De novo profiling of RNA viruses in *Anopheles* malaria vector mosquitoes from forest ecological zones in Senegal and Cambodia

Eugeni Belda[1,2,3], Ferdinand Nanfack Minkeu[1,2,4], Karin Eiglmeier[1,2], Guillaume Carissimo[5], Inge Holm[1,2], Mawlouth Diallo[6], Diawo Diallo[6], Amélie Vantaux[7], Saorin Kim[7], Igor V. Sharakhov[8], and Kenneth D. Vernick[1,2]*

[1] Unit of Insect Vector Genetics and Genomics, Department of Parasites and Insect Vectors, Institut Pasteur, Paris, France

[2] CNRS Unit of Evolutionary Genomics, Modeling, and Health (UMR2000), Institut Pasteur, Paris, France

[3] Integromics Unit, Institute of Cardiometabolism and Nutrition, Assistance Publique Hôpitaux de Paris, Pitié-Salpêtrière Hospital, Paris, France

[4] Graduate School of Life Sciences ED515, Sorbonne Universités UPMC Paris06, 4 Place Jussieu, 75252 Paris, France.

[5] Laboratory of Microbial Immunity, Singapore Immunology Network, Agency for Science, Technology and Research (A(∗)STAR), Singapore

[6] Institut Pasteur de Dakar, Dakar, Senegal

[7] Institut Pasteur of Cambodia, Phnom Penh, Cambodia

[8] Department of Entomology, Virginia Polytechnic Institute and State University, Blacksburg VA, USA

* Corresponding author
Email: kvernick@pasteur.fr

Author email:

EB e.belda@ican-institute.org
FNM ferdinand.nanfack-minkeu@pasteur.fr
KE karin.eiglmeier@pasteur.fr
GC guillaume_carissimo@immunol.a-star.edu.sg
IH holm@pasteur.fr
MD diallo@pasteur.sn
DD ddiallo@pasteur.sn
AV avantaux@pasteur-kh.org

45 SK ksaorin@pasteur-kh.org
46 IVS igor@vt.edu
47 KDV kvernick@pasteur.fr

Short Title:
Novel *Anopheles* RNA viruses

Keywords:
virus genome assembly, insect specific virus, RNA virus, *Anopheles*, malaria vector, virome

Abstract

Background

Mosquitoes are colonized by a large but mostly uncharacterized natural virome of RNA viruses. *Anopheles* mosquitoes are efficient vectors of human malaria, and the composition and distribution of the natural RNA virome may influence the biology and immunity of *Anopheles* malaria vector populations.

Results

Anopheles vectors of human malaria were sampled in forest village sites in Senegal and Cambodia, including *Anopheles funestus, Anopheles gambiae* group sp., and *Anopheles coustani* in Senegal, and *Anopheles hyrcanus* group *sp., Anopheles maculatus* group *sp.,* and *Anopheles dirus* in Cambodia. Small and long RNA sequences were depleted of mosquito host and de novo assembled to yield non-redundant contigs longer than 500 nucleotides. Analysis of the assemblies by sequence similarity to known virus families yielded 125 novel virus sequences, 39 from Senegal *Anopheles* and 86 from Cambodia. Important monophyletic virus clades in the *Bunyavirales* and *Mononegavirales* orders are found in these *Anopheles* from Africa and Asia. Small RNA size and abundance profiles were used to cluster non-host RNA assemblies that were unclassified by sequence similarity. 39 unclassified non-redundant contigs >500 nucleotides strongly matched a pattern of classic RNAi processing of viral replication intermediates, and 1566 unclassified contigs strongly matched a pattern consistent with piRNAs. Analysis of piRNA expression in *Anopheles coluzzii* after infection with O'nyong nyong virus

78 (family *Togaviridae*) suggests that virus infection can specifically alter abundance
79 of some piRNAs.
80
81 **Conclusions**
82 RNA viruses ubiquitously colonize Anopheles vectors of human malaria
83 worldwide. At least some members of the mosquito virome are monophyletic with
84 other arthropod viruses. However, high levels of collinearity and similarity of
85 Anopheles viruses at the peptide level is not necessarily matched by similarity at
86 the nucleotide level, indicating that *Anopheles* from Africa and Asia are colonized
87 by closely related but clearly diverged virome members. The interplay between
88 small RNA pathways and the virome may represent an important part of the
89 homeostatic mechanism maintaining virome members in a commensal or
90 nonpathogenic state, and host-virome interactions could influence variation in
91 malaria vector competence.

92 # Introduction

93 *Anopheles* mosquitoes are the only vectors of human malaria, which kills at least
94 400,000 persons and causes 200 million cases per year, with the greatest impact
95 concentrated in sub-Saharan Africa and South-East Asia [1]. In addition to malaria,
96 *Anopheles* mosquitoes also transmit the alphavirus O'nyong nyong (ONNV, family
97 *Togaviridae*), which is the only arbovirus known to employ *Anopheles* mosquitoes
98 as the primary vector [2, 3].
99
100 *Anopheles* mosquitoes harbor a diverse natural virome of RNA viruses [4-7]. A
101 recent survey found evidence of at least 51 viruses naturally associated with
102 *Anopheles* [2]. The *Anopheles* virome is composed mainly of insect specific viruses
103 (ISVs) that multiply only in insects, but also includes relatives of arboviruses that
104 can replicate in both insects and vertebrate cells.
105
106 Culicine mosquitoes in the genera *Aedes* and *Culex* transmit multiple arboviruses
107 such as dengue (DENV, family *Flaviviridae*) Zika (ZIKV, family *Flaviviridae*),
108 chikungunya (CHIKV, family *Togaviridae*) and others, but do not transmit human
109 malaria. This apparent division of labor between culicine and *Anopheles*
110 mosquitoes for transmission of arboviruses and Plasmodium, respectively, has led
111 to a relative lack of study about *Anopheles* viruses. *Anopheles* viruses have been
112 discovered by isolation from cultured cells exposed to mosquito extract, serology,
113 specific amplification and sequencing, and more recently, deep sequencing and de
114 novo assembly [2]. Although this work has increased the number of ISVs
115 discovered in Anopheles, it appears that there are many still unknown.
116

Here, we assembled small and long RNA sequences from wild *Anopheles* mosquitoes captured in forest ecologies in central and northern Cambodia and eastern Senegal. The sites are considered disease emergence zones, with high levels of fevers and encephalopathies of unknown origin. Sequence contig evidence of a number of novel RNA viruses and variants was detected, and potentially many unclassified viruses.

It is likely that persistent exposure to ISVs, rather than the relatively infrequent exposure to arboviruses such as ONNV, has been the main evolutionary pressure shaping *Anopheles* antiviral immunity. *Anopheles* resistance mechanisms against arbovirus infection may be quite efficient, based on their lack of virus transmission despite highly anthropophilic feeding behavior, including on viremic hosts. Nevertheless, ONNV transmission is the exception that indicates arbovirus transmission by *Anopheles* is possible, so it is a biological puzzle that transmission is apparently restricted to just one virus. Identifying the complement of natural viruses inhabiting the *Anopheles* niche will help clarify the biology underlying the apparent inefficiency of arbovirus transmission by *Anopheles*, and may suggest new tools to raise the barrier to arbovirus transmission by the more efficient *Aedes* and *Culex* vectors.

Results

Mosquito species estimation

Metagenomic sequencing of long and small fractions of RNA was carried out for four biological replicates pools of mosquitoes from Ratanakiri and Kampong Chnang provinces in central and northern Cambodia near the border with Laos, and four replicate pools from Kedougou in eastern Senegal near the border with the Republic of Guinea (Conakry). Mosquito species composition of sample pools was estimated using sequences of transcripts from the mitochondrial cytochrome c oxidase subunit 1 (COI) gene, which were compared with *Anopheles* sequences from the Barcode of Life COI-5P database (Figure 1, Additional File 1: Table S1). In the Senegal samples, the most frequent mosquito species were *Anopheles rufipes*, *Anopheles funestus, Anopheles gambiae* group sp., and *Anopheles coustani*, which are all human malaria vectors, including the recently incriminated *An. rufipes* [8]. In the Cambodia samples, the most frequent species were *Anopheles hyrcanus* group *sp., Anopheles maculatus* group *sp., Anopheles karwari, Anopheles jeyporeisis, Anopheles aconitus* and *Anopheles dirus*. All are considered human malaria vectors [9-12]. Elevated rates of human blood-feeding by a mosquito species is a prerequisite for malaria vectorial capacity [13], and therefore the main *Anopheles* species sampled for virome discovery in this study display consistently high levels of human contact in nature.

Virus discovery by de novo RNAseq assembly and classification by sequence similarity

Small and long RNA reads were de novo assembled after removal of mosquito sequences. Non-redundant contigs longer than 500 nucleotides from assemblies

of both countries, Cambodia and Senegal, were used to search the GenBank protein sequence database using BLASTX with an e-value threshold of 1e-10. This allowed identification of 125 novel assembled virus sequences, 39 from the Senegal samples (virus ID suffix "Dak", Table 1), and 86 from the Cambodia samples (virus ID suffix "Camb", Table 2), possibly pointing to higher viral diversity in mosquitoes from Cambodia. Some of the 125 virus sequences showed remote similarity by BLASTX to 24 reference viruses in GenBank that include ssRNA-negative strand viruses of the families *Orthomyxoviridae*, *Rhabdoviridae* and *Bunyaviridae*, ssRNA positive-strand viruses of the families *Virgaviridae*, *Flaviviridae* and *Bromoviridae*, dsRNA viruses of the family *Reoviridae* and multiple unclassified viruses of both ssRNA and dsRNA types (Table 3). Most of these remote similarities were with viruses characterized in a recent virus survey of 70 different arthropod species collected in China [14], which emphasizes the importance of high throughput surveys of arthropod virosphere in the identification of viruses associated with different arthropod species.

In order to place these 125 novel virus assemblies in an evolutionary context, phylogenetic trees were constructed from conserved regions of the RNA-dependent RNA polymerase gene annotated in the 125 virus sequences, along with related virus sequences from GenBank. This allowed the placement of 44 of the 125 assembled viruses in phylogenetic trees, revealing clusters of highly related viruses in the analyzed wild *Anopheles*. Notable examples include five novel virus assemblies from Cambodian *Anopheles* placed near Wuhan Mosquito Virus 1 in a monophyletic group of the Phasmavirus clade (*Bunyavirales*) (Figure 2). Also, within the order *Mononegavirales*, 14 novel *Anopheles* virus assemblies

(7 from Cambodia and 7 from Senegal) formed a monophyletic group that includes Xincheng Mosquito Virus and Shungao Fly Virus. Finally, 10 novel virus assemblies (9 from Cambodia, 1 from Senegal) formed a monophyletic group that includes Beaumont Virus and a rhabdovirus from *Culex tritaeniorhynchus* within the Dimarhabdovirus clade (Figure 3A). TBLASTX comparisons of virus sequences in these groups with the closest reference viruses in the phylogenetic trees showed high levels of collinearity and similarity at protein level that was not matched by comparable levels of similarity at the nucleotide level, indicating that populations of closely related but diverged viruses colonize *Anopheles* from widely separated geographic locations (Figure 3B).

Quantification of novel virus sequences in mosquito sample pools

In order to evaluate the prevalence of novel virus sequences across the analyzed mosquito samples, host-filtered small and long RNA reads were mapped over the 125 novel virus sequences identified by de novo sequence assembly. Based on long RNAseq reads, the abundance profiles of the 125 virus assemblies display a non-overlapping distribution across different sample pools, and virus sequences can be localized to particular sample pools from the abundance profiles (Figure 4, left panel). This probably indicates a patchy prevalence and abundance of the different viruses among individual mosquitoes, such that an individual mosquito highly infected with a given virus could potentially generate a strong signal for the virus in the sample pool. The sample pools from Cambodia share a higher fraction of common viruses, while there is less overlap in virus abundance distribution across sample pools from Senegal. The representation of virus distribution based on small RNA sequence reads displayed profiles broadly similar to the long RNA-

based abundance distribution (Figure 4, right panel). This observation may be consistent with the expectation that small RNA representation is a signature of virus double-stranded RNA (dsRNA) processing by the mosquito RNA interference (RNAi) machinery [15], and therefore was specifically examined next.

Small RNA size profiling

The processing of virus sequences by small RNA pathways of the insect host generates diagnostic patterns of small RNA read sizes from different viruses. In order to evaluate this phenomenon in the 125 novel virus assemblies characterized by sequence similarity in the analyzed sample pools, small RNA reads that mapped to each virus assembly were extracted, and their size distributions were normalized with a z-score transformation. This allowed comparison of the z-score profiles among virus assemblies by pairwise correlation analysis and hierarchical clustering. The relationship between the small RNA profiles of the different viruses could then be visualized as a heat map. The results of this analysis revealed the presence of four major groups of virus sequences based on small RNA size profiles (Figure 5). Cluster 1 consists of 7 virus assemblies generating small RNAs predominantly in the size range of 23-29 nt mapping over the positive, and to a lesser extent negative, strand. Cluster 2 includes 7 viruses, all from Senegal, and displays a similar size profile as viruses of Cluster 1 with reads in the 23-29 nt size range, but also with a higher frequency of 21 nt reads mapping over the positive and negative strands, emblematic of virus cleavage through the mosquito host RNAi pathway. Cluster 3 includes 15 viruses that exhibit the classic pattern of viruses processed by the host RNAi pathway, with predominantly reads of 21 nt in length mapping over virus positive and

236 negative strands (Additional File 2: Figure S1). Finally, Cluster 4 includes 59
237 viruses with small RNA size profiles dominated by reads of 23-29 nt mapping
238 predominantly over the negative strand of virus sequences. Because of the strong
239 strand bias of small RNAs observed, this pattern could correspond to degradation
240 products of virus RNAs, although alternatively, there appears to be size
241 enrichment in the 27-28 nt size peaks characteristic of PIWI-interacting RNAs
242 (piRNAs).
243
244 **Viral origin of unclassified transcripts by small RNA size profiling**
245 A major drawback of sequence similarity-based identification of novel viruses in
246 de novo sequence assemblies is the dependence of detection upon existing records
247 of close relatives in public databases. It was proposed that the small RNA size
248 profiles of arthropod-derived viruses detected by sequence similarity could be
249 used as signature to recruit unclassified contigs from de novo sequence
250 assemblies of potential viral origin [15]. We implemented this strategy in order to
251 identify additional sequences of putative viral origin in the set of 2114 contigs
252 with at least 100 small RNA sequence reads left unclassified by sequence
253 similarity searching.
254
255 Of these unclassified contigs, a likely viral origin is supported for 4 and 35 contigs
256 that display strong association by small RNA profile with Cluster 2 and Cluster 3,
257 respectively (Spearman correlation>0.9, Additional File 3: Figure S2). These
258 clusters display small RNA size profiles mapping to both genome strands, and
259 characteristic of classic RNAi processing of viral dsRNA replication intermediates.
260 Thus, in addition to the 125 novel virus assemblies classified by sequence

261 similarity to known viruses, 39 unclassified novel *Anopheles* virus assemblies
262 were identified, without sequence similarity to identified viruses but meeting the
263 quality criteria of non-redundant assemblies longer than 500 nucleotides. Further
264 work will be necessary to characterize the biology of these unclassified novel virus
265 assemblies.
266
267 Of the other assemblies unclassified by sequence similarity, 1566 showed strong
268 associations between their small RNA size profiles and the small RNA size profiles
269 of virus contigs detected by sequence similarity (Spearman correlation>0.9).
270 Among these, the majority were associated with Cluster 4 virus assemblies (1219
271 unclassified contigs) and to less extent with Cluster 1 (309 unclassified contigs).
272 Both clusters were characterized by a strong bias towards reads from a single
273 strand (positive for Cluster 1 and negative for Cluster 4).
274
275 To evaluate how specific these latter profiles of 1219 and 309 contigs are for virus-
276 related sequences, we designed a reconstruction control experiment using the
277 same small RNA size profiling and clustering analysis as above, but instead using
278 669 RNA contigs known to map to the mosquito reference assembly, thus strictly
279 of host origin. As above, contigs with at least 100 small RNA sequence reads were
280 used. 561 of these mosquito contigs could be grouped with small RNA size profiles
281 of virus contigs (Spearman correlation>0.9), most of them (98.21%) with Cluster
282 4 (78.6%) and Cluster 1 (19.6%) profiles.
283
284 However, many somatic piRNAs map to only one strand in Drosophila and other
285 arthropods [16, 17]. Notably, many virus-related piRNAs in *Aedes*, which are

largely ISV-derived, mainly map only to the virus strand antisense to the viral ORF [18]. In *An. coluzzii*, about half of expressed piRNAs display a strong or exclusive strand bias, which is a greater proportion of unidirectional piRNAs than *Drosophila* [19]. Until the current study, *Anopheles* piRNAs have not previously been examined for relatedness to ISVs. Overall, these results are probably most consistent with an interpretation that RNA profile Cluster 1 and Cluster 4 detect strand-biased piRNAs derived from the natural ISV virome of wild *Anopheles*. On that interpretation, the above host-sequence control contigs that share the Cluster 1 and Cluster 4 RNA profiles are most likely also piRNAs, but instead derived from endogenous host templates. Previous results showed that most *An. coluzzii* piRNAs target long-terminal repeat retrotransposons and DNA transposable elements [19]. Our current results add wild ISVs as a likely source of template for *Anopheles* piRNA production, and indicate that further work is warranted in the interpretation of small RNA profiles for discovery of unclassified viruses. Our results also suggest the possibility that piRNAs may be involved in *Anopheles* response to viruses, a phenomenon found for only *Aedes* among a wide range of arthropods, but *Anopheles* were not yet tested [17].

