Virus Research

(special issue: "Integrated Viral Sequence").

A novel insect-infecting virga/nege-like virus group and its pervasive endogenization into insect genomes

Hideki Kondo [a] *, Sotaro Chiba [b, c], Kazuyuki Maruyama [a], Ida Bagus Andika [a], Nobuhiro Suzuki [a]

[a] *Institute of Plant Science and Resources (IPSR), Okayama University, Kurashiki 710-0046, Japan*

[b] *Asian Satellite Campuses Institute, Nagoya University, Nagoya 464-8601, Japan.*

[c] *Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan.*

***Corresponding author**

Hideki Kondo

Institute of Plant Science and Resources (IPSR), Okayama University,

Kurashiki 710-0046, Japan

E-mail: hkondo@okayama.u-ac.jp

Tel. +81(86) 434-1232 / Fax. +81(86) 434-1232

MS information: Abstract, 241 words; Main text, 6187 words; Figures, 5; Supplemental tables, 3; Supplemental figures, 7.

Abstract

Insects are the host and vector of diverse viruses including those that infect vertebrates, plants, and fungi. Recent wide-scale transcriptomic analyses have uncovered the existence of a number of novel insect viruses belonging to an alphavirus-like superfamily (virgavirus/negevirus-related lineage). In this study, through an *in silico* search using publicly available insect transcriptomic data, we found numerous virus-like sequences related to insect virga/nege-like viruses. Phylogenetic analysis showed that these novel viruses and related virus-like sequences fill the major phylogenetic gaps between insect and plant virga/negevirus lineages. Interestingly, one of the phylogenetic clades represents a unique insect-infecting virus group. Its members encode putative coat proteins which contained a conserved domain similar to that usually found in the coat protein of plant viruses in the family *Virgaviridae*. Furthermore, we discovered endogenous viral elements (EVEs) related to virga/nege-like viruses in the insect genomes, which enhances our understanding on their evolution. Database searches using the sequence of one member from this group revealed the presence of EVEs in a wide range of insect species, suggesting that there has been prevalent infection by this virus group since ancient times. Besides, we present detailed EVE integration profiles of this virus group in some species of the *Bombus* genus of bee families. A large variation in EVE patterns among *Bombus* species suggested that while some integration events occurred after the species divergence, others occurred before it. Our analyses support the view that insect and plant virga/nege-related viruses might share common virus origin(s).

Keywords

Endogenous viral element; Whole genome shotgun assembly; Transcriptome shotgun assembly; Insect; Bumblebee; Plant alpha-like virus; Evolution

1. Introduction

Positive-strand RNA (+ssRNA) viruses infecting eukaryotes are divided into three major evolutionary lineages, the picornavirus-like, flavivirus-like and alphavirus-like superfamilies (Koonin and Dolja, 1993). The alphavirus-like superfamily is a dominant group of plant viruses consisting of the members of the order *Tymovirales* as well as several other families or genera; however, there are only a few taxon of vertebrate (family *Togaviridae*) and invertebrate (family *Alphatetraviridae*) viruses belonging to this group (Dolja and Koonin, 2011; Koonin et al., 2015). These viruses encode replication-associated protein(s) that contain the same order of 5'-cap methyltransferase (MET), superfamily 1 helicase (RNA helicase, HEL), and RNA-dependent RNA polymerase (RdRp) domains (Koonin and Dolja, 1993). Plant alpha-like viruses belonging to the families *Bromoviridae*, *Closteroviridae*, and *Virgaviridae*, with a segmented or non-segmented genome form a large monophyletic group (named virgavirus-related lineage) (Adams et al., 2009; Liu et al., 2009). The members of *Bromoviridae* and *Closteroviridae* families with icosahedral (or bacilliform) or filamentous particles are known to be transmitted by aphids, mealybugs, or whiteflies in a non- or semi-persistent (non-replicative) manner (Whitfield et al., 2015). Virgaviruses with rod shaped virions, except for tobamoviruses (genus *Tobamovirus*) whose vectors are unreported, are mainly transmitted by soil-inhabiting organisms (plasmodiophorid protists or nematodes) (Andika et al., 2016). Recent reports have revealed the presence of novel segmented viruses, such as citrus leprosis virus cytoplasmic type (CiLV-C, a floating genus *Cilevirus*) (Locali-Fabris et al., 2006), whose non-enveloped bacilliform particles are transmitted by false spider mites, *Brevipalpus* spp., in a possibly persistent-circulative (non-replicative) manner (Tassi et al., 2017). CiLV-C and newly discovered related viruses (genera *Cilevirus* and *Higrevirus*) are distantly related to virga-, brom-, and closteroviruses and have a large replicase-associated protein (~298 kDa) containing additional domains (cysteine protease-like and/or FtsJ-like methyltransferase, FtsJ) between the MET and HEL domains (Locali-Fabris et al., 2006; Melzer et al., 2013; Roy et al., 2013; 2015). An FtsJ-like methyltransferase domain is also present at the N-terminal of NS5 protein (RdRp) of flaviviruses

(family *Flaviviridae*), which exhibits a methyltransferase activity on 5'-capped RNA (Egloff et al, 2002).

Over the past decade, advances in next generation sequencing and metagenomics have expanded our knowledge of virus diversities, including those of insect-specific viruses whose host range is restricted to arthropods, but they are phylogenetically related to arthropod-borne viruses that are associated with disease in vertebrates (Bolling et al., 2015; Li et al., 2015; Vasilakis and Tesh, 2015). The newly recognized alpha-like insect-specific viruses (so called "Negevirus") from mosquitos and sandflies have a non-segmented RNA genome with unique spherical/elliptical virions (Nabeshima et al., 2014; Vasilakis et al., 2013). Phylogenetic analyses of negeviruses (negevirus-related lineage) suggest that they are distantly related to plant cile/higreviruses, and form two distinct clades, called "Sandewavirus" and "Nelorpivirus", whose genomes encode a large replicase-associated protein with an FtsJ-like domain (Carapeta et al., 2015; Kallies et al., 2014; Nunes et al., 2017; Vasilakis et al., 2013). More recently, a wide-range survey of RNA viruses using deep transcriptome sequencing revealed the presence of a dozen of insect alpha-like viruses related to virga/nege-like viruses (above mentioned viruses) (Shi et al., 2016). Therefore, there is still very little information about the diversity of invertebrate alpha-like viruses and their evolutionary history.

The recent availability of genome sequences of a large number of eukaryotes has led to the discovery of endogenous non-retroviral RNA virus-like elements, known as endogenous viral elements (EVEs) or non-retroviral integrated RNA viruses (Aiewsakun and Katzourakis, 2015; Chu et al., 2014; Feschotte and Gilbert, 2012; Holmes, 2011; Horie and Tomonaga, 2011). Non-retroviral viruses lack an integrase and an integration process during their life cycle. Therefore, the heritable horizontal transfer of genetic information from these type of viruses to eukaryotic genomes is likely to occur accidentally during infection (Feschotte and Gilbert, 2012). The identification of non-retroviral EVEs in eukaryotic genomes provides an interesting insight into the long-term viral evolution and host-virus interactions as well as host-virus coevolution (Aiewsakun and Katzourakis, 2015; Feschotte and Gilbert, 2012). To date, a large number of non-retroviral EVEs related to dsRNA viruses and

negative stranded (−)ssRNA viruses, such as members of families *Totiviridae* and *Partitiviridae*, and order *Mononegavirales*, have been widely discovered in plant, insect, and/or fungal genomes (Chiba et al., 2011; Katzourakis and Gifford, 2010; Kondo et al., 2013a; Liu et al., 2010; Taylor and Bruenn, 2009). In contrast, only a limited number of non-retroviral EVEs related to (+)ssRNA viruses are found in plant and insect genomes (Aiewsakun and Katzourakis, 2015). These EVEs are mainly associated with flaviviruses (Katzourakis and Gifford, 2010) and some plant alphavirus-like viruses (Chiba et al., 2011; Cui and Holmes, 2012; da Fonseca et al., 2016; Kondo et al., 2013b).

In this study, we conducted an *in silico* search using public transcriptomic data to further explore the occurrence of alpha-like viruses (virga/negevirus linage) in a wide range of insect species. We found numerous insect transcriptome shotgun assembly (TSA) accessions related to insect virga/nege-like viruses. We further expanded the investigation of non-retroviral EVE sequences related to insect virga/nege-like viruses in the genome of divergent insect species, and characterized their integration pattern in the genome of some insect species. Our analyses suggest a close evolutionary relationship between insect and plant virga/nege-like viruses.

2. Materials and Methods

2.1. Database searches and sequence analyses

BLAST (tblastn) searches were conducted against genome sequence databases available in the National Center for Biotechnology Information (NCBI) (whole-genome shotgun contigs, WGS; non-human, non-mouse expressed sequence tags, EST; transcriptome shotgun assembly, TSA) (http://www.ncbi.nlm.nih.gov/).

Open reading frames (ORFs) encoded insect virus-like TSAs or EVE candidates were identified using EnzymeX version 3.3.3 (http://nucleobytes.com/enzymex/index.html). Sequence similarities were calculated using the BLAST program available from NCBI. The putative insect EVEs that matched viral proteins with E-values < 0.001 were extracted, and a possible ORF was

restored by adding Ns as unknown sequences to obtain continuous amino acid sequences (Xs, edited residues). Putative insect EVE encoded peptides were then used to screen the Genbank non-redundant (nr) database in a reciprocal tblastn search. Transposable element sequences were identified using the Censor (http://www.girinst.org/censor/index.php) (Kohany et al., 2006). Detailed methods for EVE analyses have been previously described (Kondo et al., 2015).

2.2. Insect materials

Bombus terrestris and *Bombus ignitus* individuals were collected at Naganuma Town, Hokaido, Japan (43˚ north latitude, 141˚ east longitude) and kindly provided by Drs. F. Tanaka and Y. Hashimoto (Agriculture Research Department, Hokkaido Research Organization, Japan) or were collected at Takahashi City, Okayama, Japan (34˚ north latitude, 133˚ east longitude) and kindly provided by Dr. M. Tobikawa (Okayama Prefectural Takahashi Agricultural Extension Center, Japan). *Bombus hypocrite* individuals were kindly provided by N. Abe and T. Ayabe (Institute for Ecosystem & Firefly Breeding, Itabashi-Ward, Tokyo, Japan).

2.3. Nucleic acid extraction, PCR amplification and DNA sequencing

Total genomic DNA was purified using the DNeasy ® Blood and Tissue Kit (Qiagen). Purified DNA samples were stored at 4°C until use. To amplify the insect EVEs from the DNA samples by PCR, primer pairs were designed based on the virus-related (EVE candidate) sequences and their flanking sequences (see Table S1). The sequences of the primers used in this analysis are available upon request. PCR amplification was performed using Quick Taq HS DyeMix (TOYOBO.) The PCR conditions were an initial denaturation at 94°C for 2 min, followed by 35 cycles at 94°C for 10 s, 58°C for 30 s, and 68°C for 1 min, then finishing with final extension at 68°C for 10 min. PCR products were separated by agarose gel electrophoresis. Purified PCR fragments were sequenced using an ABI3100 DNA sequencer (Applied Biosystems, Foster City, CA, USA). For selected EVE

fragments, PCR products were cloned in pGEM T-easy (Promega) and sequenced. The obtained sequences were assembled and analyzed using the DNA processing software packages, AutoAssembler (Applied Biosystems) and GENETYX (SDC, Tokyo, Japan).

