Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Insect Science

Endogenous viral elements in mosquito genomes: current knowledge and outstanding questions

Umberto Palatini, Claudia A Contreras, Laila Gasmi and Mariangela Bonizzoni

Integrations from non-retroviral RNA viruses (nrEVEs) have been identified across several taxa, including mosquitoes. Amongst all Culicinae species, the viral vectors *Aedes aegypti* and *Aedes albopictus* stand out for their high number of nrEVEs. In addition, *Aedes* nrEVEs are enriched in piRNA clusters and generate piRNAs that can silence incoming viral genomes. As such, nrEVEs represent a new form of inherited antiviral immunity. To propel this discovery into novel transmission-blocking vector control strategies, a deeper understanding of nrEVE biology and evolution is essential because differences in the landscape of nrEVEs have been identified in wild-caught mosquitoes, the piRNA profile of nrEVEs is not homogeneous and nrEVEs outside piRNA clusters exist and are expressed at the mRNA level. Here we summarise current knowledge on nrEVEs in mosquitoes and we point out the many unanswered questions and potentials of these genomic elements.

Address
Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9, 27100 Pavia, Italy.

Corresponding author: Bonizzoni, Mariangela (m.bonizzoni@unipv.it)

Current Opinion in Insect Science 2022, **49**:22–30

This review comes from a themed issue on **Parasites/parasitoids/biological control**

Edited by **James Bryan Whitfield** and **Jean-Michel Drezen**

https://doi.org/10.1016/j.cois.2021.10.007

Introduction

Host-virus coevolution is fuelled by the selective pressures the two partners impose on each other leading to reciprocal and adaptive genetic changes. In this arms-race, the transfer of genetic material from viruses to eukaryotic cells has long been recognised as an important mechanism shaping host genome diversity and evolution. As an example, 8% of the human genome is composed of endogenous viral elements (EVEs) from retroviruses, deriving from millions of years of coevolution. Advances in genome sequencing and bioinformatics have shown that EVEs arise not only from retroviruses and DNA viruses but also from non-retroviral (nr) RNA viruses. Among invertebrates, *Aedes* spp. mosquitoes, which are the primary vectors for arboviruses world-wide, stand out for the number of nrEVEs and the genomic sites of nrEVE occurrence, raising questions on the mechanisms of viral integration and their potential biological role. Here, we summarise our current knowledge on EVEs in mosquito genomes. We start by describing the distribution of EVEs across the currently available mosquito genomes and in wild-caught mosquitoes. Next, we review data on the genomic sites of integration events, from which hypotheses regarding the function of nrEVEs have been formulated. We close by discussing experiments supporting an immunity role for nrEVEs and their use in genetic-based control strategies. Throughout the review, we highlight outstanding questions, including methodological approaches to annotate viral integrations.

An overview of mosquitoes and their viral flora

Mosquitoes belong to the Culicidae family, a monophyletic group including 4500 species divided into the Anophelinae and Culicinae subfamilies. Anophelinae and Culicinae separated roughly 180–250 million years ago and present remarkable differences in their genome size, primarily due to the larger content of transposable elements (TEs) of Culicinae (Table 1) [1–4,5•,6•,7]. Additionally, Anophelinae comprise primarily vectors for protozoan parasites, while Culicinae are mostly vectors of nematodes (i.e. *Dirophilaria immitis*) and arboviruses [8,9]. The mosquito species *Aedes aegypti* and *Aedes albopictus* are the arboviral vectors with the highest public health relevance worldwide based on their geographic distribution, adaptation to live in close proximity to humans and the fact that they are efficient vectors for a large number of arboviruses, including Dengue, Zika, Yellow Fever and Chikungunya viruses [10,11,12,13]. Despite differences in vector competence with respect to arboviruses, both Anophelinae and Culicinae mosquitoes support replication of Insect-Specific Viruses (ISVs) [14,15]. ISVs isolated from mosquitoes belong to the same taxonomic categories as arboviruses, such as the *Flaviviridae*, *Togaviridae*, *Reoviridae*, and *Rhabdoviridae* families and the *Bunyavirales* order [14,16,17]. *Bunyavirales* ISVs and some *Flaviviridae* ISVs such as Cell Fusing Agent Virus (CFAV), Kamiti River virus (KRV), Culex flavivirus, Aedes flavivirus, Quang Binh virus, Nienokoue virus, Nakiwogo virus, Xishuangbanna virus and Palm Creek virus are phylogenetically ancestral to arboviruses.

Table 1

Comparison of the more recent Culicinae genome assemblies

Species	Assembly	Genome size (Mbp)	% TE	% LTR	References
Aedes aegypti	AaegL5	1278.7	62	11.7	[3,4]
Aedes albopictus	AaloF1	2538.4	55	5.6	[6•,7]
Culex quinquefasciatus	CpipJ2	579.1	44	10	[3,5•]
Anopheles albimanus	AalbS2	173.3	2.7	52	[3,5•]
Anopheles arabiensis	AaraD1	246.6	9.8	22	[3,5•]
Anopheles atroparvus	AatrE1	225.3	4.9	44.4	[3,5•]
Anopheles christyi	AchrA1	172.7	2.2	35.5	[3,5•]
Anopheles coluzzii	AcolM1	224.4	11	17	[3,5•]
Anopheles culicifacies	AculA1	203	3.8	11.8	[3,5•]
Anopheles darlingi	AdarC3	137	2.6	43.7	[3,5•]
Anopheles dirus	AdirW1	216.3	6.4	34.2	[3,5•]
Anopheles epiroticus	AepiE1	223.5	6.1	30.7	[3,5•]
Anopheles farauti	AfarF2	183.1	2.2	33.3	[3,5•]
Anopheles funestus	AfunF1	211	5.5	47.4	[3,5•]
Anopheles gambiae	AgamP4	281.4	20	29.3	[3,5•]
Anopheles koliensis	AKwgs3	151.1	0.1	43.8	[3,5•]
Anopheles maculatus	AmacM1	141.9	4.4	19.3	[3,5•]
Anopheles melas	AmelC2	224.2	7.8	21.3	[3,5•]
Anopheles merus	AmerM2	288.1	9.8	23.5	[3,5•]
Anopheles minimus	AminM1	201.8	4.8	41.7	[3,5•]
Anopheles punctulatus	APwgs2	146.2	0.4	34.5	[3,5•]
Anopheles quadriannulatus	AquaS1	283.8	7.9	19.9	[3,5•]
Anopheles sinensis	AsinC2	220.8	10	30.2	[3,5•]
Anopheles stephensi	AsteS1	225.4	5.7	35.5	[3,5•]

Besides ancestral, or classic insect-specific flaviviruses (cISFs), numerous ISFs have been shown to be phylogenetically closer to mosquito-borne flaviviruses and hence have been called dual host associated ISFs (dISFs) [18,19].