O'nyong nyong alphavirus infection influences expression of piRNAs in *Anopheles coluzzii*

piRNAs are endogenous small noncoding RNAs of about 24-30 nt that ensure genome stability by protecting it from invasive transposable elements such as retrotransposons and repetitive or selfish sequences [17]. In addition, in *Aedes* mosquito cells, piRNAs can probably mediate responses to arboviruses or ISVs [17, 18, 20, 21]. *Anopheles* mosquitoes express piRNAs from genomic piRNA

clusters [19, 22], but piRNA involvement in response or protection to virus infection in *Anopheles* has not been reported to our knowledge. To examine the potential that *Anopheles* piRNAs could be involved in response to viruses, we challenged *An. coluzzii* mosquitoes with the alphavirus, ONNV by feeding an infectious bloodmeal, and sequenced small RNAs expressed during the primary infection at 3 d post-bloodmeal. Mosquitoes fed a normal bloodmeal were used as the control condition.

Analysis of the small RNA expression data using Cuffdiff and DESeq2 detected 86 potential significantly differentially expressed transcripts between ONNV infected mosquitoes and normal bloodmeal controls (Additional File 4: Table S2). Filtering for appropriate length of contiguous expressed region for piRNA <40 nt, and high abundance of expression in ONNV and control samples taken together, yielded two annotated piRNA candidates. The candidates were both downregulated after ONNV infection as compared to uninfected controls (p=5e-5, q=6.7e-3, locus XLOC_012931, coordinates UNKN:19043685-19043716; and p=9.5e-4, q=0.046, locus XLOC_012762, coordinates UNKN:13088289-13088321; Figure 7).

Discussion

The current study contributes to a growing body of work defining the deep diversity of the invertebrate virosphere [14, 23, 24]. Because mosquitoes transmit viral infections of humans and animals, there is particular interest in discovery of ISVs comprising the mosquito virome [6, 25-27]. Here, we sampled *Anopheles* mosquitoes from two zones of forest exploitation in Africa and Asia, considered disease emergence zones with likely zoonotic exposure of the human and domestic animal populations. Using assembly quality criteria of non-redundant contigs at least 500 nt in length, we identified 125 novel RNA virus assemblies by sequence similarity to known virus families, and an additional 39 high-confidence virus assemblies that were unclassified by sequence similarity, but display characteristic products of RNAi processing of replication intermediates. Finally, 1566 unclassified contigs possessed comparable assembly quality, and lacked a strong RNAi processing signature, but displayed a signature consistent with piRNA origin. This latter group will require additional work to filter bona fide virus-derived piRNA sequences, which have been previously reported in *Aedes* mosquitoes [17, 18, 20, 21], from other potential sources of piRNAs such as retrotransposons and DNA transposable elements, as well as possible physical degradation.

Nevertheless, taken together at least 164 novel and non-redundant virus assemblies, and possibly many more, were identified in wild *Anopheles* mosquitoes in the current report. Small and long RNAs were sequenced from pools of 5-10 mosquitoes. Pooled sample analysis obscures the distribution and abundance of viruses among individuals in the population. Individual mosquito

353 analysis will likely become a research focus as sequencing costs drop. However,
354 some insight about virus distribution can be gained from comparison of sample
355 pools collected from the same site, for example Senegal or Cambodia. The
356 abundance heat map shown in Figure 5 indicates that virus diversity is high in the
357 population, and evenness is relatively low among sample pools from the same site.
358 This suggests that the number of viruses per individual is probably also low, with
359 a patchy distribution among individuals. This expectation is consistent with a
360 small number of individual mosquitoes with RNAs deep sequenced and de novo
361 assembled in our laboratory, which identifies <5 distinct viruses per individual.
362
363 The dynamics of the virome may thus be different from the bacterial microbiome,
364 in which tens of taxa are typically present per individual, and microbial diversity
365 is thought to lead to homeostasis or resilience of the microbiota as an ecosystem
366 within the host [28, 29]. By comparison, very little is known about the function of
367 the mosquito virome within the host. At least three important topics are worth
368 exploring. First, unlike the bacterial microbiota, the stability and resilience over
369 time of the viral assemblage in an individual mosquito is unknown. Members of
370 the virome could persist in individual host populations over time in commensal
371 form, or the uneven and patchy viral distribution observed among sample pools
372 could be a consequence of successive waves of epidemic infection peaks and
373 valleys passing through local populations. The commensal or epidemic models
374 could have distinct biological implications for the potential influence of the
375 virome, including on host immunity and competence for transmission of
376 pathogens.
377

Second, the individual and population-level effect of ISV carriage on vector competence for pathogen transmission is a key question. In the current study, the predominant host species sampled are *Anopheles* vectors of human malaria, and in Africa, some of these species are also vectors of ONNV. ISVs have not been tested for influence on *Plasmodium* or ONNV infection in *Anopheles*, to our knowledge. ISVs could affect host immunity and malaria susceptibility, or even cause temporary vector population reduction during a putative ISV epidemic. A similar concept may apply to ISV interactions with the mosquito host for arbovirus transmission [26]. We identified relatives of Phasi Charoen-like virus (PCLV) in *Anopheles* from Senegal and Cambodia. PCLV relatives also infect *Aedes*, where they were observed to reduce the replication of ZIKV and DENV arboviruses [30]. Palm Creek virus, an insect specific flavivirus, causes reduced replication of the West Nile virus and Murray Valley encephalitis arboviruses in *Aedes* cells [31]. In any case, ISV co-infection of mosquito vectors with *Plasmodium* and/or arboviruses in nature is highly probable as a general case, because all *Anopheles* sample pools in the current work were ISV-positive, so more research is warranted.

Third, characterization of the arthropod virome may shed light on the evolution of mosquito antiviral immune mechanisms, as well as the evolution of pathogenic arboviruses. ISV replication is restricted to insect cells, but the potential of most mosquito-associated viruses for transmission to humans or other vertebrates is currently unknown, because few studies of host range and transmission have been done. Some viruses may have a host range restricted to only *Anopheles*. For example, Anopheles cypovirus and Anopheles C virus replicate and are

maintained by vertical transmission in *An. coluzzii*, but were not able to infect *Ae. aegypti* in exposure experiments [4]. Both of these viruses were able to replicate in *Anopheles stephensi* after exposure, but Anopheles C virus was not stably maintained and disappeared after several generations. Thus, these two viruses may be *Anopheles*-specific, and possibly restricted only to certain *Anopheles* species.

It is likely that the main evolutionary pressure shaping mosquito antiviral mechanisms in general is their persistent exposure in nature to members of the natural virome, rather than the probably less frequent exposure to vertebrate-pathogenic arboviruses. Maintenance of bacterial microbiome commensals in the non-pathogenic commensal state requires active policing by basal host immunity [32]. By analogy, the maintenance of persistent ISVs as non-pathogenic may also result from a dialog with host immunity. Presumably, the same antiviral mechanisms used in basal maintenance of ISVs are also deployed against arboviruses when encountered, which are often in the same families as members of the insect virome [2]. Knowledge of the mechanisms that allow *Anopheles* to carry a natural RNA virome, but apparently reject arboviruses, may provide new tools to raise the barrier to arbovirus transmission by the more efficient *Aedes* and *Culex* vectors.

In addition to the canonical immune signaling pathways, piRNAs can be involved in antiviral protection, although this research is just beginning [18, 33]. One function of genomic piRNA clusters appears to be storage of a molecular archive of genomic threats such as transposable elements, linked to an effector

mechanism to inactivate them. This is analogous to bacterial molecular memory mediated by the CRISPR/Cas system. We identified two candidate piRNAs that are downregulated upon ONNV infection in *An. coluzzii*. Involvement of piRNAs during viral infection has not been previously demonstrated in *Anopheles*. piRNA monitoring of the virome may be part of the normal basal management of ISVs, which could potentially be pathogenic if not controlled, but more work is required to draw these connections.

The current report shows that the *Anopheles* virome is complex and diverse, and can be influenced by the geography of mosquito species. This is exemplified by the fact that some viruses are restricted to Senegalese *Anopheles* and others to *Anopheles* from Cambodia (Table3). Similar results were seen in *Ae. aegypti,* where five ISVs were specific to the Australian host population, while six others were found only in the Thai host population [34]. Differences in the *Anopheles* virome across geography could be explained by climate, environmental conditions, breeding sites, and mosquito bloodmeal sources, among other factors. The presence in this study of such a large number of novel and unclassified virus assemblies highlights the fact that the malaria vector virome is understudied. The same observation has been made during metagenomics surveys in *Drosophila*, *Aedes* and *Culex* [24, 35, 36] among other arthropods, indicating that the vast majority of insect viruses are not yet discovered.

Methods

Sample collections

Mosquitoes were collected in Cambodia in Kres village, Ratanakiri province (sample pools Cam5−02 and Cam10−02) and Cheav Rov village, Kampong Chnang province (sample pools Cam5−01 and Cam10−01). The majority of inhabitants are engaged in forest-related activities (agriculture, logging and hunting) and may spend the night in forest plots during the harvest period. Vegetation varies from evergreen forest to scattered forest, and the dry season typically runs from November to May and the rainy season from June to October. In Senegal, sampling sites were located in the department of Kedougou in southeastern Senegal. Kedougou lies in a transition zone between dry tropical forest and the savanna belt, and includes the richest and most diverse fauna of Senegal. Recent arbovirus outbreaks include Chikungunya in 2009-2010, Yellow Fever in 2011, Zika in 2010, and Dengue in 2008-2009.

Permission to collect mosquitoes was obtained by Institut Pasteur Cambodia from authorities of Ratanakiri and Kampong Chnang, and by Institut Pasteur Dakar from authorities of Kedougou. Wild mosquitoes visually identified as *Anopheles* spp. at the collection site (non-*Anopheles* were not retained) were immediately transferred into RNAlater stabilization reagent kept at 4°C, and then returned to the laboratory and stored at -80°C until RNA extraction.

RNA extraction, library construction, and sequencing

Total RNA was extracted from four pools of mosquitoes from each of Senegal and Cambodia (Senegal sample pools: 5 mosquitoes, Dak5-03, Dak5-04, 10

474 mosquitoes, Dak10-03, Dak10-04; Cambodia sample pools: 5 mosquitoes, Cam5-
475 01, Cam5-02, 10 mosquitoes, Cam10-01, Cam10-02) using the Nucleospin RNA kit
476 (Macherey-Nagel) following the supplied protocol. Library preparation and
477 sequencing steps were performed by Fasteris (Plan-les-Ouates, Switzerland,
478 www.fasteris.com). Long RNA libraries from the eight mosquito pools were made
479 from total RNA depleted of ribosomal RNA by treatment with RiboZero (Illumina,
480 San Diego, CA). Libraries were multiplexed and sequenced on a single lane of the
481 Illumina HiSeq 2500 platform (Illumina, San Diego, CA) by the paired-ends
482 method (2x125 bp), generating on average 36 million high-quality read pairs per
483 library. Small RNA libraries with insert size 18-30 nt were generated from the
484 same eight mosquito pools as above, multiplexed and sequenced in duplicate (two
485 technical replicates per pool) in two lanes of the Illumina HiSeq2500 platform
486 (Illumina, San Diego, CA) by the single-end method (1x50 bp) generating on
487 average 34 million reads of high-quality small RNA reads per library.
488
489 **Pre-processing of long and small RNA libraries**
490 Cutadapt 1.13 [37] was used for quality filtering and adaptor trimming of reads
491 from long and small RNA libraries. Low-quality 3' ends of long RNA reads were
492 trimmed by fixing a phred quality score of 15, and reads smaller than 50 bp after
493 quality filtering and adaptor trimming were removed. In the case of small RNA
494 libraries, reads shorter than 15 bp after quality filtering and adaptor trimming
495 were removed.
496
497 In order to filter sequences originating in the mosquito host, sequences passing
498 the above quality filter step were mapped against a custom database consisting of

499 24 *Anopheles* genomes available in Vectorbase in February 2016 [38]. Bowtie 1.2.0
500 [39] was used to map small RNA libraries with two mismatches allowed, whereas
501 the BWA-MEM algorithm from BWA-0.7.12 [40] with default parameters was used
502 to map long RNA libraries. Sequence reads that did not map against *Anopheles*
503 genomes, herein referred to as non-host processed reads, were retained and used
504 for de novo assembly and subsequent binning of virus transcripts.
505
506 **Estimation of Anopheles species composition of mosquito sample pools**
507 Quality-filtered long RNA read pairs were mapped with SortMeRNA [41] against a
508 custom database of *Anopheles* sequences of the mitochondrial cytochrome c
509 oxidase subunit 1 gene (COI-5P database) extracted from the Barcode of Life
510 database [42]. 98% identity and 98% alignment coverage thresholds were fixed
511 for the operational taxonomic unit (OTU) calling step of SortMeRNA. OTU counts
512 were collapsed at species level and relative abundances of *Anopheles* species with
513 at least 100 reads and 1% frequency in the sample pool were represented as
514 piecharts using the ggplots2 R package.
515
516 **De novo sequence assembly and identification of virus contigs by sequence**
517 **similarity**
518 Processed reads from each country (Cambodia and Senegal) were combined and
519 de novo assembled using different strategies for long and small RNA libraries.
520 Small RNA reads were assembled using the Velvet/Oases pipeline [43] using a
521 range of k-mer values from 13 to 35. Long RNA reads were assembled using both
522 the Velvet/Oases pipeline with a range of k-mer values from 11 to 67 and Trinity
523 [44].

Contigs produced by parallel assembly of Cambodia and Senegal processed reads were filtered in order to remove trans-self chimeric sequences using custom shell scripts, and the resulting contigs were merged with cd-hit-est [45] (95% nucleotide identity over 90% alignment length) in order to generate a final set of non-redundant contig sequences. Non-redundant contigs longer than 500 nucleotides were compared against the GenBank protein sequence reference database using BLASTX [46] with an e-value threshold of 1e-10, and the results were imported into MEGAN6 in order to classify contigs taxonomically using the LCA algorithm [47]. Contigs of viral origin were further manually curated by comparing their sequence with that of the closest virus reference genomes by using Artemis Comparison Tool [48].

Structural and functional annotation of virus assemblies

Assembled contigs of viral origin were annotated as follows: ORFs were predicted with MetaGeneMark [49], and functionally annotated using Prokka [50] with Virus kingdom as primary core reference database for initial BLASTP searches and including also as reference Hidden Markov Models (HMMS) of virus protein families defined in vFam database [51]. Also, protein sequences of predicted ORFs were processed with the Blast2GO pipeline [52], that generates functional annotation of proteins from BLASTP results against the virus subdivision of GenBank as well as Gene Ontology annotations from top BLASTP results. Prediction of InterPro signatures over viral proteins was also carried out with the InterProScan tool integrated in Blast2GO. The results of the different strategies of structural and functional annotation were integrated and manually curated with

549 Artemis [53].

550

551 **Prediction of unclassified contigs of viral origin by small RNA size profiling**

552 In order to recruit contigs of potential viral origin from the pool of unclassified
553 transcripts, we use the approach of Aguiar and collaborators [15]. This approach
554 uses the size profiles of small RNA reads that maps over positive and negative
555 strands of viruses detected by sequence similarity as a signature to identify
556 unclassified transcripts by sequence similarity of potential viral origin. For this
557 purpose, processed small RNA reads were re-mapped over virus contigs and
558 unclassified contigs by sequence similarity using bowtie 1.2.0 [39] allowing at
559 most one mismatch. From the mapped small RNA reads over each contig, the small
560 RNA size profiles were defined as the frequency of each small RNA read of size
561 from 15 to 35 nucleotides that map over the positive and negative strand of the
562 reference sequence. To compute these small RNA size profiles, reads mapped over
563 positive and negative strands of each reference sequence were extracted with
564 Samtools [54], and the size of small RNA reads were computed with the Infoseq
565 program of the EMBOSS package [55]. Custom shell scripts were used to parse
566 Infoseq output to a matrix representing the frequency of reads of different sizes
567 and polarity across virus/unclassified contigs. This matrix was further processed
568 in R (version 3.3.2). In order to normalize the small RNA size profiles, a z-score
569 transformation is applied over the read frequencies of each contig
570 (virus/unclassified). The similarity between small RNA size profiles of virus and
571 unclassified contigs is computed as the Pearson correlation coefficient of the
572 corresponding z-score profiles, and the relationship between small RNA size
573 profiles of virus/unclassified contigs was defined from this similarity values using

574 UPGMA as linkage criterion with the R package Phangorn [56]. These relationships
575 were visualized as heatmaps of the z-score profiles in R with gplots package
576 (version 3.0.1) using the UPGMA dendrogram as the clustering pattern of
577 virus/unclassified sequences. Unclassified contigs with a Pearson correlation
578 coefficient of at least 0.9 with virus contigs and coming from the same mosquito
579 sample pool were regrouped into clusters.
580
581 **Phylogenetic analyses**
582 In order to place the new virus sequences characterized in the present study into
583 an evolutionary context, the peptide sequences of RNA dependent RNA
584 polymerase ORFs detected in the annotation step were aligned with the
585 corresponding homologs in reference positive-sense and negative-sense single-
586 strand RNA viruses (ssRNA) and double strand RNA viruses (dsRNA) using MAFFT
587 v7.055b with the E-INS-i algorithm [57]. Independent alignments were generated
588 for all ssRNA and dsRNA viruses and for different virus families (Bunya-
589 Arenavirus, Monenegavirus, Orthomyxovivirus, Flavivirus, Reovirus). The
590 resulting alignments were trimmed with TrimAI [58] in order to remove highly
591 variable positions, keeping the most conserved domains for phylogenetic
592 reconstruction. Phylogenetic trees were reconstructed by maximum likelihood
593 with RAxML [59] with the WAG+GAMMA model of amino acid substitution and
594 100 bootstrap replicates. Phylogenetic trees were visualized with the R package
595 Ape [60].
596
597 **ONNV infection and candidate piRNA gene regulation**

598 Infection of *An. coluzzii* with ONNV, library preparations, and sequencing were
599 described [61]. Briefly, small RNA sequence reads from 2 pools of 12 mosquitoes
600 each fed an ONNV-infected bloodmeal (unfed mosquitoes removed), and 2 control
601 pools of 12 mosquitoes each fed an uninfected normal bloodmeal were mapped to
602 the *An. gambiae* PEST AgamP4 genome assembly using STAR version 2.5 with
603 default parameters [62]. The resulting SAM files were analyzed using
604 featureCounts [63] with default parameters to count mapped small RNAs
605 overlapping with previously annotated *An. coluzzii* piRNA genes in 187 genomic
606 piRNA clusters, in the file, GOL21-bonafide-piRNAs-24-29nt.fastq, from [19].
607 featureCounts considers a small RNA sequence read as overlapping a piRNA
608 feature if at least one base of the small RNA read overlaps the piRNA feature. Small
609 RNA sequence reads are not counted if they overlap more than one piRNA feature.
610 piRNAs in *An. coluzzii* are annotated by George et al. [19] as novel genes (denoted
611 XLOC loci) as well as piRNAs produced from loci within existing genes of the *An.*
612 *gambiae* PEST reference (AGAP loci). The Cuffdiff function in Cufflinks version
613 2.2.1 and DESeq2 version 1.20.0 were used to count and test for significant
614 differential expression levels between ONNV infected and control uninfected
615 samples, yielding 86 piRNA features that were potentially differentially
616 represented in the small RNA sequences between the ONNV and control treatment
617 conditions (Additional File 4: Table S2). The 86 candidates were filtered for a)
618 length of the contiguous region expressed in small RNA less than 40 nt, and b) in
619 the upper 10% of small RNA sequence read depth in all sequence samples
620 combined.