2.4. Phylogenetic analysis

Restored and estimated amino acid sequences of EVE candidates and possible viral elements in transcriptomic databases were aligned with MAFFT version 6 under the default parameters (Katoh and Toh, 2008) (http://mafft.cbrc.jp/alignment/server). Ambiguously aligned regions were removed using Gblocks 0.91b (Talavera and Castresana, 2007), with the stringency levels lowered for all parameters. The maximum likelihood (ML) phylogenetic trees were estimated using PhyML 3.0 (Guindon et al., 2010), with automatic model selection using smart model selection (SMS) (http://www.atgc-montpellier.fr/phyml-sms/). The branch support values were estimated by the approximate likelihood ratio test (aLRT) with a Shimodaira–Hasegawa–like (SH-like) algorithm (Anisimova and Gascuel, 2006). A coat protein (CP) neighbor-joining (NJ) tree was constructed by MAFFT. Trees were represented with midpoint rooting and visualized using FigTree (version 1.3.1) (http://tree.bio.ed.ac.uk/software/figtree/).

3. Results

3.1. Identification of virga/negevirus-like sequences in insect transcriptome shotgun assembly data

Previous studies have reported the presence of novel virga/nege-like viruses in insects (Shi et al., 2016). In order to further explore the possible occurrence of this or related virus groups in a wide range of insect species, we took advantage of publicly accessible TSA sequence libraries in NCBI, as previously used to identify novel virus-like sequences in plant or fungal species (Jo et al., 2016; Kondo et al., 2016; Mushegian et al., 2016; Nibert et al., 2016). We performed tblastn searches using

a replicase sequence from representative members of the virga/nege-like virus group [such as CiLV-C and Hubei virga-like virus 2 (HVLV2, an insect-infecting virus)] (Locali-Fabris et al., 2006; Shi et al., 2016) as the queries against insect TSA databases. As expected, the searches yielded numerous virus-related sequences with significant E-values ($E=0.0\sim e^{-50}$) (data not shown). Notably, a number of virus-related sequences were found in TSA data of many insect species belonging to the order Hymenoptera, that previously were not known to harbor this virus group. For further analyses, TSA accessions shorter than 9.0 kb were excluded, and only one TSA sequence from nearly identical accessions was chosen. We selected 27 distinct TSA accessions with potentially complete or near complete genome (~13 kb) and no annotation to viral origins, except for a house fly TSA sequence (*Musca domestica*, nearly identical to HVLV11) (Shi et al., 2016) (Table S1). Among them, 12 and 6 TSA accessions are mainly derived from recent meta-transcriptome of hymenopterans and dipterans (order Diptera), respectively (Misof et al., 2014; Peters et al., 2017). All replicase-like ORFs in the TSAs, except for some accessions (*Centris flavifrons*, *Megastigmus spermotrophus*, *Monosapyga clavicornis* and *Corydalus cornutus*), appeared to be intact. Therefore, those virus-like TSAs most likely represent *bona fide* viruses infecting the insect specimen(s) used for transcriptome analyses, and are likely not from insect chromosomal sequences; this is because, EVEs are usually fragmented, and their ORFs are often disrupted by stop codons and frame-shift mutations occurring during the course of evolution (Aswad and Katzourakis, 2016; Kondo et al., 2015).

To understand the phylogenetic relationships among newly found virga/nege-like virus-related TSA accessions and other reported viruses, an ML tree was constructed using PhyML 3.0 (Guindon et al., 2010), based on their replicase sequences. As shown in Fig. 1 (color online only), the phylogenetic tree indicated that the majority of TSA-derived sequences, together with recently discovered invertebrate viruses (Shi et al., 2016), fall between the virga/bromo/closterovirus (virgavirus-related linage, highlighted by green box) and recently discovered negevirus (negevirus-related linage, highlighted by red box) clades (see Table S2 for viruses), filling the major phylogenetic gaps. One TSA sequence from a thrips (*Gynaikothrips ficorum*, order Thysanoptera) was positioned in the negevirus clade (Sandewavirus/Nelorpivirus) (Kallies et al., 2014; Nunes et al.,

2017; Vasilakis et al., 2013) (Fig. 1A, in red-highlighting box), whose members have two structural protein genes encoding a predicted glycoprotein in ORF2 and a predicted membrane protein SP24 (ORF3, pfam16504), respectively (Fujita et al., 2017; Kuchibhatla et al., 2014; O'Brien et al., 2017) (Fig. S1A). The thrips TSA sequence is also closely related to HVLV7 (Shi et al., 2016) and mite-transmitted plant viruses in the genera *Cilevirus* and *Higrevirus* (Locali-Fabris et al., 2006; Melzer et al., 2013) (Fig. S1A), and another plant virus, blueberry necrotic ring blotch virus (genus *Blunervirus*) (Quito-Avila et al., 2013). Note that we also discovered a number of negevirus-like sequence fragments in the NCBI Expressed Sequence Tag (EST) database derived from *Drosophila melanogaster* Schneider L2 cell culture (Berkeley Drosophila Genome Project), and we reconstructed them as a possible genome-complete contig (named DmeEST). The DmeEST sequence is most likely to encode a replicase and two structural protein-like genes associated with those of other negeviruses (Fig. 1 and S1A). In particular, the replicase and SP24 orthologue, but not a putative glycoprotein of DmeEST, showed significant amino acid sequence identity (=46%) to that of an invertebrate virus (Wuhan house centipede virus 1). The finding of this virus candidate in the Schneider L2 EST library is reminiscent of infection of Sf9 and Sf21 cell lines from the fall armyworm (*Spodoptera frugiperda*) with a novel rhabdovirus (Sf-rhabdovirus) (Ma et al., 2014). However, in our preliminary study, the negevirus-related sequences could not be detected by RT-PCR nor genomic PCR in Schneider L2 cells available in Japan (H.K. unpublished results), and no closely related EVEs were found from the *D. melanogaster* WGS library (data not shown). Therefore, further study is necessary to confirm the nege-like virus infection in particular cell lines of *D. melanogaster*.

Three virus-like TSA sequences from the red fire ant (*Solenopsis invicta*), and two springtail species (*Sminthurus viridis* and *Anurida maritim*, order Collembola) also contained SP24 homologs, but they form distinct floating clades with or without insect viruses (Fig. 1). Eight new detected virus-like sequences, four from hymenopteran, three from dipteran, and one from lepidopteran TSA libraries, formed a large monophyletic clade (tentatively named “invertebrate virus group A”) with novel insect and nematode viruses such as HVLV15, Culex negev-like virus 1 (CNLV1), and Xingshan nematode virus 1 (Shi et al., 2016; 2017) (Fig. 1, highlighted by grey box, and Table S1).

Although the members of this clade have a diverse genome structure with 3–8 ORFs, a putative glycoprotein and a putative small transmembrane protein(s) (tentatively named SP-like protein) or multiple copies of SP-like proteins were encoded by most of these viruses in the 3′ proximal genome region (Shi et al., 2016) (Fig. S1B). The putative two structural proteins were also encoded by most viral-like TSA sequences in this clade (data not shown), further supporting their close relationship.

Of particular interest were five TSA accessions, three from hymenopteran species, a wasp (*Argochrysis armilla*, AarTSA), and two bees (*C. flavifrons*, CflTSA; *Coelioxys conoidea*, CcoTSA), and two from a fruit fly (*Ceratitis capitata*, CcaTSA1) and whitefly (*Bemisia tabaci*, order Hemiptera, BtaTSA1), formed a monophyletic clade with HVLV1 and HVLV2 (named "insect tobamo-like group") (Fig. 1, highlighted by blue box). The genome of HVLV1 and HVLV2 only encodes a replicase and a putative coat protein (tentatively named CP-like, CPL) containing a conserved tobacco mosaic virus (TMV)-coat superfamily domain (accession cl20208) (Shi et al., 2016) (Fig. 2A, color online only). Similarly, these TSA sequences, except for AarTSA and CcaTSA1, have two non-overlapping ORFs with moderate levels of amino acid sequence identities (25~48%) among their putative replicase and CPL proteins (Fig. 2A, S1C and S2). The CPLs of those four TSAs, and HVLV1 and HVLV2, are closely related to each other and together form a single clade in a neighbor-joining tree (Fig. S2C, highlighted by blue box). Interestingly, an invertebrate virus (Beihai charybdis crab virus 1), which is closely related to plant virgaviruses (Shi et al., 2016) (Fig. 1), also encodes virgavirus-related CPL, but its replicase lacks an additional region including an FtsJ-like domain between the MET and HEL domains (data not shown). Note that a mirid bug virus (Adelphocoris suturalis-associated virus 1, ASV1) (Li et al., 2017) and the house fly TSA sequence (*M. domestica* TSA), which was positioned together at a separate clade, also contain CPL conserved sequences in their 3′ proximal ORF (Fig. 2B and S2). Moreover, distantly related CPL sequences were also found in some other virus-like sequences such as the whitefly (*B. tabaci* TSA2), a hemiptern (*Okanagana villos*), and the dobsonfly (*Corydalus cornutus*, order Megaloptera) TSAs (data not shown), and their replicase sequence formed distinct small clades together with novel insect viruses (Fig. 1). Thus, these observations further support the close evolutionary relationship between plant viruses from the virga-

related lineage and certain alpha-like insect-infecting viruses although there are variations between their genome organization (Shi et al., 2016).

3.2. Virga/negevirus-like sequences in insect genomes

To gain insight into the long-term history and evolution of virga/nege-related insect viruses, we carried out a tblastn search on insect whole-genome shotgun (WGS) assemblies available in NCBI, using a replicase sequence of virga/nege-related insect viruses as queries. First, in preliminary searches using several virus sequences, such as a negevirus (Negev virus, Nelorpivirus group) and HVLV9 as queries, we observed that the search using HVLV1 (a member of the insect tobamo-like group) as a query yielded relatively a lower E-value and longer sequence coverage (data not shown). Therefore, for further comprehensive tblastn analysis, we focused on using HVLV1 (or alternatively a related TSA, AarTSA) replicase sequence as a query.

The tblastn search identified a large number of EVE candidates in the genome of divergent insect species representing five major insect orders, Hymenoptera, Diptera, Lepidoptera, Hemiptera, and Thysanoptera (Fig. 3, color online only and Table S3). Several virga/negevirus-related sequences (named “virgavirus replicase-like sequences, VRLSs”) identified in this study have been listed by a previous pioneering work (Cui and Holmes, 2012); however, detailed analysis has not been conducted for most of these sequences. As shown in Fig. 3, the VRLSs identified in the genome of insects belonging to the Hymenoptera order, especially in the *Bombus* genus (*Bombus impatiens* and *B. terrestris*, the common eastern and buff-tailed bumblebees), cover wide portions of the HVLV1 replicase sequences that correspond to the regions of MET, FtsJ, HEL and RdRp domains, while those domains found from the genome of other insect orders were distributed disproportionately, such as the majority of VRLSs in the genome of some dipteran species, including *Aedes*, *Anopheles*, and *Drosophila* spp., correspond to HEL and/or RdRp domains (Fig. 3 and Table S3). Only one insect species (*Frankliniella occidentalis*, thrips) belonging to the Thysanoptera order is currently available

in the database, and its genome contained VRLSs that correspond to the MET and RdRp domain regions (Fig. 3 and Table S3).