Number and type of viral integrations in mosquito genomes

The first nrEVEs of mosquitoes were serendipitously discovered in the genomes of *Ae. albopictus* and *Ae. aegypti* and bore similarity to CFAV and KRV [20,21]. The expansion of Next Generation Sequencing (NGS) and bioinformatic approaches allowed us and others to probe mosquito genomes for the presence of viral integrations [9,22,23,24••,25,26,27,28••,29••,30] (Table 2). The number of EVEs identified in Culicidae is not concordant across studies because different genome assemblies, viral databases, bioinformatic algorithms and thresholds (i.e. minimum length of the putative EVE) were used to annotate viral integrations. This heterogeneity makes comparative analyses difficult. Having a common framework for EVEs annotation would be a great advance to obtain reproducible results and gain insights into their biology [31,32]. For instance, a range of between 17–29 nrEVEs with similarity to flaviviruses were annotated in the latest *Ae. aegypti* assembly (AaegL5) within the same four genomic regions (Figure 1) [28••,30]. The difference in nrEVE number resulted from the parameters used to filter and assign blast hit results to either different integrations from the same virus [30] or to group them into a unique longer integration based on the contiguity of blast hits on the mosquito genome [28••]. Without a clear understanding of the mechanisms of integration, the substrate (i.e. viral DNA fragments or other sources) and the relative role of TEs in favouring integrations and/or nrEVE rearrangements, we support a conservative approach, which avoids mechanistic interpretations of blast hit results [28••]. Overall, BLASTx-based methods are emerging as preferable and powerful strategies to identify nrEVEs, especially if a high stringency is used (10^{-3} or lower), followed by steps to refine predictions and eliminate false positive hits due to the presence of low complexity sequences. This strategy can detect EVEs regardless of their coding potential.

Despite differences in the exact number of EVEs, all studies are concordant in finding integrations exclusively from non-retroviral RNA viruses, which are ~10-fold more abundant in the genomes of *Ae. aegypti* and *Ae. albopictus* than in any other tested mosquito genomes (Table 2). In Anophelinae, the largest number of nrEVEs (64) was identified in the genome of *An. gambiae* and included predominantly integrations from unclassified viruses [25]. *Anopheles stephensi* and *An. minimus* are the only Anophelinae harbouring each a single, and not orthologous, integration from ISFs [9,23]. Excluding nrEVEs from viruses whose phylogeny is still uncertain, the majority of *Aedes* nrEVEs appear to derive from viruses of the *Flaviviridae* and *Rhabdoviridae* families. Except for an integration with similarity to Dengue virus 1 (DENV1) in *Ae. aegypti*, flavivirus-like integrations of both *Aedes* spp. genomes are from a limited number of

Table 2

Most-up to date list of studies describing nrEVEs in mosquito genomes, including methodology used and number of viral integrations identified

Subfamily	Species	Genome assembly	Method	Viruses in database	EVEs (in piRNA clusters[a])	Reference
Culicinae	*Aedes albopictus*	AaloF1	tBLASTx	RNA (424), DNA (1)	72 (9)	[9]
		AaloF1	BLASTx	RNA (na), DNA (na)	502 (3)	[25] [b]
		AaloF1	tBLASTx	RNA (na)	33	[29] [d]
		Canu_80X (C6/36)	tBLASTn	RNA (1933)	276	[26]
		Canu_80X (C6/36)	tBLASTx	RNA (na)	91	[29] [d]
		AalbF2	BLASTx	RNA (1563)	456 (138)	[6]
Culicinae	*Ae. aegypti*	AaegL1	tBLASTn	RNA (66), DNA (21)	57	[22]
		AaegL3	tBLASTx	RNA (424), DNA (1)	122 (54)	[9]
		AaegL3	tBLASTx	RNA (na)	29	[29] [d]
		Aag2	BLASTx	RNA (na)	472 (256)	[24]
		Aag2	BLASTx	RNA (na), DNA (na)	273 (117)	[25] [b]
		Aag2	tBLASTx	RNA (na)	68	[29] [d]
		AaegL5	tBLASTn	RNA (1933)	276	[26]
		AaegL5	ORF prediction	na	129	[27]
		AaegL5	BLASTx	RNA (1563)	252 (188)	[28]
		AaegL5	BLASTx	RNA (na)	29	[30] [c]
Culicinae	*Culex quinquefasciatus*	CpipJ1	tBLASTn	RNA (66), DNA (21)	1	[22]
		CpipJ2	tBLASTx	RNA (424), DNA (1)	1	[9]
		CpipJ2	tBLASTn	RNA (1933)	28	[26]
		CpipJ2	tBLASTx	RNA (na)	20	[29] [d]
Anophelinae	*Anopheles arabiensis*	AaraD1	tBLASTx	RNA (424), DNA (1)	5	[9]
		AaraD1	tBLASTx	RNA (50)	0	[23] [c]
		AaraD1	BLASTx	RNA (na), DNA (na)	16 (6)	[25] [b]
		AaraD1	tBLASTx	RNA (na)	2	[29] [d]
Anophelinae	*An. gambiae*	AgamP3	tBLASTn	RNA (66), DNA (21)	0	[22]
		AgamP4	tBLASTx	RNA (424), DNA (1)	0	[9]
		AgamP4	tBLASTx	RNA (50)	0	[23] [c]
		AgamP4	BLASTx	RNA (na), DNA (na)	64 (7)	[25] [b]
		AgamP4	tBLASTn	RNA (1933)	24	[26]
		AgamP4	tBLASTx	RNA (na)	2	[29] [d]
Anophelinae	*An. stephensi*	AsteI2	tBLASTx	RNA (424), DNA (1)	0	[9]
		AsteI2	tBLASTx	RNA (50)	0	[23] [c]
		AsteI2	BLASTx	RNA (na), DNA (na)	23 (5)	[25] [b]
		AsteI2	tBLASTx	RNA (na)	2	[29] [d]
		AsteS1	tBLASTx	RNA (50)	0	[23] [c]
		AsteS1	tBLASTx	RNA (na)	3	[29] [d]
Anophelinae	*An. christy*	AchrA1	tBLASTx	RNA (424), DNA (1)	0	[9]
		AchrA1	tBLASTx	RNA (50)	0	[23] [c]
		AchrA1	tBLASTx	RNA (na)	0	[29] [d]
Anophelinae	*An. coluzzi*	AcolM1	tBLASTx	RNA (424), DNA (1)	0	[9]
		AcolM1	tBLASTx	RNA (50)	0	[23] [c]
		AcolM1	tBLASTx	RNA (na)	1	[29] [d]
Anophelinae	*An. melas*	AmelC1	tBLASTx	RNA (50)	0	[23] [c]
		AmelC2	tBLASTx	RNA (424), DNA (1)	0	[9]
		AmelC2	tBLASTx	RNA (na)	0	[29] [d]
Anophelinae	*An. merus*	AmerM1	tBLASTx	RNA (50)	0	[23] [c]
		AmerM2	tBLASTx	RNA (424), DNA (1)	2	[9]
		AmerM2	tBLASTx	RNA (na)	2	[29] [d]
Anophelinae	*An. quadrianulatus A*	AquaS1	tBLASTx	RNA (424), DNA (1)	2	[9]
		AquaS1	tBLASTx	RNA (50)	0	[23] [c]
		AquaS1	tBLASTx	RNA (na)	2	[29] [d]
Anophelinae	*An. epiroticus*	AepiE1	tBLASTx	RNA (424), DNA (1)	7	[9]
		AepiE1	tBLASTx	RNA (50)	0	[23] [c]
		AepiE1	tBLASTx	RNA (na)	3	[29] [d]
Anophelinae	*An. maculatus B*	AmacM1	tBLASTx	RNA (424), DNA (1)	2	[9]
		AmacM1	tBLASTx	RNA (50)	0	[23] [c]
		AmacM1	tBLASTx	RNA (na)	0	[29] [d]
Anophelinae	*An. culicifacies*	AculA1	tBLASTx	RNA (424), DNA (1)	0	[9]
		AculA1	tBLASTx	RNA (50)	0	[23] [c]
		AculA1	tBLASTx	RNA (na)	0	[29] [d]
Anophelinae	*An. minimus A*	AminM1	tBLASTx	RNA (424), DNA (1)	2	[9]
		AminM1	tBLASTx	RNA (50)	1	[23] [c]
		AminM1	tBLASTx	RNA (na)	3	[29] [d]