Declarations

Ethics approval and consent to participate

There were no human or animal subjects. Permission to collect wild mosquitoes was obtained by Institut Pasteur Cambodia from authorities of Ratanakiri and Kampong Chnang, Cambodia; and by Institut Pasteur Dakar from authorities of Kedougou, Senegal.

Consent for publication

Not applicable.

Availability of data and material

All sequence files are available from the EBI European Nucleotide Archive database (http://www.ebi.ac.uk/ena/) under study accession number [REQUESTED], and sample accession numbers: [REQUESTED]). All assembled sequences are available from NCBI (accession numbers: [REQUESTED]).

Competing interests

The authors declare that they have no competing interests.

Funding

This work received financial support to KDV from the European Commission, Horizon 2020 Infrastructures #731060 Infravec2; European Research Council, Support for frontier research, Advanced Grant #323173 AnoPath; and French Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" #ANR-10-LABX-62-IBEID, and support to IVS from USDA National Institute of

646 Food and Agriculture Hatch project #223822. The funders had no role in study
647 design, data collection and analysis, decision to publish, or preparation of the
648 manuscript.

649

650 **Authors' contributions**

651 Conceived and designed the experiments: EB, FNM, KE, GC, IH, MD, DD, AV, SK,
652 IVS, KDV

653 Performed the experiments: EB, FNM, KE, GC, IH, MD, DD, AV, SK

654 Analysed the data: EB, FNM, KE, IVS, KDV

655 Wrote the manuscript: EB, FNM, KE, IVS, KDV

656 All authors read and approved the final manuscript.

657

658 **Acknowledgements**

659 We acknowledge the assistance of Allan Dickerman, Jiyoung Lee and Song Li,
660 Virginia Polytechnic Institute and State University, and of Silke Jensen, Clermont
661 Université, Clermont-Ferrand, France, for helpful advice and assistance on piRNA
662 analysis.

References

1. World Health Organization: **World Malaria Report 2017**. In. Geneva:World Health Organization; 2017.

2. Nanfack Minkeu F, Vernick KD: **A Systematic Review of the Natural Virome of Anopheles Mosquitoes**. *Viruses* 2018, **10**(5).

3. Rezza G, Chen R, Weaver SC: **O'nyong-nyong fever: a neglected mosquito-borne viral disease**. *Pathog Glob Health* 2017, **111**(6):271-275.

4. Carissimo G, Eiglmeier K, Reveillaud J, Holm I, Diallo M, Diallo D, Vantaux A, Kim S, Menard D, Siv S *et al*: **Identification and Characterization of Two Novel RNA Viruses from Anopheles gambiae Species Complex Mosquitoes**. *PloS one* 2016, **11**(5):e0153881.

5. Colmant AMG, Hobson-Peters J, Bielefeldt-Ohmann H, van den Hurk AF, Hall-Mendelin S, Chow WK, Johansen CA, Fros J, Simmonds P, Watterson D *et al*: **A New Clade of Insect-Specific Flaviviruses from Australian Anopheles Mosquitoes Displays Species-Specific Host Restriction**. *mSphere* 2017, **2**(4).

6. Fauver JR, Grubaugh ND, Krajacich BJ, Weger-Lucarelli J, Lakin SM, Fakoli LS, 3rd, Bolay FK, Diclaro JW, 2nd, Dabire KR, Foy BD *et al*: **West African Anopheles gambiae mosquitoes harbor a taxonomically diverse virome including new insect-specific flaviviruses, mononegaviruses, and totiviruses**. *Virology* 2016, **498**:288-299.

7. Villinger J, Mbaya MK, Ouso D, Kipanga PN, Lutomiah J, Masiga DK: **Arbovirus and insect-specific virus discovery in Kenya by novel six genera multiplex high-resolution melting analysis**. *Mol Ecol Resour* 2017, **17**(3):466-480.

8. Tabue RN, Awono-Ambene P, Etang J, Atangana J, C AN, Toto JC, Patchoke S, Leke RG, Fondjo E, Mnzava AP *et al*: **Role of Anopheles (Cellia) rufipes (Gough, 1910) and other local anophelines in human malaria transmission in the northern savannah of Cameroon: a cross-sectional survey**. *Parasites & vectors* 2017, **10**(1):22.

9. Alam MS, Khan MG, Chaudhury N, Deloer S, Nazib F, Bangali AM, Haque R: **Prevalence of anopheline species and their Plasmodium infection status in epidemic-prone border areas of Bangladesh**. *Malar J* 2010, **9**:15.

10. Durnez L, Mao S, Denis L, Roelants P, Sochantha T, Coosemans M: **Outdoor malaria transmission in forested villages of Cambodia**. *Malar J* 2013, **12**:329.

11. Marasri N, Overgaard HJ, Sumarnrote A, Thanispong K, Corbel V, Chareonviriyaphap T: **Abundance and distribution of Anopheles mosquitoes in a malaria endemic area along the Thai-Lao border**. *Journal of vector ecology : journal of the Society for Vector Ecology* 2017, **42**(2):325-334.

12. St Laurent B, Oy K, Miller B, Gasteiger EB, Lee E, Sovannaroth S, Gwadz RW, Anderson JM, Fairhurst RM: **Cow-baited tents are highly effective in sampling diverse Anopheles malaria vectors in Cambodia**. *Malar J* 2016, **15**(1):440.

13. Service MW: **Mosquito ecology: Field sampling methods, 2nd ed.**, 2nd ed. edn. London: Chapman & Hall; 1993.

14. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, Qin XC, Xu J, Holmes EC, Zhang YZ: **Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses**. *Elife* 2015, **4**.

15. Aguiar ER, Olmo RP, Paro S, Ferreira FV, de Faria IJ, Todjro YM, Lobo FP, Kroon EG, Meignin C, Gatherer D *et al*: **Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host**. *Nucleic acids research* 2015, **43**(13):6191-6206.

16. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM *et al*: **Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies**. *Cell* 2009, **137**(3):509-521.

17. Lewis SH, Quarles KA, Yang Y, Tanguy M, Frezal L, Smith SA, Sharma PP, Cordaux R, Gilbert C, Giraud I *et al*: **Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements**. *Nat Ecol Evol* 2018, **2**(1):174-181.

18. Palatini U, Miesen P, Carballar-Lejarazu R, Ometto L, Rizzo E, Tu Z, van Rij RP, Bonizzoni M: **Comparative genomics shows that viral integrations are abundant and express piRNAs in the arboviral vectors Aedes aegypti and Aedes albopictus**. *BMC Genomics* 2017, **18**(1):512.

19. George P, Jensen S, Pogorelcnik R, Lee J, Xing Y, Brasset E, Vaury C, Sharakhov IV: **Increased production of piRNAs from euchromatic clusters and genes in Anopheles gambiae compared with Drosophila melanogaster**. *Epigenetics Chromatin* 2015, **8**:50.

20. Leger P, Lara E, Jagla B, Sismeiro O, Mansuroglu Z, Coppee JY, Bonnefoy E, Bouloy M: **Dicer-2- and Piwi-mediated RNA interference in Rift Valley fever virus-infected mosquito cells**. *J Virol* 2013, **87**(3):1631-1648.

21. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM: **Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma**. *PLoS Pathog* 2012, **8**(1):e1002470.

22. Castellano L, Rizzi E, Krell J, Di Cristina M, Galizi R, Mori A, Tam J, De Bellis G, Stebbing J, Crisanti A *et al*: **The germline of the malaria mosquito produces abundant miRNAs, endo-siRNAs, piRNAs and 29-nt small RNAs**. *BMC Genomics* 2015, **16**:100.

23. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS *et al*: **Redefining the invertebrate RNA virosphere**. *Nature* 2016.

24. Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G, Quintana JF, Brouqui JM, Bayne EH, Longdon B, Buck AH *et al*: **The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster**. *PLoS Biol* 2015, **13**(7):e1002210.

25. Cook S, Chung BY, Bass D, Moureau G, Tang S, McAlister E, Culverwell CL, Glucksman E, Wang H, Brown TD *et al*: **Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts**. *PloS one* 2013, **8**(11):e80720.

26. Hall RA, Bielefeldt-Ohmann H, McLean BJ, O'Brien CA, Colmant AM, Piyasena TB, Harrison JJ, Newton ND, Barnard RT, Prow NA *et al*: **Commensal Viruses of Mosquitoes: Host Restriction, Transmission, and Interaction with Arboviral Pathogens**. *Evol Bioinform Online* 2016, **12**(Suppl 2):35-44.

27. Huhtamo E, Cook S, Moureau G, Uzcategui NY, Sironen T, Kuivanen S, Putkuri N, Kurkela S, Harbach RE, Firth AE *et al*: **Novel flaviviruses from mosquitoes: mosquito-specific evolutionary lineages within the phylogenetic group of mosquito-borne flaviviruses**. *Virology* 2014, **464-465**:320-329.

28. Broderick NA, Buchon N, Lemaitre B: **Microbiota-induced changes in drosophila melanogaster host gene expression and gut morphology**. *MBio* 2014, **5**(3):e01117-01114.

29. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R: **Diversity, stability and resilience of the human gut microbiota**. *Nature* 2012, **489**(7415):220-230.

30. Schultz MJ, Frydman HM, Connor JH: **Dual Insect specific virus infection limits Arbovirus replication in Aedes mosquito cells**. *Virology* 2018, **518**:406-413.

31. Hobson-Peters J, Yam AW, Lu JW, Setoh YX, May FJ, Kurucz N, Walsh S, Prow NA, Davis SS, Weir R *et al*: **A new insect-specific flavivirus from northern Australia suppresses replication of West Nile virus and Murray Valley encephalitis virus in co-infected mosquito cells**. *PloS one* 2013, **8**(2):e56534.

32. Buchon N, Broderick NA, Lemaitre B: **Gut homeostasis in a microbial world: insights from Drosophila melanogaster**. *Nat Rev Microbiol* 2013, **11**(9):615-626.

33. Whitfield ZJ, Dolan PT, Kunitomi M, Tassetto M, Seetin MG, Oh S, Heiner C, Paxinos E, Andino R: **The Diversity, Structure, and Function of Heritable Adaptive Immunity Sequences in the Aedes aegypti Genome**. *Curr Biol* 2017, **27**(22):3511-3519 e3517.

34. Zakrzewski M, Rasic G, Darbro J, Krause L, Poo YS, Filipovic I, Parry R, Asgari S, Devine G, Suhrbier A: **Mapping the virome in wild-caught Aedes aegypti from Cairns and Bangkok**. *Sci Rep* 2018, **8**(1):4690.

35. Frey KG, Biser T, Hamilton T, Santos CJ, Pimentel G, Mokashi VP, Bishop-Lilly KA: **Bioinformatic Characterization of Mosquito Viromes within the Eastern United States and Puerto Rico: Discovery of Novel Viruses**. *Evol Bioinform Online* 2016, **12**(Suppl 2):1-12.

36. Atoni E, Wang Y, Karungu S, Waruhiu C, Zohaib A, Obanda V, Agwanda B, Mutua M, Xia H, Yuan Z: **Metagenomic Virome Analysis of Culex Mosquitoes from Kenya and China**. *Viruses* 2018, **10**(1).

37. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads**. *EMBnetjournal* 2011, **17**(1):3.

38. Giraldo-Calderon GI, Emrich SJ, MacCallum RM, Maslen G, Dialynas E, Topalis P, Ho N, Gesing S, VectorBase C, Madey G *et al*: **VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases**. *Nucleic acids research* 2015, **43**(Database issue):D707-713.

39. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome**. *Genome Biol* 2009, **10**(3):R25.

40. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM**. *arXiv preprint arXiv:13033997* 2013.

41. Kopylova E, Noe L, Touzet H: **SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data**. *Bioinformatics* 2012, **28**(24):3211-3217.

42. Ratnasingham S, Hebert PD: **bold: The Barcode of Life Data System (http://www.barcodinglife.org)**. *Mol Ecol Notes* 2007, **7**(3):355-364.

43. Schulz MH, Zerbino DR, Vingron M, Birney E: **Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels**. *Bioinformatics* 2012, **28**(8):1086-1092.

44. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome**. *Nature biotechnology* 2011, **29**(7):644-652.

45. Li W, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences**. *Bioinformatics* 2006, **22**(13):1658-1659.

46. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic acids research* 1997, **25**(17):3389-3402.

47. Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S, Ruscheweyh HJ, Tappu R: **MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data**. *PLoS Comput Biol* 2016, **12**(6):e1004957.

48. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J: **ACT: the Artemis Comparison Tool**. *Bioinformatics* 2005, **21**(16):3422-3423.

49. Zhu W, Lomsadze A, Borodovsky M: **Ab initio gene identification in metagenomic sequences**. *Nucleic acids research* 2010, **38**(12):e132.

50. Seemann T: **Prokka: rapid prokaryotic genome annotation**. *Bioinformatics* 2014, **30**(14):2068-2069.

51. Skewes-Cox P, Sharpton TJ, Pollard KS, DeRisi JL: **Profile hidden Markov models for the detection of viruses within metagenomic sequence data**. *PloS one* 2014, **9**(8):e105067.

52. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite**. *Nucleic acids research* 2008, **36**(10):3420-3435.

53. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation**. *Bioinformatics* 2000, **16**(10):944-945.

54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.

55. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite**. *Trends Genet* 2000, **16**(6):276-277.

56. Schliep KP: **phangorn: phylogenetic analysis in R**. *Bioinformatics* 2011, **27**(4):592-593.

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

58. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T: **trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses**. *Bioinformatics* 2009, **25**(15):1972-1973.

59. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. *Bioinformatics* 2014, **30**(9):1312-1313.

60. Paradis E, Claude J, Strimmer K: **APE: Analyses of Phylogenetics and Evolution in R language**. *Bioinformatics* 2004, **20**(2):289-290.

61. Carissimo G, Pain A, Belda E, Vernick KD: **Highly focused transcriptional response of Anopheles coluzzii to O'nyong nyong arbovirus during the primary midgut infection**. *BMC Genomics* 2018, **19**(1):526.

62. Dobin A, Gingeras TR: **Mapping RNA-seq Reads with STAR**. *Curr Protoc Bioinformatics* 2015, **51**:11 14 11-19.

63. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features**. *Bioinformatics* 2014, **30**(7):923-930.

Figure Legends

Figure 1. Taxonomic profile of *Anopheles* sample pools. Relative abundances of *Anopheles* species were computed from mapping of long-RNA reads over mitochondrial cytochrome C oxidase subunit I gene sequences (COI-5P) from Barcode of Life Database. Taxa represented by >100 sequence reads and 1% frequency in the sample pool were plotted in pie charts. White wedges represent the proportion of sequence matches present at less than 1% frequency. All data are presented in tabular form in Additional File: Table S1.

Figure 2. Phylogenetic tree of reference and novel virus assemblies from the *Bunyaviridae* family. Novel viruses characterized from Cambodia and Senegal *Anopheles* sample pools (red labels) are placed within the Phasmavirus clade and in a basal position of the Phebovirus-Tenuivirus clade.

Figure 3. Phylogenetic tree of reference and novel virus assemblies from the *Mononegavirales* order. A) Novel virus assemblies characterized from Cambodia and Senegal *Anopheles* sample pools (red labels) are predominantly placed within the Dimarhabdovirus clade and as close relative of the Nyamivirus clade. **B)** In this latter group close to Nyamivirus, the novel virus assemblies identified are close relatives of Xincheng mosquito virus, sharing a high degree of genome collinearity based on TBLASTX comparisons of novel and reference Xinxeng mosquito reference sequences.