Interestingly, distinct VRLSs corresponding to the HEL domain region were found to exist in very high hit numbers in the genome of several lepidopteran species, such as butterflies (*Heliconius* spp., HelVRLSs, HcyVRLSs, HhiVRLSs, HtiVRLSs and many others), the diamondback moth (*Plutella xylostella*, PxyVRLSs), and the fall armyworm (*S. frugiperda*, SfrVRLSs), and also in a hemipteran (the plant hopper, *Nilaparvata lugens*, NluVRLSs), as already described by Lazareva et al. (2015) (Fig. 3, Table S3 and data not shown). We also found some TSAs were likely derived from retrotransposons carrying HEL-like domains in a lepidopteran species, *Spodoptera litura* (SliTSA) and in a cave cricket, *Ceuthophilus* sp. (CeuTSA, order Orthoptera) (Fig. S3). In the *S. litura* TSA, a retrotransposon-like sequence possessed a HEL-like domain closely related to that of SfrVRLSs (Fig. 4, color online only), and is most likely inserted into an alpha amylase gene of the *S. litura* genome (see Fig. S3). Other similar retrotransposon-related TSA sequences containing a HEL-like domain, such as a TSA from a hemipteran (*N. lugens*), were also found (Fig. S3 and data not shown).

ML trees based on HEL or RdRp domain regions showed that most discovered VRLSs in the genome of hymenopteran, lepidopteran, hemipteran, and dipteran species are related to HVLV1 or HVLV2 (insect tobamo-like group), while some VRLSs, such as those from a thysanopteran (*F. occidentalis*, FocVRLS), a lepidopteran (*Calycopis cecrops*, CceVRLS), and mosquitos (AaeVRLS2, AfaVRLS) appear to be distantly related to other insect virga/nege-like viruses (Fig. 4), suggesting divergent virus origins of the VRLSs among insect orders. All HEL-like domains found in VRLS lepidopteran, except for CceVRLS, form a monophylogenetic cluster, while those from non-hemipteran species (plant hopper), as mentioned above, are placed within the insect tobamo-like virus clade (Fig. 4). The AfaVRLS is closely related to known insect viruses (83% amino acid sequence identity to that of the HVLV21 HEL domain); thus, this VRLS might result from a relatively more recent insertion into the host genome. Intriguingly, the virga/nege-related insect viruses or virus-like

TSA sequences are not found in the majority of insect species inheriting VRLSs, thus may have had ancient infection(s) to acquire the source of fossil viral elements.

3.3. Integration profile of virgavirus-like sequences in bumblebees

The extensive endogenization of VRLSs in the bumblebee species (Fig. 3) prompted us to further characterize their integration pattern in the genome of some species of the *Bombus* genus. The WGS assemblies of two species, *B. impatiens* and *B. terrestris*, which are both key commercial and natural pollinators, are publicly available in the NBCI database (Sadd et al., 2015). *B. impatiens* and *B. terrestris* belong to two distantly related sub-genera, *Pyrobombus* and *Bombus*, within the genus *Bombus*, respectively (Cameron et al., 2007). First, we wanted to verify the *Bombus* VRLS profiles revealed via database search, using PCR amplification and sequencing of *B. terrestris* genomic DNA. PCR analysis on the *B. terrestris* genome corroborated the VRLS profile according to the database, as all listed VRLS sequences (BteVRLSs) were amplified (Fig. 5, color online only, S4 and S5). Furthermore, the absence of a VRLS corresponding to the MTR or RdRp domain region that exists in the *B. impatiens* database (BimVRLS9 or BimVRLS11) was confirmed by PCR using *B. terrestris* DNA (data not shown). Thus by comparing the data set of two databases, it appears that *B. impatiens* and *B. terrestis* have contrasting pattern of VRLS integration (Fig. 5). The *B. terrestris* genome contains VRLSs corresponding to the HEL and RdRp domains (BteVRLS2) and the RdRp domain (BteVRLS4) that are not found in *B. impatiens*, but likely lacks VRLSs corresponding to the MTR domain (BimVRLS9) and the RdRp domain (BimVRLS11) that exist in *B. impatiens* (Fig. 5 and Fig. S4). This suggests that the integration events might have occured after the divergence of *Pyrobombus* and *Bombus* sub-genera, approximately 18 million year ago (Cameron et al., 2007; Hines, 2008).

We further extended VRLS analysis to two other native bumblebee species in Japan, *B. hypocrita* and *B. ignitus*, which are species closely related to *B. terrestris* (Cameron et al., 2007; Hines, 2008). Genomic PCR and sequencing revealed that, in general, the integration patterns of VRLSs in *B. hypocrite* and *B. ignitus* appear to resemble those in *B. terrestis*, although two VRLSs

(VRLS1 and VRLS7) were not amplified from *B. hypocrita* or *B. ignitus* genomic DNA samples (Fig. 5, S4 and S5). Sequence deletions cause minor differences in VRLSs between *B. hypocrita* and *B. ignitus* (see VRLS2 and VRLS3, Fig. 5 and S6). VRLSs integrated into the same genome position were highly similar among species (95–99% nucleotide sequences), whereas amino acid sequence identities of VRLSs among those in different genome positions (between VRLS2 and VRLS1 or VRLS3) were only moderately related (56–58% or 64–71%) (Fig. S7), although they are from the same bee species. This highlights the possibility that the VRLSs integrated into the ancestral lineage of the sub-genera *Bombus*, at least for *B. terrestis*, *B. hypocrita*, and *B. ignitus*, whose divergence is roughly estimated at around 8 million year ago (Hines, 2008), are derived from two related viruses (see Fig. 4). Notably, some variants of BteVRLS2, BigVRLS1, and BigVRLS2 were found in *B. terrestis* and *B. ignitus*, with minor nucleotide sequence variations including a few nucleotide insertions/deletions (Fig. S6). This suggests that these VRLSs might be duplicated in the *Bombus* genome after single/multiple integration events. Interestingly, virgavirus-related sequences corresponding to the CP-like gene (VCLS) were discovered in the genome of *B. terrestis* and *B. impatiens* (BteVCLS3 and BimVCLS11), but they appear to be integrated in different genome positions along with respective upstream VRLS regions (Fig. 5). The VCLS3 sequences were also identified in the *B. hypocrita* and *B. ignitus* genome (BhyVCLS3 and BigVCLS3) with high amino acid sequence identity (84–89%) (Fig. 5, S6 and S7). However, using *Bombus* VCLSs as queries for the tblastn searches resulted in no significant hits against any other WGS libraries of hymenopterans or other insects. It is noteworthy that searches for CP-like sequences using ASaV or HVLV9 CPL as queries also yielded some hits in the insect WGS libraries (data not shown).

4. Discussion

4.1. The evolutionary perspective of virga/nege-like viruses in insects

By tblastn searches, we discovered at least 27 virga/negevirus-like sequences from insect TSAs, mainly derived from hymenopteran and dipteran species, all of which appear to consist of a

complete genome sequence (>9.0~13 kb lengths) (Fig. 2, S1 and Table S1). Most of the TSA-derived sequences, together with recently discovered invertebrate viruses (Shi et al., 2016), filled major phylogenetic gaps between virgavirus- and negevirus-related lineages in an ML tree based on replicase sequences (Fig. 1). Moreover, reverse tblastn analysis using virga/nege-like viruses in the TSA library of various insect species yielded several significant match with ones less than 9.0 kb in length, which may represent partial virus genome sequences (data not shown). However, these were not further characterized in this study. Therefore, the occurrence of insect virga/nege-like viruses most likely extends to broader range of insects and other invertebrate species.

Our analysis shows a close phylogenetic relationship between insect-infecting virga/nege-like viruses and their plant-infecting counterparts. Moreover, some insect virga/nege-like viruses or TSA accessions contain a tobamovirus-like genome structure with a CP-like gene (Fig. 2 and S1C). Although the associations between insect tobamo-like viruses and plants are not known, these observations raise the possibility of cross-kingdom virus transfer between insects and plants in ancient times. All known insect-transmitted plant (+)ssRNA viruses are non-replicative in their insect vectors (majority aphids, stylet-born or circulating manner) (Whitfield et al., 2015) (Fig. 1), suggesting the existences of host-specific barriers for plant (+)ssRNA viruses to cross infection to insects. A number of plant (–)ssRNA viruses (tospoviruses and rhabdoviruses) that infect their insect vectors (thrips, aphids, plant- and leafhoppers), form an enveloped virion with a prominent viral glycoprotein spike(s), which is believed to play an important role during viral entrance into insect host cells (Dietzgen et al., 2016; Whitfield et al., 2015). Interestingly, many insect virga/nege-like viruses and related-TSA accessions, but not most plant (+)ssRNA viruses, appear to encode a putative glycoprotein gene(s) (Kuchibhatla et al., 2014; Shi et al., 2016) (Fig. 1, 2 and S1). An insect RNA virus (flock house virus, family *Nodaviridae*, picorna-like superfamily) was shown to replicate in the plant cells (Dasgupta et al., 2001). However, the virus requires an expression of plant virus movement protein for cell-to-cell and subsequent systemic transport in the plant (Dasgupta et al., 2001). Thus, it is suggested that acquisition of a gene(s) responsible for viral movement might be essential for certain insect viruses to fully adapt to the plant hosts. It is possible that this event(s) occurred during the course of evolution of

cile/higrevirus and blunevirus lineages (Lazareva et al., 2017) (Fig. S1). Furthermore, the genome segmentation(s) of their ancestral genomes may also have occurred (Fig. 1 and S1), and this is possibly linked to *Brevipalpus* mite vectors, as previously proposed (Kondo et al., 2017).

4.2. Endogenization of virga/nege-like viruses into the insect genomes

In this study, we further characterized EVE candidates derived from virga/nege-like viruses in insect genomes, several of which were not described previously by Cui and Holmes (2012) (Fig. 3). In the HEL- or RdRp-based tree (Fig. 4), most EVE candidates (VRLSs) fell within a clade of the insect tobamo-like group, and they were nested with virus-like TSAs from the same insect order, indicating their close relationship. However, the presence of some VRLSs that fell outside of the clade of the insect tobamo-like group suggested the presence of unseen insect virga/nege-like viruses or virus groups in related insect species (Fig. 4). Thus, our data strongly support the view that insect EVE candidates related to virga/nege-like viruses might be the footprints of ancient insect RNA viruses, but not plant RNA viruses, as previously suggested by Cui and Holmes (2012).

