



HAL
open science

The diversity of endogenous viral elements in insects

Clément Gilbert, Carole Belliardo

► **To cite this version:**

Clément Gilbert, Carole Belliardo. The diversity of endogenous viral elements in insects. *Current Opinion in Insect Science*, 2022, 49, pp.48-55. 10.1016/j.cois.2021.11.007 . hal-03799718

HAL Id: hal-03799718

<https://hal.science/hal-03799718>

Submitted on 10 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The diversity of endogenous viral elements in insects

Clément Gilbert^{1*} & Carole Belliardo^{2,3}

¹ Université Paris-Saclay, CNRS, IRD, UMR Évolution, Génomes, Comportement et Écologie, 91198 Gif-sur-Yvette, France

² Université Côte d'Azur, INRAE, CNRS, Institut Sophia Agrobiotech, 06903 Sophia Antipolis, France

³ MYCOPHYTO, 50 Avenue de la Plaine, 06250, Mougins, France

*Corresponding author: clement.gilbert@egce.cnrs-gif.fr

Abstract

We provide an overview of the currently known diversity of viral sequences integrated into insect genomes. Such endogenous viral elements (EVE) have so far been annotated in at least eight insect orders and can be assigned to at least three families of large double-stranded (ds) DNA viruses, at least 22 families of RNA viruses, and three families of single-stranded DNA viruses. The study of these EVE has already produced important insights into insect-virus interactions, including the discovery of a new form of adaptive antiviral immunity. Insect EVE diversity will continue to increase as new insect genomes and exogenous viruses are sequenced, which will continue to make paleovirology a vibrant research field in this group of animals in the years to come.

Highlights

- Insect endogenous viral elements (EVE) belong to at least 28 viral families.
- The number of EVE is correlated to insect genome size.
- So far most insect EVE derive from ss(-)RNA viruses
- EVE studies inform us on insect-virus interactions, including antiviral immunity.
- EVE diversity will further grow with the number of insect genomes and metagenomes.

Declarations of interest: none.

Introduction

Endogenous viral elements (EVE) are viral genomes or fragments of viral genomes integrated into the genomes of their hosts [1]. Early studies of EVE were mainly focused on endogenous retroviruses (*Retroviridae*) and caulimoviruses (*Caulimoviridae*) which are abundant in vertebrate and plant genomes, respectively [2,3]. A few reports of EVE derived from other viral families were published in the late 1990's and early 2000's including some in insects [4–7], followed by larger scale studies of EVE originating from multiple viral families in various eukaryote lineages [1,8–10]. From then on, it quickly became apparent that any type of virus could undergo accidental chromosomal integration in any eukaryotic host and that annotating and analyzing EVE could yield important insights into the ancestry of viral families, virus host ranges and host-virus interactions [11]. Molecular dating using EVE as calibration points revealed unexpectedly deep origins of current viruses. For example, *Bornaviridae* and *Hepadnaviridae* were inferred to circulate in vertebrates for 300 and 100 million years, respectively, and insect large dsDNA viruses were shown to be as old as 300 million years [12–14]. Functional studies unveiled spectacular cases of EVE domestication, with vertebrate syncytin genes involved in placentation derived from multiple retroviral lineages [15], and gene/protein delivery agents from parasitoid wasps derived from several large dsDNA viruses [16]. Molecular domestication of EVE was also shown to fuel a variety of antiviral immune functions in eukaryotes, with the characterization of *Bornaviridae*- and *Retroviridae*-derived antiviral proteins [17,18] as well as of *Retroviridae*-derived interferon inducible enhancers in mammals [19]. Another remarkable case of EVE-derived antiviral immunity was recently reported in *Aedes* mosquitoes, whereby EVE-derived piRNA limit replication of viruses with complementary genomes [20,21]. It has been proposed that this form of heritable adaptive antiviral immunity may have induced a strong selective pressure on vertically-transmitted insect-specific viruses in the mosquito germline, in turn triggering the emergence of horizontally transmitted arboviruses [20].

Deciphering the role of EVE in mosquito antiviral immunity heavily relied on a comprehensive EVE annotation [22,23]. Such annotation was made possible not only thanks to high-quality mosquito genome assemblies but also because our knowledge of the insect virosphere has recently expanded, owing to several large-scale viral metagenomic studies [24–28]. One of these studies reported 1,445 RNA viruses from transcriptomes of over 220 invertebrate species [24]. These viruses were assigned to 16 large clades merging previously defined virus families, and five new clades were introduced to accommodate viruses showing extreme divergence to previously known viruses. These findings were then again considerably augmented by the characterization of another 1,213 RNA viruses belonging to at least 40 viral families, that were recovered in transcriptomes of more than 600 insect species from 32 orders [25]. In addition to these large-scale efforts, many viral metagenomics studies targeting species of interest such as *Drosophila melanogaster* [29], *Apis mellifera* [30] or *Aedes* mosquitoes [31] also contributed to expanding the catalogue of insect viruses. While these studies challenge the classical way viruses are characterized, they also offer unprecedented opportunities to better understand virus evolution and ecology, as well as the way by which host-virus interactions have shaped eukaryote evolution, including through viral endogenization [32,33]. Here we aim at complementing previous reviews on insect EVE [34,35] by providing an overview of the diversity of EVE that has so far been unearthed from insect genomes, putting this diversity into perspective of that of insect exogenous viruses. We present separately EVE deriving from large dsDNA viruses and those deriving from other viruses because these two types of EVE have generally been reported in separate publications owing to the different methods employed to characterize them.