Figure 4. Viral abundance profiles across mosquito sample pools based on small and long RNA sequence mapping. Heatmap of log2-transformed reads per

kilobase per million reads (RPKM) abundance values of novel virus assemblies identified from Cambodia and Senegal pools based on long and small RNA sequence libraries. Broadly similar viral abundance profiles are observed for different pools based on small and long RNA sequence data. Representation of particular viruses is uneven among pools, possibly indicating inter-individual mosquito differences for virus carriage.

Figure 5. Small RNA size profiles of novel virus assemblies from Cambodia and Senegal sample pools. Hierarchical clustering of novel virus assemblies based on Pearson correlation of z-score transformed small RNA size profiles (the frequency of small RNA reads of size 15 to 35 nucleotides that maps over the positive and negative strand of the reference sequence). Four main clusters were defined based on these small RNA size profiles, among which the classical siRNA size profile (21 nt reads mapping over positive and negative strand) is represented in the Cluster 3.

Supporting information

Additional File 1: Table S1. *Anopheles* mosquito taxa represented in the collections from Senegal and Cambodia, as detected by comparison to *Anopheles* sequences from the Barcode of Life COI-5P database. Data corresponds to pie charts of *Anopheles* taxa by country and sample pool depicted in Figure 1.

Additional File 2: Figure S1. Small RNA size profiles (A) and coverage profiles

925 **(B) of 15 novel virus assemblies with classic RNAi processing pattern.** Virus
926 assemblies shown are in Figure 5, Cluster 3, and are classified by sequence
927 similarity to known virus assemblies. Red vertical bars represent reads mapped
928 over the positive strand of reference viral sequence, and blue bars represent reads
929 mapped over the negative strand.
930
931 **Additional File 3: Figure S2. Small RNA size profiles of contigs left**
932 **unclassified by sequence similarity grouping.** Unclassified contigs that display
933 strong association by small RNA profile with Figure 5, Cluster 2 and Cluster 3. Red
934 bars represent reads mapped over the positive strand of reference viral sequence,
935 and blue bars represent reads mapped over the negative strand.
936
937 **Additional File 4: Table S2. Anopheles coluzzii piRNAs potentially**
938 **differentially represented in the small RNA sequences between the ONNV**
939 **and control treatment conditions.**

940 **Table 1. Summary of virus assemblies, Senegal *Anopheles* sample pools.**
941

Reference virus	NCBI classification reference virus	Closest relative	Assembled sequence	Length
DsRNA virus environmental sample clone mill.culi_contig84	Viruses; dsRNA viruses; environmental samples.	gi|766989332|gb|AJT39580.1| proline-alanine-rich protein [dsRNA virus environmental sample]	PrAlaRichProt_EnvVirDak	1345
Homalodisca vitripennis reovirus segment S3	Viruses; dsRNA viruses; Reoviridae; Sedoreovirinae; Phytoreovirus; unclassified Phytoreovirus	gi|226423326|ref|YP_002790886.1| major core protein [Homalodisca vitripennis reovirus]	CP_HVreovirusDak	5674
Daeseongdong virus 1 strain A12.2708/ROK/2012	Viruses; unclassified viruses.	gi|959121745|ref|YP_009182191.1| putative RNA-dependent RNA polymerase [Daeseongdong virus 1].	RdRP_DaeseondongVirDak	2530
Ixodes scapularis associated virus 2 isolate A1, partial genome	Viruses; unclassified viruses.	gi|669132782|gb|AII01812.1| hypothetical protein, partial [Ixodes scapularis associated virus 2]	HP1.1_IxodesVirDak	2820
			HP1.2_IxodesVirDak	2561
Uncultured virus isolate acc_7.4	Viruses; environmental samples.	gi|545716017|gb|AGW51759.1| RNA-dependent RNA polymerase-like protein [uncultured virus]	RdRP_UncVir1Dak	1488
Uncultured virus isolate acc_1.3	Viruses; environmental samples.	gi|545716010|gb|AGW51755.1| RNA-dependent RNA polymerase-like protein, partial [uncultured virus]	RdRP_UncVir2Dak	2011
American dog tick phlebovirus isolate FI3	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Bunyaviridae; Phlebovirus; unclassified Phlebovirus.	gi|734669629|gb|AJA31764.1| nucleocapsid, partial [American dog tick phlebovirus]	NuclCap1.1_ADTphlebovirusDak	1105
			NuclCap1.2_ADTphlebovirusDak	1148
Culex tritaeniorhynchus rhabdovirus RNA, complete genome, strain:TY	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; unclassified Rhabdoviridae.	gi|700895640|ref|YP_009094323.1| large protein [Culex tritaeniorhynchus rhabdovirus]	LP_CulexRhabdovDak	1526
Phasi Charoen-like virus	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Bunyaviridae; unclassified Bunyaviridae.	gi|664682120|gb|AIF71032.1| nucleocapsid [Phasi Charoen-like virus]	NuclCap1.1_PCLVDak	1187
			NuclCap1.2_PCLVDak	1104
			NuclCap1.3_PCLVDak	1125
			NuclCap1.4_PCLVDak	1144

		gi|870898373|gb|AKP18600.1| glycoprotein [Phasi Charoen-like virus]	GP_PCLVDak	3887
		gi|664682116|gb|AIF71030.1| RNA-dependent RNA polymerase [Phasi Charoen-like virus]	RdRP_PCLVDak	6711
Wellfleet Bay virus isolate 10-280-G segment 4	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Orthomyxoviridae; Quaranjavirus; unclassified Quaranjavirus.	gi|727361119|ref|YP_009110683.1| nucleoprotein [Wellfleet Bay virus]	Nuclprot1.1_WBvirDak	1973
			Nuclprot1.2_WBvirDak	3252
Wuhan Mosquito Virus 9 strain JX1-13	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.	gi|752455731|gb|AJG39214.1| ORF1 [Wuhan Mosquito Virus 9]	ORF1_Wuhan9virDak	1574
Wuhan Mosquito Virus 1 strain WT3-15	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.	gi|752455880|gb|AJG39296.1| glycoprotein precursor [Wuhan Mosquito Virus 1].	GP_Wuhan1virDak	3547
Wuhan Spider Virus strain SYZZ-2	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.	gi|752455826|gb|AJG39269.1| RNA-dependent RNA polymerase [Wuhan Spider Virus]	RdRP1.1_WSVDak	998
			RdRP1.2_WSVDak	2083
			RdRP1.3_WSVDak	1070
Xincheng Mosquito Virus strain XC1-6	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.	gi|752455743|gb|AJG39224.1| ORF1 [Xincheng Mosquito Virus]	ORF1_XinchengVirDak	947
		gi|752455744|gb|AJG39225.1| ORF2 [Xincheng Mosquito Virus]	ORF2_XinchengVirDak	1726
		gi|752455745|gb|AJG39226.1| glycoprotein [Xincheng Mosquito Virus]	GP_XinchengVirDak	5993
		gi|752455746|gb|AJG39227.1| RNA-dependent RNA polymerase [Xincheng Mosquito Virus]	RdRP1.1_XinchengVirDak	11707
			RdRP1.2_XinchengVirDak	11722
			RdRP1.3_XinchengVirDak	11710
			RdRP1.4_XinchengVirDak	11694
			RdRP1.5_XinchengVirDak	11728
			RdRP1.6_XinchengVirDak	11716
			RdRP1.7_XinchengVirDak	6128

Xinzhou Mosquito Virus strain XC3-5	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.	gi|752455830|gb|AJG39271.1| RNA-dependent RNA polymerase [Xinzhou Mosquito Virus].	RdRP1.1_XinzhouVirDak	7527
			RdRP1.2_XinzhouVirDak	7524
Sunn-hemp mosaic virus	Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Virgaviridae; Tobamovirus.	gi|12643499|sp|P89202.2|RDRP_SHMV RecName: Full=Replicase large subunit]	RdRP_SHMVDak	1216
Omono River virus	Viruses; dsRNA viruses	gi|307933351|dbj|BAJ21511.1| RNA-dependent RNA polymerase [Omono River virus]	RdRP_OmonoVirDak	613
Jurona virus	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; Vesiculovirus.	gi|701219310|ref|YP_009094377.1| polymerase [Jurona virus]	RdRP_JuronaVirDak	818
Beaumont virus strain 6	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; unclassified Rhabdoviridae.	gi|550631504|gb|AGX86091.1| RNA-dependent RNA polymerase, partial [Beaumont virus]	RdRP_BeaumontVirDak	805

942

943

944 **Table 2. Summary of virus assemblies, Cambodia *Anopheles* sample pools.**
945

Reference virus	NCBI classification reference virus	Closest relative	Assembled sequence	Length
uncultured virus	Viruses; environmental samples.	RNA-dependent RNA polymerase-like protein, partial [uncultured virus] (KF298266.1)	vcambTR48403_c0_g1_i2	857
			RdRP1.2_UncVir2Camb	797
			RdRP1.3_UncVir2Camb	2177
			RdRP1.4_UncVir2Camb	2574
			RdRP1.5_UncVir2Camb	2722
Culex tritaeniorhynchus rhabdovirus RNA, complete genome; NC_025384	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; unclassified Rhabdoviridae.	gi|700895639|ref|YP_009094322.1| glycoprotein [Culex tritaeniorhynchus rhabdovirus]	GP_CulexRhabdovCamb	2866
		gi|700895640|ref|YP_009094323.1| large protein [Culex tritaeniorhynchus rhabdovirus]	LP1.1_CulexRhabdovCamb	755
			LP1.2_CulexRhabdovCamb	1454
Phasi Charoen-like virus	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Bunyaviridae; unclassified Bunyaviridae	gi|870898373|gb|AKP18600.1| glycoprotein [Phasi Charoen-like virus]	GP1.1_PCLVCamb	642
			GP1.2_PCLVCamb	933
		gi|870898376|gb|AKP18601.1| Nucleocapsid [Phasi Charoen-like virus]	NuclCap1.1_PCLVCamb	1104
			NuclCap1.2_PCLVCamb	533
			NuclCap1.3_PCLVCamb	535
			NuclCap1.4_PCLVCamb	2157
Bivens Arm virus isolate UF 10	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; Tibrovirus	gi|751997168|gb|AJG05818.1| nucleoprotein N [Bivens Arm virus]	NProt_BivArmsVirCamb	516
Jurona virus	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; Vesiculovirus.	gi|701219310|ref|YP_009094377.1| polymerase [Jurona virus] >gnl|... 176 2e-43	RdRP_JuronaVirCamb	1329
Puerto Almendras virus isolate LO-39	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; unclassified Rhabdoviridae.	gi|701219331|ref|YP_009094394.1| L protein [Puerto Almendras vir... 213 2e-54	Lprot1.1_PAvirCamb	1869
			Lprot1.2_PAvirCamb5	3895

		gi|701219327|ref|YP_009094389.1| N protein [Puerto Almendras virus]	Nprot_PAvirCamb	4449
Beaumont virus strain 6	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; unclassified Rhabdoviridae.	gi|550631504|gb|AGX86091.1| RNA-dependent RNA polymerase, partial [Beaumont virus]	RdRP1.1_BeaumontVirCamb	586
			RdRP1.2_BeaumontVirCamb	633
			RdRP1.3_BeaumontVirCamb	594
			RdRP1.4_BeaumontVirCamb	1359
			RdRP1.5_BeaumontVirCamb	1606
			RdRP1.6_BeaumontVirCamb	1141
			RdRP1.7_BeaumontVirCamb	1667
Wellfleet Bay virus isolate 10-280-G segment 4	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Orthomyxoviridae; Quaranjavirus; unclassified Quaranjavirus.	gi|727361119|ref|YP_009110683.1| nucleoprotein [Wellfleet Bay virus]	Nuclprot1.1_WBvirCamb	1011
			Nuclprot1.2_WBvirCamb	1139
			Nuclprot1.3_WBvirCamb	2942
Xinzhou Mosquito Virus strain XC3-5	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.	gi|752455830|gb|AJG39271.1| RNA-dependent RNA polymerase [Xinzhou Mosquito Virus]	RdRP_XinzhouVirCamb	8129
Wuhan Mosquito Virus 1 strain WT3-15	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.	gi|752455822|gb|AJG39267.1| RNA-dependent RNA polymerase [Wuhan Mosquito Virus 1]	RdRP1.1_Wuhan1virCamb	3576
			RdRP1.2_Wuhan1virCamb	2929
			RdRP1.3_Wuhan1virCamb	3943
			RdRP1.4_Wuhan1virCamb	686
			RdRP1.5_Wuhan1virCamb	518
			RdRP1.6_Wuhan1virCamb	6431
			RdRP1.7_Wuhan1virCamb	6435
			GP1.1_Wuhan1virCamb	523

		gi|752455880|gb|AJG39296.1|glycoprotein precursor [Wuhan Mosquito Virus 1]	GP1.2_Wuhan1virCamb	1127
			GP1.3_Wuhan1virCamb	1282
			GP1.4_Wuhan1virCamb	2434
			GP1.5_Wuhan1virCamb	2231
			GP1.6_Wuhan1virCamb	2205
			GP1.7_Wuhan1virCamb	2219
		gi|752455945|gb|AJG39330.1| nucleopasid protein [Wuhan Mosquito Virus 1]	NuclCap1.1_Wuhan1virCamb	645
			NuclCap1.2_Wuhan1virCamb	735
			NuclCap1.3_Wuhan1virCamb	629
			NuclCap1.4_Wuhan1virCamb	546
			NuclCap1.5_Wuhan1virCamb	549
			NuclCap1.6_Wuhan1virCamb	1209
			NuclCap1.7_Wuhan1virCamb	1259
			NuclCap1.8_Wuhan1virCamb	1015
			NuclCap1.9_Wuhan1virCamb	3081
			NuclCap1.10_Wuhan1virCamb	1473
			NuclCap1.11_Wuhan1virCamb	1791
			NuclCap1.12_Wuhan1virCamb	2147
Wuhan Mosquito Virus 9 strain JX1-13	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.	gi|752455734|gb|AJG39217.1| glycoprotein [Wuhan Mosquito Virus 9]	GP1.1_Wuhan9virCamb	658
			GP1.2_Wuhan9virCamb	924
			GP1.3_Wuhan9virCamb	2429
			ORF1.1_Wuhan9virCamb	1872

		gi|752455731|gb|AJG39214.1| ORF1 [Wuhan Mosquito Virus 9]	ORF1.2_Wuhan9virCamb	1625
			ORF1.3_Wuhan9virCamb	1202
Xincheng Mosquito Virus strain XC1-6	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses	gi|752455743|gb|AJG39224.1| ORF1 [Xincheng Mosquito Virus]	ORF1_XinchengVirCamb	1329
		gi|752455745|gb|AJG39226.1| glycoprotein [Xincheng Mosquito Virus]	GP1.1_XinchengVirCamb	509
			GP1.2_XinchengVirCamb	953
			GP1.3_XinchengVirCamb	1635
			GP1.4_XinchengVirCamb	1298
			GP1.5_XinchengVirCamb	1313
			GP1.6_XinchengVirCamb	3076
			GP1.7_XinchengVirCamb	1314
			GP1.8_XinchengVirCamb	2660
			GP1.9_XinchengVirCamb	1757
		gi|752455746|gb|AJG39227.1| RNA-dependent RNA polymerase [Xincheng Mosquito Virus]	RdRP1.1_XinchengVirCamb	925
			RdRP1.2_XinchengVirCamb	904
			RdRP1.3_XinchengVirCamb	991
			RdRP1.4_XinchengVirCamb	1065
			RdRP1.5_XinchengVirCamb	1354
			RdRP1.6_XinchengVirCamb	2062
			RdRP1.7_XinchengVirCamb	3974
Nienokoue virus isolate B51/CI/2004	Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae	gi|655454925|ref|YP_009041466.1| polyprotein [Nienokoue virus]	PolProt1.1_FlavivirusCamb	1008
			PolProt1.2_FlavivirusCamb	2193
			PolProt1.3_FlavivirusCamb	1010

			PolProt1.4_FlavivirusCamb	10610
Tobacco streak virus isolate pumpkin segment RNA1	Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae; Ilarvirus	gi|254554401|gb|ACT67442.1| replicase [Tobacco streak virus]	Replicase_TSvirCamb	1565
Oat golden stripe virus RNA1	Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Virgaviridae; Furovirus	gi|9635455|ref|NP_059511.1| replicase [Oat golden stripe virus]	Replicase_OatGSvirCamb	1661

946
947

948 **Table 3. Similarity of Senegal and Cambodia virus assemblies by BLASTX to**

949 **24 reference viruses in GenBank.** 10 targets are shared, 9 are Senegal-specific,

950 and 5 are Cambodia-specific.

