**Supplementary information**

**Host influence on the eukaryotic virome of sympatric mosquitoes and abundance of diverse viruses with a broad host range**

**Côme Morel1, Patricia Gil1, Antoni Exbrayat1, Etienne Loire1, Florian Charriat1, Baptiste Prepoint1, Celine Condachou1, Assane G. Fall2, Geoffrey Gimonneau1,2, Biram Biteye2, Momar Talla Seck2, Marc Eloit3,4 and Serafin Gutierrez1\***

*1ASTRE, Cirad, INRAe, Univ Montpellier, Montpellier, France.*

*2Institut Sénégalais de Recherches Agricoles/Laboratoire National de l’Elevage et de Recherches Vétérinaires (ISRA), Dakar-Hann, Senegal.*

*3Institut Pasteur, Université Paris Cité, Pathogen Discovery Laboratory, Paris, France.*

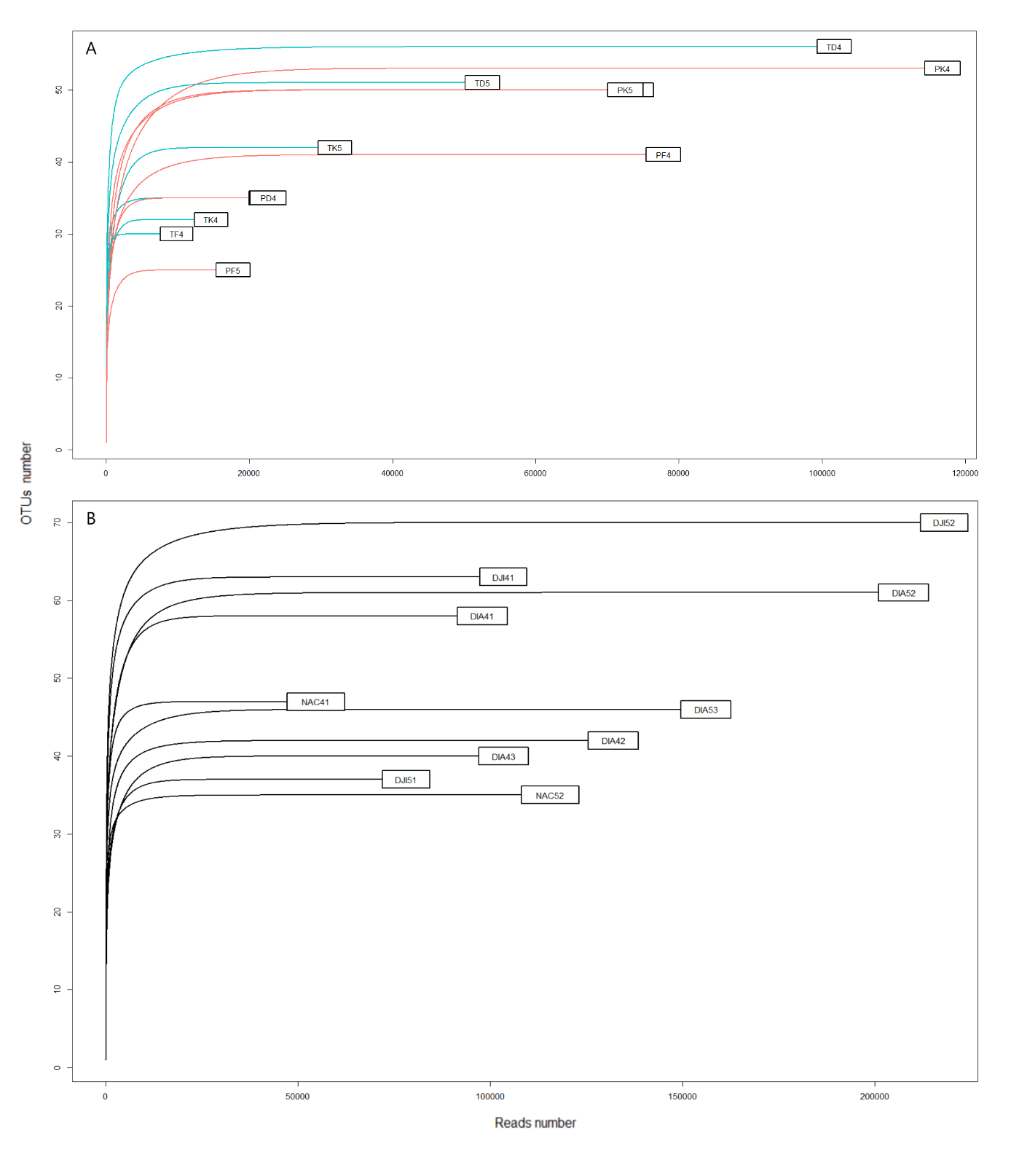
*4Institut Pasteur, Université Paris Cité, The WOAH (OIE) Collaborating Center for the detection and identification in humans of emerging animal pathogens, Paris, France.*