Large dsDNA viruses

Large dsDNA viruses currently known to circulate in insects belong to three families of eukaryotic nucleo-cytoplasmic large DNA viruses (NCLDVs) (*Ascoviridae*, *Iridoviridae*, *Poxviridae*) and three families of viruses characterized by a circular genome and rod-shaped nucleocapsids (*Baculoviridae*, *Hytrosaviridae*, *Nudiviridae*) [36]. While irido-, pox- and nudi- viruses are widespread in several insect orders, baculoviruses have only been reported in Lepidoptera, Hemiptera and Diptera and hytrosaviruses are only known in Diptera and Hymenoptera. Ascoviruses are known to infect only lepidopteran larvae, mainly from the Noctuidae family [36]. Large dsDNA viruses are often excluded from EVE annotation studies because they possess many accessory genes that underwent recurrent host-to-virus and virus-to-virus horizontal transfer [37]. EVE annotation typically involves a first similarity search between viral proteins and genomes of interest using blast or related algorithms. This search is followed by a second, reciprocal similarity search between all viral protein-like sequences identified in the first search and a database including both viral and non-viral proteins. The reciprocal search ensures that viral-like proteins identified in the first search on viral-only proteins are not more similar to non-viral proteins. In the case of host sequences showing similarity to large dsDNA virus accessory proteins, it is often difficult to rule out non-viral origin, even when in-depth phylogenetic analyses involving dense sampling of host species are performed. Thus, most studies reporting large dsDNA virus EVE have so far focused on the detailed characterization of a single virus in one or a few host species. These include the discovery of multiple independent EVE domestication events underpinning the emergence of gene and/or protein delivery agents playing a role in host immune-suppression in parasitoid wasps (Ichneumonoidea and Cynipoidea) (Figure 1). So far, six domesticated large dsDNA viruses have been reported in parasitoid wasps, two of which are related to alphanudiviruses (in *Fopius arisanus* and *Venturia canescens*) [38,39], two are of unclear origin but possibly related to NCLDVs (in Campopleginae and Banchinae) [38,40–42], one is distantly related to hytrosaviruses (in *Leptopilina* spp.) [43] and one shares common ancestry with betanudiviruses (in microgastroid wasps) [16]. Multigene, large dsDNA virus-derived EVE with no evidence for domestication were also reported in a number of species, including three alphanudivirus-related EVE in the aphid *Aphis glycines* [44], the planthopper *Nilaparvata lugens* [45], and the chalcid parasitoid wasp *Eurytoma brunniventris* [46]. A betanudivirus-like EVE was also found in the *Eurytemora brunniventris* genome [46]. More recently, two larger-scale studies reported robust evidence for the presence of multigene alphanudivirus-like EVE in several other aphid species and in the whitefly *Bemisia tabaci*, as well as for hytrosavirus-like EVE in the wasps *Cotesia vestalis* and *Dolichomitus* sp [41,47]. In total, considering that all EVE found in aphids likely derive from independent integration events, at least 15 independent endogenization events involving multiple (>10) genes of large dsDNA viruses have so far been reported in insects (Figure 1).

In addition to these multi-gene EVE, 20 single-gene putative EVE similar to large dsDNA viruses have been reported in ants [48], including one showing similarity to the RNA polymerase RP0147, a core poxvirus gene shared by invertebrate and vertebrate poxviruses [49], and 16 showing similarity to PIF1, PIF2 or PIF3 genes, which are core baculovirus genes shared with hytrosaviruses [50]. These genes are all located in scaffolds longer than 10 kb, most of which also contain one or more host genes, indicating they are part of the ant genome. Yet, given the distant amino acid similarity of these putative EVE to viral proteins (31% on average) and their single-gene nature, definitive conclusion on their origin must await further analyses. This is also true for a number of putative single-gene EVE found in six dipterans, two lepidopterans and five other arthropod species, for which no definitive evidence of viral origin could be found [47]. The parasitoid killing factor genes, recently discovered in lepidopteran genomes may also be added to the list of EVE deriving from large dsDNA viruses [51], but more work is needed to assess whether these genes were indeed horizontally transferred from viruses to their hosts or whether viruses repeatedly captured them from their

hosts. Thus, we feel that robust evidence of endogenization exists only for three out of the six families of large dsDNA viruses known to infect insects and in only two insect orders (Hemiptera and Hymenoptera). These patterns are at least in part due to a bias in searching efforts, but we note that while whole-genome sequences of 240 arthropod species were screened for endogenous nudiviruses in [47], including 197 insect species from 14 orders, strongly-supported EVE have only been retrieved from hemipterans and hymenopterans. On the virus side, most insect large dsDNA EVE so far characterized derive from nudiviruses (11/15; Figure 1). Some properties of these viruses may facilitate endogenization, including the fact that they replicate into host cell nuclei and that one exogenous nudivirus is known both to replicate specifically in the gonads of its hosts and to undergo chromosomal integration in a host cell line [52].