Reference virus	Viral taxonomy	Senegal Libraries	Cambodia Libraries
Culex tritaeniorhynchus rhabdovirus RNA, complete genome	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; unclassified Rhabdoviridae.		
Phasi Charoen-like virus	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Bunyaviridae; unclassified Bunyaviridae.		
Uncultured virus isolate acc_1.3	Viruses; environmental samples.		
Wellfleet Bay virus isolate 10-280-G segment 4	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Orthomyxoviridae; Quaranjavirus; unclassified Quaranjavirus.		
Wuhan Mosquito Virus 1 strain WT3-15	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.		
Wuhan Mosquito Virus 9 strain JX1-13	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.		
Xincheng Mosquito Virus strain XC1-6	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.		
Xinzhou Mosquito Virus strain XC3-5	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.		
Beaumont virus strain 6	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; unclassified Rhabdoviridae.		
Jurona virus	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; Vesiculovirus.		
Omono River virus	Viruses; dsRNA viruses		
American dog tick phlebovirus isolate FI3	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Bunyaviridae; Phlebovirus; unclassified Phlebovirus.		
Daeseongdong virus 1 strain A12.2708/ROK/2012	Viruses; unclassified viruses.		
DsRNA virus environmental sample clone mill.culi_contig84	Viruses; dsRNA viruses; environmental samples.		
Homalodisca vitripennis reovirus segment S3	Viruses; dsRNA viruses; Reoviridae; Sedoreovirinae; Phytoreovirus; unclassified Phytoreovirus		
Ixodes scapularis associated virus 2 isolate A1, partial genome	Viruses; unclassified viruses.		
Sunn-hemp mosaic virus	Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Virgaviridae; Tobamovirus.		
Uncultured virus isolate acc_7.4	Viruses; environmental samples.		
Wuhan Spider Virus strain SYZZ-2	Viruses; ssRNA viruses; ssRNA negative-strand viruses; unclassified ssRNA negative-strand viruses.		
Nienokoue virus isolate B51/CI/2004	Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae		
Oat golden stripe virus RNA1	Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Virgaviridae; Furovirus		
Puerto Almendras virus isolate LO-39	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; unclassified Rhabdoviridae.		
Tobacco streak virus isolate pumpkin segment RNA1	Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae; Ilarvirus		
Bivens Arm virus isolate UF 10	Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; Tibrovirus		

951

FIGURE 1

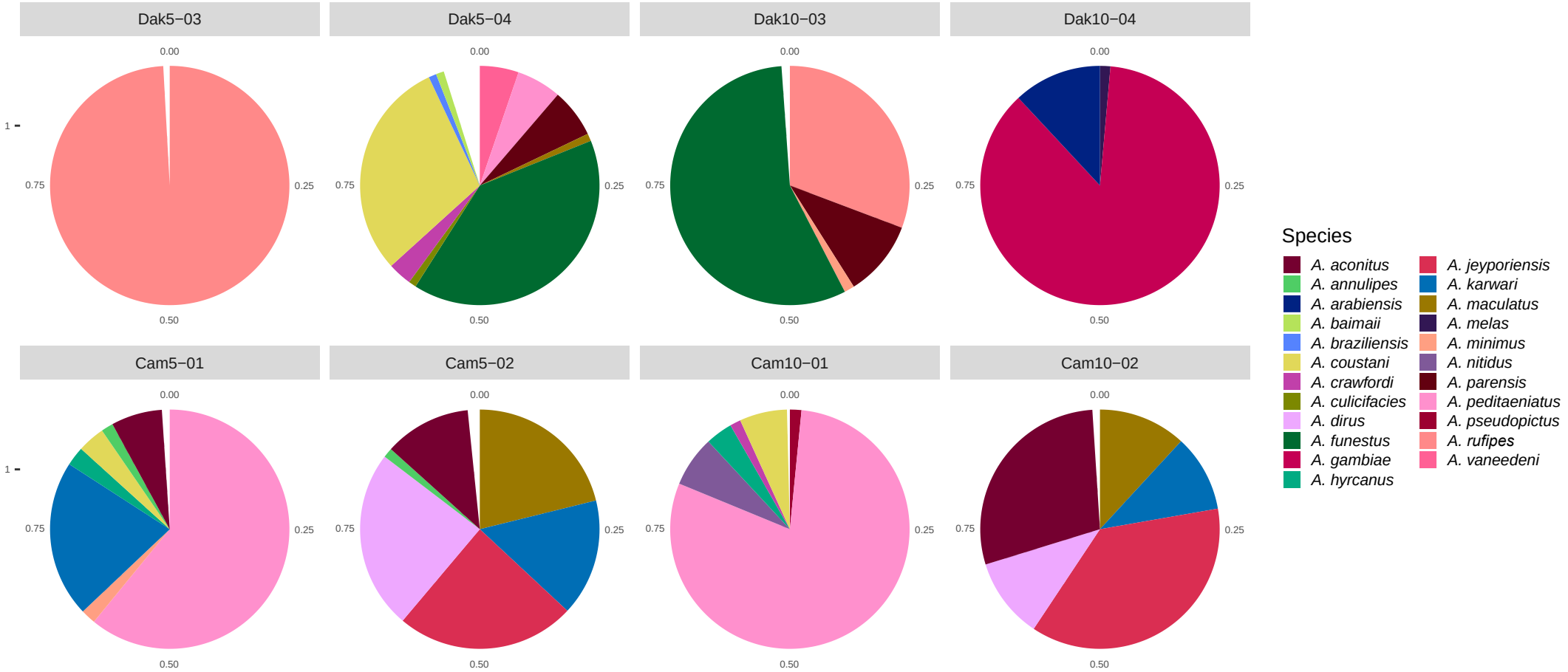

Dak5−03
Dak5−04
Dak10−03
Dak10−04
Cam5−01
Cam5−02
Cam10−01
Cam10−02
0.00
0.25
0.50
0.75
1
Species
A. aconitus
A. annulipes
A. arabiensis
A. baimaii
A. braziliensis
A. coustani
A. crawfordi
A. culicifacies
A. dirus
A. funestus
A. gambiae
A. hyrcanus
A. jeyporiensis
A. karwari
A. maculatus
A. melas
A. minimus
A. nitidus
A. parensis
A. peditaeniatus
A. pseudopictus
A. rufipes
A. vaneedeni

FIGURE 2

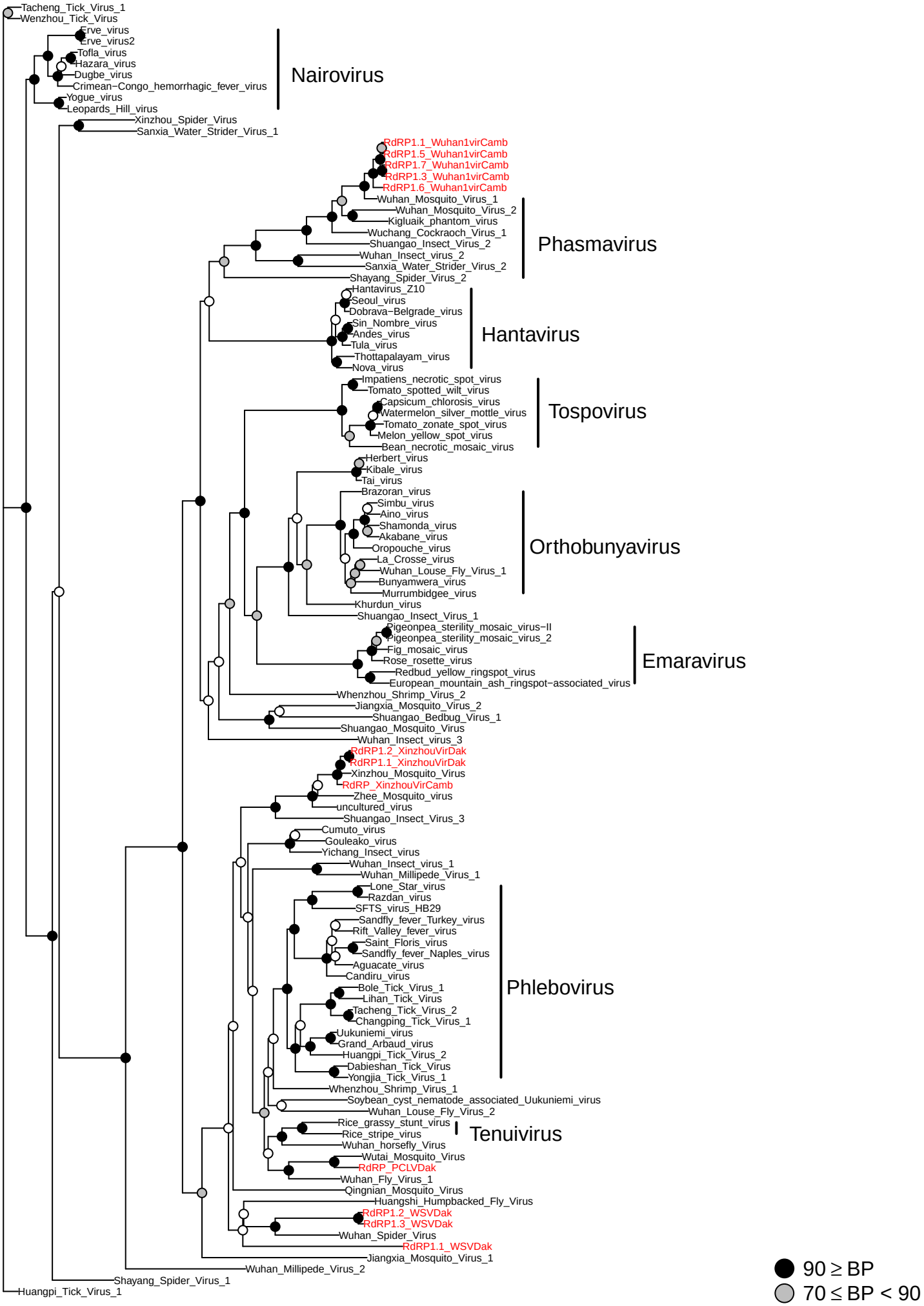
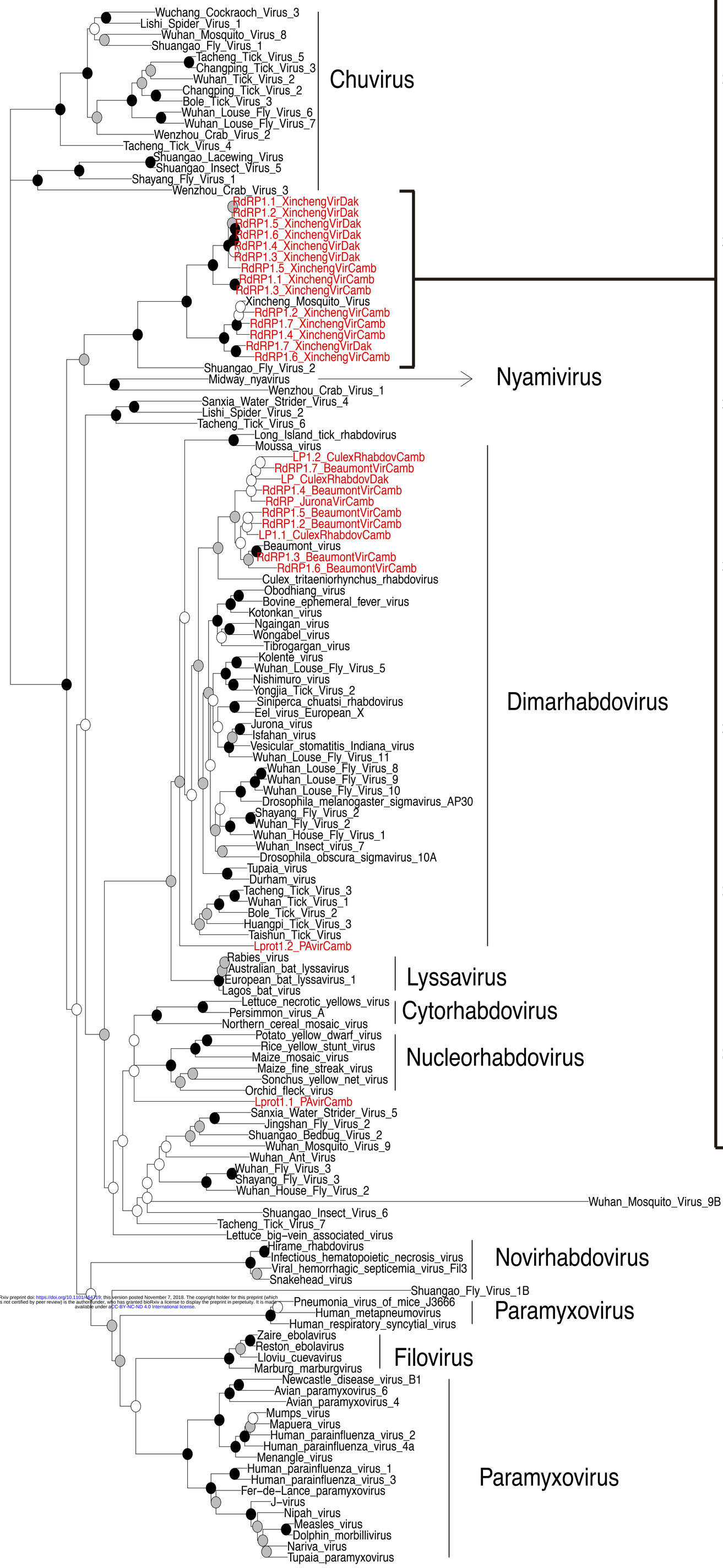
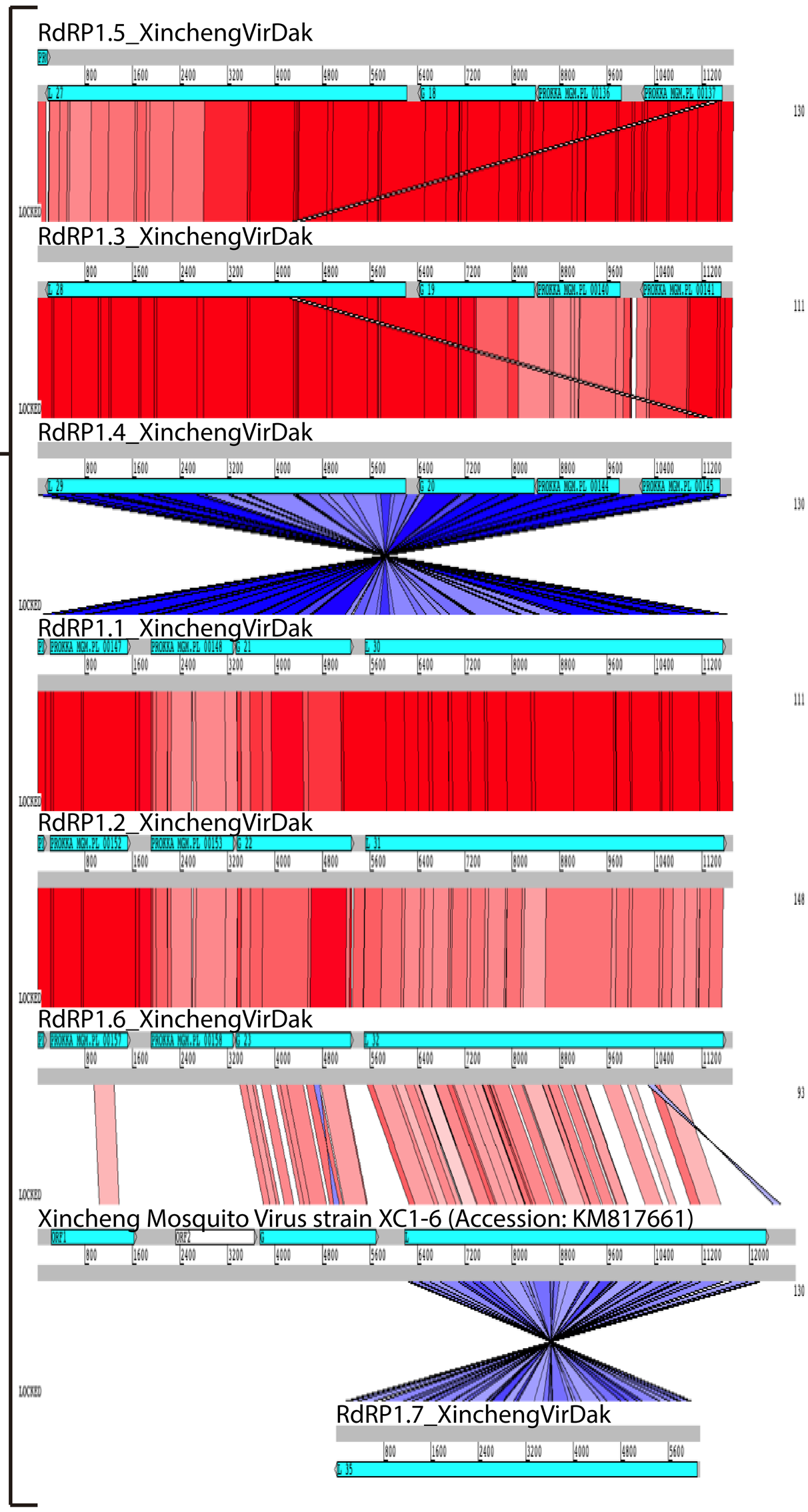