Many of the insect EVE candidates were located adjacent to or flanked by transposable elements or repeat sequences (see Fig. 5 for bumblebee EVEs), suggesting that integrations might have occured due to non-homologous recombination between an RNA of the progenitor virus and that of a retrotransposon during reverse transcription of the retrotransposon (Ballinger et al., 2012; Geuking et al., 2009; Horie et al., 2010). Interestingly, the transcription of several transposable elements (i.e., putative RNA-directed DNA polymerases) was induced by a plant rhabdovirus or plant reovirus infection in their vector insects (plant/leafhoppers) (Martin et al., 2017). The current and previous studies (Lazareva et al., 2015) (Fig. S3), indicate that the virgavirus-like HEL domain sequence is located within long interspersed elements, a class of non-LTR retrotransposons, in the genome of lepidopteran species such as *H. melpomene* and *P. xylostella*. It has been proposed that the high copy number of EVEs in the genome of some lepidopterans is associated with the acquisition of viral HEL domains by retrotransposons (Lazareva et al., 2015). Likewise, a similar manner of

acquisition of virus-related sequences, such as those from nodaviruses and two (–)ssRNA viruses (phleboviruses and rhabdoviruses), might have occurred in LTR-retrotransposons of insects (*Drosophila*) and nematodes (*Caenorhabditis elegans* and *Bursaphelenchus xylophilus*) (Ballinger et al., 2012; Cotton et al., 2016; Malik et al., 2000). The acquisition of viral HEL domains by retrotransposons may be associated with the RNA silencing suppression activity of the HEL-like domains as demonstrated by Lazareva et al. (2015), and is probably linked to the counteraction of the RNA silencing-based defense against retrotransposons in insect cells (Morozov et al., 2017).

Insect specific viruses, including negeviruses, are most likely to be maintained in certain insects *via* vertical (and/or venereal) transmission (Vasilakis and Tesh, 2015), suggesting that the viruses could invade host germ cells where the conversion of viral RNA to viral-derived complementary DNA (cDNA) fragments occurs prior to their endogenizations into the host genome (Olson and Bonizzoni, 2017). On the contrary, vertical transmission efficiency of plant (+)ssRNA viruses, including virga-related viruses, is generally low and the majority of them are horizontally transmitted by vectors in a non-replicative manner (Andika et al., 2016; Whitfield et al., 2015), although several of them are efficiently transmitted vertically through seeds or pollen (Hull, 2002). There is a biological barrier(s) *via* the RNA silencing mechanism that protects plant germ cells (the shoot apical meristem) against viral invasion (Foster et al., 2002; Martín-Hernández and Baulcombe, 2008). Therefore, the inability of viruses to persistently invade the plant meristem tissue may account for the limited endogenization of alpha-like or other (+)ssRNA viruses into plant genomes. In line with this view, there are a large number of plant EVEs related to the persistent dsRNA virus group (plant partitiviruses, expanded picorna-like superfamily) (Chiba et al., 2011; Liu et al., 2010), which likely invade plant germ cells and consequently are transmitted vertically through seeds or pollen.

The honeybee, *Apis mellifera* is known as a host to several viruses belonging to the order *Picornavirales* (Tehel et al., 2016). Among them, deformed wing virus (family *Iflaviridae*) is a major threat to the honeybee, possibly through contributing to colony losses and declines (Wilfert et al., 2016). Recent studies suggest a role of the antiviral RNAi pathway in pollinator bees (Maori et al.,

2009; Niu et al., 2016) and other agriculturally important insects, such as plant virus insect vectors (Lan et al., 2016; Li et al., 2013; Xu et al., 2012). Honeybees carrying an EVE related to Israeli acute paralysis virus (family *Dicistoviridae, Picornavirales*), which is present only in 30% of tested bee populations, are resistant to further viral challenge (Maori et al., 2007). Although the biological significance of EVEs from virga/nege-like viruses is still unknown, recent research progress in model insects, *D. melanogaster* and *Aedes* mosquitoes, suggests the importance of EVEs or viral-derived cDNA fragments as one of the source elements for the production of antiviral small interfering RNAs (Goic et al., 2013; 2016; Nag et al., 2016; Tassetto et al., 2017). Interestingly, *Aedes* EVEs derived from non-retrovirus RNA viruses are preferentially integrated in piRNA (a class of small RNAs known to silence transposons) clusters and produce piRNA-like small RNAs in a mostly antisense orientation (Olson and Bonizzoni, 2017; Palatini et al., 2017; Suzuki et al., 2017; Whitfield et al., 2017). This suggests a functional link between EVEs and antiviral RNAi activity, as previously described for EVEs in mammalian lineages (Miesen et al., 2016; Parrish et al., 2015). Therefore, small RNAs derived from EVEs and/or viral cDNA fragments, as well as their potential contribution to antiviral RNAi immunity in agriculturally important insects, are worthy of future investigation.

Acknowledgements

We would like to thank S. Hisano, M. Fujita, and H. Nishimura for their helpful technical assistance, Drs. F. Tanaka, Y. Hashimoto, M. Tobikawa, N. Abe, and T. Ayabe for providing or collecting bumblebee specimens, Drs. M. Hirai, S. Sonoda, and T. Tamada for their kind advices, and the two reviewers for their valuable suggestions and comments. This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (KAKENHI Grant numbers 15K07312, 16H06436, 16H06429 and 16K21723), as well as by Yomogi Inc. and the Ohara Foundation for Agriculture Research.

References

Adams, M.J., Antoniw, J.F., Kreuze, J., 2009. *Virgaviridae*: a new family of rod-shaped plant viruses. Arch. Virol. 154, 1967–1972.

Aiewsakun, P., Katzourakis, A., 2015. Endogenous viruses: Connecting recent and ancient viral evolution. Virology 479, 26–37.

Andika, I.B., Kondo, H., Sun, L.Y., 2016. Interplays between soil-borne plant viruses and RNA silencing-mediated antiviral defense in roots. Front. Microbiol. 7, 1458.

Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst. Biol. 55, 539–552.

Aswad, A., Katzourakis, A., 2016. Paleovirology: The study of endogenous viral elements. *Virus Evolution: Current Research and Future Directions*, 273–292.

Ballinger, M.J., Bruenn, J.A., Taylor, D.J., 2012. Phylogeny, integration and expression of sigma virus-like genes in *Drosophila*. Mol. Phylogenet. Evol. 65, 251–258.

Bolling, B.G., Weaver, S.C., Tesh, R.B., Vasilakis, N., 2015. Insect-specific virus discovery: significance for the arbovirus community. Viruses 7, 4911–4928.

Cameron, S.A., Hines, H.M., Williams, P.H., 2007. A comprehensive phylogeny of the bumble bees (*Bombus*). Biol. J. Linnean Soc. 91, 161–188.

Carapeta, S., do Bem, B., McGuinness, J., Esteves, A., Abecasis, A., Lopes, Â., de Matos, A.P., Piedade, J., António P.G. de Almeida, A.P.G., Parreira, R., 2015. Negeviruses found in multiple species of mosquitoes from southern Portugal: isolation, genetic diversity, and replication in insect cell culture. Virology 483, 318–328.

Chiba, S., Kondo, H., Tani, A., Saisho, D., Sakamoto, W., Kanematsu, S., Suzuki, N., 2011. Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. PLoS Pathog. 7, e1002146.

Chu, H., Jo, Y., Cho, W.K., 2014. Evolution of endogenous non-retroviral genes integrated into plant genomes. Curr. Plant Biol. 1, 55–59.

Cotton, J.A., Steinbiss, S., Yokoi, T., Tsai, I.J., Kikuchi, T., 2016. An expressed, endogenous Nodavirus-like element captured by a retrotransposon in the genome of the plant parasitic nematode *Bursaphelenchus xylophilus*. Sci. Rep. 6, 39749.

Cui, J., Holmes, E.C., 2012. Endogenous RNA viruses of plants in insect genomes. Virology 427, 77-79.

da Fonseca, G.C., de Oliveira, L.F.V., de Morais, G.L., Abdelnor, R.V., Nepomuceno, A.L., Waterhouse, P.M., Farinelli, L., Margis, R., 2016. Unusual RNA plant virus integration in the soybean genome leads to the production of small RNAs. Plant Sci. 246, 62–69.

Dasgupta, R., Garcia, B.H., Goodman, R.M., 2001. Systemic spread of an RNA insect virus in plants expressing plant viral movement protein genes. Proc. Natl. Acad. Sci. U. S. A. 98, 4910–4915.

Dietzgen, R.G., Mann, K.S., Johnson, K.N., 2016. Plant virus-insect vector interactions: current and potential future research directions. Viruses 8, 303.

Dolja, V.V., Koonin, E.V., 2011. Common origins and host-dependent diversity of plant and animal viromes Valerian. Curr. Opin. Virol. 1, 322–331.

Egloff, M. P., Benarroch, D., Selisko, B., Romette, J. L., Canard, B., 2002. An RNA cap (nucleoside-2'-*O*-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. EMBO J. 21, 2757–2768.

Feschotte, C., Gilbert, C., 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat. Rev. Genet. 13, 283–296.

Foster, T.M., Lough, T.J., Emerson, S.J., Lee, R.H., Bowman, J.L., Forster, R.L.S., Lucas, W.J., 2002. A surveillance system regulates selective entry of RNA into the shoot apex. Plant Cell 14, 1497–1508.

Fujita, R., Kuwata, R., Kobayashi, D., Bertuso, A.G., Isawa, H., Sawabe, K., 2017. Bustos virus, a new member of the negevirus group isolated from a Mansonia mosquito in the Philippines. Arch. Virol. 162, 79–88.

Geuking, M.B., Weber, J., Dewannieux, M., Gorelik, E., Heidmann, T., Hengartner, H., Zinkernagel, R.M., Hangartner, L., 2009. Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. Science 323, 393–396.

Goic, B., Stapleford, K.A., Frangeul, L., Doucet, A.J., Gausson, V., Blanc, H., Schemmel-Jofre, N., Cristofari, G., Lambrechts, L., Vignuzzi, M., Saleh, M.C., 2016. Virus-derived DNA drives mosquito vector tolerance to arboviral infection. Nature Commun. 7, 12410.

Goic, B., Vodovar, N., Mondotte, J.A., Monot, C., Frangeul, L., Blanc, H., Gausson, V., Vera-Otarola, J., Cristofari, G., Saleh, M.C., 2013. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model *Drosophila*. Nature Immunol. 14, 396–403.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Hines, H.M., 2008. Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera : Apidae : *Bombus*). Syst. Biol. 57, 58–75.

Holmes, E.C., 2011. The evolution of endogenous viral elements. Cell Host & Microbe 10(4), 368-377.

Horie, M., Tomonaga, K., 2011. Non-retroviral fossils in vertebrate genomes. Viruses-Basel 3(10), 1836-1848.

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta1, K., Jern, P., Gojobori, T., Coffin, J. M. Tomonaga, K., 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature 463, 84.

Hull, R., 2002. Matthews Plant Virology. 4th ed. ed. Academic Press, San Diego Ca, U.S.A.

Jo, Y., Choi, H., Kim, S.M., Kim, S.L., Lee, B.C., Cho, W.K., 2016. Integrated analyses using RNA-Seq data reveal viral genomes, single nucleotide variations, the phylogenetic relationship, and recombination for *Apple stem grooving virus*. BMC Genomics 17, 579.