RNA and small DNA viruses

The diversity of RNA and small DNA viruses for which partial or complete genomes have been sequenced in insects is extensive, with RNA viruses belonging to at least 40 families merged in 22 larger clades [24,25] and small DNA viruses from at least five families and many unclassified lineages [27,53]. Many studies of EVE deriving from RNA and small DNA viruses have been performed in insects, most of which focus on one or a few insect species. To provide an overview of the diversity of insect EVE and how it compares to that of exogenous viruses, we have chosen to illustrate the results of a large-scale EVE characterization in insects [54] and to discuss these results in light of other studies focusing on fewer insect species. We assigned EVE retrieved by ter Horst et al. [54] from 37 insect genomes to the 22 clades of RNA viruses proposed in Shi et al. [24] and to small DNA virus families [55], as well as to corresponding viral Classes of the International Committee on Taxonomy of Viruses (ICTV) [55]. To be conservative, we only kept EVE located in scaffolds longer than 10,000 bp and we merged EVE belonging to the same clade and located less than 50 bp away from each other. We retained 1,876 EVE from species belonging to six insect orders (Figure 2), which together with a few EVE retrieved in Isoptera [26] and Thysanoptera [56] brings to at least eight the number of insect orders in which EVE were found. These EVE fall into 14 of the 22 clades of RNA viruses proposed in [24] and three families of ssDNA viruses (Figure 2; supplementary table 1). The nine clades of RNA viruses totaling more than 10 EVE include members of 22 virus families. Importantly, EVE are distantly related to exogenous viruses (mean amino acid identity = 31% [SD=11%]), with only 12 EVE being more than 80% identical to their cognate exogenous virus (supplementary table 1).

The numbers reported in Figure 2 should not be seen as definitive because large-scale EVE screening depends on parameter tuning of similarity searches and on the quality of genome assemblies. However, the general trends outlined in Figure 2 are in line with other studies focusing on one or a few insect species. The number of EVE clades/families per species varies between 0 (in *Drosophila erecta*, *D. melanogaster*, *D. simulans* and *Apis mellifera*) to more than 250 in *Aedes* mosquitoes. The low number and diversity of EVE in *Drosophila* species is in agreement with a study reporting only one dsRNA partitivirus-like EVE segregating at low frequency in European *D. melanogaster* populations [29]. Also, in agreement with the results of [54] illustrated in Figure 2, all EVE screenings performed in *Aedes* mosquitoes revealed high numbers of highly diverse EVE [22,23,57,58], contrasting with low EVE prevalence in anopheline mosquitoes [59]. For example, in *A. aegypti*, Russo et al. [57] retrieved 242 EVE belonging to nine of the viral groups considered here, compared to 262 from 10 viral groups in Figure 2. The large number of EVE detected in the pea aphid *Acyrtosiphon pisum* is largely driven by *Parvoviridae* (n=114 in Figure 2), echoing earlier findings [10,53]. The high diversity of EVE in the ter Horst et al. [54] dataset is upheld by another large-scale EVE screening study of 20 ant species, in which EVE were assigned to eight of the RNA viral groups considered here and two families of small DNA viruses [48]. Overall, the number of EVE appears

strongly correlated to genome size in insects (Pearson's $r=0.72$, $p\text{-value}=3.8\times 10^{-07}$ for filtered EVE dataset; Pearson's $r=0.7$, $p\text{-value}=9.3\times 10^{-07}$ for unfiltered dataset; supplementary table 2), a trend reminiscent to that observed for transposable elements (TEs) and other repeats [60,61]. The majority of insect EVE annotated in [54] derive from negative sense RNA viruses ($n = 934$ versus 450, 258 and 233 for positive sense RNA viruses, double-stranded RNA viruses and small DNA viruses, respectively) (Figure 2). However, this overall trend is heavily driven by EVE landscapes from the two *Aedes* species (which total 534 EVE deriving from negative sense viruses) and EVE deriving from negative sense RNA viruses outnumber other EVEs in only 15 out of the 35 species included in the study.

The number of independent endogenization events in insect genomes cannot be directly inferred from the number of EVE as many EVE underwent post-endogenization duplications, likely mediated by (TEs) [22,56,57,59,62]. One way this has occurred is by amplification of retrotransposons that have co-opted virus-derived protein domains [63,64]. In fact, most *Chuviridae* EVE likely result from retrotransposition of long terminal repeat (LTR) elements containing co-opted chuvirus-like glycoprotein domains [64], hence their separation from other EVE in Figure 2. Likewise, many Hepe-Virga EVE observed in two lepidopteran species (Figure 2) were produced by retrotransposition-mediated amplification of *Virgaviridae* helicase-bearing retrotransposons [56,63]. That said, EVE clades/families are generally widespread among insects (eight of twelve are found in at least ten species) and relatively few insect EVE have been shown to be shared at orthologous loci among species. This indicates that endogenization of most viral groups considered here occurred independently and repeatedly along the insect tree.

A general trend emerging from the annotation of RNA and small DNA virus-derived EVE is that they tend to be enriched near retrotransposons, suggesting that endogenization often occurs through template-switching during retrotransposition. Furthermore, EVE are enriched in piRNA clusters in various species, setting endogenization as a possible source of heritable antiviral piRNA in several insect and other arthropod lineages [22,23,54,57,59]. While support for this hypothesis exists in *Aedes* mosquitoes, which possess an extended set of PIWI proteins [20,21], the ubiquity of EVE-derived piRNA-mediated antiviral immunity among other insects remains to be assessed [65]. Antiviral piRNA production is not the only function for which EVE have been domesticated in insects, as several EVE shared at orthologous loci between species were shown to evolve under purifying selection, pointing towards protein-level domestication. This is true for a *Rhabdoviridae*-derived EVE shared between worldwide populations of *Aedes aegypti* [62], for a *Phasmaviridae*- and a *Virgaviridae*-derived EVE shared between *Sophophora* and *Drosophila* flies [9,66] and for a *Mesoniviridae*-like EVE shared among widely-divergent dipterans species [57]. Based on the divergence time of their respective hosts, such orthologous EVE provide direct evidence for the antiquity of their cognate viral families, which circulated among insects millions of years ago [9,57,62,66]. Finally, there is strong evidence supporting the role of two densovirus-derived EVE in the ability of the pea aphid to produce winged offspring [67]. Remarkably, the two EVE are closely related to a mutualist exogenous densovirus found in rosy apple aphids, in which production of winged individuals is only possible in the presence of the virus. This suggested that the two viral genes endogenized in the pea aphid retained their function after endogenization [67].