FIGURE 3

A



B



90 ≥ BP

60 ≤ BP < 90

BP < 60

FIGURE 4

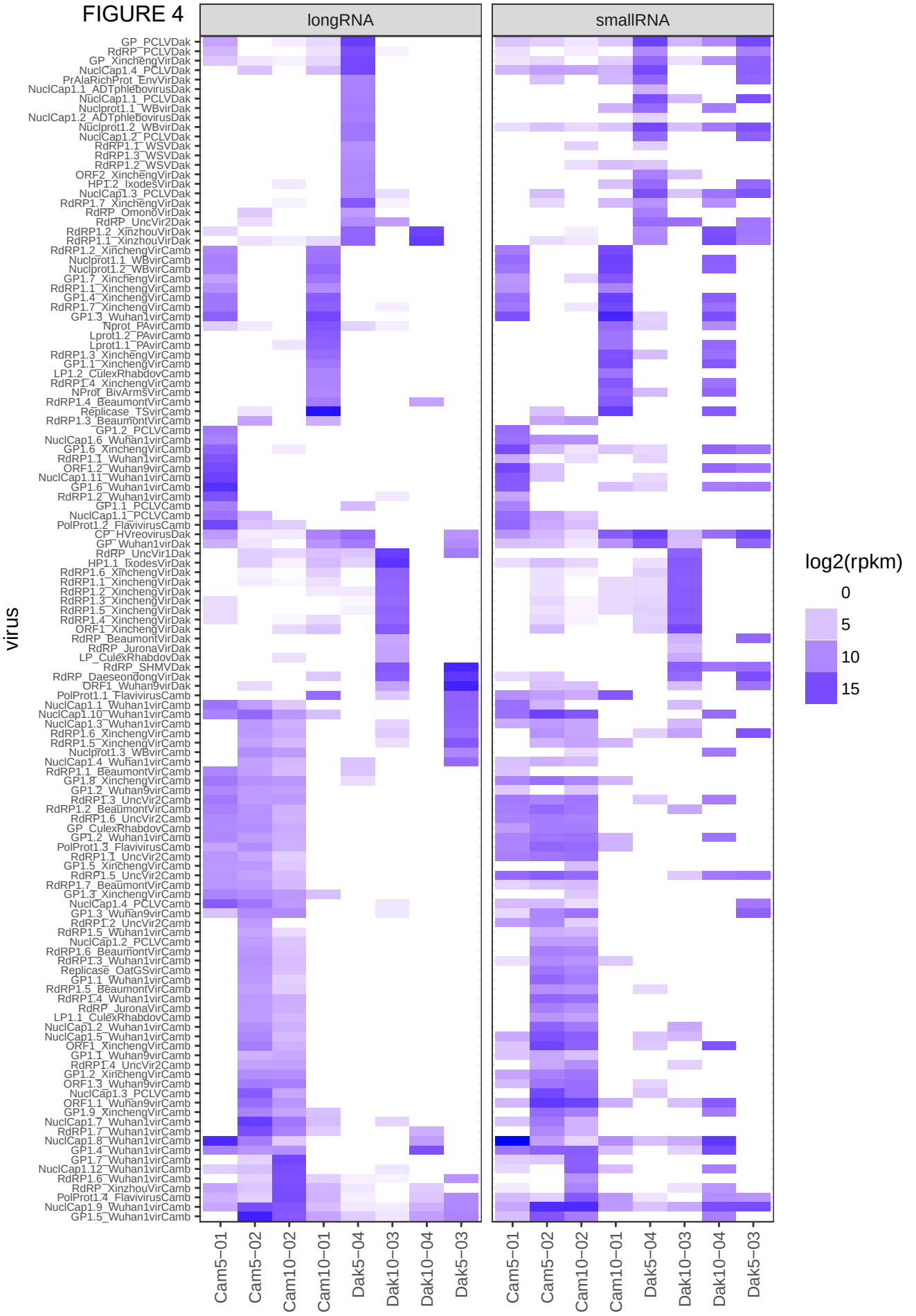
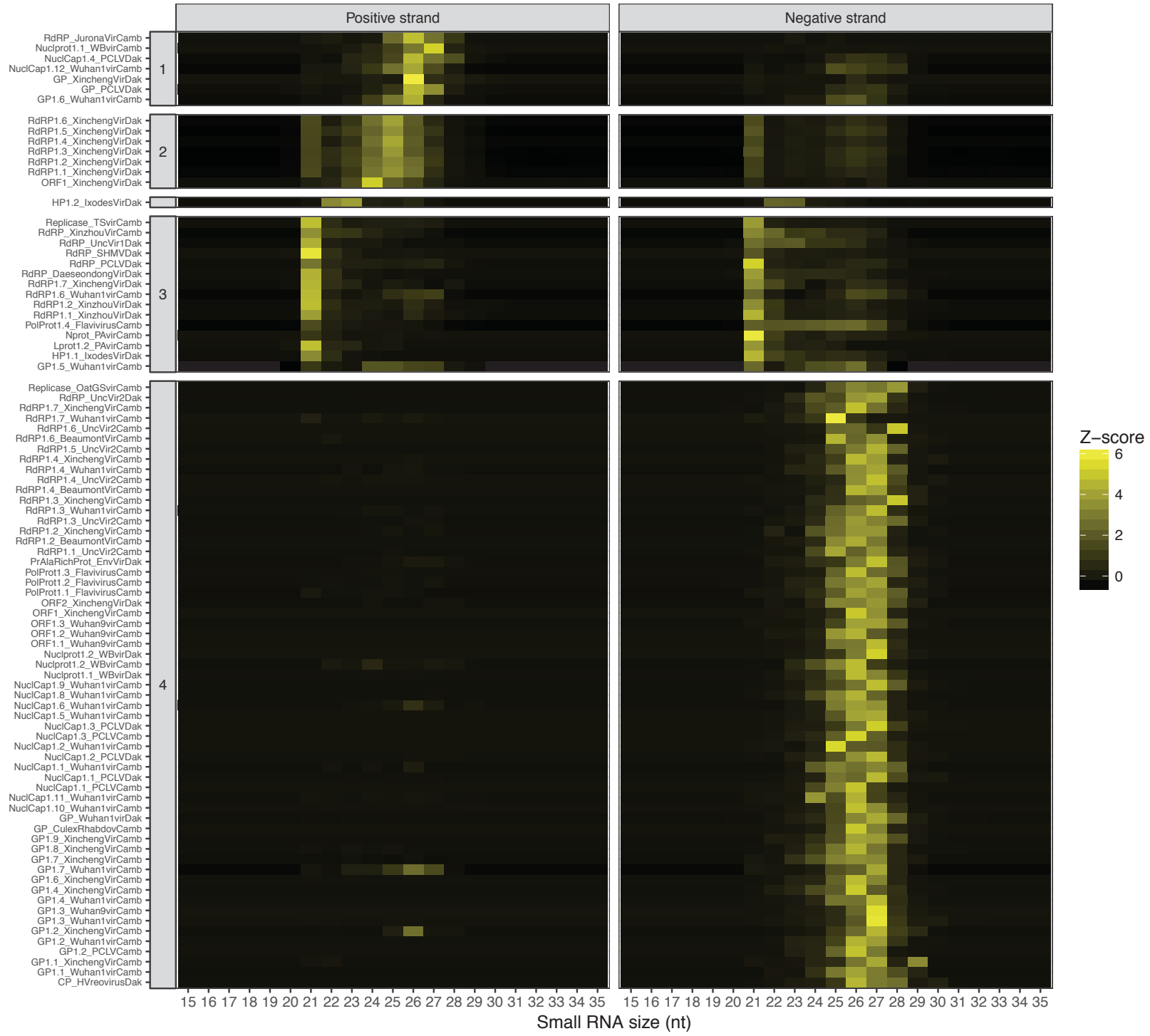

longRNA
smallRNA
virus
log2(rpkm)
0
5
10
15
GP_PCLVDak
RdRP_PCLVDak
GP_XinchengVirDak
NuclCap1.4_PCLVDak
PrAlaRichProt_EnvVirDak
NuclCap1.1_ADTphlebovirusDak
NuclCap1.1_PCLVDak
Nuclprot1.1_WBvirDak
NuclCap1.2_ADTphlebovirusDak
Nuclprot1.2_WBvirDak
NuclCap1.2_PCLVDak
RdRP1.1_WSVDak
RdRP1.3_WSVDak
RdRP1.2_WSVDak
ORF2_XinchengVirDak
HP1.2_IxodesVirDak
NuclCap1.3_PCLVDak
RdRP1.7_XinchengVirDak
RdRP_OmonoVirDak
RdRP_UncVir2Dak
RdRP1.2_XinzhouVirDak
RdRP1.1_XinzhouVirDak
RdRP1.2_XinchengVirCamb
Nuclprot1.1_WBvirCamb
Nuclprot1.2_WBvirCamb
GP1.7_XinchengVirCamb
RdRP1.1_XinchengVirCamb
GP1.4_XinchengVirCamb
RdRP1.7_XinchengVirCamb
GP1.3_Wuhan1virCamb
Nprot_PAvirCamb
Lprot1.2_PAvirCamb
Lprot1.1_PAvirCamb
RdRP1.3_XinchengVirCamb
GP1.1_XinchengVirCamb
LP1.2_CulexRhabdovCamb
RdRP1.4_XinchengVirCamb
NProt_BivArmsVirCamb
RdRP1.4_BeaumontVirCamb
Replicase_TSvirCamb
RdRP1.3_BeaumontVirCamb
GP1.2_PCLVCamb
NuclCap1.6_Wuhan1virCamb
GP1.6_XinchengVirCamb
RdRP1.1_Wuhan1virCamb
ORF1.2_Wuhan9virCamb
NuclCap1.11_Wuhan1virCamb
GP1.6_Wuhan1virCamb
RdRP1.2_Wuhan1virCamb
GP1.1_PCLVCamb
NuclCap1.1_PCLVCamb
PolProt1.2_FlavivirusCamb
CP_HVreovirusDak
GP_Wuhan1virDak
RdRP_UncVir1Dak
HP1.1_IxodesVirDak
RdRP1.6_XinchengVirDak
RdRP1.1_XinchengVirDak
RdRP1.2_XinchengVirDak
RdRP1.3_XinchengVirDak
RdRP1.5_XinchengVirDak
RdRP1.4_XinchengVirDak
ORF1_XinchengVirDak
RdRP_BeaumontVirDak
RdRP_JuronaVirDak
LP_CulexRhabdovDak
RdRP_SHMVDak
RdRP_DaeseondongVirDak
ORF1_Wuhan9virDak
PolProt1.1_FlavivirusCamb
NuclCap1.1_Wuhan1virCamb
NuclCap1.10_Wuhan1virCamb
NuclCap1.3_Wuhan1virCamb
RdRP1.6_XinchengVirCamb
RdRP1.5_XinchengVirCamb
Nuclprot1.3_WBvirCamb
NuclCap1.4_Wuhan1virCamb
RdRP1.1_BeaumontVirCamb
GP1.8_XinchengVirCamb
GP1.2_Wuhan9virCamb
RdRP1.3_UncVir2Camb
RdRP1.2_BeaumontVirCamb
RdRP1.6_UncVir2Camb
GP_CulexRhabdovCamb
GP1.2_Wuhan1virCamb
PolProt1.3_FlavivirusCamb
RdRP1.1_UncVir2Camb
GP1.5_XinchengVirCamb
RdRP1.5_UncVir2Camb
RdRP1.7_BeaumontVirCamb
GP1.3_XinchengVirCamb
NuclCap1.4_PCLVCamb
GP1.3_Wuhan9virCamb
RdRP1.2_UncVir2Camb
RdRP1.5_Wuhan1virCamb
NuclCap1.2_PCLVCamb
RdRP1.6_BeaumontVirCamb
RdRP1.3_Wuhan1virCamb
Replicase_OatGSvirCamb
GP1.1_Wuhan1virCamb
RdRP1.5_BeaumontVirCamb
RdRP1.4_Wuhan1virCamb
RdRP_JuronaVirCamb
LP1.1_CulexRhabdovCamb
NuclCap1.2_Wuhan1virCamb
NuclCap1.5_Wuhan1virCamb
ORF1_XinchengVirCamb
GP1.1_Wuhan9virCamb
RdRP1.4_UncVir2Camb
GP1.2_XinchengVirCamb
ORF1.3_Wuhan9virCamb
NuclCap1.3_PCLVCamb
ORF1.1_Wuhan9virCamb
GP1.9_XinchengVirCamb
NuclCap1.7_Wuhan1virCamb
RdRP1.7_Wuhan1virCamb
NuclCap1.8_Wuhan1virCamb
GP1.4_Wuhan1virCamb
GP1.7_Wuhan1virCamb
NuclCap1.12_Wuhan1virCamb
RdRP1.6_Wuhan1virCamb
RdRP_XinzhouVirCamb
PolProt1.4_FlavivirusCamb
NuclCap1.9_Wuhan1virCamb
GP1.5_Wuhan1virCamb
Cam5−01
Cam5−02
Cam10−02
Cam10−01
Dak5−04
Dak10−03
Dak10−04
Dak5−03
Cam5−01
Cam5−02
Cam10−02
Cam10−01
Dak5−04
Dak10−03
Dak10−04
Dak5−03

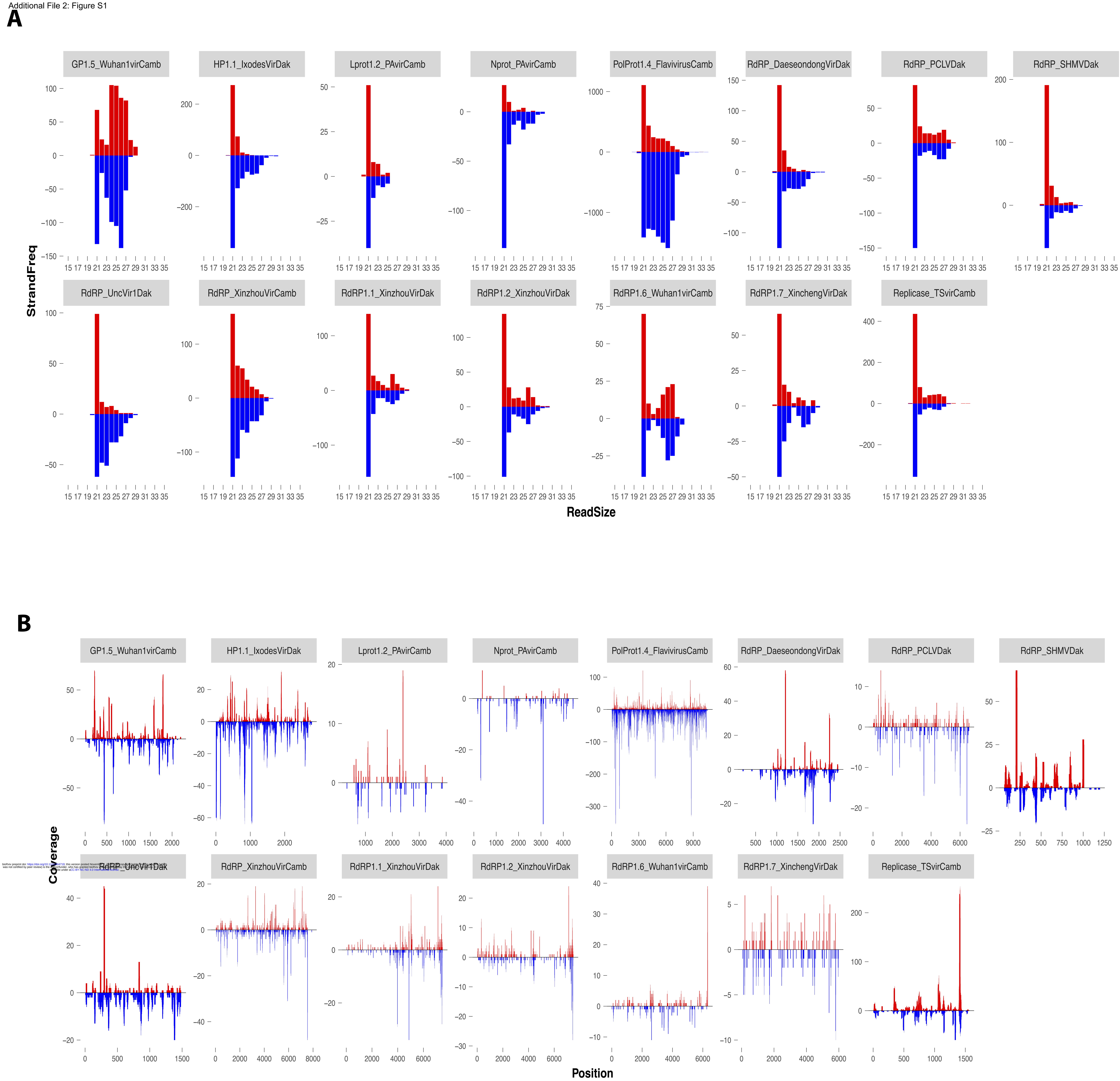
FIGURE 5


Positive strand
Negative strand
RdRP_JuronaVirCamb
Nuclprot1.1_WBvirCamb
NuclCap1.4_PCLVDak
NuclCap1.12_Wuhan1virCamb
GP_XinchengVirDak
GP_PCLVDak
GP1.6_Wuhan1virCamb
1
RdRP1.6_XinchengVirDak
RdRP1.5_XinchengVirDak
RdRP1.4_XinchengVirDak
RdRP1.3_XinchengVirDak
RdRP1.2_XinchengVirDak
RdRP1.1_XinchengVirDak
ORF1_XinchengVirDak
2
HP1.2_IxodesVirDak
Replicase_TSvirCamb
RdRP_XinzhouVirCamb
RdRP_UncVir1Dak
RdRP_SHMVDak
RdRP_PCLVDak
RdRP_DaeseondongVirDak
RdRP1.7_XinchengVirDak
RdRP1.6_Wuhan1virCamb
RdRP1.2_XinzhouVirDak
RdRP1.1_XinzhouVirDak
PolProt1.4_FlavivirusCamb
Nprot_PAvirCamb
Lprot1.2_PAvirCamb
HP1.1_IxodesVirDak
GP1.5_Wuhan1virCamb
3
Replicase_OatGSvirCamb
RdRP_UncVir2Dak
RdRP1.7_XinchengVirCamb
RdRP1.7_Wuhan1virCamb
RdRP1.6_UncVir2Camb
RdRP1.6_BeaumontVirCamb
RdRP1.5_UncVir2Camb
RdRP1.4_XinchengVirCamb
RdRP1.4_Wuhan1virCamb
RdRP1.4_UncVir2Camb
RdRP1.4_BeaumontVirCamb
RdRP1.3_XinchengVirCamb
RdRP1.3_Wuhan1virCamb
RdRP1.3_UncVir2Camb
RdRP1.2_XinchengVirCamb
RdRP1.2_BeaumontVirCamb
RdRP1.1_UncVir2Camb
PrAlaRichProt_EnvVirDak
PolProt1.3_FlavivirusCamb
PolProt1.2_FlavivirusCamb
PolProt1.1_FlavivirusCamb
ORF2_XinchengVirDak
ORF1_XinchengVirCamb
ORF1.3_Wuhan9virCamb
ORF1.2_Wuhan9virCamb
ORF1.1_Wuhan9virCamb
Nuclprot1.2_WBvirDak
Nuclprot1.2_WBvirCamb
Nuclprot1.1_WBvirDak
NuclCap1.9_Wuhan1virCamb
NuclCap1.8_Wuhan1virCamb
NuclCap1.6_Wuhan1virCamb
NuclCap1.5_Wuhan1virCamb
NuclCap1.3_PCLVDak
NuclCap1.3_PCLVCamb
NuclCap1.2_Wuhan1virCamb
NuclCap1.2_PCLVDak
NuclCap1.1_Wuhan1virCamb
NuclCap1.1_PCLVDak
NuclCap1.1_PCLVCamb
NuclCap1.11_Wuhan1virCamb
NuclCap1.10_Wuhan1virCamb
GP_Wuhan1virDak
GP_CulexRhabdovCamb
GP1.9_XinchengVirCamb
GP1.8_XinchengVirCamb
GP1.7_XinchengVirCamb
GP1.7_Wuhan1virCamb
GP1.6_XinchengVirCamb
GP1.4_XinchengVirCamb
GP1.4_Wuhan1virCamb
GP1.3_Wuhan9virCamb
GP1.3_Wuhan1virCamb
GP1.2_XinchengVirCamb
GP1.2_Wuhan1virCamb
GP1.2_PCLVCamb
GP1.1_XinchengVirCamb
GP1.1_Wuhan1virCamb
CP_HVreovirusDak
4
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
Small RNA size (nt)
Z−score
6
4
2
0