Table 2 (*Continued*)

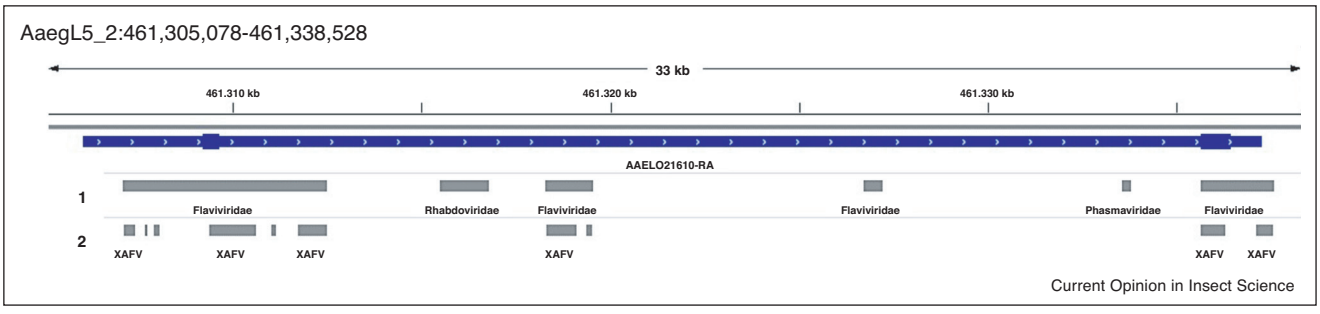
Subfamily	Species	Genome assembly	Method	Viruses in database	EVEs (in piRNA clusters[a])	Reference
Anophelinae	*An. funestus*	AfunF1	tBLASTx	RNA (424), DNA (1)	8	[9]
		AfunF1	tBLASTx	RNA (50)	0	[23] [c]
		AfunF1	tBLASTx	RNA (na)	2	[29] [d]
Anophelinae	*An. dirus*	AdirW1	tBLASTx	RNA (424), DNA (1)	4	[9]
		AdirW1	tBLASTx	RNA (50)	0	[23] [c]
		AdirW1	tBLASTx	RNA (na)	7	[29] [d]
Anophelinae	*An. farauti*	AfarF1	tBLASTx	RNA (50)	0	[23] [c]
		AfarF1	tBLASTx	RNA (424), DNA (1)	7	[9]
		AfarF1	tBLASTx	RNA (na)	2	[29] [d]
Anophelinae	*An. atroparvus*	AatrE1	tBLASTx	RNA (424), DNA (1)	3	[9]
		AatrE1	tBLASTx	RNA (50)	0	[23] [c]
		AatrE1	tBLASTx	RNA (na)	1	[29] [d]
Anophelinae	*An. sinensis*	AsinS2	tBLASTx	RNA (424), DNA (1)	2	[9]
		AsinS2	tBLASTx	RNA (50)	2	[23] [c]
		AsinS2	tBLASTx	RNA (na)	2	[29] [d]
		AsinC2	tBLASTx	RNA (50)	2	[23] [c]
		AsinC2	tBLASTx	RNA (na)	1	[29] [d]
Anophelinae	*An. albimanus*	AalbS1	tBLASTx	RNA (424), DNA (1)	0	[9]
		AalbS1	tBLASTx	RNA (50)	0	[23] [c]
		AalbS1	tBLASTx	RNA (na)	3	[29] [d]
Anophelinae	*An. darlingi*	AdarC3	tBLASTx	RNA (424), DNA (1)	0	[9]
		AdarC3	tBLASTx	RNA (50)	0	[23] [c]
		AdarC3	tBLASTx	RNA (na)	0	[29] [d]
Anophelinae	*An. koliensis*	AKwgs3	tBLASTx	RNA (50)	0	[23] [c]
		AKwgs3	tBLASTx	RNA (na)	1	[29] [d]
Anophelinae	*An. nili*	Anili1	tBLASTx	RNA (50)	0	[23] [c]
		Anili1	tBLASTx	RNA (na)	2	[29] [d]
Anophelinae	*An. punctulatus*	APwgs2	tBLASTx	RNA (50)	0	[23] [c]
		APwgs2	tBLASTx	RNA (na)	0	[29] [d]