Conclusion and perspectives

The diversity of insect EVE so far includes viruses related to three large dsDNA virus families, at least 14 of the 22 large clades of RNA viruses proposed in Shi et al. [24] and three families of ssDNA viruses. This diversity is not exceptional among arthropods [54,57,68] but is much higher than in vertebrates, in which EVE from only two families of non-retroviral RNA viruses and four families of

DNA viruses have been found to date [1,69,70]. Yet, the diversity of insect EVE is inherently underestimated given that most EVE are distantly related to current exogenous viruses and many could be placed in their own new family, some of which may have gone extinct. Furthermore, the catalogue of insect EVE will increase rapidly in the future, as new viruses identified by large-scale metagenomics are included as bait in similarity searches. The insect EVE landscape thus appears highly diverse and dynamic, and its study will undoubtedly yield new important contributions to our understanding of insect-virus interactions in the next few years. The development of new EVE screening pipelines will facilitate screening for the presence of multigene EVE derived from *Poxviridae* and *Iridoviridae*, which like nudiviruses are widespread in insects, and deciphering the history of large dsDNA virus domestication in both hymenopteran and non-hymenopteran insects. It will also be interesting to measure the proportion of EVE which have an impact on host fitness and the functions they may fulfill, two questions that will benefit from population genetics approaches, as illustrated by the case of *Aedes* mosquitoes [23,71]. A key question is to what extent EVE-derived antiviral immunity plays a role in natural mosquito populations, whether the evolution of this pathway has triggered the emergence of arboviruses [20] and whether variations in it affect vector competence [34]. Finally, the EVE landscape appears highly variable among insect species and characterizing the factors underpinning this variability – besides genome size and quality of genome assemblies [48] – will be a stimulating task.

Figure Legends

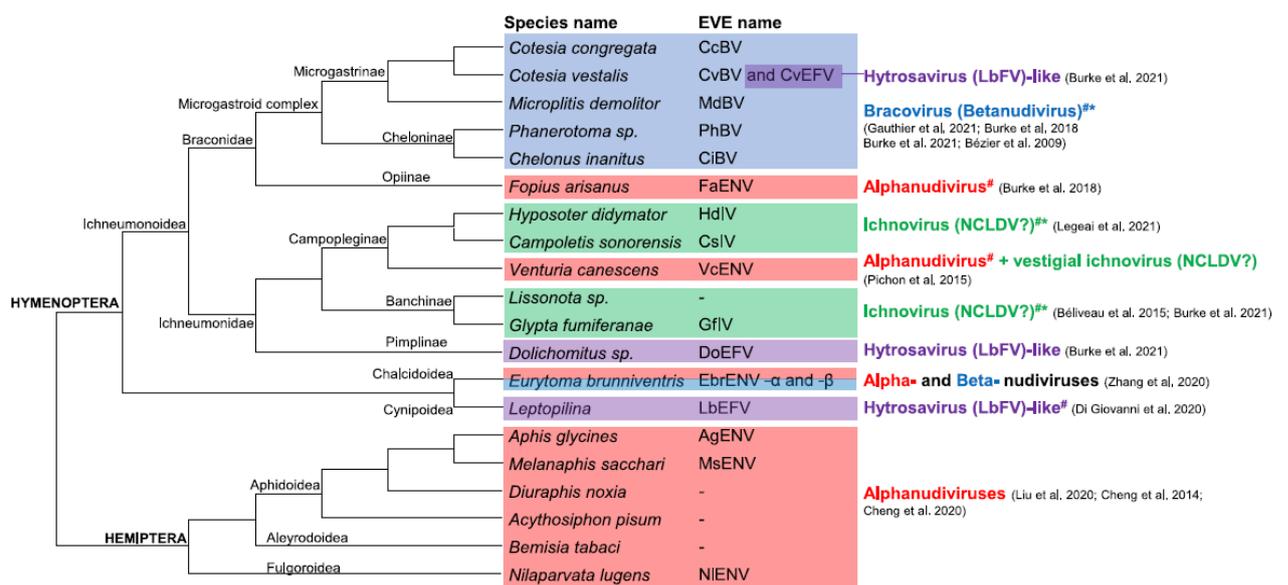


Figure 1. Diversity of endogenous viral elements (EVE) derived from large double-stranded DNA viruses. The tree of insects in which EVE were identified was built after [72,73]. Branch lengths are not proportional to divergence. References in which each EVE was characterized are indicated.

*indicates that EVE present in the group of species delineated by colored rectangles are orthologous or likely orthologous. #indicates domesticated EVE.