Kallies, R., Kopp, A., Zirkel, F., Estrada, A., Gillespie, T.R., Drosten, C., Junglen, S., 2014. Genetic characterization of goutanap virus, a novel virus related to negeviruses, cileviruses and higreviruses. Viruses 6, 4346–4357.

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9, 286–298.

Katzourakis, A., Gifford, R.J., 2010. Endogenous viral elements in animal genomes. Plos Genet. 6, e1001191.

Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7, 474.

Kondo, H., Chiba, S., Suzuki, N., 2015. Detection and analysis of non-retroviral RNA virus-like elements in plant, fungal, and insect genomes. In: Uyeda, I., Masuta, C. (Eds.), Plant Virology Protocols - New approaches to detect viruses and host responses, Third Edition ed. Springer Protocol, Humana Press Inc, New York, pp. 73–88.

Kondo, H., Chiba, S., Toyoda, K., Suzuki, N., 2013a. Evidence for negative-strand RNA virus infection in fungi. Virology 435, 201–209.

Kondo, H., Hirano, S., Chiba, S., Andika, I.B., Hirai, M., Maeda, T., Tamada, T., 2013b. Characterization of burdock mottle virus, a novel member of the genus *Benyvirus*, and the identification of benyvirus-related sequences in the plant and insect genomes. Virus Res. 177, 75–86.

Kondo, H., Hirota, K., Maruyama, K., Andika, I.B., Suzuki, N., 2017. A possible occurrence of genome reassortment among bipartite rhabdoviruses. Virology 508, 18–25.

Kondo, H., Hisano, S., Chiba, S., Maruyama, K., Andika, I.B., Toyoda, K., Fujimori, F., Suzuki, N., 2016. Sequence and phylogenetic analyses of novel totivirus-like double-stranded RNAs from field-collected powdery mildew fungi. Virus Res. 213, 353–364.

Koonin, E.V., Dolja, V.V., 1993. Evolution and taxonomy of positive-strand RNA viruses - Implications of comparative-analysis of amino-acid-sequences. Crit. Rev. Biochem. Mol. Biol. 28, 375–430.

Koonin, E.V., Dolja, V.V., Krupovic, M., 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology 479, 2–25.

Kuchibhatla, D.B., Sherman, W.A., Chung, B.Y.W., Cook, S., Schneider, G., Eisenhaber, B., Karlin, D.G., 2014. Powerful sequence similarity search methods and in-depth manual analyses can identify remote homologs in many apparently "orphan" viral proteins. J. Virol. 88, 10–20.

Lan, H.H., Wang, H.T., Chen, Q., Chen, H.Y., Jia, D.S., Mao, Q.Z., Wei, T.Y., 2016. Small interfering RNA pathway modulates persistent infection of a plant virus in its insect vector. Sci. Rep. 6, 20699

Lazareva, E., Lezzhov, A., Vassetzky, N., Solovyev, A., Morozov, S., 2015. Acquisition of full-length viral helicase domains by insect retrotransposon-encoded polypeptides. Front. Microbiol. 6, 1447.

Lazareva, E.A., Lezzhov, A.A., Komarova, T.V., Morozov, S.Y., Heinlein, M., Solovyev, A.G., 2017. A novel block of plant virus movement genes. Mol. Plant Pathol. 18, 611–624.

Li, C.X., Shi, M., Tian, J.H., Lin, X.D., Kang, Y.J., Chen, L.J., Qin, X.C., Xu, J.G., Holmes, E.C., Zhang, Y.Z., 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. Elife 4, e05378.

Li, J.M., Andika, I.B., Shen, J.F., Lv, Y.D., Ji, Y.Q., Sun, L.Y., Chen, J.P., 2013. Characterization of rice black-streaked dwarf virus- and rice stripe virus-derived siRNAs in singly and doubly infected insect vector *Laodelphax striatellus.* Plos One 8, e66007.

Li, X.L., Xu, P.J., Yang, X.M., Yuan, H., Chen, L.Z., Lu, Y.H., 2017. The genome sequence of a novel RNA virus in *Adelphocoris suturalis*. Arch. Virol. 162, 1397–1401.

Liu, H.Q., Fu, Y.P., Jiang, D.H., Li, G.Q., Xie, J., Peng, Y.L., Yi, X.H., Ghabrial, S.A., 2009. A novel mycovirus that is related to the human pathogen hepatitis E virus and rubi-like viruses. J. Virol. 83, 1981–1991.

Liu, H.Q., Fu, Y.P., Jiang, D.H., Li, G.Q., Xie, J.T., Cheng, J.S., Peng, Y.L., Ghabrial, S.A., Yi, X.H., 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. J. Virol. 84, 11876–11887.

Locali-Fabris, E.C., Freitas-Astua, J., Souza, A.A., Takita, M.A., Astua-Monge, G., Antonioli-Luizon, R., Rodrigues, V., Targon, M., Machado, M.A., 2006. Complete nucleotide sequence, genomic organization and phylogenetic analysis of *Citrus leprosis virus* cytoplasmic type. J. Gen. Virol. 87, 2721–2729.

Ma, H.L., Galvin, T.A., Glasner, D.R., Shaheduzzaman, S., Khan, A.S., 2014. Identification of a novel rhabdovirus in *Spodoptera frugiperda* cell lines. J. Virol. 88, 6576–6585.

Malik, H.S., Henikoff, S., Eickbush, T.H., 2000. Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. 10, 1307–1318.

Maori, E., Paldi, N., Shafir, S., Kalev, H., Tsur, E., Glick, E., Sela, I., 2009. IAPV, a bee-affecting virus associated with Colony Collapse Disorder can be silenced by dsRNA ingestion. Insect Mol. Biol. 18, 55–60.

Maori, E., Tanne, E., Sela, I., 2007. Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. Virology 362, 342–349.

Martin, K.M., Barandoc-Alviar, K., Schneweis, D.J., Stewart, C.L., Rotenberg, D., Whitfield, A.E., 2017. Transcriptomic response of the insect vector, *Peregrinus maidis*, to Maize mosaic rhabdovirus and identification of conserved responses to propagative viruses in hopper vectors. Virology 509, 71–81.

Martín-Hernández, A., Baulcombe, D., 2008. *Tobacco rattle virus* 16-kilodalton protein encodes a suppressor of RNA silencing that allows transient viral entry in meristems. J Virol 82, 4064–4071.

Melzer, M.J., Simbajon, N., Carillo, J., Borth, W.B., Freitas-Astua, J., Kitajima, E.W., Neupane, K.R., Hu, J.S., 2013. A cilevirus infects ornamental hibiscus in Hawaii. Arch. Virol. 158, 2421–2424.

Miesen, P., Joosten, J., van Rij, R.P., 2016. PIWIs Go Viral: Arbovirus-derived piRNAs in vector mosquitoes. Plos Pathogens 12, e1006017.

Misof, B., Liu, S.L., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspock, U., Aspock, H., Bartel, D., Blanke, A., Berger, S., Bohm, A., Buckley, T.R., Calcott, B., Chen, J.Q., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S.C., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y.Y., Li, Z.Y., Li, J.G., Lu, H.R., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G.L., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y.X., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schutte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W.H., Su, X., Szucsich, N.U., Tan, M.H., Tan, X.M., Tang, M., Tang, J.B., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X.L., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G.X., Xie, Y.L., Yang, S.Z., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W.W., Zhang, Y.H., Zhao, J., Zhou, C.R., Zhou, L.L., Ziesmann, T., Zou, S.J., Li, Y.R., Xu, X., Zhang, Y., Yang, H.M., Wang, J., Kjer, K.M., Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767.

Morozov, S.Y., Lazareva, E.A., Solovyev, A.G., 2017. RNA helicase domains of viral origin in proteins of insect retrotransposons: possible source for evolutionary advantages. PeerJ 5, e3673.

Mushegian, A., Shipunov, A., Elena, S.F., 2016. Changes in the composition of the RNA virome mark evolutionary transitions in green plants. BMC Biol. 14, 68.

Nabeshima, T., Inoue, S., Okamoto, K., Posadas-Herrera, G., Yu, F.X., Uchida, L., Ichinose, A., Sakaguchi, M., Sunahara, T., Buerano, C.C., Tadena, F.P., Orbita, I.B., Natividad, F.F., Morita, K., 2014. Tanay virus, a new species of virus isolated from mosquitoes in the Philippines. J. Gen. Virol. 95, 1390–1395.

Nag, D.K., Brecher, M., Kramer, L.D., 2016. DNA forms of arboviral RNA genomes are generated following infection in mosquito cell cultures. Virology 498, 164–171.

Nibert, M.L., Pyle, J.D., Firth, A.E., 2016. A +1 ribosomal frameshifting motif prevalent among plant amalgaviruses. Virology 498, 201–208.

Niu, J., Smagghe, G., De Coninck, D. I., Van Nieuwerburgh, F., Deforce, D., Meeus, I., 2016. *In vivo* study of Dicer-2-mediated immune response of the small interfering RNA pathway upon systemic infections of virulent and avirulent viruses in *Bombus terrestris*. Insect Biochem. Mol. Biol. 70, 127–137.

Nunes, M.R.T., Contreras-Gutierrez, M.A., Guzman, H., Martins, L.C., Barbirato, M.F., Savit, C., Balta, V., Uribe, S., Vivero, R., Suaza, J.D., Oliveira, H., Neto, J.P.N., Carvalho, V.L., da Silva, S.P., Cardoso, J.F., de Oliveira, R.S., Lemos, P.D., Wood, T.G., Widen, S.G., Vasconcelos, P.F.C., Fish, D., Vasilakis, N., Tesh, R.B., 2017. Genetic characterization, molecular epidemiology, and phylogenetic relationships of insect-specific viruses in the taxon Negevirus. Virology 504, 152–167.

O'Brien, C.A., McLean, B.J., Colmant, A.M.G., Harrison, J.J., Hall-Mendelin, S., van den Hurk, A.F., Johansen, C.A., Watterson, D., Bielefeldt-Ohmann, H., Newton, N.D., Schulz, B.L., Hall, R.A., Hobson-Peters, J., 2017. Discovery and characterisation of castlerea virus, a new species of negevirus isolated in Australia. Evol. Bioinform. 13, 1176934317691269.

Olson, K.E., Bonizzoni, M., 2017. Nonretroviral integrated RNA viruses in arthropod vectors: an occasional event or something more? Curr. Opin. Insect Sci. 22, 45–53.

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

Parrish, N.F., Fujino, K., Shiromoto, Y., Iwasaki, Y.W., Ha, H., Xing, J.C., Makino, A., Kuramochi-Miyagawa, S., Nakano, T., Siomi, H., Honda, T., Tomonaga, K., 2015. piRNAs derived from ancient viral processed pseudogenes as transgenerational sequence-specific immune memory in mammals. RNA 21, 1691–1703.

Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., Diez, P.A., Heraty, J., Kjer, K.M., Klopfstein, S., Meier, R., Polidori, C., Schmitt, T., Liu, S.L., Zhou, X., Wappler, T., Rust, J., Misof, B., Niehuis, O., 2017. Evolutionary history of the Hymenoptera. Curr. Biol. 27, 1013–1018.