*5Ecole nationale vétérinaire d’Alfort, University of Paris-Est, Maisons-Alfort, France.*

\*Corresponding author : Serafin Gutierrez

Mailing Address : [serafin.gutierrez@cirad.fr](mailto:serafin.gutierrez@cirad.fr)

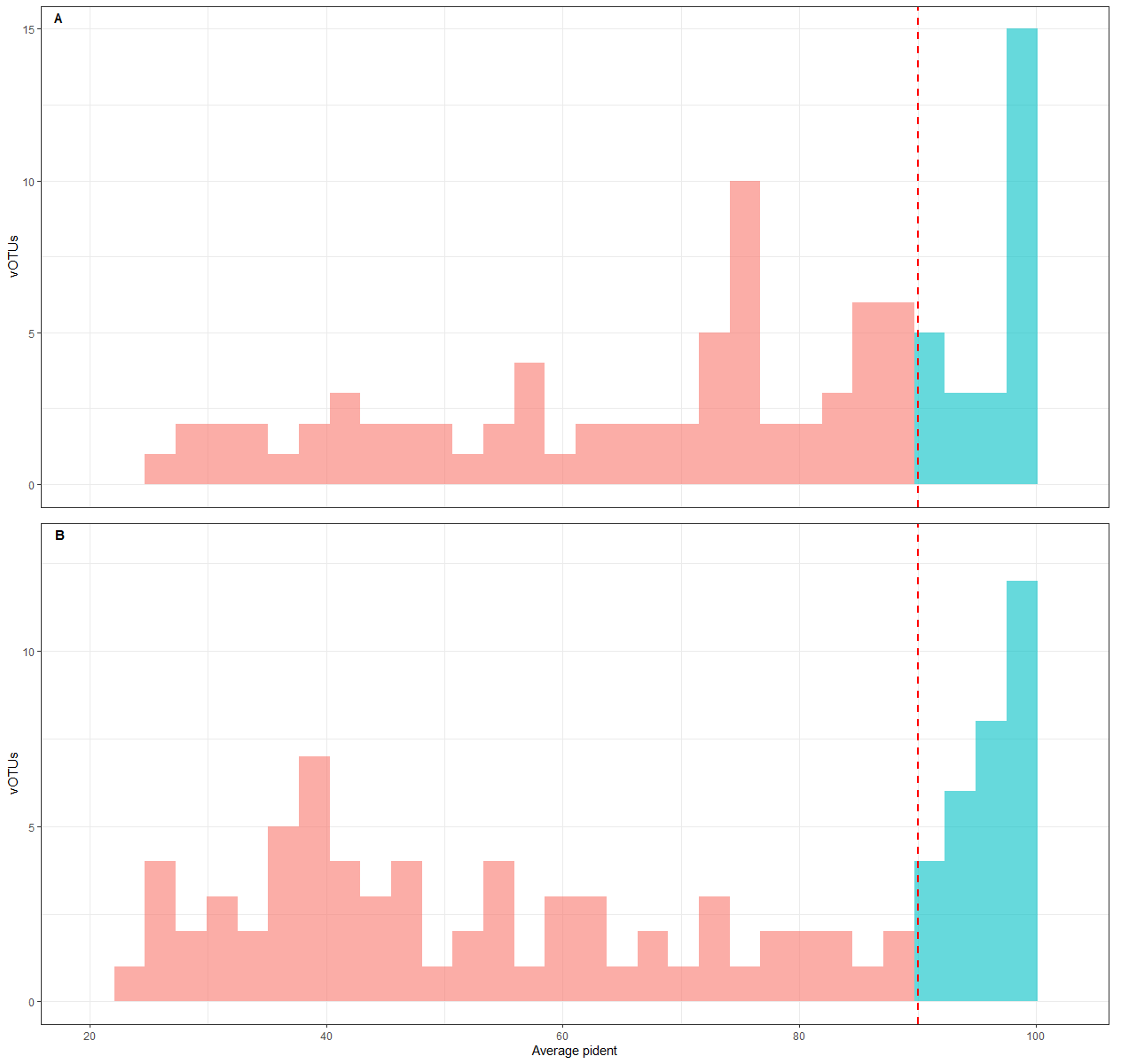
Telephone number : +33 4 67 59 39 11

****

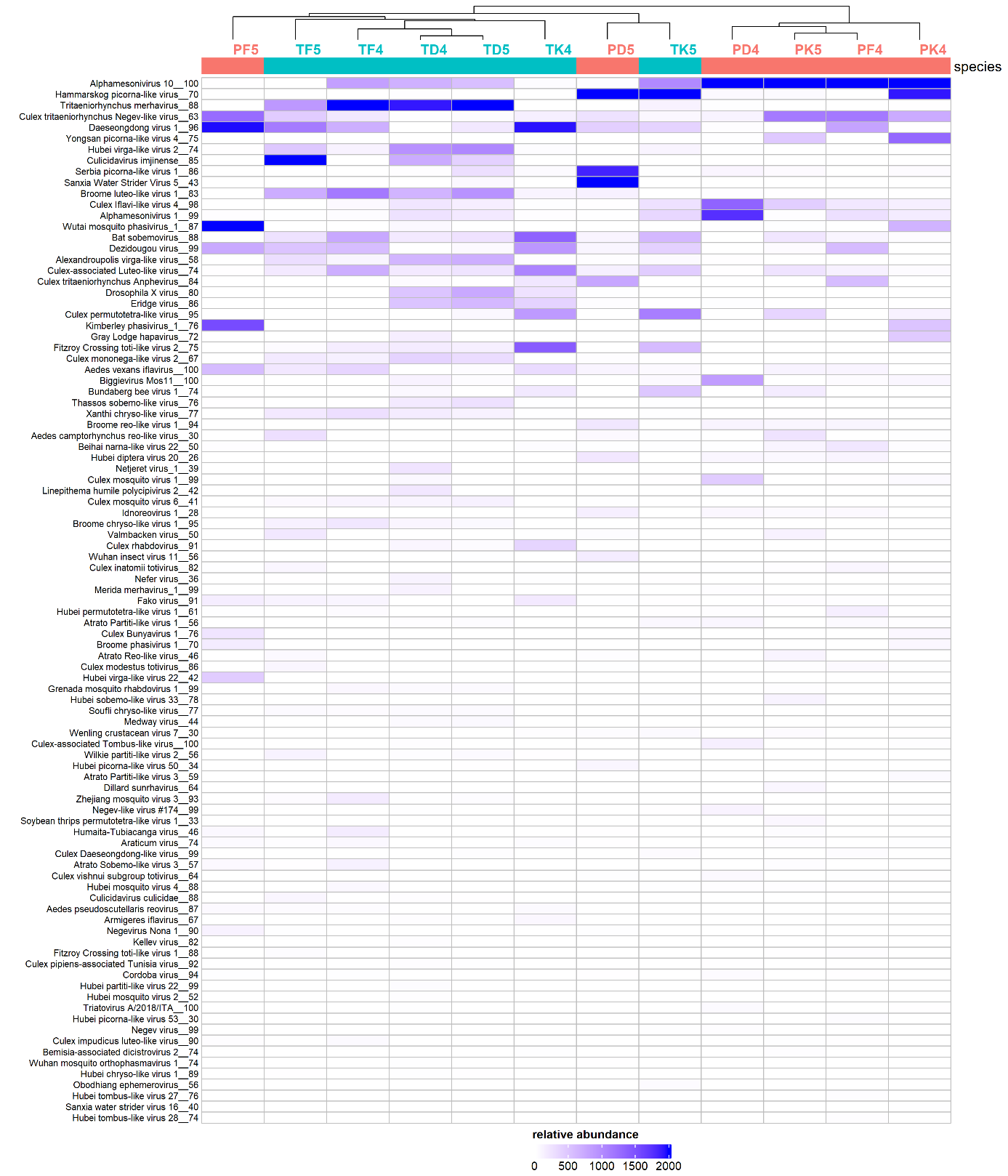
**Figure S1 :** Rarefaction curve analysis of each library (see Table 1 for explanation of acronyms). The x-axis shows the number of virus-like reads, and the y-axis the number of viral taxonomic units (VTUs) per library. **(A)** Libraries of *Culex poicilipes* (red) and *Culex tritaeniorhynchus* (blue). **(B)** Libraries of *Aedes vexans*.