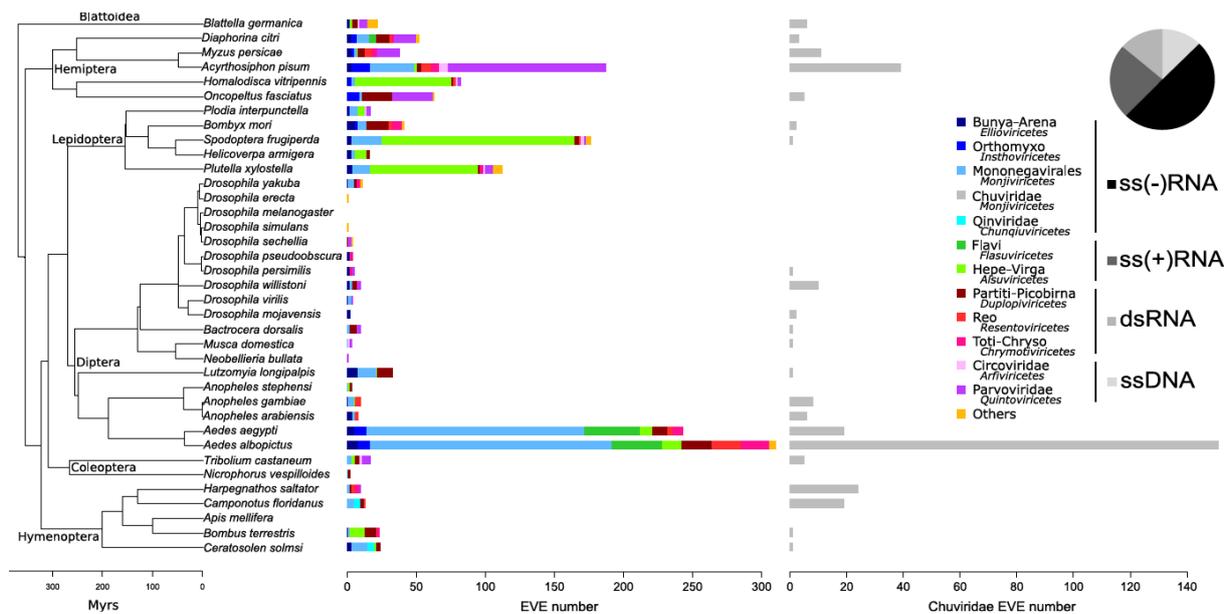


Figure 2. Diversity of endogenous viral elements (EVE) derived from RNA and small DNA viruses in insects. The tree and divergence times of insects in which EVE were searched in [54] was built after [74]. Numbers of EVE for each insect species are taken from Datasets S1 and S2 provided in [54]. Only EVE present in scaffolds longer than 10,000 bp were retained. EVE belonging to the same viral clade and separated by less than 50 bp were conservatively merged. Viral clades or families for which less than 10 EVE were identified in total were lumped in the “Others” category, which includes the following clades and families: *Astro-Poty*, *Benyviridae*, *Bidnaviridae*, *Geminiviridae*, *Kitaviridae*, *Luteo-Sobemo*, *Mayoviridae*, *Narna-Levi*, *Picornica-Calici*, *Tombus-Noda*. *Heliconius melpomene* is not included because the whole genome sequence used in [54] for this species could not be retrieved. EVE were classified by matching information of the closest viral blastx hit (GenBank Accession no. and/or viral protein and/or viral species) and the classification proposed in [24]. The Mono-Chu clade was split in Mononegavirales and *Chuviridae* because many putative *Chuviridae*-derived EVE may in fact correspond to glycoproteins of retrotransposons [64]. The corresponding viral Class proposed by the ICTV classification is shown in italics below each clade. Features of filtered EVE retained to build this figure are described in Supplementary Table 1.

Supplementary Table 1. Features of endogenous viral elements retained to build Figure 2. EVE were retrieved from Dataset 1 and 2 in ter Horst et al. (2018). Only EVE located on scaffolds longer than 10,000bp were retained. EVE were further merged if they were located less than 50 bp from each other on the same scaffold and belonged to the same virus clade. EVE for which taxonomy could not be retrieved in NCBI using the taxonomizr R package were assigned to a virus clade based on the supplementary table provided in Shi et al. (2016).

Supplementary Table 2: Number of EVE and genome size of insect species included in ter Horst et al. (2018). These values were used to assess whether genome size is correlated to EVE number in insects. We tested correlation with both the number of EVE provided in ter Horst et al. (2018) (“EVE not merged”) and the number of EVE we have retained here after removing all EVE located on scaffolds smaller than 10000 bp and after merging EVE belonging to the same viral clade and located less than 50 bp away from each other.

Acknowledgements

We apologize to our colleagues whose relevant work and original articles could not be cited owing to constraints on the number of references. We thank two reviewers and Richard Cordaux, Jonathan Filée and Alice Russo for their comments on an earlier version of this manuscript. C.G. acknowledges funding from Agence Nationale de la Recherche, project ANR-18-CE02-0021-01 TranspHorizon.