[a] Number of viral integrations in piRNA clusters is reported when availble.
[b] Search for viral integrations included ssDNA and RNA viruses.
[c] Search for viral integrations was limited to Flavivirus.
[d] Search for viral integrations was limited to Chuviruses.

cISFs and do not appear to be orthologs [6•,28••,30]. Rhabdovirus-like integrations are more variable; they show similarity to both viruses infecting vertebrates (Bas Congo, Chandipura, Adelaide River, Bovine ephemeral fever and Berrimah viruses) and arthropod-restricted viruses (i.e. Parage aegeria sigmavirus, Culex tritaeniorhynchus rhabdovirus and Ohlosdorf virus) [6•,28••,30,33–35].

In several Anophelinae and in both *Aedes* species, the presence of viral integrations from the recently described *Chuviridae* viral family [36] is notable because Chuviridae-like nrEVEs encompass primarily viral glycoprotein sequences and are associated with, if not completely embedded within, long-terminal repeat (LTRs) of, mostly, the Bel/Pao family [6•,28••,29••]. These results suggest the occurrence of a specific interaction between

Figure 1



nrEVEs annotations vary across studies.
Integrative Genomics Browser visualization of a locus on the *Aedes aegypti* chromosome 2 where a different number and variety of nrEVEs have been annotated; 1 and 2 refer to nrEVES from Crava *et al.* [28••] and Spader *et al.* [30], respectively.

Chuviruses and Bel/Pao TEs, which resulted in the endogenization of viral *env* genes [28••,29••]. Phylogenetic analyses of both Bel/Pao TEs and Chuviridae-like nrEVEs support a single integration event before the split between Anophelinae and Culicinae or few endogenization events associated with an ancestral Pao element, followed by extensive recombination events as several Chuviridae-like nrEVEs were identified sharing high sequence similarity and occurring both linked to different clades of Pao TEs and not ('solo glycoproteins') [29••]. The estimate of the exact number of endogenization events is uncertain because alignments of the regions flanking integrations did not identify orthologous nrEVEs across mosquito species [29••], contrasting with the hypothesis of a unique endogenization event into the ancestor of the Culicidae family. Acquisition of the *env* gene by TEs can result in the gain of infectious properties and a retrovirus-like behaviour [37]. Whether Bel/Pao TEs, which endogenized Chuviridae-like glycoproteins, became retroviruses has not been demonstrated yet, but a total of 13 complete and potentially active elements, called Anakin, have been characterised in the genomes of *Ae. aegypti*, *Cx. quinquefasciatus* and the *Ae. albopictus* C6/36 cell line thus representing an ideal substrate to test this hypothesis [29••].

Lessons from viral integrations of wild-caught mosquitoes

Both *Ae. aegypti* and *Ae. albopictus* are invasive species for which detailed historical-records and extensive population genetic studies have allowed to define the route and timing of their global expansion from their respective native ranges [38–40,41•]. This knowledge represents the ideal scenario to unravel the origin and evolution of nrEVEs. Are endogenization events ancestral and sporadic or frequent? Are nrEVEs maintained in mosquito genomes by drift suggesting they are dispensable sequences, or do they harbour signs of selection suggesting a functional role?

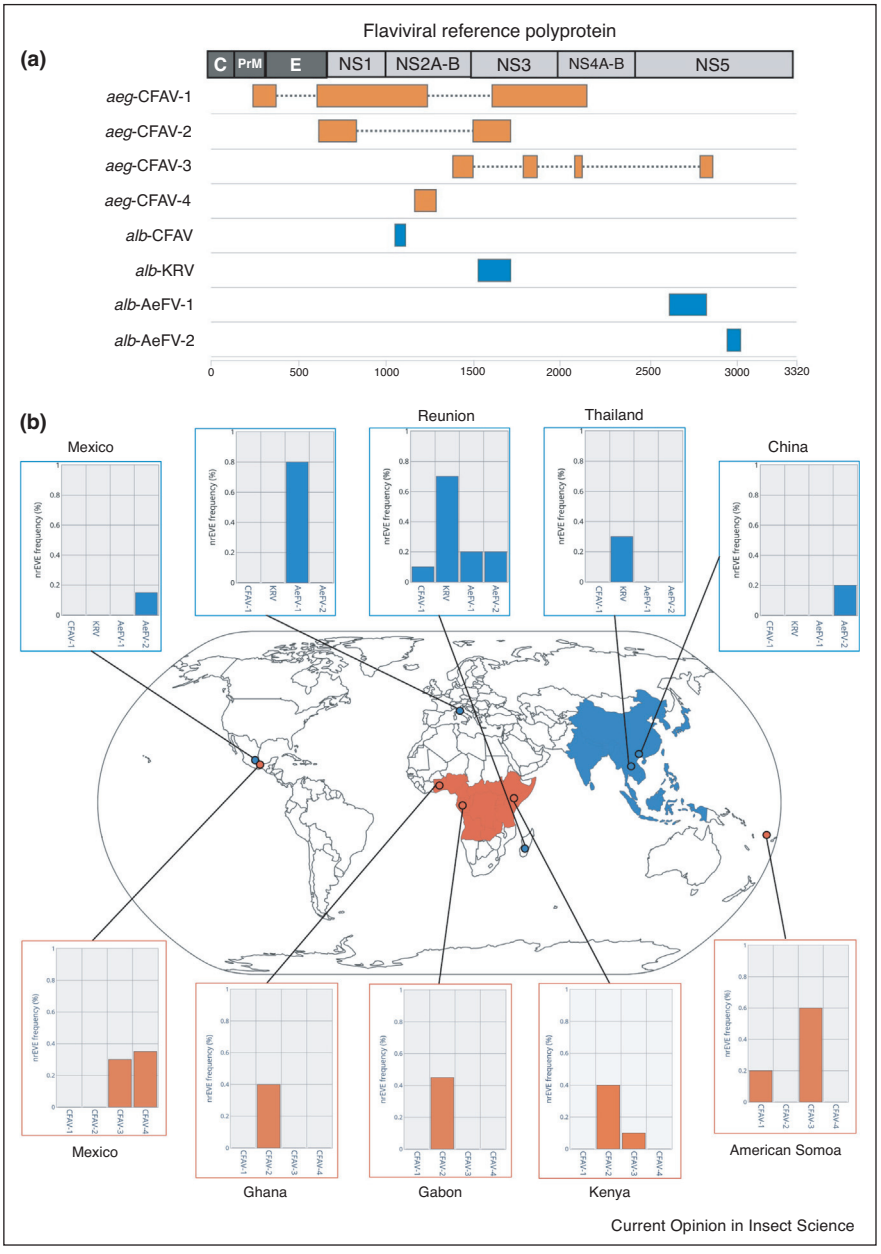
Data on the distribution of nrEVEs in wild mosquitoes are still limited, but differences between *Ae. aegypti* and *Ae. albopictus* were already revealed. For instance, flavivirus-like nrEVEs are highly conserved across *Ae. aegypti* samples, but variably distributed and less prevalent across *Ae. albopictus* wild mosquitoes; the opposite trend was observed for rhabdovirus-like nrEVEs [28••,30]. Whether differences in the number and prevalence of nrEVEs from different viruses reflect virus-host lineage-specific interactions or are dependent on the ecological niches and geographic origin of the two species is still unanswered, but the level of sequence polymorphism seen in nrEVEs in comparison to mosquito genes supports the conclusion that not all viral integrations are dispensable repetitive elements maintained by drift [42].