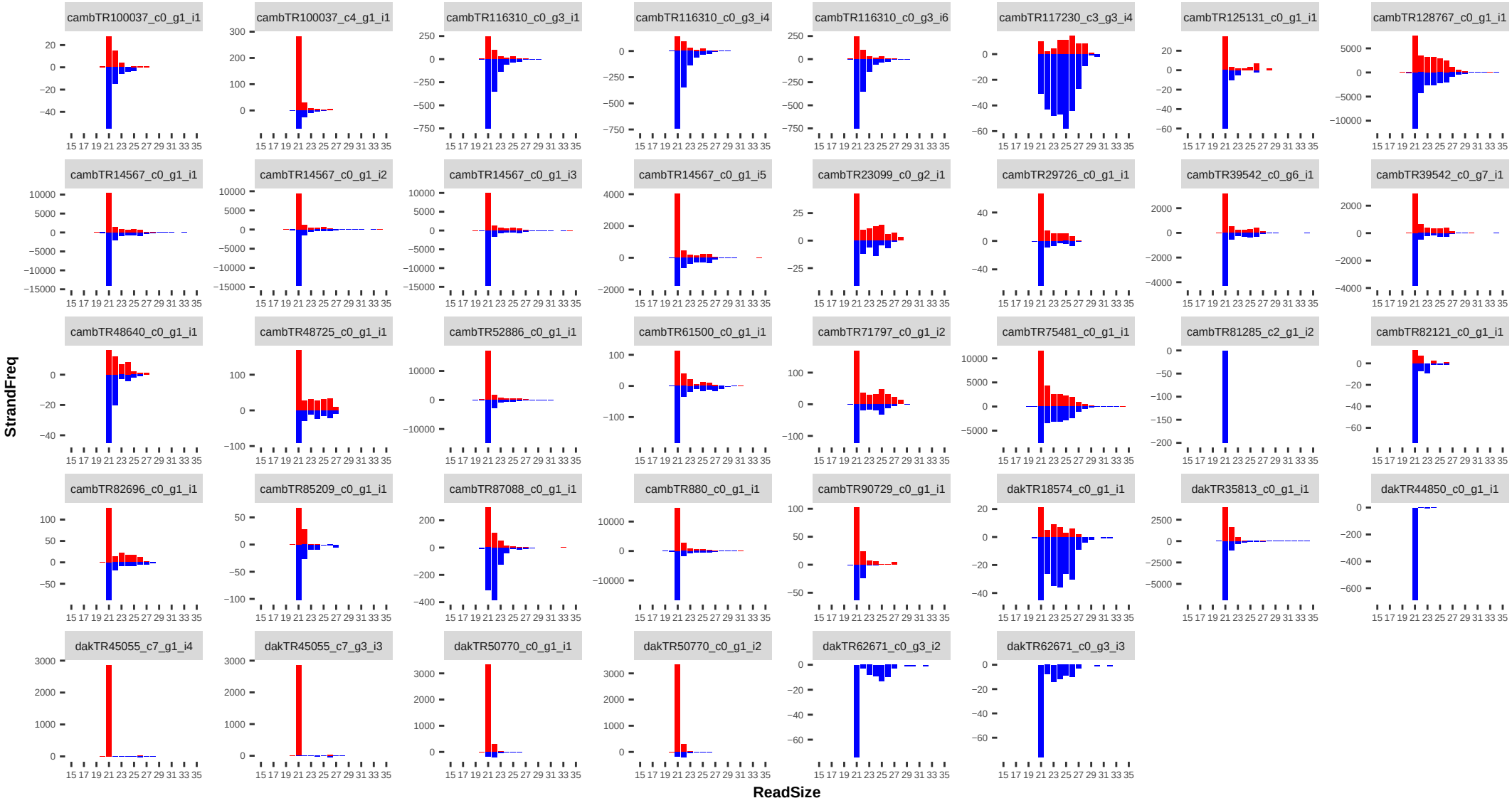
Additional File 1: Table S1

	Cam10-01	Cam10-02	CamS-01	CamS-02	Dak10-03	Dak10-04	DakS-03	DakS-04	processid	sampleid	bin_uri	sp. name
A. aconitus	0	0,287185	0,0688732	0,116965	0	0	0,00034265	0	GBDP1898-06;GBDP1895-06;GBDP1897-06;GBDP1904-06;GBDP1893-06;GBDP1900-06;GBMIN35311-13;GBANO744-12;GBDP1901-06;GBDP1903-06;GBDP1899-06;GBDP1223-06;GBDP1896-06;GBDP1894-06;GBDP1902-06	DQ000258;DI	BOLD:ACE38	Anopheles aconitus
A. albitarsis	0,00035901	0	0	0	0	0	0	0	MBIIA146-12;MBIE344-09;MBIE171-09;MBIE178-09;MBII250-09;GBDP14652-13;GBMIN12948-13;GBDP14641-13;GBDP14656-13;MBIIA155-12;MBIIA199-12;GBDP14628-13;GBDP2012-06;CYTC4996-12;CYTC4999-12;MBIE162-09;MBII252-09;MBIK145-10;MBII239-09;GBANO974-14;MBIE163-09;MBII223-09;MBIE03	MT25_100;5	BOLD:ABZ73	Anopheles albitarsis H;Anophele
A. annularis	0	0	0,00145927	0	0	0	0	0	MAMOS168-12;GBANO742-12;MAMOS173-12;GBDP1874-06;MAMOS174-12	NIBGE MOS	BOLD:AAR32	Anopheles annularis A;Anophele
A. annulipes	0	0,00536866	0,0162575	0,0129462	0,0036595	0,00010373	0,0047579	0,00321318	MOAV029-15;MOAV030-15;MOAV034-15;MOAV028-15;MOAV033-15	VAITC4309B;'	BOLD:AAB22	Anopheles annulipes
A. arabiensis	0	0	0	0	0	0,119129	0	0	GBDP6364-09;GBDP3804-07;GBDP0529-06;GBDP3805-07;GBDP3806-07;GBDP3822-07;GBDP3823-07;GBDP3836-07;GBDP6361-09;GBDP6357-09;GBDP6356-09;GBDP6365-09;GBDP6508-09;GBDP3815-07;GBDP3818-07;GBDP3817-07;GBDP6360-09;GBDP3814-07;GBDP3824-07;GBDP3799-07;GBDP3800-07;GBDP38	DQ465258;D	;BOLD:AAA3	Anopheles arabiensis
A. baimaii	0	0	0	0	0	0	0	0,0107677	GBDP5474-08;GBDP5312-08;GBDP5313-08;GBDP5332-08;GBDP5301-08;GBDP5318-08;GBDP5368-08;GBDP5404-08;GBDP5391-08;GBDP5302-08;GBDP5343-08;GBDP5375-08;GBDP5401-08;GBDP5360-08;GBDP5339-08;GBDP5370-08;GBDP5361-08;GBDP5546-08;GBDP5337-08;GBDP5326-08;GBDP5347-08;GBDP53	AJ877403;AJ877565;AJ877	BOLD:AAA5	Anopheles baimaii
A. barbirostris	0	0	0	0	0	0	0,00058087	0	GBDP4367-08;GBDP4689-08;GBDP6779-09;GBDP6771-09;GBDP6793-09;GBDP17076-15;GBDP6788-09;GBDP17078-15;GBDP6777-09;GBDP6773-09;GBDP6795-09;GBDP6786-09;GBDP4366-08;GBDP4691-08;GBDP4363-08;GBDP6798-09;GBDP6784-09;GBDP4362-08;GBDP6782-09;GBDP4683-08;GBDP4690-08;GBDP	AB362235;AE	BOLD:AAA56	Anopheles barbirostris
A. braziliensis	0	0	0,00044806	0	0	0,00012603	0,00070161	0,0104202	GBDP4781-08;GBDP4765-08;GBDP4813-08;GBDP4776-08;GBANO474-12;GBDP4791-08;GBDP4790-08;GBDP4778-08;GBDP4770-08;GBANO978-14;GBDP4808-08;GBDP4809-08;GBDP4786-08;GBDP4767-08;GBDP4780-08;GBDP4758-08;GBDP4763-08;GBDP4768-08;GBDP4802-08;GBDP2044-06;GBDP4785-08;GBDP4	DQ913857;D	;BOLD:AAA5	Anopheles braziliensis
A. campestris	0,00046072	0	0	0	0	0	0	0	GBANO043-12;GBANO037-12;GBDP4679-08;GBANO049-12;GBANO052-12;GBANO034-12;GBANO030-12;GBANO749-12;GBANO035-12;GBANO051-12;GBANO750-12;GBDP4680-08;GBANO044-12;GBANO050-12;GBANO028-12;GBDP4682-08;GBANO042-12;GBANO029-12;GBANO046-12;GBANO040-12;GBDP4681-0	AB436119;AE	BOLD:AAA56	Anopheles campestris
A. coustani	0,0640974	0	0,0369955	0	0	0	0	0,297003	GBANO1016-14;GBANO1013-14;GBANO1007-14;GBANO995-14;GBANO1011-14;GBANO1009-14;GBANO990-14;GBANO1008-14;GBANO1003-14;GBANO996-14;GBANO1002-14;GBANO999-14;GBANO1004-14;GBANO1014-14;GBANO997-14;GBANO1017-14;GBANO1006-14;GBANO1000-14;GBANO992-14;GBANO99	KM097027;K	BOLD:AAN94	Anopheles coustani
A. crawfordi	0,0145607	0	0	0	0	0	0	0,0321671	GBANO1873-15;GBANO1106-14;GBANO1895-15;GBANO1902-15;GBANO1103-14;GBANO1886-15;GBANO1872-15;GBANO1888-15;GBANO1894-15;GBANO1887-15;GBANO1875-15;GBANO1889-15;GBANO1901-15;GBANO1089-14;GBANO1871-15;GBANO1898-15;GBANO1877-15;GBANO1105-14;GBANO1884-15;GB	AB779184;AE	BOLD:AAJ28	Anopheles crawfordi
A. culicifacies	0	0	0	0,00032493	0	0	0	0,0102792	GBDP17116-15;GBDP4059-07;GBDP17112-15;MAMOS167-12;LGEN122-14;GBDP1811-06;GBDP17113-15;MAMOS145-12;GBDP17110-15;GBDP1089B-12;GBDP6188-09;MAMOS171-12;GBDP6187-09;GBDP17108-15;MAMOS1330-12;GBDP17103-15;GBDP17104-15;MAMOS1614-13;GBDP10897-12;GBDP17111-15;GI	KP197036;DC	;BOLD:AAR3	Anopheles culicifacies;Anophele
A. darlingi	0	0,00156377	0	0	0	0	0	0,00332902	GBDP2391-06;GBANO810-12;GBDP10923-12;GBDP2377-06;GBDP2385-06;GBDP2394-06;GBDP2398-06;GBANO809-12;GBDP10704-12;GBDP10933-12;GBDP2043-06;GBDP2381-06;GBDP2389-06;GBDP2403-06;GBDP12019-12;GBDP10930-12;GBDP7970-09;GBDP2396-06;GBDP2410-06;GBDP2401-06;GBANO482-12;	DQ298224;JX	;BOLD:AAA2	Anopheles darlingi
A. dirus	0	0,108862	0	0,241997	0	0	0	0,00769049	GBDP5436-08;GBDP5463-08;GBDP5486-08;GBDP5488-08;GBDP5489-08;GBDP5513-08;GBDP5544-08;GBDP5460-08;GBDP5509-08;GBDP5519-08;GBDP5459-08;GBDP5516-08;GBDP5558-08;GBDP5462-08;GBDP5443-08;GBDP5520-08;GBDP5526-08;GBDP5529-08;GBDP5475-08;GBDP5532-08;GBDP5484-08;GBDP54	AJ877441;AJI	BOLD:ABZ2	Anopheles dirus;Anopheles diru
A. dravidicus	0	0,00308572	0	0,00685737	0	0	0	0	MAMOS1613-13	NIBGE MOS	BOLD:ABY95	Anopheles sp nr dravidicus YML
A. fluviatilis	0	0	0	0	0,0005741	0	0,00143259	0,00114828	GBDP2256-06;GBANO370-12;GBDP1879-06;GBDP5971-09;GBANO462-12;GBDP5969-09;GBMIN12970-13;GBDP5972-09;GBMIN12964-13;GBDP4549-08;GBMIN12979-13;GBMIN12977-13;GBANO359-12;GBANO380-12;GBDP4390-08;GBDP6179-09;GBMIN12971-13;GBDP5990-09;GBMIN12978-13;GBDP5993-09;GBD	DQ154158;GI	;BOLD:AAA8	Anopheles fluviatilis;Anopheles
A. fragilis	0	0	0	0	0	0	0	0,00539895	GBDP17132-15;GBDP17137-15;GBDP17135-15;GBDP17133-15;GBDP17134-15;GBDP17136-15	KF564689;KF	BOLD:ACV94	Anopheles fragilis
A. funestus	0	0	0	0	0,564078	0	0	0,400681	GBMIN14779-13;GBMIN14681-13;GBMIN14665-13;GBMIN14676-13;GBMIN14674-13;GBMIN14688-13;GBMIN14697-13;GBMIN14750-13;GBMIN14780-13;GBMIN14771-13;GBMIN14776-13;GBMIN14748-13;GBMIN14691-13;GBMIN14683-13;GBMIN14672-13;GBMIN14678-13;GBMIN14689-13;GBMIN14764-13;GI	JQ424643;JQ	BOLD:AAE40	Anopheles funestus;Anopheles
A. gambiae	0	0	0	0	0	0,865717	0	0	GBDP3753-07;GBDP3766-07;GBDP3786-07;GBDP6350-09;GBDP3768-07;GBDP3788-07;GBDP3770-07;GBDP3764-07;GBDP3765-07;GBDP3780-07;GBDP3787-07;GBDP3759-07;GBDP3750-07;GBDP3779-07;GBDP3761-07;GBDP6355-09;GBDP3821-07;GBDP6347-09;GBDP3771-07;GBDP1229-06;GBDP3224-07;GBDP37	DQ465333;DI	;BOLD:AAA3	Anopheles gambiae
A. hyrcanus	0,036451	0	0,0258228	0	0	0	0	0	GBANO586-12;GBANO582-12;GBANO575-12;GBANO588-12;GBMIN17608-13;GBANO581-12;GBANO593-12;GBDP4985-08;GBANO580-12;GBANO583-12;GBMIN17609-13;GBANO576-12;GBDP14482-13;GBANO591-12;GBDP5706-09;GBANO584-12;GBANO592-12;GBANO587-12;GBDP14480-13;GBDP14481-13;GBAN	HQ197444;HI	;BOLD:ABZ6	Anopheles hyrcanus
A. insulaeflorum	0	0	0	0	0,00061369	0	0,00190577	0	GBDP10903-12	GQ259191		Anopheles insulaeflorum
A. jeyporiensis	0	0,370471	0	0,241745	0,00433541	0	0	0,0059982	GBDP2257-06;GBANO764-12;GBANO747-12;GBDP4082-07;GBMIN35199-13;GBANO766-12;GBDP2255-06;GBDP4081-07;GBDP10907-12;GBDP4080-07;GBANO765-12	DQ154159;JC	;BOLD:AAC5	Anopheles jeyporiensis
A. karwari	0	0,103915	0,211861	0,157841	0	0	0	0	GBDP17141-15;GBDP17140-15;GBDP17139-15	KF564711;KF	BOLD:AAE53	Anopheles karwari
A. leesoni	0	0	0	0	0,00032805	0	0	0	GBDP1224-06	AY423056	BOLD:AAR32	Anopheles leesoni
A. lesteri	0,00041884	0	0	0	0	0	0	0	GBDP7889-09;GBDP4893-08;GBDP4914-08;GBDP13491-13;GBANO1021-14;GBDP4870-08;GBDP4898-08;GBDP4901-08;GBDP4909-08;GBDP4885-08;GBDP4896-08;GBDP4892-08;GBDP4911-08;GBDP14456-13;GBDP13293-13;GBDP4888-08;GBDP4913-08;GBDP4869-08;GBDP4915-08;GBDP4880-08;GBDP4923-08;GB	FJ556998;EU	BOLD:AAA57	Anopheles lesteri
A. letifer	0	0	0	0	0	0	0	0,00351536	GBDP17146-15;GBDP17147-15	KF564694;KF	BOLD:AAA85	Anopheles letifer
A. longirostris	0	0	0	0	0	0,00027727	0	0,00693	GBDP8910-10;GBDP8940-10;GBDP8943-10;GBDP8926-10;GBDP8885-10;GBDP8932-10;GBDP8930-10;GBDP8901-10;GBDP8911-10;GBDP8882-10;GBDP8916-10;GBDP8891-10;GBDP8924-10;GBDP8922-10;GBDP8918-10;GBDP8928-10;GBDP8929-10;GBDP8902-10;GBDP8912-10;GBDP8938-10;GBDP8899-10;GBDP89	GU247086;GU247056;GU2	Anopheles longirostris D NWB-2	
A. maculatus	0	0,118328	0	0,211197	0	0	0	0,0101583	GBDP5998-09;GBDP17149-15;GBANO694-12;GBDP10904-12;GBDP17150-15;GBDP2371-06;GBDP17153-15;GBDP17151-15;GBDP17152-15;GBDP17148-15	EU256336;KN	;BOLD:AAC3	Anopheles maculatus
A. maculipennis	0,00032311	0	0	0,00748462	0	0	0	0	GBDP1188-06;GBDP0597-06;GBDP0594-06;GBDP0593-06;CULBE160-14;ACMC314-04;GBDP0598-06;CULBE165-14;GBANO652-12;GBDP4314-07;GBANO1870-15;CULBE158-14;GBDP2239-06;CULBE156-14;CULBE163-14;CULBE166-14;GBDP2238-06;GBANO1022-14;CULBE159-14;CULBE161-14;GBDP0596-06;GBDP05	AY258190;AF	BOLD:AAA96	Anopheles maculipennis;Anophe
A. melas	0	0	0	0	0	0,0140465	0	0	GBDP3223-07;GBDP3222-07	DQ792579;D	BOLD:AAA34	Anopheles melas
A. minimus	0	0	0,0201091	0	0,0137076	0	0	0	GBDP10902-12;GBDP4386-08;GBDP6196-09;GBDP6189-09;GBDP4387-08;GBMIN35203-13;GBANO378-12;GBDP10894-12;GBANO622-12;GBMIN35200-13;GBDP10893-12;GBANO375-12;GBDP6193-09;GBDP6192-09;GBDP1226-06;GBDP4383-08;GBDP6194-09;GBDP1225-06;GBMIN35202-13;GBDP4380-08;GBDP615	GQ259189;EF	BOLD:ACF27	Anopheles minimus;Anopheles
A. nitidus	0,068579	0	0	0	0	0	0	0	GBANO1056-14;GBANO1057-14;GBANO1919-15;GBANO1058-14;GBANO1914-15;GBANO1926-15;GBANO1115-14;GBANO1920-15;GBANO1912-15;GBANO1922-15;GBANO1909-15;GBANO1913-15;GBANO1917-15;GBANO1060-14;GBANO1916-15;GBANO1918-15;GBANO1925-15;GBANO1910-15;GBANO1908-15;GB	AB781762;AE	BOLD:AAA57	Anopheles nitidus
A. nivipes	0,00073297	0	0,00080979	0	0	0	0	0	GBANO696-12	JN596974		Anopheles nivipes
A. parensis	0	0	0	0	0,102774	0	0	0,065976	GBMIN14741-13;GBDP1227-06;GBMIN14734-13;GBMIN14661-13;GBMIN14738-13;GBMIN14743-13;GBMIN14659-13;GBMIN14654-13;GBMIN14733-13;GBMIN14650-13;GBMIN14742-13;GBMIN14656-13;GBMIN14740-13;GBMIN14744-13;GBMIN14732-13;GBMIN14660-13;GBMIN14736-13;GBMIN14653-13;GBM	JQ424719;AY	BOLD:AAE40	Anopheles parensis
A. peditaeniatus	0,795789	0	0,608502	0	0	0	0	0,0603555	GBANO1065-14;GBDP8466-10;GBMIN14561-13;GBMIN14617-13;GBMIN14622-13;GBDP8470-10;GBMIN14611-13;GBMIN14621-13;MAMOS190-12;GBDP10899-12;GBANO692-12;MAMOS184-12;GBMIN14614-13;GBMIN14552-13;GBMIN14551-13;MAMOS185-12;GBMIN14566-13;GBDP8467-10;GBMIN14548-13;G	AB781776;AE	BOLD:AAD30	Anopheles peditaeniatus
A. pseudopictus	0,015521	0	0,00669618	0	0	0	0	0,00835528	GBDP4984-08;GBDP5704-09;GBANO1928-15;GBDP5705-09	AM773815;F	BOLD:ABZ68	Anopheles pseudopictus
A. pulcherrimus	0	0	0	0	0,00109163	0	0	0,000916613	GBMIN17591-13;GBMIN17592-13;MAMOS159-12;MAMOS169-12;MAMOS156-12;MAMOS158-12;MAMOS161-12;GBMIN17610-13;GBMIN17611-13;MAMOS163-12;MAMOS160-12;MAMOS162-12	JX255712;JX	BOLD:ACB27	Anopheles pulcherrimus
A. pursati	0,00146295	0	0,00109753	0	0	0	0	0	GBANO1031-14;GBANO1024-14;GBANO1086-14;GBANO1026-14;GBANO1032-14;GBANO1033-14;GBANO1085-14;GBANO1087-14;GBANO1083-14;GBANO1028-14;GBANO1029-14;GBANO1088-14;GBANO1084-14;GBANO1027-14;GBANO1025-14;GBANO1030-14	AB826089;AE	BOLD:ACQ53	Anopheles pursati
A. quadriannulatus	0	0	0	0	0	0,00046146	0	0	GBDP3221-07	DQ792581	BOLD:AAA34	Anopheles quadriannulatus
A. rangeli	0	0	0	0	0	0	0	0,00091157	GBANO468-12;GBANO789-12;GBDP12050-12;GBANO602-12;GBANO788-12;GBANO467-12;GBANO469-12;GBANO791-12;GBANO787-12;GBANO466-12;GBMIN12951-13;GBANO470-12;GBMIN12936-13;GBANO790-12	HM022392;JN	BOLD:AAB05	Anopheles rangeli
A. sinensis	0,00124455	0	0,00106876	0	0	0	0	0	GBANO1969-15;GBDP13425-13;GBDP13292-13;GBDP17168-15;GBDP13335-13;GBDP13389-13;GBDP17164-15;GBANO1036-14;GBDP17167-15;GBDP13368-13;GBDP13310-13;GBDP13313-13;GBDP13342-13;ACMC017-04;GBDP17166-15;GBDCU715-13;GBANO1970-15;GBANO989-14;GBMIN17526-13;GBDP13367-1	KM502246;A	BOLD:AAA53	Anopheles sinensis
A. rufipes	0	0	0	0	0,307121	0	0,990278	0	GBANO1018-14	KM097029	BOLD:AAN94	Anopheles sp.
A. stephensi	0	0,00073589	0	0,00199194	0	0	0	0	MAMOS053-12;MAMOS910-12;MAMOS165-12;GBDP1679-06;MAMOS122-12;MAMOS1324-12;MAMOS121-12;GBDP5708-09;MADIP309-10;MAMOS051-12;MAMOS893-12;MADIP281-10;MAMOS123-12;MAMOS1744-13;MAMOS1297-12;MAMOS1227-12;MAMOS070-12;GBDP17170-15;LGEN121-14;GBDP17177-1	NIBGE MOS	BOLD:AAC57	Anopheles stephensi
A. subpictus	0	0,00048502	0	0,00064985	0	0	0	0	MADIP280-10;GBDP8491-10;MAMOS636-12;MADIP282-10;GBANO1938-15;GBDP6183-09;GBDP1669-06;MADIP278-10;GBANO1941-15;MADIP308-10;MAMOS1090-12;GBANO1971-15;MAMOS233-12;GBDP4087-07;MAMOS992-12;GBANO1936-15;GBDP8496-10;GBANO1937-15;MAMOS1123-12;MAMOS141-12;M	NIBGE DIP	BOLD:AAD26	Anopheles subpictus;Anopheles
A. triannulatus	0	0	0	0	0	0,00011827	0	0,00235701	GBMIN11395-13;GBDP12058-12;GBMIN11381-13;GBDP12060-12;GBMIN11426-13;GBANO463-12;GBMIN11382-13;GBMIN11399-13;GBMIN11383-13;GBMIN11438-13;GBMIN11392-13;GBMIN11393-13;GBMIN11427-13;GBMIN11423-13;GBMIN11380-13;GBANO782-12;GBMIN11417-13;GBMIN11378-13;GBMIN1	KC167679;JF	BOLD:AAA97	Anopheles triannulatus
A. vaneedeni	0	0	0	0	0,00171663	0	0	0,0524283	GBMIN14719-13;GBMIN14707-13;GBMIN14720-13;GBMIN14629-13;GBMIN14640-13;GBMIN14639-13;GBMIN14714-13;GBMIN14626-13;GBMIN14630-13;GBMIN14637-13;GBMIN14709-13;GBMIN14715-13;GBMIN14721-13;GBMIN14710-13;GBMIN14636-13;GBMIN14635-13;GBDP1228-06;GBMIN14713-13;GBM	JQ424763;JQ	BOLD:AAE40	Anopheles vaneedeni

