinareoviridae (n=7)
unclassified Reovirales (n=15)
Cystoviridae (n=2)
Hepeviridae (n=3)
Togaviridae (n=3)
Virgaviridae (n=9)
unclassified Martellivirales (n=30)
Gammaflexiviridae (n=7)
Tymoviridae (n=1)
unclassified Tymovirales (n=5)
Flaviviridae (n=1)
Sinhaliviridae (n=1)
unclassified Nodamuvirales (n=1)
Tombusviridae (n=3)
unclassified Kitrinoviricota (n=1)
Narnaviridae (n=4)
Fiersviridae (n=3)
unclassified Leviviricetes (n=1)
Botourmiaviridae (n=1)
Discoviridae (n=16)
Peribunyaviridae (n=1)
Phasmaviridae (n=1)
Phenuiviridae (n=5)
unclassified Bunyavirales (n=24)
Orthomyxoviridae (n=2)
Mymonaviridae (n=32)
Rhabdoviridae (n=3)
unclassified Mononegavirales (n=11)
unclassified Monjiviricetes (n=2)
Yueviridae (n=2)
unclassified Negarnaviricota (n=37)
Curvulaviridae (n=12)
Fusariviridae (n=2)
Hypoviridae (n=15)
Partitiviridae (n=200)
Picobirnaviridae (n=1)
unclassified Durnavirales (n=3)
Caliciviridae (n=1)
Dicistroviridae (n=18)
Iflaviridae (n=9)
Picornaviridae (n=2)
Polycipiviridae (n=2)
unclassified Picornavirales (n=21)
Potyviridae (n=2)
unclassified Patatavirales (n=15)
Hadakaviridae (n=1)
unclassified Pisuviricota (n=44)
unclassified Orthornavirae (n=21)
Caulimoviridae (n=3)
unclassified Artverviricota (n=1)
Polymycoviridae (n=30)
unclassified Riboviria (n=30)
Hytrosaviridae (n=79)
Nudiviridae (n=19)
Anelloviridae (n=10)
unclassified (n=193)
Log2 Read Counts
0 10 20 30
Virus type
Eukaryotic virus
Bacteriophage
unknown
Kingdom
Heunggongvirae
Loebvirae
Orthornavirae
Pararnavirae
Shotokuvirae
unclassified
Phylum
Artverviricota
Cossaviricota
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


Rhabdoviridae
Flaviviridae
Phylum
Duplornaviricota
Kitrinoviricota
Lenarviricota
Negarnaviricota
Pisuviricota
Pisuviricota (dsRNA)
Reverse transcriptases
Neri et al. Cell (2022)
Blackflies
Catalytic domain order
ABC
CAB


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