Using a viral database of 167 viruses previously identified as components of the mosquito virome, nrEVEs absent in the respective reference genomes were identified in both *Ae. aegypti* and *Ae. albopictus* wild-caught mosquitoes [6•,28••,42,43] (Figure 2). In *Ae. aegypti* samples, these newly identified integrations are composed of fragments, which are not contiguous on the corresponding viral genomes, suggesting rearrangements before integration [28••] (Figure 2a). In *Ae. albopictus*, the newly identified nrEVEs are shorter than 700 bp and correspond to a portion of a single viral open reading frame (ORF) [43,44••] (Figure 2a). In both cases, newly identified nrEVEs include integrations from cISFs and Aedes Anphenvirus, have >70% sequence similarity to currently circulating viruses and are found mostly in one or few populations [28••,43,44••] (Figure 2b). These results support the hypothesis that endogenization events are rare and some have occurred within the past 500 years when *Ae. aegypti* left Africa through the trade of slaves. These findings also support the hypothesis that substrates for integration are episomal viral DNA fragments that are formed from defective viral particles after viral infections of mosquitoes [45] and that endogenization events may involve short (<350 bp) and long rearranged sequences, which may be further shuffled after integration (Figure 2a) [28••]. In support of the hypothesis that endogenization events are rare, infection of *Ae. albopictus* with Chikungunya virus did not result in any integration event [46]. The exact frequency of integration events could be determined using one of the available *Aedes* cell-lines [47–49] and testing for multiple viral species, infection doses and passages. This experimental approach should be complemented with parallel analyses of the virome and the EVEome of wild caught mosquitoes to understand whether the occurrence and/or the frequency of integrations is viral lineage specific, correlated with the prevalence of environmental viruses and/or their route of infection (i.e. sexual transmission between male and females; vertical transmission to the progeny and/or hematophagous-dependent acquisition and transmission).

Genomic context of viral integrations and nrEVE functions

In the genome of *Aedes* spp. mosquitoes, nrEVEs are not distributed casually, but are statistically enriched in piRNA clusters and produce antisense primary piRNAs, suggesting that nrEVEs-piRNAs can target incoming viral nucleic acids [9,24••,25]. This hypothesis was recently validated in *Ae. aegypti* Aag2 cells and adult females through the identification of new CFAV integrations; the newly identified integration in adult *Ae. aegypti* is composed of different viral fragments, of which the NS2 gene coding region is shared with the integration detected in the genome of Aag2 cells [50••,51]. In both cases, CFAV replication was controlled through the targeting of viral RNA by piRNA derived from the NS2

Figure 2



Wild-caught *Aedes* mosquitoes harbour different viral integrations than the reference genome.
Novel viral integrations identified in the genomes of wild-caught *Aedes aegypti* (orange) and *Aedes albopictus* (blue) mosquitoes. **(a)** Structure of the newly identified nrEVEs in comparison to an archetypical Flavivirus genome. Viral integrations composed of fragments that are not contiguous on the viral genomes are shown with dotted lines. **(b)** Frequency of newly identified integrations in geographic populations. Blue and orange areas on the map show the native area of *Ae. albopictus* and *Ae. aegypti*, respectively.

region of the viral integration [50••,51]. This observation is interesting considering that the piRNA profile of nrEVEs is not homogeneous along the corresponding viral genome sequence and also varies in somatic versus germline cells, implicating sequence and tissue-specificity [28••,46]. Whether the different piRNAs expression profile of nrEVEs has functional implications should be further investigated considering the above-mentioned results and the redundancy of nrEVE. Redundancy is the presence of nrEVEs with high sequence similarity or mapping to the same regions of corresponding viral genomes; some of these nrEVEs may represent independent integrations whereas others are the results of rearrangements after integration. In adult mosquitoes,

nrEVE-derived piRNAs did not abrogate, but limited CFAV load exclusively in ovaries [51], suggesting nrEVEs may not have a major antiviral role against CFAV. Whether nrEVE antiviral activity can be modulated by the initial CFAV viral loads, the genetic background of mosquitoes, by controlling for the sequence-specificity with respect to cognate viruses or nrEVE location within somatic piRNA clusters should be further investigated given that mosquito immunity is a multifactorial process [52,53] and that in *Aedes* mosquitoes, the piRNA and siRNA pathways interact to control viral infections [51].

piRNA clusters of *Aedes* genomes are enriched not only of nrEVEs, but also of LTR retrotransposons [28••]. LTR retrotransposons replication cycle includes a cytoplasmatic stage where recombination with foreign RNA molecules may occur via copy-choice mechanism followed by reverse transcription [54]. Upon infection with RNA viruses, hybrid sequences composed of LTRs and vDNA fragments have been identified in *Aedes* cells and mosquitoes supporting the hypothesis that these episomal fragments are the substrate for integrations, in line with nrEVEs enrichment next to LTRs [28••,32,50••]. However, this may not be the only mechansim of integration as nrEVEs that are outside piRNA clusters are predominantly flanked by Class II transposons or no repetitive sequences as observed also for newly identified nrEVEs [6•,24••,28••]. nrEVEs mapping outside piRNA clusters can encompass complete viral ORFs and are expressed at the mRNA level [6•,28••,42]. Functional characterization of these nrEVEs could reveal new modes of action such as *in trans* competition with incoming viruses either at the mRNA or protein levels [55–58]. Additionally, mRNAs expressed from nrEVEs could also be degraded by the mosquito RNAi machinery into small interfering RNAs that could directly act on corresponding viral sequences. Unravelling nrEVEs function and regulation is essential to design genetic manipulation of nrEVEs either as direct antiviral effectors or as sites for integration of effectors.