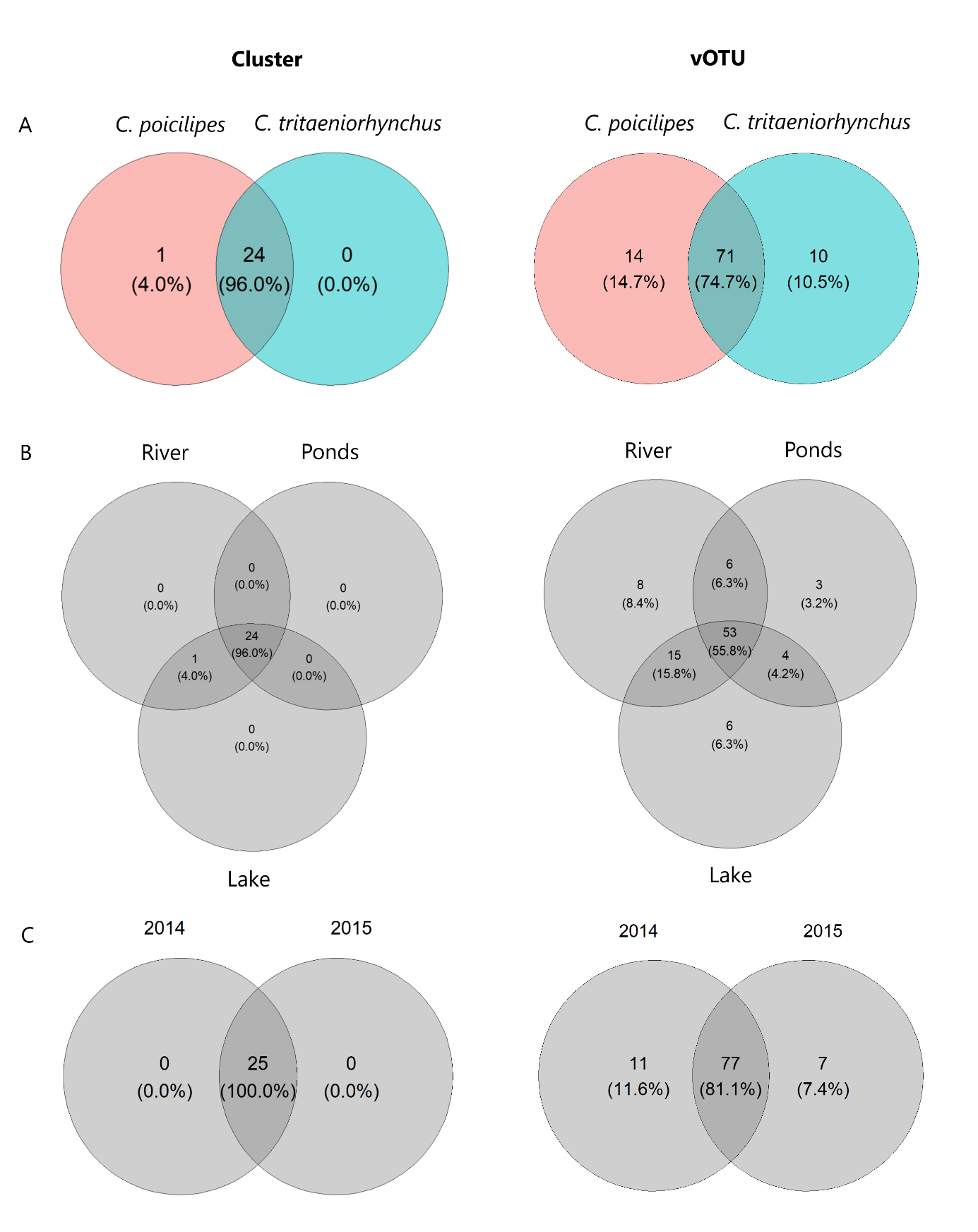
**Figure S2 :** Distribution of viral operational taxonomic units (vOTUs) among orders (external donut chart) and phyla (inner pie chart) found in (A) *Culex poicilipes* and *Culex tritaeniorhynchus*, and (B) *Aedes vexans*. Percentages between brackets represent the proportion of all vOTUs in each order or phyla. The term “*Incertae* *sedis*” stands for taxa whose classification is still undefined at the phylum level.