Quito-Avila, D.F., Brannen, P.M., Cline, W.O., Harmon, P.F., Martin, R.R., 2013. Genetic characterization of *Blueberry necrotic ring blotch virus*, a novel RNA virus with unique genetic features. J. Gen. Virol. 94, 1426–1434.

Roy, A., Choudhary, N., Guillermo, L.M., Shao, J., Govindarajulu, A., Achor, D., Wei, G., Picton, D.D., Levy, L., Nakhla, M.K., Hartung, J.S., Brlansky, R.H., 2013. A novel virus of the genus *Cilevirus* causing symptoms similar to citrus leprosis. Phytopathology 103, 488–500.

Roy, A., Hartung, J.S., Schneider, W.L., Shao, J., Leon, G., Melzer, M.J., Beard, J.J., Otero-Colina, G., Bauchan, G.R., Ochoa, R., Brlansky, R.H., 2015. Role Bending: Complex relationships between viruses, hosts, and vectors related to citrus leprosis, an emerging disease. Phytopathology 105, 1013–1025.

Sadd, B.M., Barribeau, S.M., Bloch, G., de Graaf, D.C., Dearden, P., Elsik, C.G., Gadau, J., Grimmelikhuijzen, C.J.P., Hasselmann, M., Lozier, J.D., Robertson, H.M., Smagghe, G., Stolle, E., Van Vaerenbergh, M., Waterhouse, R.M., Bornberg-Bauer, E., Klasberg, S., Bennett, A.K., Caamara, F., Guigo, R., Hoff, K., Mariotti, M., Munoz-Torres, M., Murphy, T., Santesmasses, D., Amdam, G.V., Beckers, M., Beye, M., Biewer, M., Bitondi, M.M.G., Blaxter, M.L., Bourke, A.F.G., Brown, M.J.F., Buechel, S.D., Cameron, R., Cappelle, K., Carolan, J.C., Christiaens, O., Ciborowski, K.L., Clarke, D.F., Colgan, T.J., Collins, D.H., Cridge, A.G., Dalmay, T., Dreier, S., du Plessis, L., Duncan, E., Erler, S., Evans, J., Falcon, T., Flores, K., Freitas, F.C.P., Fuchikawa, T., Gempe, T., Hartfelder, K., Hauser, F., Helbing, S., Humann, F.C., Irvine, F., Jermiin, L.S., Johnson, C.E., Johnson, R.M., Jones, A.K., Kadowaki, T., Kidner, J.H., Koch, V., Kohler, A., Kraus, F.B., Lattorff, H.M.G., Leask, M., Lockett, G.A., Mallon, E.B., Antonio, D.S.M., Marxer,

M., Meeus, I., Moritz, R.F.A., Nair, A., Napflin, K., Nissen, I., Niu, J., Nunes, F.M.F., Oakeshott, J.G., Osborne, A., Otte, M., Pinheiro, D.G., Rossie, N., Rueppell, O., Santos, C.G., Schmid-Hempel, R., Schmitt, B.D., Schulte, C., Simoes, Z.L.P., Soares, M.P.M., Swevers, L., Winnebeck, E.C., Wolschin, F., Yu, N., Zdobnov, E.M., Aqrawi, P.K., Blankenburg, K.P., et al., 2015. The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. 16, 76.

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., Buchmann, J., Wang, W., Xu, J., Holmes, E.C., Zhang, Y.Z., 2016. Redefining the invertebrate RNA virosphere. Nature 7634, 539–543.

Shi, M., Neville, P., Nicholson, J., Eden, J.S., Imrie, A., Holmes, E.C., 2017. High resolution meta-transcriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. J. Virol. 91, e00680-17.

Suzuki, Y., Frangeul, L., Dickson, L.B., Blanc, H., Verdier, Y., Vinh, J., Lambrechts, L., Saleh, M.C., 2017. Uncovering the repertoire of endogenous flaviviral elements in *Aedes* mosquito genomes. J. Virol. 91, e00571-17.

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577.

Tassetto, M., Kunitomi, M., Andino, R., 2017. Circulating immune cells mediate a systemic RNAi-based adaptive antiviral response in *Drosophila*. Cell 169, 314–325.

Tassi, A.D., Garita-Salazar, L.C., Amorim, L., Novelli, V.M., Freitas-Astua, J., Childers, C.C., Kitajima, E.W., 2017. Virus-vector relationship in the Citrus leprosis pathosystem. Exp. Appl. Acarol. 71, 227-241.

Taylor, D.J., Bruenn, J., 2009. The evolution of novel fungal genes from non-retroviral RNA viruses. BMC Biol. 7, 88.

Tehel, A., Brown, M.J.F., Paxton, R.J., 2016. Impact of managed honey bee viruses on wild bees. Curr. Opin. Virol. 19, 16–22.

Vasilakis, N., Forrester, N.L., Palacios, G., Nasar, F., Savji, N., Rossi, S.L., Guzman, H., Wood, T.G., Popov, V., Gorchakov, R., Gonzalez, A.V., Haddow, A.D., Watts, D.M., da Rosa, A., Weaver, S.C., Lipkin, W.I., Tesh, R.B., 2013. Negevirus: a proposed new taxon of insect-specific viruses with wide geographic distribution. J. Virol. 87, 2475–2488.

Vasilakis, N., Tesh, R.B., 2015. Insect-specific viruses and their potential impact on arbovirus transmission. Curr. Opin. Virol. 15, 69–74.

Whitfield, A.E., Falk, B.W., Rotenberg, D., 2015. Insect vector-mediated transmission of plant viruses. Virology 479, 278–289.

Whitfield, Z.J., Dolan, P.T., Kunitomi, M., Tassetto, M., Seetin, M.G., S, O., Jeiner, C., P., 2017. Long-read assembly of the *Aedes aegypti* genome reveals the nature of heritable adaptive immunity sequences. bioRxiv, 127498.

Wilfert, L., Long, G., Leggett, H.C., Schmid-Hempel, P., Butlin, R., Martin, S.J.M., Boots, M., 2016. Deformed wing virus is a recent global epidemic in honeybees driven by *Varroa* mites. Science 351, 594-597.

Xu, Y., Huang, L.Z., Fu, S., Wu, J.X., Zhou, X.P., 2012. Population diversity of rice stripe virus-derived sirnas in three different hosts and RNAi-based antiviral immunity in *Laodelphgax striatellus*. Plos One 7, e46238.

Figure legends

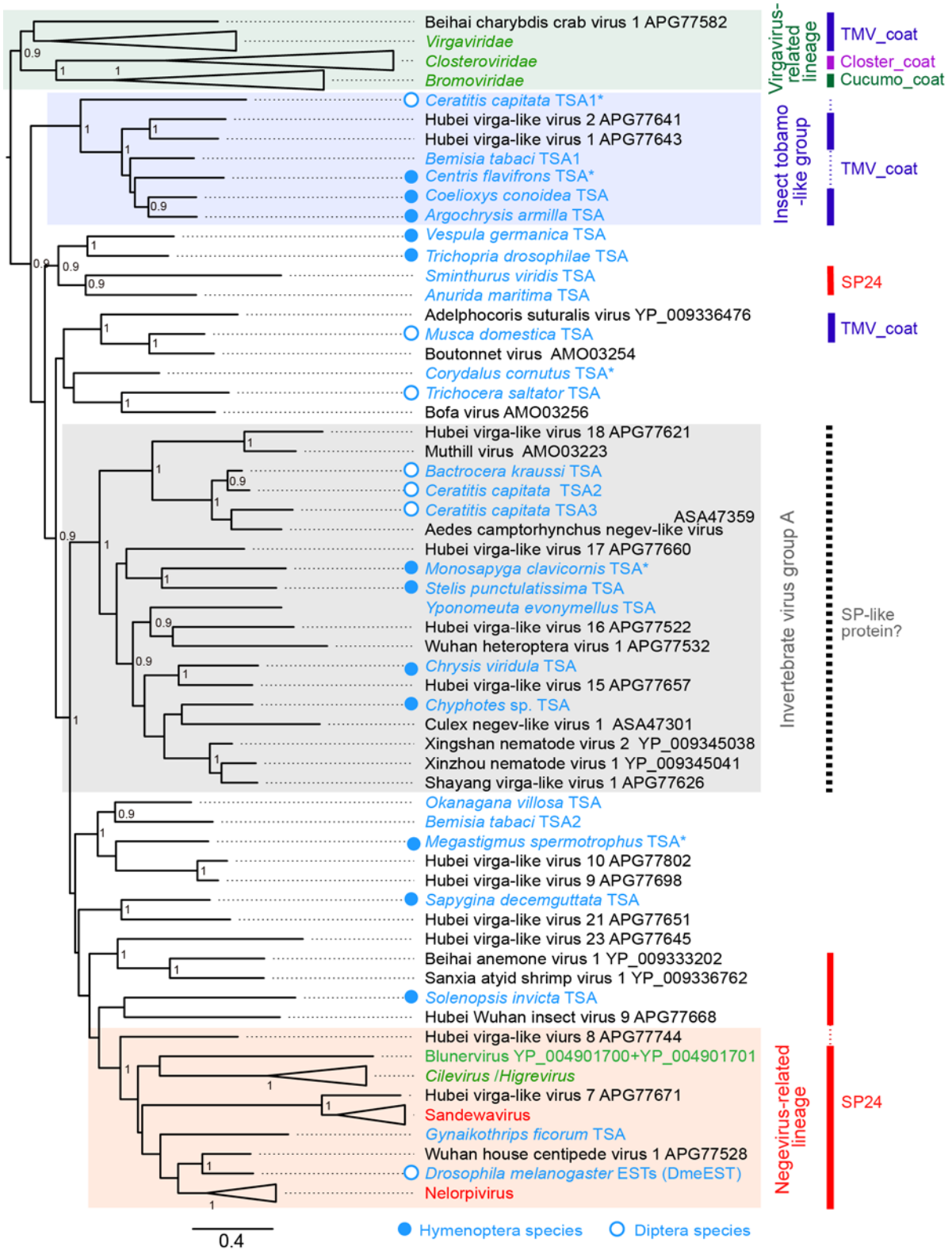

Fig. 1. Molecular phylogenetic analysis of the replicase sequences of insect virga/nege-like viruses and virus-like transcriptome shotgun assemblies (TSAs). A maximum likelihood (ML) phylogenetic

tree was constructed using PhyML 3.0, based on the multiple amino acid sequence alignment of the replicase protein or its candidate sequences. A model LG+I+G+F was selected as best-fit model for the alignment. The tree is rooted using the midpoint rooting method. Insect virus-like TSA sequences (Table S1) or an assembled expressed sequence tags (DmeEST) derived from *Drosophila melanogaster* Schneider L2 cell culture (Fig. S1C) are shown in blue text. TSAs derived from hymenopteran and dipteran species are marked with filled and open circles, respectively. Vertical colored lines and dashed line indicate that those virus or virus-like sequences encode structural or putative structural proteins containing the pfam of known coat proteins (TMV_coat, pfam00721; Closter_coat, pfam01785; Cucumo_coat, pfam00760; SP24, pfam16504) or conserved small proteins (SP24 or SP-like, see text) (Kuchibhatla et al., 2014; Shi et al., 2016). The type members of plant-infecting viruses (families *Bromoviridae*, *Closteroviridae* and *Virgaviridae*, genera *Cilevirus* and *Higrevirus*) and insect-infecting viruses (the proposed groups “Sandewavirus” and “Nelorpivirus”) are displayed as collapsed triangles. Virus names and their GenBank accession (or reference sequence) numbers of the sequences are listed in Table S2. Asterisk indicates TSAs whose replicase-like ORFs appeared to be not intact. Numbers at the nodes indicate highly supported values (aLRT > 0.9).