References

1. Katzourakis A, Gifford RJ: **Endogenous viral elements in animal genomes.** *PLOS Genetics* 2010, **6**:e1001191.
2. Diop SI, Geering ADW, Alfama-Depauw F, Loaec M, Teycheney P-Y, Maumus F: **Tracheophyte genomes keep track of the deep evolution of the Caulimoviridae.** *Sci Rep* 2018, **8**:572.
3. Johnson WE: **Origins and evolutionary consequences of ancient endogenous retroviruses.** *Nat Rev Microbiol* 2019, **17**:355–370.
4. Maori E, Tanne E, Sela I: **Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes.** *Virology* 2007, **362**:342–9.
5. Bejarano ER, Khashoggi A, Witty M, Lichtenstein C: **Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution.** *Proceedings of the National Academy of Sciences* 1996, **93**:759–764.
6. Delaroque N, Maier I, Knippers R, Müller DG: **Persistent virus integration into the genome of its algal host, *Ectocarpus siliculosus* (Phaeophyceae).** *Journal of General Virology* 1999, **80**:1367–1370.
7. Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, Belhouchet M, Lemasson J-J, de Micco P, de Lamballerie X: **Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes.** *Journal of General Virology* 2004, **85**:1971–1980.
8. Belyi VA, Levine AJ, Skalka AM: **Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old.** *J Virol* 2010, **84**:12458–62.
9. Ballinger MJ, Bruenn JA, Hay J, Czechowski D, Taylor DJ: **Discovery and evolution of bunyavirids in arctic phantom midges and ancient bunyavirid-like sequences in insect genomes.** *Journal of Virology* 2014, **88**:8783–94.
10. Liu H, Fu Y, Jiang D, Li G, Xie J, Cheng J, Peng Y, Ghabrial SA, Yi X: **Widespread Horizontal Gene Transfer from Double-Stranded RNA Viruses to Eukaryotic Nuclear Genomes.** *J Virol* 2010, **84**:11876–11887.
11. Feschotte C, Gilbert C: **Endogenous viruses: insights into viral evolution and impact on host biology.** *Nat Rev Genet* 2012, **13**:283–96.
12. Lauber C, Seitz S, Mattei S, Suh A, Beck J, Herstein J, Börold J, Salzburger W, Kaderali L, Briggs JAG, et al.: **Deciphering the Origin and Evolution of Hepatitis B Viruses by Means of a Family of Non-enveloped Fish Viruses.** *Cell Host & Microbe* 2017, **22**:387-399.e6.
13. Kawasaki J, Kojima S, Mukai Y, Tomonaga K, Horie M: **100-My history of bornavirus infections hidden in vertebrate genomes.** *Proc Natl Acad Sci USA* 2021, **118**:e2026235118.
14. Theze J, Bezier A, Periquet G, Drezen JM, Herniou EA: **Paleozoic origin of insect large dsDNA viruses.** *Proc Natl Acad Sci U S A* 2011, **108**:15931–5.
15. Cornelis G, Funk M, Vernochet C, Leal F, Tarazona OA, Meurice G, Heidmann O, Dupressoir A, Miralles A, Ramirez-Pinilla MP, et al.: **An endogenous retroviral envelope syncytin and its cognate**

receptor identified in the viviparous placental *Mabuya* lizard. *Proc Natl Acad Sci USA* 2017, **114**:E10991–E11000.

16. Gauthier J, Boulain H, van Vugt JJFA, Baudry L, Persyn E, Aury J-M, Noel B, Bretaudeau A, Legeai F, Warris S, et al.: **Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization.** *Commun Biol* 2021, **4**:104.

*Together with ref 40, this study provides a high-quality assembly of parasitic wasp genomes which reveals the detailed genome structure of a domesticated large dsDNA virus.

17. Armezzani A, Varela M, Spencer T, Palmarini M, Arnaud F: **“Ménage à Trois”: The Evolutionary Interplay between JSRV, enJSRVs and Domestic Sheep.** *Viruses* 2014, **6**:4926–4945.

18. Horie M, Tomonaga K: **Paleovirology of bornaviruses: What can be learned from molecular fossils of bornaviruses.** *Virus Research* 2019, **262**:2–9.

19. Chuong EB, Elde NC, Feschotte C: **Regulatory evolution of innate immunity through co-option of endogenous retroviruses.** *Science* in press,

20. 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**.

**Together with ref 21, this study provides functional evidence supporting the role of EVE-derived piRNA in antiviral immunity, and thus for adaptive immunity, in *Aedes aegypti*.

21. Suzuki Y, Baidaliuk A, Miesen P, Frangeul L, Crist AB, Merklings 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.** *Current Biology* 2020, **30**:3495-3506.e6.

22. 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.** *Current Biology* 2017, **27**:3511-3519.e7.

23. 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.

*Together with ref 71, this is the most comprehensive population-level analysis of EVE, revealing that a few contribute to adaptive evolution in *Aedes* mosquitoes.

24. 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, doi:10.1038/nature20167.

25. Wu H, Pang R, Cheng T, Xue L, Zeng H, Lei T, Chen M, Wu S, Ding Y, Zhang J, et al.: **Abundant and Diverse RNA Viruses in Insects Revealed by RNA-Seq Analysis: Ecological and Evolutionary Implications.** *mSystems* 2020, **5**.