Conclusions

Detailed analysis of genome assemblies and next-generation sequencing data from different mosquito species showed the widespread genomic distribution of nrEVEs. Current studies are concordant in depicting a larger number of nrEVEs in *Aedes* versus all other Culicinae and a trend for nrEVEs to be associated to TEs within piRNA clusters. In general, *Aedes* genomes harbour a higher percentage and a larger diversity of TEs than Anophelinae, suggesting that either transposition and endogenization rates are higher in *Aedes* species or these mosquitoes can better tolerate accumulation of repetitive sequences, nrEVEs included. Differences in the details of nrEVE annotation among studies are becoming relevant, and should be resolved, in light of the recent findings supporting biological roles for nrEVEs. nrEVEs have been shown to encode for piRNAs that are able to limit cognate viral infections; other nrEVEs encompass complete viral ORF expressed at the mRNA level; lastly, a complete BEL/Pao transposons with Chuviridae-like glycoprotein has been identified bioinformatically. Which, among the hundreds of nrEVEs of *Aedes* genomes, are functionally active and how? How specific would the nrEVEs antiviral activity be? These are some of the outstanding questions to be answered to propel the genetic engineering of nrEVEs for transmission-blocking vector control strategies.

Conflict of interest statement

Nothing declared.

Acknowledgements

We are grateful to current and past members of the Bonizzoni lab for fruitful discussion.

Funding was provided to MB by the Human Frontier Science Program Research grant (RGP0007/2017), the Italian Ministry of Education, University and Research FARE-MIUR project R1623HZAH5 and the European Research Council Consolidator Grant (ERC-CoG 682394) under the European Union's Horizon 2020 Programme. The Department of Biology and Biotechnology 'L. Spallanzani' also received financial support from the tIalian Ministry of Education, University and Research (MIUR) through the Dipartimenti d'Eccellenza Program (2018–2022). Funding sources had no role in study design; in the collection, analysis, and interpretation of data; in the writing of this review; and in the decision to submit the manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, Bartholomay L, Bidwell S, Caler E, Camara F *et al.*: **Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics**. *Science (80-)* 2010, **330**:86-88.

2. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, Amon J, Arcà B, Arensburger P, Artemov G *et al.*: **Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes**. *Science* 2015, **347**:1258522.

3. Lewis SH, Salmela H, Obbard DJ: **Duplication and diversification of dipteran Argonaute genes, and the evolutionary divergence of Piwi and Aubergine**. *Genome Biol Evol* 2016, **8**:507-518.

4. Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, Glassford WJ, Herre M, Redmond SN, Rose NH *et al.*: **Improved reference genome of *Aedes aegypti* informs arbovirus vector control**. *Nature* 2018, **563**:501-507.

5. • Melo ES, de, Wallau GL: **Mosquito genomes are frequently invaded by transposable elements through horizontal transfer**. *PLoS Genet* 2020, **16**:e1008946

Applying a comparative genomics approach, the authors characterized transposable elements (TEs) in mosquito genomes. TEs influence the genome size and variability and are assumed to be responsible for the production and genome integration of viral DNA from RNA viruses.

6. • Palatini U, Masri RA, Cosme LV, Koren S, Thibaud-Nissen F, Biedler JK, Krsticevic F, Johnston JS, Halbach R, Crawford JE *et al.*: **Improved reference genome of the arboviral vector *Aedes albopictus***. *Genome Biol* 2020, **21**:215

This article reports several insights from the latest *Aedes albopictus* reference genome, including the reference list of nrEVEs and their genomic position. In addition, newly identified nrEVEs in wild-mosquitoes, not present in the reference genome, are described.

7. Marconcini M, Hernandez L, Iovino G, Houé V, Valerio F, Palatini U, Pischedda E, Crawford JE, White BJ, Lin T *et al.*: **Polymorphism analyses and protein modelling inform on functional specialization of Piwi clade genes in the arboviral vector *Aedes albopictus***. *PLoS Negl Trop Dis* 2019, **13**:e0007919.

8. Sieglaff DH, Dunn WA, Xie XS, Megy K, Marinotti O, James AA: **Comparative genomics allows the discovery of *cis*-regulatory elements in mosquitoes**. *Proc Natl Acad Sci U S A* 2009, **106**:3053-3058.

9. 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**:512.

10. Souza-Neto JA, Powell JR, Bonizzoni M: ***Aedes aegypti* vector competence studies: a review**. *Infect Genet Evol* 2019, **67**:191-209.

11. da Fonseca Júnior DP, Serpa LLN, Barbosa GL, Pereira M, Holcman MM, Voltolini JC, Marques GRAM: **Vectors of arboviruses in the state of São Paulo: 30 years of *Aedes aegypti* and *Aedes albopictus***. *Rev Saude Publica* 2019, **53**:84.

12. Weetman D, Kamgang B, Badolo A, Moyes CL, Shearer FM, Coulibaly M, Pinto J, Lambrechts L, McCall PJ: ***Aedes* mosquitoes and Aedes-borne arboviruses in Africa: current and future threats**. *Int J Environ Res Public Health* 2018, **15**.

13. Kraemer MUG, Reiner RC, Brady OJ, Messina JP, Gilbert M, Pigott DM, Yi D, Johnson K, Earl L, Marczak LB *et al.*: **Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus***. *Nat Microbiol* 2019, **4**:854-863.

14. Blitvich B, Firth A: **Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization**. *Viruses* 2015, **7**:1927-1959.