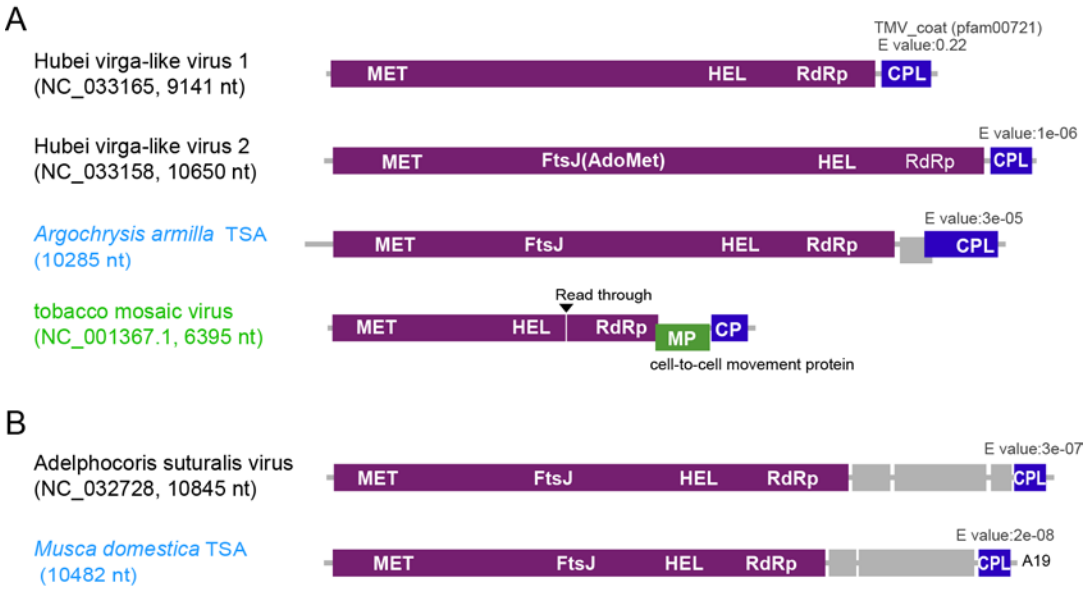

Fig. 2. Putative genome structure of insect virga/nege-like viruses and virus-like TSAs. Schematic representations of virga/nege-like insect TSA accessions from *Argochrysis armilla* (a member of the insect tobamo-like group in Fig. 1) (A) and *Musca domestica* (B) are shown together with a plant virus (tobacco mosaic virus, TMV, genus *Tobamovirus*) and related insect virga/nege-like viruses. Conserved domains using the NCBI Conserved Domain Database are shown (methyltransferase, MET; FtsJ-like methyltransferase, FtsJ; RNA helicase, HEL; RNA-dependent RNA polymerase, RdRp; and coat protein-like, CPL). Grey boxes indicate ORFs encoding proteins of unknown function. The E-values against putative conserved domains of CP (TMV_coat, pfam00721) are shown above each open reading frame.

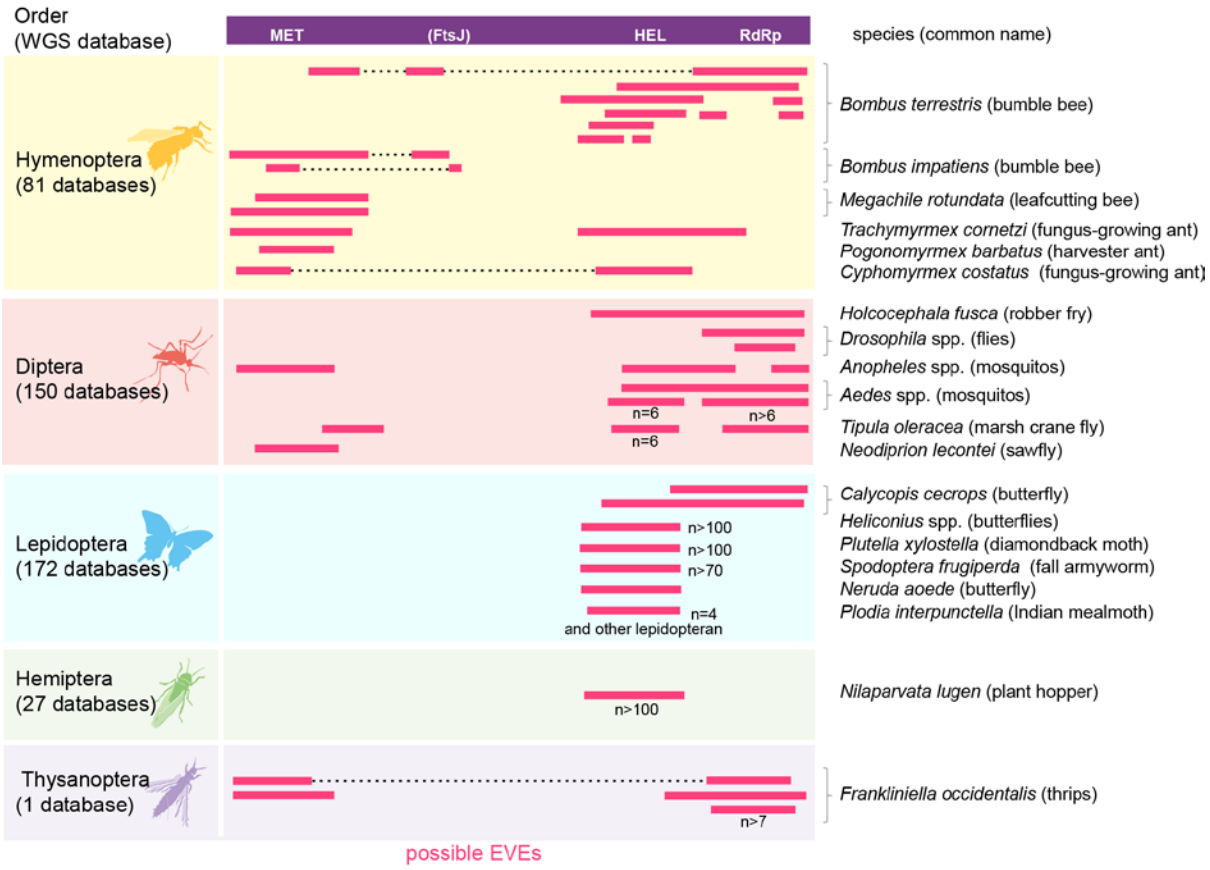

Fig. 3. Endogenous viral elements (EVEs) related to virga/nege-like virus in the insect genomes. A schematic diagram of replicase sequence Hubei virga-like virus 1 (see Fig. 2A) used as a query is shown at the top. EVE candidates (virgavirus replicase-like sequences, VRLSs) are positioned according to the corresponding sequences in the query. VRLSs are grouped with respect to major insect orders, Hymenoptera, Diptera, Lepidoptera, Hemiptera, and Thysanoptera (cartoons of the representative insect are presented). The tblastn hit number of VRLSs are shown below or above the VRLSs, but not in the case of a single hit. Insect species (their common name) from which whole genome shotguns (WGSs) are derived are shown on the right.

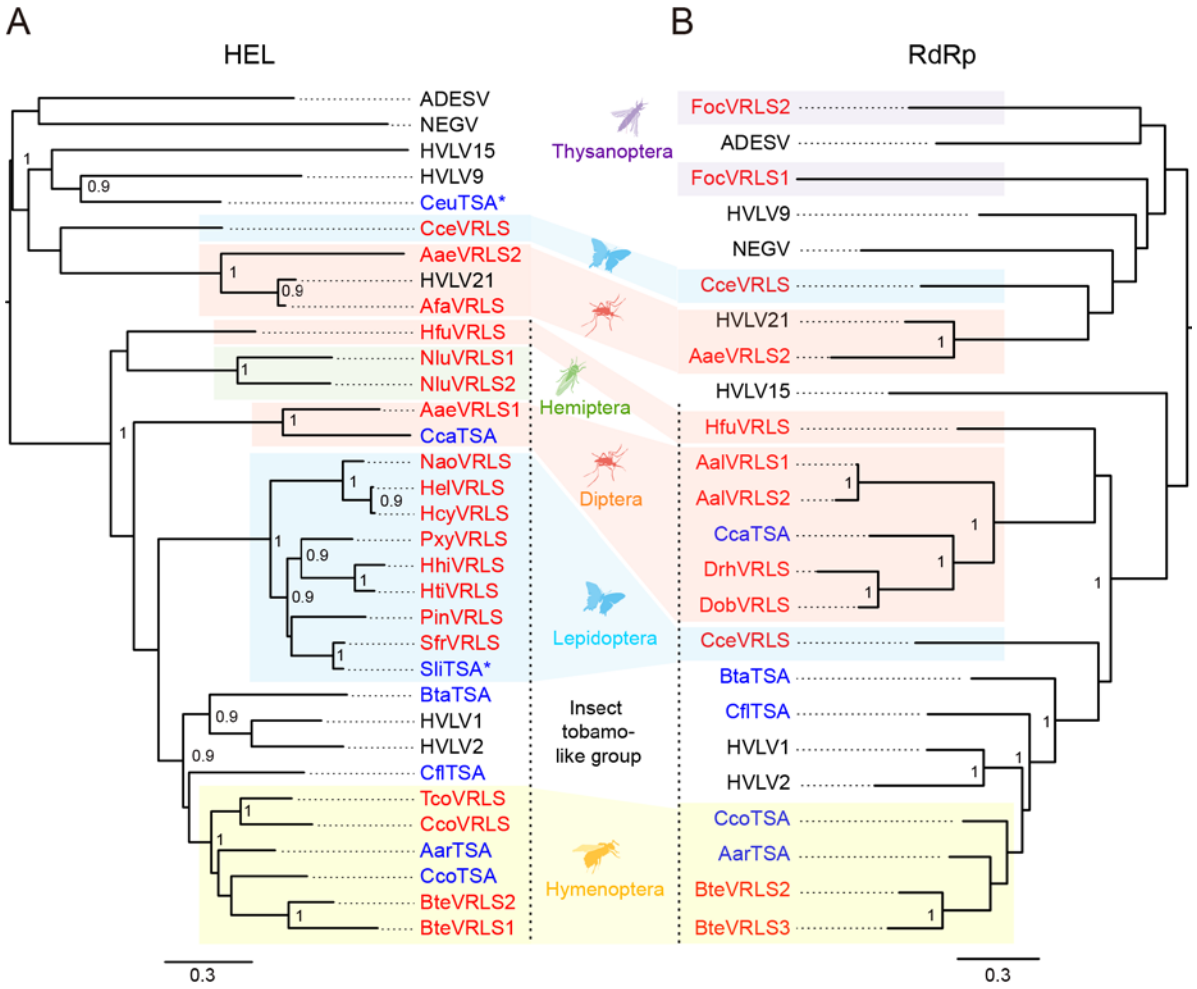

Fig. 4. Molecular phylogenetic analyses of the replicase sequences of insect virga/nege-like viruses and their EVE candidates (virgavirus replicase-like sequences, VRLSs). ML phylogenetic trees were constructed using PhyML 3.0 based on the multiple amino acid sequence alignment of the RNA helicase (HEL) (A), and RNA-dependent RNA polymerase (RdRp) (B) domains, with their flanking regions. The best-fit model LG+I+G+F was selected for both alignments. Insect virga/negevirus-related TSAs and VRLSs derived from five major insect orders (illustrated by cartoons) are listed in Table S1 and S3. TSAs derived from retrotransposons carrying HEL-like domains are marked with asterisk (Fig. S3). The insect tobamo-like group is indicated with vertical dashed lines. Numbers at the nodes indicate highly supported values (aLRT > 0.9).