**The most comprehensive viral metagenomics study performed in insects so far.

26. 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**.
27. Rosario K, Mettel KA, Benner BE, Johnson R, Scott C, Youssef-Vanegas SZ, Baker CCM, Cassill DL, Storer C, Varsani A, et al.: **Virus discovery in all three major lineages of terrestrial arthropods highlights the diversity of single-stranded DNA viruses associated with invertebrates.** *PeerJ* 2018, **6**:e5761.
28. Käfer S, Paraskevopoulou S, Zirkel F, Wieseke N, Donath A, Petersen M, Jones TC, Liu S, Zhou X, Middendorf M, et al.: **Re-assessing the diversity of negative strand RNA viruses in insects.** *PLoS Pathog* 2019, **15**:e1008224.
29. Wallace MA, Coffman KA, Gilbert C, Ravindran S, Albery GF, Abbott J, Argyridou E, Bellosta P, Betancourt AJ, Colinet H, et al.: **The discovery, distribution and diversity of DNA viruses associated with *Drosophila melanogaster* in Europe.** *Virus Evolution* 2021, doi:10.1093/ve/veab031.
30. Galbraith DA, Fuller ZL, Ray AM, Brockmann A, Frazier M, Gikungu MW, Martinez JFI, Kapheim KM, Kerby JT, Kocher SD, et al.: **Investigating the viral ecology of global bee communities with high-throughput metagenomics.** *Sci Rep* 2018, **8**:8879.
31. Shi C, Zhao L, Atoni E, Zeng W, Hu X, Matthijssens J, Yuan Z, Xia H: **Stability of the Virome in Lab- and Field-Collected *Aedes albopictus* Mosquitoes across Different Developmental Stages and Possible Core Viruses in the Publicly Available Virome Data of *Aedes* Mosquitoes.** *mSystems* 2020, **5**.
32. Obbard DJ: **Expansion of the metazoan virosphere: progress, pitfalls, and prospects.** *Current Opinion in Virology* 2018, **31**:17–23.
33. Bonning BC: **The Insect Virome: Opportunities and Challenges.** In *Insect Molecular Virology: Advances and Emerging Trends*. . Caister Academic Press; 2019.
34. Olson KE, Bonizzoni M: **Nonretroviral integrated RNA viruses in arthropod vectors: an occasional event or something more?** *Current Opinion in Insect Science* 2017, **22**:45–53.
35. Blair CD, Olson KE, Bonizzoni M: **The Widespread Occurrence and Potential Biological Roles of Endogenous Viral Elements in Insect Genomes.** In *Insect Molecular Virology: Advances and Emerging Trends*. . Caister Academic Press; 2019.
36. Williams T, Bergoin M, van Oers MM: **Diversity of large DNA viruses of invertebrates.** *Journal of Invertebrate Pathology* 2017, **147**:4–22.
37. Theze J, Takatsuka J, Nakai M, Arif B, Herniou EA: **Gene acquisition convergence between entomopoxviruses and baculoviruses.** *Viruses* 2015, **7**:1960–74.
38. Pichon A, Bézier A, Urbach S, Aury J-M, Jouan V, Ravallec M, Guy J, Cousserans F, Thézé J, Gauthier J, et al.: **Recurrent DNA virus domestication leading to different parasite virulence strategies.** *Sci Adv* 2015, **1**:e1501150.
39. Burke GR, Simmonds TJ, Sharanowski BJ, Geib SM: **Rapid Viral Symbiogenesis via Changes in Parasitoid Wasp Genome Architecture.** *Molecular Biology and Evolution* 2018, **35**:2463–2474.

40. Legeai F, Santos BF, Robin S, Bretaudeau A, Dikow RB, Lemaitre C, Jouan V, Ravallec M, Drezen J-M, Tagu D, et al.: **Genomic architecture of endogenous ichnoviruses reveals distinct evolutionary pathways leading to virus domestication in parasitic wasps.** *BMC Biol* 2020, **18**:89.

41. Burke GR, Hines HM, Sharanowski BJ: **The presence of ancient core genes reveals endogenization from diverse viral ancestors in parasitoid wasps.** *Genome Biology and Evolution* 2021, doi:10.1093/gbe/evab105.

*Together with ref 47, this study performs a comprehensive multi-species search for EVE deriving from large dsDNA viruses.

42. Beliveau C, Cohen A, Stewart D, Periquet G, Djoumad A, Kuhn L, Stoltz D, Boyle B, Volkoff AN, Herniou EA, et al.: **Genomic and Proteomic Analyses Indicate that Banchine and Campoplegine Polydnviruses Have Similar, if Not Identical, Viral Ancestors.** *Journal of Virology* 2015, **89**:8909–21.

43. Di Giovanni D, Lepetit D, Guinet B, Bennetot B, Boulesteix M, Couté Y, Bouchez O, Ravallec M, Varaldi J: **A Behavior-Manipulating Virus Relative as a Source of Adaptive Genes for *Drosophila* Parasitoids.** *Molecular Biology and Evolution* 2020, **37**:2791–2807.

**The first report of a domesticated large dsDNA virus in a species not belonging to the Ichneumonoidea superfamily of parasitoid wasps.

44. Liu S, Coates BS, Bonning BC: **Endogenous viral elements integrated into the genome of the soybean aphid, *Aphis glycines*.** *Insect Biochemistry and Molecular Biology* 2020, **123**:103405.

45. Cheng R-L, Xi Y, Lou Y-H, Wang Z, Xu J-Y, Xu H-J, Zhang C-X: **Brown Planthopper Nudivirus DNA Integrated in Its Host Genome.** *Journal of Virology* 2014, **88**:5310–5318.

46. Zhang Y, Wang J, Han G-Z: **Chalcid wasp paleoviruses bridge the evolutionary gap between bracoviruses and nudiviruses.** *Virology* 2020, **542**:34–39.

47. Cheng R-L, Li X-F, Zhang C-X: **Nudivirus Remnants in the Genomes of Arthropods.** *Genome Biology and Evolution* 2020, **12**:578–588.

48. Flynn PJ, Moreau CS: **Assessing the Diversity of Endogenous Viruses Throughout Ant Genomes.** *Front Microbiol* 2019, **10**:1139.

**Together with refs 54 and 57, this large-scale EVE screening reveals a large diversity of EVE in multiple insect species.

49. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL: **Poxvirus Orthologous Clusters: toward Defining the Minimum Essential Poxvirus Genome.** *J Virol* 2003, **77**:7590–7600.

50. Garavaglia MJ, Miele SAB, Iserte JA, Belaich MN, Ghiringhelli PD: **The ac53, ac78, ac101, and ac103 Genes Are Newly Discovered Core Genes in the Family Baculoviridae.** *Journal of Virology* 2012, **86**:12069–12079.