15. Tree MO, McKellar DR, Kieft KJ, Watson AM, Ryman KD, Conway MJ: **Insect-specific flavivirus infection is restricted by innate immunity in the vertebrate host**. *Virology* 2016, **497**:81-91.

16. Öhlund P, Lundén H, Blomström AL: **Insect-specific virus evolution and potential effects on vector competence**. *Virus Genes* 2019, **55**:127-137.

17. Roundy CM, Azar SR, Rossi SL, Weaver SC, Vasilakis N: **Insect-specific viruses**. In *Advances in Virology*. Edited by Kielian M, Mettenleiter TC, Roossinck MJ. Academic Press; 2017:119-146.

18. Guzman H, Contreras-Gutierrez MA, Travassos da Rosa APA, Nunes MRT, Cardoso JF, Popov VL, Young KI, Savit C, Wood TG, Widen SG *et al.*: **Characterization of three new insect-specific flaviviruses: their relationship to the mosquito-borne flavivirus pathogens**. *Am J Trop Med Hyg* 2018, **98**:410-419.

19. Kobayashi D, Watanabe M, Faizah AN, Amoa-Bosompem M, Higa Y, Tsuda Y, Sawabe K, Isawa H: **Discovery of a novel flavivirus (*Flaviviridae*) from the horse fly, *Tabanus rufidens* (Diptera: *Tabanidae*): the possible coevolutionary relationships between the classical insect-specific flaviviruses and host dipteran insects**. *J Med Entomol* 2021, **58**:880-890.

20. Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, Belhouchet M, Lemasson JJ, de Micco P, de Lamballerie X: **Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes**. *J Gen Virol* 2004, **85**:1971-1980.

21. Roiz D, Vázquez A, Seco MPS, Tenorio A, Rizzoli A: **Detection of novel insect flavivirus sequences integrated in *Aedes albopictus* (Diptera: Culicidae) in Northern Italy**. *Virol J* 2009, **6**:93.

22. Katzourakis A, Gifford RJ: **Endogenous viral elements in animal genomes**. *PLoS Genet* 2010, **6**.

23. Lequime S, Lambrechts L: **Discovery of flavivirus-derived endogenous viral elements in *Anopheles* mosquito genomes supports the existence of Anopheles-associated insect-specific flaviviruses**. *Virus Evol* 2017, **3**.

24. •• 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* fenome**. *Curr Biol* 2017, **27**:3511-3519.e7
The authors presented a useful bioinformatic pipeline for the annotation of viral integrations in genome assemblies. This is the first study on nrEVEs that showed the advantages of using the BLASTx algorithm over tBLASTx.

25. ter Horst AM, Nigg JC, Dekker FM, Falk BW: **Endogenous viral elements are widespread in arthropod genomes and commonly give rise to PIWI-interacting RNAs**. *J Virol* 2019, **93**: e02124-18.

26. Russo AG, Kelly AG, Tuipulotu DE, Tanaka MM, White PA: **Novel insights into endogenous RNA viral elements in *Ixodes scapularis* and other arbovirus vector genomes**. *Virus Evol* 2019, **5**:1-18.

27. Aguiar ERGR, de Almeida JPP, Queiroz LR, Oliveira LS, Olmo RP, de Faria IJ, da S, Imler J-L, Gruber A, Matthews BJ, Marques JT: **A single unidirectional piRNA cluster similar to the *flamenco* locus is the major source of EVE-derived transcription and small RNAs in *Aedes aegypti* mosquitoes**. *RNA* 2020, **26**:581-594.

28. •• Crava CM, Varghese FS, Pischedda E, Halbach R, Palatini U, Marconcini M, Gasmi L, Redmond S, Afrane Y, Ayala D *et al.*: **Population genomics in the arboviral vector *Aedes aegypti* reveals the genomic architecture and evolution of endogenous viral elements**. *Mol Ecol* 2021, **30**:1594-1611 http://dx.doi.org/10.1111/mec.15798
This paper provides a detailed examination of nrEVEs in *Ae. aegypti* and their genomic context and organization. The authors describe the variable landscape of known nrEVEs in mosquitoes from different geographical locations and show signs of positive selection. In addition, they identify new nrEVEs derived from currently circulating insect-specific viruses whose endogenized sequences are rearranged.

29. •• Dezordi FZ, Vasconcelos CRDS, Rezende AM, Wallau GL: **In and outs of *Chuviridae* endogenous viral elements: origin of a potentially new retrovirus and signature of ancient and ongoing arms race in mosquito genomes**. *Front Genet* 2020, **11**:542437
A study on endogenized *Chuviridae* sequences in the genome of *Aedes* and other mosquitoes. The relationship between nrEVEs from Chuviruses and Pao family retroelements is investigated and a possibly functional Pao-chuvirus hybrid, termed Anakin, is described.

30. Spadar A, Phelan JE, Benavente ED, Campos M, Gomez LF, Mohareb F, Clark TG, Campino S: **Flavivirus integrations in *Aedes aegypti* are limited and highly conserved across samples from different geographic regions unlike integrations in *Aedes albopictus***. *Parasit Vectors* 2021, **14**:332.

31. Kojima S, Yoshikawa K, Ito J, Nakagawa S, Parrish NF, Horie M, Kawano S, Tomonaga K: **Virus-like insertions with sequence signatures similar to those of endogenous nonretroviral RNA viruses in the human genome**. *Proc Natl Acad Sci U S A* 2021, **118**:e2010758118.

32. Blair CD, Olson KE, Bonizzoni M: **The widespread occurrence and potential biological roles of endogenous viral elements in insect genomes**. *Curr Issues Mol Biol* 2020, **34**:13-30.

33. MacLachlan NJ, Dubovi EJ: **Rhabdoviridae**. *Fenner's Veterinary Virology*. edn 5. Elsevier; 2017:357-372.

34. Longdon B, Murray GGR, Palmer WJ, Day JP, Parker DJ, Welch JJ, Obbard DJ, Jiggins FM: **The evolution, diversity, and host associations of rhabdoviruses**. *Virus Evol* 2015, **1**:vev014.

35. Shahhosseini N, Lühken R, Jöst H, Jansen S, Börstler J, Rieger T, Krüger A, Yadouleton A, de Mendonça Campos R, Cirne-Santos CC *et al.*: **Detection and characterization of a novel rhabdovirus in *Aedes cantans* mosquitoes and evidence for a mosquito-associated new genus in the family *Rhabdoviridae***. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis* 2017, **55**:260-268.

36. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, Qin X-C, Li J, Cao J-P, Eden J-S *et al.*: **Redefining the invertebrate RNA virosphere**. *Nature* 2016, **540**:539-543.

37. Hayward A: **Origin of the retroviruses: when, where, and how?** *Curr Opin Virol* 2017, **25**:23-27.

38. Bonizzoni M, Gasperi G, Chen X, James AA: **The invasive mosquito species *Aedes albopictus*: current knowledge and future perspectives**. *Trends Parasitol* 2013, **29**:460-468.

39. Vega-Rúa A, Marconcini M, Madec Y, Manni M, Carraretto D, Gomulski LM, Gasperi G, Failloux A-B, Malacrida AR: **Vector competence of *Aedes albopictus* populations for chikungunya virus is shaped by their demographic history**. *Commun Biol* 2020, **3**:326.

40. Powell JR, Tabachnick WJ: **History of domestication and spread of *Aedes aegypti*-a review**. *Mem Inst Oswaldo Cruz* 2013, **108 (Suppl)**:11-17.

41. • Powell JR, Gloria-Soria A, Kotsakiozi P: **Recent history of *Aedes aegypti*: vector genomics and epidemiology records**. *Bioscience* 2018, **68**:854-860

This article contains a detailed analyses of the global invasion hystory of *Aedes aegypti* and the concurrent spread of vector-borne viral diseases.

42. Pischedda E, Scolari F, Valerio F, Carballar-Lejarazu R, Catapano PL, Waterhouse RM, Bonizzoni M: **Insights Into an unexplored component of the mosquito repeatome: distribution and variability of viral sequences integrated into the genome of the arboviral vector *Aedes albopictus***. *Front Genet* 2019, **10**:93.

43. Palatini U: **An improved genome for the Asian tiger mosquito *Aedes albopictus* and its applications in studying endogenous viral sequences**. *PhD Thesis*. 2021 http://dx.doi.org/10.13140/RG.2.2.25280.05121.

44. •• Pischedda E, Crava C, Carlassara M, Zucca S, Gasmi L, Bonizzoni M: **ViR: a tool to solve intrasample variability in the prediction of viral integration sites using whole genome sequencing data**. *BMC Bioinformatics* 2021, **22**:45 http://dx.doi.org/10.1186/s12859-021-03980-5

Here the authors describe a bioinformatic tool and framework for the detection of new viral integrations in whole genome sequencing data from single or pooled individuals collected in the wild or under experimental conditions. This method can be used in non-model organisms as long as a reference genome is available.

45. Goic B, Stapleford KA, Frangeul L, Doucet AJ, Gausson V, Blanc H, Schemmel-Jofre N, Cristofari G, Lambrechts L, Vignuzzi M *et al.*: **Virus-derived DNA drives mosquito vector tolerance to arboviral infection**. *Nat Commun* 2016, **7**:12410.

46. Marconcini M, Pischedda E, Houé V, Palatini U, Lozada-Chávez N, Sogliani D, Failloux A-B, Bonizzoni M: **Profile of small RNAs, vDNA forms and viral integrations in late chikungunya virus infection of *Aedes albopictus* mosquitoes**. *Viruses* 2021, **13**.

47. Walker T, Jeffries CL, Mansfield KL, Johnson N: **Mosquito cell lines: history, isolation, availability and application to assess the threat of arboviral transmission in the United Kingdom**. *Parasit Vectors* 2014, **7**:382.

48. Roberts GC, Zothner C, Remenyi R, Merits A, Stonehouse NJ, Harris M: **Evaluation of a range of mammalian and mosquito cell lines for use in Chikungunya virus research**. *Sci Rep* 2017, **7**:14641.

49. Brackney DE, Scott JC, Sagawa F, Woodward JE, Miller NA, Schilkey FD, Mudge J, Wilusz J, Olson KE, Blair CD *et al.*: **C6/36 *Aedes albopictus* cells have a dysfunctional antiviral RNA interference response**. *PLoS Negl Trop Dis* 2010, **4**:e856.

50. •• Tassetto M, Kunitomi M, Whitfield ZJ, Dolan PT, Sánchez-Vargas I, Garcia-Knight M, Ribiero I, Chen T, Olson KE, Andino R: **Control of RNA viruses in mosquito cells through the acquisition of vDNA and endogenous viral elements**. *eLife* 2019, **8**

This study demonstrates for the first time an antiviral activity for EVE-derived piRNAs. The authors showed that *Ae. aegypti* EVEs express antisense piRNA which upon binding with Piwi4 can reduce viral replication and promote tolerance to infection.

51. Suzuki Y, Baidaliuk A, Miesen P, Frangeul L, Crist AB, Merkling SH, Fontaine A, Lequime S, Moltini-Conclois I, Blanc H *et al.*: **Non-retroviral endogenous viral element limits cognate virus replication in *Aedes* aegypti ovaries**. *Curr Biol* 2020, **30**:3495-3506.e6.

52. Lee WS, Webster JA, Madzokere ET, Stephenson EB, Herrero LJ: **Mosquito antiviral defense mechanisms: a delicate balance between innate immunity and persistent viral infection**. *Parasit Vectors* 2019, **12**.

53. Drost HG, Sanchez DH: **Becoming a selfish clan: recombination associated to reverse-transcription in LTR retrotransposons**. *Genome Biol Evol* 2019, **11**:3382-3392.

54. Samuel GH, Adelman ZN, Myles KM: **Antiviral immunity and virus-mediated antagonism in disease vector mosquitoes**. *Trends Microbiol* 2018, **26**:447-461.

55. Frank JA, Feschotte C: **Co-option of endogenous viral sequences for host cell function**. *Curr Opin Virol* 2017, **25**:81-89.

56. Fujino K, Horie M, Honda T, Merriman DK, Tomonaga K: **Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome**. *Proc Natl Acad Sci U S A* 2014, **111**:13175-13180.

57. Honda T, Tomonaga K: **Endogenous non-retroviral RNA virus elements evidence a novel type of antiviral immunity**. *Mob Genet Elements* 2016, **6**:1-6.

58. Horie M, Tomonaga K: **Paleovirology of bornaviruses: what can be learned from molecular fossils of bornaviruses**. *Virus Res* 2019, **262**:2-9.