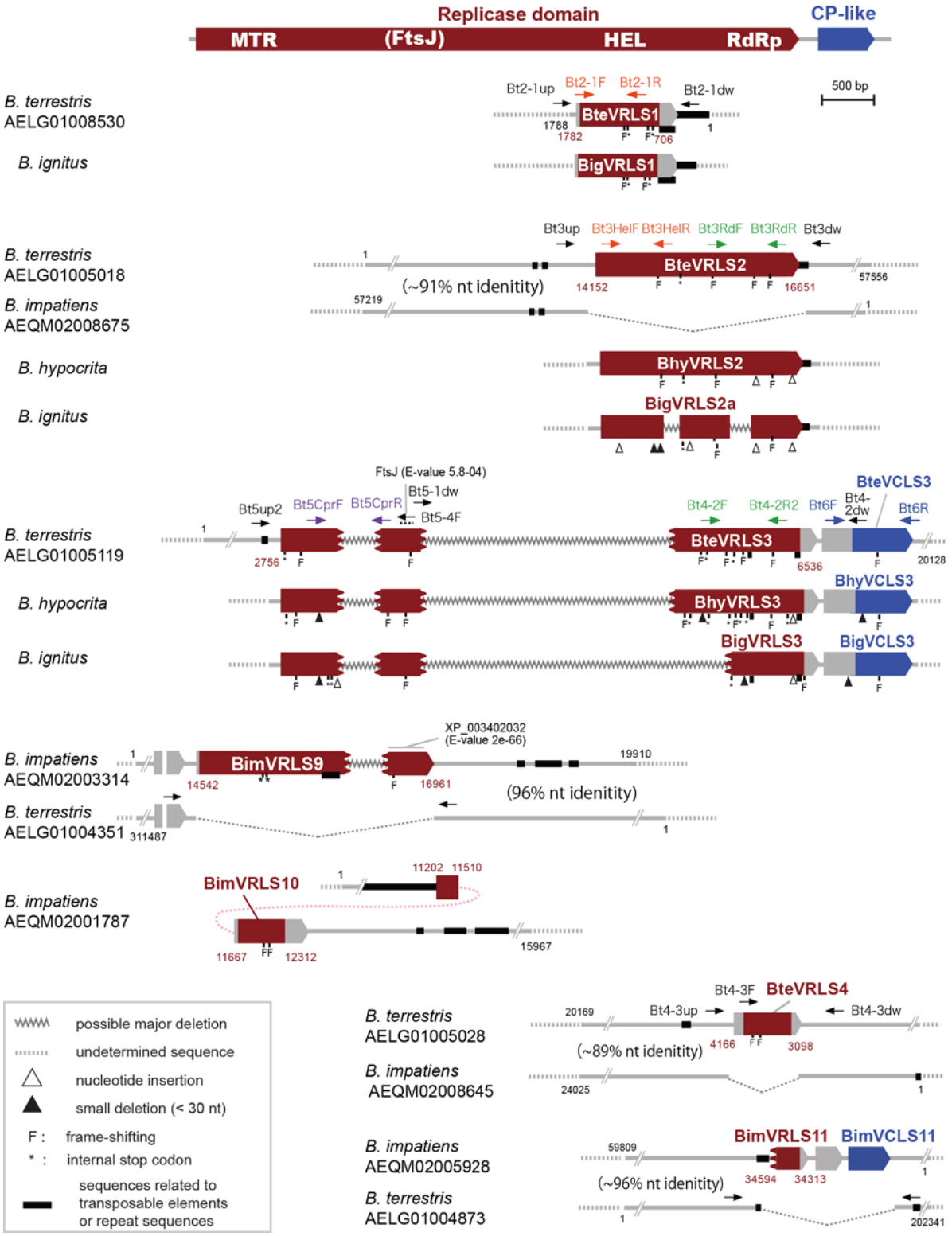

Fig. 5. Virgavirus replicase- or coat protein-related sequences (VRLS/VCLS) in the genome of some species of the *Bombus* genus. A schematic diagram of the replicase and coat protein-like (CP-like) sequence of Hubei virga-like virus 1 used as a query is shown at the top. VRLSs/VCLSs are

positioned according to the corresponding sequences in the query, and the potential coding regions of *Bombus* EVEs (VRLS/VCLSs) are shown as dark red or light blue colored boxes. BimVRLS1, BimVRLS/VCLS3, BteVRLS2 and BteVRLS4 are probably not present in either the *B. terrestris* or *B. impatiens* WGSs. The nucleotide identities of WGS sequences flanking to their counterparts in other species that lack similar VRLS/VCLSs are shown in parenthesis. BimVRLS2 is comprised of two virus-like sequence fragments placed in inverse arrangement. Bte/Bhy/BigVRLS3 have a large internal deletion. The WGS-assembled sequences and undetermined sequences are represented with solid and dashed thin-lines, respectively. Arrows and the text above indicate the position of the primer used for PCR amplification. The symbols referring to the nucleotide mutations are shown within the box.

Supplementary Tables

Table S1. Plant virga/nege-like virus sequences identified in the insect transcriptome shotgun assembly (TSA) database.

Table S2. GenBank/Refseq accession numbers of replicase or its domain sequences presented in Fig. 1.

Table S3. Virga/negevirus replicase-related EVE candidates found in insect genomes.

Supplementary Figure legends

Fig. S1. Putative genome structure of insect virga/nege-related viruses and virus-like TSAs. (A-C) Schematic representations of virus-like insect TSAs (or an EST contig) related to nege- and cileviruses (A), Hubei virga-like virus 15 (B) and Hubei virga-like virus 1 (C, see also Fig. 3A). A virus-like sequence contig (named DmeEST) assembled from *Drosophila melanogaster* ESTs derived from Schneider L2 cell culture is shown in (A). List of the *D. melanogaster* ESTs for DmeEST construction: GenBank accession, BI638919, BG635657, BI638482, AI532995, AI534252, BG637461, BI141725, BI639062, BI638906, AI544395, BG640747, AI533274, BI636597, BG635757, BG641116, AI542032, BI142081, AI532201, BI635191, BI639542, AI542337, AI533639, AI533639, BI141910, BG640612, AI543524, AI530962, BG635737, AI532854, BI636200, AI534776, AI532083, AI532473, AI542186, AI532583, AI542621, AI534897, AI530958, BI636474, AI533948, BI640491, BI639403, BI635232, BG635707, BI636250, AI541916, AI542073, AI531793, BG635866, BG635772, EL872030, EL872030, AI546576, AI532485, BI639356, AI534952, BI635674, BI640165, AI544345, BG640392, AI534985, BI639679, BI632687, BI641883, BI142109, BG635714, AI541862, BI636567, BI640187, AI534771, BI638192, BI641563, BG640363, BI635234, AI542213, BI637338, AI533282, BI636119, BI636730, BI636683, BI641632, BI636623, BI637714, AI533430, BI632540, BI633908, BI638733, BI640451, BG637012, BG640577, BG635825,

BG640481, EL884146, BI141775 and AI531240.Two *Centris flavifrons* TSA sequences were used to construct CflTSA (C).

Fig. S2. Sequence identities among insect tobamo-like viruses. (**A and B**) Replicase (A) and coat protein like (CPL) (B) amino acid sequences of insect viruses and virus-like TSA accessions were compared in pairs using BLAST. (**C**) The neighbor-joining (NJ) tree based on CP and CP-like (CPL) amino acid sequences of plant virgaviruses (highlighted by green boxes), insect virga-like viruses and virus-like TSAs (highlighted by red boxes).

Fig. S3 Schematic representations of insect TSAs derived from putative retrotransposons carrying HEL-like domains (highlighted by purple) in a lepidopteran species, *Spodoptera litura* (A), a cave cricket (*Ceuthophilus* sp.) (B), and a hemipteran (*N. lugens*) (C). The E-values against putative HEL domains are shown above each ORF. The retrotransposon are inserted into a putative alpha amylase coding region (yellow green) in the *S. litura* TSA.

Fig. S4. PCR amplification of virga/negevirus-related EVEs in the genome of the *Bombus* genus. Genomic DNA from *B. terrestris*, *B. hypocrite*, and *B. ignitus* were used for amplification of virgavirus replicase- or coat protein-related sequences (VRLSs/VCLSs). Primer positions for each VRLS/VCLS are shown in Fig. 5 and S1.

Fig. S5. Virgavirus replicase-related sequences (VRLS) identified in the *Bombus terrestris* genome. A schematic diagram of the replicase sequence of Hubei virga-like virus 1 used as a query is shown at the top. VRLSs are positioned according to the corresponding sequences in the query and the potential coding regions of *Bombus* EVEs are shown as dark red colored boxes. The WGS-assembled sequences and undetermined sequences are represented with solid and dashed thin-lines, respectively.

Arrows and the texts above indicate the position of primer used for the PCR amplification. The symbols referring to nucleotide mutations are shown below the grey line.

Fig. S6. Multiple alignments of the *Bombus* EVEs (EVE1, VRLS1; EVE2, VRLS2; EVE3, VRLS3/VCLS3, see Fig. 5). The *Bombus* EVEs and their variants sequenced in this study were labeled with “asterisks” and “a/b”, respectively.

Fig. S7. Sequence identities among virga/negevirus-related EVEs. Nucleotide (A) and amino acid (B) sequences of the selected *Bombus* EVEs (EVE1, VRLS1; EVE2, VRLS2; EVE3, VRLS3/VCLS3, EVE9, VRLS9; EVE10, VRLS10; EVE11, VRLS/VCLS11, see Fig. 5) were compared in pairs using BLAST. The *Bombus* EVEs and their variants sequenced in this study were labeled with “asterisks” and “a/b”, respectively. Sequence identity scores are shown in %. A wasp virgavirus-like TSA (*Argochrysis armilla*, AarTSA) was also included in B.