51. Gasmi L, Sieminska E, Okuno S, Ohta R, Coutu C, Vatanparast M, Harris S, Baldwin D, Hegedus DD, Theilmann DA, et al.: **Horizontally transmitted parasitoid killing factor shapes insect defense to parasitoids.** *Science* 2021, **373**:535–541.

52. Gauthier J, Drezen J-M, Herniou EA: **The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps.** *Parasitology* 2018, **145**:713–723.
53. Francois S, Filloux D, Roumagnac P, Bigot D, Gayral P, Martin DP, Froissart R, Ogliastro M: **Discovery of parvovirus-related sequences in an unexpected broad range of animals.** *Sci Rep* 2016, **6**:30880.
54. ter Horst AM, Nigg JC, Falk BW: **Endogenous viral elements are widespread in arthropod genomes and commonly give rise to piRNAs.** 2018, doi:10.1101/396382.
- *A detailed investigation of the role of an EVE in piRNA-derived antiviral immunity in a non-mosquito insect, revealing that EVE-derived piRNA are not involved in the production of antiviral secondary piRNA.
55. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Dempsey DM, Dutilh BE, Harrach B, Harrison RL, Hendrickson RC, et al.: **Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020).** *Arch Virol* 2020, **165**:2737–2748.
56. Kondo H, Chiba S, Maruyama K, Andika IB, Suzuki N: **A novel insect-infecting virga/nege-like virus group and its pervasive endogenization into insect genomes.** *Virus Research* 2019, **262**:37–47.
57. Russo AG, Kelly AG, Enosi Tuipulotu D, Tanaka MM, White PA: **Novel insights into endogenous RNA viral elements in Ixodes scapularis and other arbovirus vector genomes.** *Virus Evolution* 2019, **5**:vez010.
58. 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.
59. 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**.
60. Lower SS, Johnston JS, Stanger-Hall KF, Hjelman CE, Hanrahan SJ, Korunes K, Hall D: **Genome Size in North American Fireflies: Substantial Variation Likely Driven by Neutral Processes.** *Genome Biology and Evolution* 2017, **9**:1499–1512.
61. Sessegolo C, Burlet N, Haudry A: **Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies.** *Biology Letters* 2016, **12**:20160407.
62. Fort P, Albertini A, Van-Hua A, Berthomieu A, Roche S, Delsuc F, Pasteur N, Capy P, Gaudin Y, Weill M: **Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality.** *Molecular Biology and Evolution* 2012, **29**:381–90.
63. Morozov SY, Lazareva EA, Solovyev AG: **RNA helicase domains of viral origin in proteins of insect retrotransposons: possible source for evolutionary advantages.** *PeerJ* 2017, **5**:e3673.
- *Together with ref 64 EVE have been co-opted and amplified by retrotransposons in insects.
64. Dezordi FZ, Vasconcelos CR dos S, 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.

65. Nigg JC, Kuo Y-W, Falk BW: **Endogenous Viral Element-Derived Piwi-Interacting RNAs (piRNAs) Are Not Required for Production of Ping-Pong-Dependent piRNAs from Diaphorina citri Densovirus.** *mBio* 2020, **11**.
66. Kirsip H, Abroi A: **Protein Structure-Guided Hidden Markov Models (HMMs) as A Powerful Method in the Detection of Ancestral Endogenous Viral Elements.** *Viruses* 2019, **11**:320.
67. Parker BJ, Brisson JA: **A Laterally Transferred Viral Gene Modifies Aphid Wing Plasticity.** *Current Biology* 2019, **29**:2098-2103.e5.
68. Metegnier G, Becking T, Chebbi MA, Giraud I, Moumen B, Schaack S, Cordaux R, Gilbert C: **Comparative paleovirological analysis of crustaceans identifies multiple widespread viral groups.** *Mob DNA* 2015, **6**:16.
69. Aswad A, Aimola G, Wight D, Roychoudhury P, Zimmermann C, Hill J, Lassner D, Xie H, Huang M-L, Parrish NF, et al.: **Evolutionary History of Endogenous Human Herpesvirus 6 Reflects Human Migration out of Africa.** *Molecular Biology and Evolution* 2021, **38**:96–107.
70. Cui J, Zhao W, Huang Z, Jarvis ED, Gilbert MTP, Walker PJ, Holmes EC, Zhang G: **Low frequency of paleoviral infiltration across the avian phylogeny.** *Genome Biol* 2014, **15**:539.
71. Houé V, Gabiane G, Dauga C, Suez M, Madec Y, Mousson L, Marconcini M, Yen P-S, de Lamballerie X, Bonizzoni M, et al.: **Evolution and biological significance of flaviviral elements in the genome of the arboviral vector *Aedes albopictus*.** *Emerging Microbes & Infections* 2019, **8**:1265–1279.
72. Sharanowski BJ, Ridenbaugh RD, Piekarski PK, Broad GR, Burke GR, Deans AR, Lemmon AR, Moriarty Lemmon EC, Diehl GJ, Whitfield JB, et al.: **Phylogenomics of Ichneumonoidea (Hymenoptera) and implications for evolution of mode of parasitism and viral endogenization.** *Molecular Phylogenetics and Evolution* 2021, **156**:107023.
73. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al.: **Phylogenomics resolves the timing and pattern of insect evolution.** *Science* 2014, **346**:763–7.
74. Kumar S, Stecher G, Suleski M, Hedges SB: **TimeTree: A Resource for Timelines, Timetrees, and Divergence Times.** *Molecular Biology and Evolution* 2017, **34**:1812–1819.