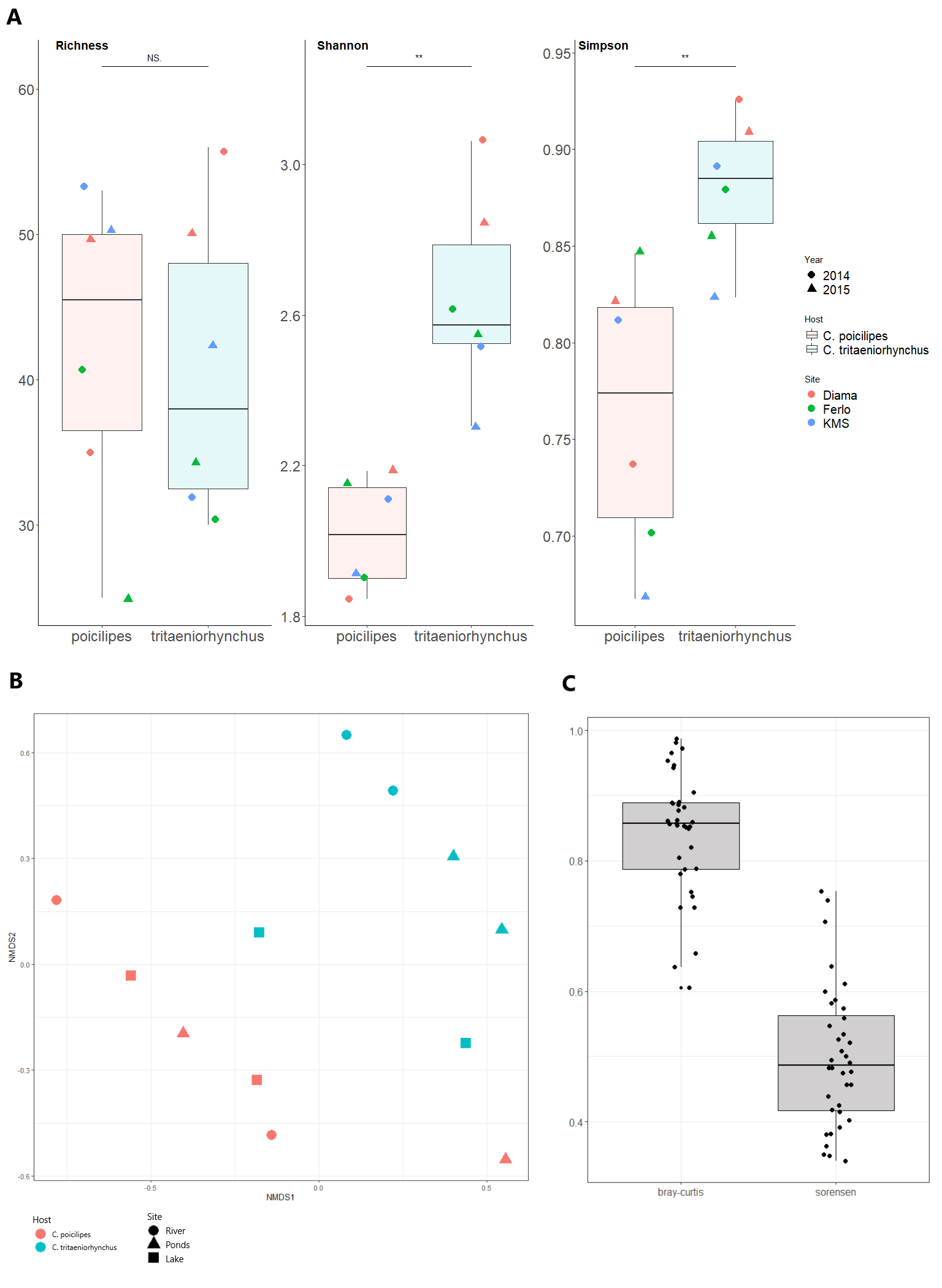
**Figure S3 :** Average percent identities at the amino acid level from all the contigs of each viral operational taxonomic units (vOTUs) with their best hit in the viromes of (A) *Culex* mosquitoes and (B) *Aedes vexans*. The bars in blue indicate an average percent identity higher than 90% and thus vOTUs likely including sequences of the virus species found as their best hit. The red bars represent vOTUs with less than 90% identity to their best hit and thus probably involving sequences of new virus species.