Additional File 2: Figure S1



Additional File 3: Figure S2



Additional File 4: Table S2

XLOC_014032	XLOC_014032	AGAP000958	X:18370380-18381700	Uninfected	ONNVirus	OK	8.40155	97.8132	3.5413	3.18141	0.00015	0.0136402	yes
XLOC_003170	XLOC_003170	AGAP013159	2R:2206678-2207523	Uninfected	ONNVirus	OK	5.15807	0	#NAME?	#NAME?	0.0002	0.0164039	yes
XLOC_005339	XLOC_005339	AGAP001766	2R:9595495-9599152	Uninfected	ONNVirus	OK	0.970736	30.7241	4.98415	4.72935	0.0002	0.0164039	yes
XLOC_007766	XLOC_007766	AGAP012104	3L:37739157-37740909	Uninfected	ONNVirus	OK	1.24432	21.2666	4.09516	3.63244	0.0002	0.0164039	yes
XLOC_013018	XLOC_013018	-	UNKN:21393264-21393309	Uninfected	ONNVirus	OK	155175	25359.5	-2.6133	-2.74495	0.0002	0.0164039	yes
XLOC_013265	XLOC_013265	AGAP000512	X:9067174-9086241	Uninfected	ONNVirus	OK	6.95754	43.4702	2.64338	2.85258	0.0002	0.0164039	yes
XLOC_003066	XLOC_003066	-	2L:49358643-49359203	Uninfected	ONNVirus	OK	83.4691	6.45484	-3.69279	-3.54425	0.00025	0.0186741	yes
XLOC_003118	XLOC_003118	AGAP001193	2R:1214535-1215158	Uninfected	ONNVirus	OK	11.0928	94.5411	3.09131	2.66549	0.00025	0.0186741	yes
XLOC_005856	XLOC_005856	AGAP002725	2R:26248499-26263468	Uninfected	ONNVirus	OK	0.647863	27.4159	5.40318	4.64829	0.00025	0.0186741	yes
XLOC_009067	XLOC_009067	AGAP012316	3L:40151687-40152337	Uninfected	ONNVirus	OK	63.347	493.51	2.96173	2.82786	0.00025	0.0186741	yes
XLOC_013634	XLOC_013634	AGAP000161	X:2822717-2830085	Uninfected	ONNVirus	OK	18.6049	140.804	2.91994	2.91297	0.00025	0.0186741	yes
XLOC_000374	XLOC_000374	AGAP005471	2L:16024035-16067199	Uninfected	ONNVirus	OK	3.90664	31.7458	3.02257	2.84024	0.0003	0.0205721	yes
XLOC_009185	XLOC_009185	AGAP007807	3R:1269162-1276293	Uninfected	ONNVirus	OK	4.76588	28.6619	2.58832	2.85996	0.0003	0.0205721	yes
XLOC_012530	XLOC_012530	AGAP012733	UNKN:25901560-25901634	Uninfected	ONNVirus	OK	58151.9	615148	3.40303	2.98813	0.0003	0.0205721	yes
XLOC_012723	XLOC_012723	-	UNKN:13287-13355	Uninfected	ONNVirus	OK	14193.8	154814	3.44721	3.08318	0.0003	0.0205721	yes
XLOC_013099	XLOC_013099	AGAP000153	X:2466315-2477593	Uninfected	ONNVirus	OK	3.62868	34.2534	3.23873	3.06399	0.0003	0.0205721	yes
XLOC_004039	XLOC_004039	AGAP002899	2R:29098112-29099864	Uninfected	ONNVirus	OK	1.92092	34.7746	4.17816	3.39816	0.00035	0.0228758	yes
XLOC_009111	XLOC_009111	AGAP012367	3L:41132772-41134888	Uninfected	ONNVirus	OK	2.74448	15.4118	2.48943	2.60117	0.00035	0.0228758	yes
XLOC_012913	XLOC_012913	-	UNKN:18155787-18227928	Uninfected	ONNVirus	OK	131005	1.16409e+06	3.15151	2.77219	0.00035	0.0228758	yes
XLOC_000179	XLOC_000179	AGAP005048	2L:8818558-8821078	Uninfected	ONNVirus	OK	2.54486	13.4692	2.40401	2.62317	0.0004	0.0253515	yes
XLOC_006567	XLOC_006567	AGAP004089	2R:49539478-49541388	Uninfected	ONNVirus	OK	2.67665	19.1152	2.83622	2.81591	0.0004	0.0253515	yes
XLOC_007549	XLOC_007549	AGAP011679	3L:31282268-31291171	Uninfected	ONNVirus	OK	1.0905	7.18386	2.71978	2.81192	0.00045	0.0272804	yes
XLOC_008857	XLOC_008857	AGAP011940	3L:35411761-35414327	Uninfected	ONNVirus	OK	5.06537	30.3991	2.58529	2.73761	0.00045	0.0272804	yes
XLOC_013939	XLOC_013939	AGAP000779	X:13969856-13978437	Uninfected	ONNVirus	OK	6.96099	44.2185	2.66729	2.8695	0.00045	0.0272804	yes
XLOC_007976	XLOC_007976	AGAP010394	3L:2612104-2635382	Uninfected	ONNVirus	OK	0.649364	6.80942	3.39043	3.07317	0.0005	0.0290486	yes
XLOC_010550	XLOC_010550	AGAP007741	3R:174262-175359	Uninfected	ONNVirus	OK	2.30264	27.8833	3.59804	3.14014	0.0005	0.0290486	yes
XLOC_014019	XLOC_014019	AGAP000932	X:17600516-17626487	Uninfected	ONNVirus	OK	1.55142	84.0543	5.75966	3.97556	0.0005	0.0290486	yes
XLOC_000668	XLOC_000668	AGAP006103	2L:26687470-26689494	Uninfected	ONNVirus	OK	7.42012	52.2227	2.81516	2.6142	0.00055	0.0302717	yes
XLOC_010196	XLOC_010196	AGAP009634	3R:37363064-37364183	Uninfected	ONNVirus	OK	0	4.24369	inf	#NAME?	0.00055	0.0302717	yes
XLOC_012896	XLOC_012896	-	UNKN:17601427-17601819	Uninfected	ONNVirus	OK	456.588	73.4723	-2.63562	-2.60231	0.00055	0.0302717	yes
XLOC_013015	XLOC_013015	-	UNKN:21005541-21006053	Uninfected	ONNVirus	OK	141.232	28.3442	-2.31694	-2.5282	0.00055	0.0302717	yes
XLOC_012808	XLOC_012808	-	UNKN:14354764-14354960	Uninfected	ONNVirus	OK	1224.08	236.904	-2.36932	-2.68354	0.0006	0.0325948	yes
XLOC_012788	XLOC_012788	-	UNKN:13913203-13913289	Uninfected	ONNVirus	OK	69965.9	13685.6	-2.35399	-2.32151	0.0007	0.0375397	yes
XLOC_012827	XLOC_012827	-	UNKN:14872675-14872943	Uninfected	ONNVirus	OK	304.222	57.8351	-2.39511	-2.59354	0.00075	0.039712	yes
XLOC_011959	XLOC_011959	AGAP012443	UNKN:1191148-1197633	Uninfected	ONNVirus	OK	8.41662	71.4601	3.08582	2.63358	0.0008	0.04183	yes
XLOC_003276	XLOC_003276	AGAP001496	2R:5638254-5642700	Uninfected	ONNVirus	OK	1.20935	13.2958	3.45867	2.92038	0.00085	0.0438957	yes
XLOC_012868	XLOC_012868	-	UNKN:16650281-16650527	Uninfected	ONNVirus	OK	353.849	78.9974	-2.16326	-2.41936	0.0009	0.0448179	yes
XLOC_012877	XLOC_012877	-	UNKN:16852075-16852325	Uninfected	ONNVirus	OK	318.32	63.365	-2.32872	-2.58497	0.0009	0.0448179	yes
XLOC_012924	XLOC_012924	-	UNKN:18747438-18747661	Uninfected	ONNVirus	OK	912.636	161.017	-2.50282	-2.50507	0.0009	0.0448179	yes
XLOC_003345	XLOC_003345	AGAP001649	2R:7281481-7283417	Uninfected	ONNVirus	OK	1.92372	12.5225	2.70256	2.4525	0.00095	0.0462076	yes
XLOC_012762	XLOC_012762	-	UNKN:13088289-13088321	Uninfected	ONNVirus	OK	656738	151656	-2.11451	-2.38627	0.00095	0.0462076	yes