**Figure S4 :** Distribution and relative abundance of viral operational taxonomic units (vOTUs) in the libraries of *Culex poicilipes* and *Culex tritaeniorhynchus*. Library names are indicated on top of the heatmap (see Table 1 for explanation of acronyms), along with a hierarchical clustering, and library colour indicates mosquito species (red : *Culex poicilipes*, blue : *Culex tritaeniorhynchus*). Tile colour stands for read abundance ; the more abundant a cluster, the warmer the colour. The vOTUs are ranked following total read abundance.



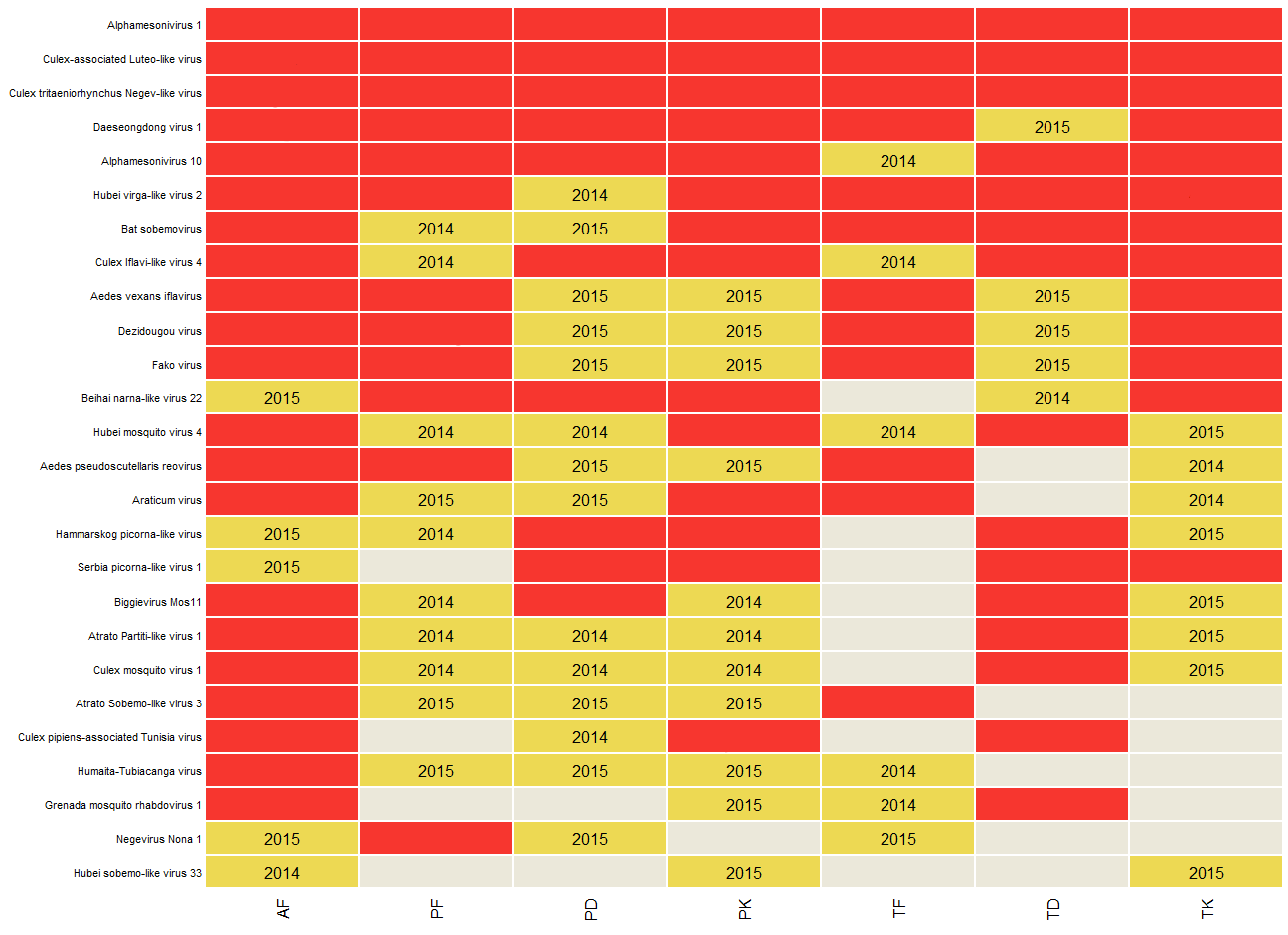
**Figure S5 :** Distribution of clusters (left) and viral operational taxonomic units (right) between mosquito species (A), sites (B) and years (C). Numbers between brackets stand for the proportion of each group among the total number of taxa.



**Figure S6.** A : Distribution of cluster richness, Shannon and Simpson indices between libraries of *Culex poicilipes* (in red) and libraries of *Culex tritaeniorhynchus* (in blue). Dot color indicates the habitat while dot shape represents year. The significance of the comparison between distributions of the two species is shown above boxplots (Wilcoxon Mann-Whitney test). B : Non-metric multidimensional scale with Bray-Curtis dissimilarities obtained from the viromes of the two *Culex* species. Dot color indicates mosquito species and dot shape represents habitat. C : Comparison of Sorensen (for presence-absence data) and Bray-Curtis (for abundance data) dissimilarities between libraries. Each point therefore represents a dissimilarity index value between two libraries belonging each to a different mosquito species.



**Figure S7 :** Heatmap showing the presence (green) or absence (blue) of clusters in the different libraries. Libraries are ranked on the x axis following a hierarchical clustering (dendrogram available on top of the heatmap). Library names are coloured following mosquito species, with libraries from *Aedes vexans* shown in yellow, *Culex poicilipes* in red and *Culex tritaeniorhynchus* in blue (see Table 1 for explanation of acronyms). To facilitate visualization of shared clusters, the heatmap is separated into a top panel with the clusters only present in either the *Aedes* or the *Culex* species, and a bottom panel with the shared clusters.

****

**Figure S8 :** Distribution of the highly-conserved viral operational taxonomic units (vOTUs) over sites, mosquito species and years. The vOTUs are named after their best hit on the y axis. Each combination of mosquito species and site is presented on the x axis. Labels for the mosquito/site combinations are coded with the first letter standing for mosquito species (A for *Aedes vexans*, P for *Culex poicilipes* and T for *Culex tritaeniorhynchus*), and the second letter for site (F for the Ferlo Region (Ponds), D for the Diama village (River) and K for the Keur Momar Sarr village (Lake)). Tile color stands for number of years with detection in a mosquito/site combination (red: two years, yellow: one year, gray: no detection). The year of detection is provided within the tile whenever the virus was detected only one year.

**Table S1:** Literature analysis of the host range (Host) and geographic range (Country) of known generalist viruses found in this study. The "Weighted\_Avg\_contigs\_p\_id" columns provide the average value of the identity percentages associated with the contigs. The column "Blastn" indicates other virus names given to a virus species in the NCBI database (e.g., different virus strains or naming errors during submission).

**Table S2:** Output of the homology and taxonomy search for the Culex dataset. The "Best-hit" column provides the accession with the lowest e-value found by Diamond. The “vOTU” column contains the vOTU names. The “cluster” to “genus” columns provide the vOTU taxonomy as stated by the ICTV. All columns from "sum\_reads" to "max coverage" provide information on the contigs of each vOTU (column fields described below).

*sum\_reads = total number of reads for each vOTU*

*Avg\_match\_length = length of the alignment on the reference*

*n\_Contigs = number of contigs associated with each vOTU*

*Avg\_contigs\_length = average length of the contigs of each vOTU*

*Sum\_contigs\_length = sum of all the contig lengths for each vOTU*

*Avg\_Coverage\_contigs = average coverage of the alignment on the contigs*

*Avg\_Coverage\_subject = average coverage on the reference (alignment length / accession length)*

*Pond\_Avg\_contigs\_p\_id = average percent identity at the amino-acid level of the contigs of a vOTU, weighted by the length of the alignments*

*min\_contigs\_p\_id = minimum percent identity at the amino-acid level of the contigs of a vOTU with their best-hits*

*max\_contig\_p\_id = maximum percent identity at the amino-acid level of the contigs of a vOTU with their best-hits*

*avg\_contigs p\_id = average percent identity at the amino-acid level of the contigs with their best-hits*

*min\_coverage = minimum read coverage*

*max\_coverage = maximum read coverage*

*Avg\_read\_depth = average read depth for each vOTU*

**Table S3:** Output of the homology and taxonomy search for the Aedes dataset. The "Best-hit" column provides the accession with the lowest e-value found by Diamond. The “vOTU” column contains the vOTU names. The “cluster” to “genus” columns provide the vOTU taxonomy as stated by the ICTV. All columns from "sum\_reads" to "max coverage" provide information on the contigs of each vOTU (column fields described in the legend of Tab. S2).