A novel and diverse family of filamentous DNA viruses associated with parasitic wasps

Benjamin Guinet, Matthieu Leobold, Elisabeth Herniou, Pierrick Bloin, Nelly Burlet, Justin Bredlau, Vincent Navratil, Marc Ravallec, Rustem Uzbekov, Karen Kester, et al.

Virus Evolution, 2024, **10(1)**, veae022
DOI: https://doi.org/10.1093/ve/veae022
Advance Access Publication 2 March 2024
Research Article

A novel and diverse family of filamentous DNA viruses associated with parasitic wasps

Benjamin Guinet,[1,†,§,††] Matthieu Leobold,[2,§,‡‡] Elisabeth A. Herniou,[2,§§] Pierrick Bloin,[2,‡,***] Nelly Burlet,[1] Justin Bredlau,[3,†††] Vincent Navratil,[4,5,6,‡‡‡] Marc Ravallec,[7,§§§] Rustem Uzbekov,[8,9,****] Karen Kester,[3,††††] Dawn Gundersen Rindal,[10] Jean-Michel Drezen,[2,‡‡‡‡] Julien Varaldi,[1,**,*,§§§§] and Annie Bézier[2,**,*****]

[1]LBBE, UMR CNRS 5558, Universite Claude Bernard Lyon 1, 43 bd du 11 novembre 1918, Villeurbanne CEDEX F-69622, France, [2]Institut de Recherche sur la Biologie de l'Insecte, UMR 7261 CNRS-Université de Tours, 20 Avenue Monge, Parc de Grandmont, Tours 37200, France, [3]Department of Biology, Virginia Commonwealth University, 1000 W. Cary Street, Room 126, Richmond, VA 23284-9067, USA, [4]PRABI, Rhône-Alpes Bioinformatics Center, Université Lyon 1, 43 bd du 11 novembre 1918, Villeurbanne CEDEX 69622, France, [5]UMS 3601, Institut Français de Bioinformatique, IFB-Core, 2 rue Gaston Crémieu, Évry CEDEX 91057, France, [6]European Virus Bioinformatics Center, Leutragraben 1, Jena 07743, Germany, [7]Diversité, génomes et interactions microorganismes insectes (DGIMI), UMR 1333 INRA, Université de Montpellier 2, 2 Place Eugène Bataillon cc101, Montpellier CEDEX 5 34095, France, [8]Laboratory of Cell Biology and Electron Microscopy, Faculty of Medicine, Université de Tours, 10 bd Tonnelle, BP 3223, Tours CEDEX 37032, France, [9]Faculty of Bioengineering and Bioinformatics, Moscow State University, Leninskye Gory 73, Moscow 119992, Russia and [10]USDA-ARS Invasive Insect Biocontrol and Behavior Laboratory, Beltsville, MD 20705, USA

[†]Present Adresses: Centre for Palaeogenetics, Stockholm, Sweden. Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden.

[‡]Present Adresses: Laboratoire d'Écologie et Diversité des Insectes Forestiers (ÉcoDIF), Centre de Foresterie des Laurentides (CFL), 1055 rue du P.E.P.S., C.P. 10,380 Québec, G1V 4C7, Canada.

[§]These authors contributed equally to this work.

[**]These authors contributed equally to this work.

[††]https://orcid.org/0000-0002-9922-2118

[‡‡]https://orcid.org/0000-0003-3392-4835

[§§]https://orcid.org/0000-0001-5362-6056

[***]https://orcid.org/0009-0009-3260-9017

[†††]https://orcid.org/0000-0001-6722-9764

[‡‡‡]https://orcid.org/0000-0001-9974-1877

[§§§]https://orcid.org/0000-0002-1781-8082

[****]https://orcid.org/0000-0002-9336-5484

[††††]https://orcid.org/0000-0001-6052-7806

[‡‡‡‡]https://orcid.org/0000-0001-7578-635X

[§§§§]https://orcid.org/0000-0002-2100-1542

[*****]https://orcid.org/0000-0002-2528-4260

[*]Corresponding author: E-mail: julien.varaldi@univ-lyon1.fr

Abstract

Large dsDNA viruses from the *Naldaviricetes* class are currently composed of four viral families infecting insects and/or crustaceans. Since the 1970s, particles described as filamentous viruses (FVs) have been observed by electronic microscopy in several species of Hymenoptera parasitoids but until recently, no genomic data was available. This study provides the first comparative morphological and genomic analysis of these FVs. We analyzed the genomes of seven FVs, six of which were newly obtained, to gain a better understanding of their evolutionary history. We show that these FVs share all genomic features of the *Naldaviricetes* while encoding five specific core genes that distinguish them from their closest relatives, the Hytrosaviruses. By mining public databases, we show that FVs preferentially infect Hymenoptera with parasitoid lifestyle and that these viruses have been repeatedly integrated into the genome of many insects, particularly Hymenoptera parasitoids, overall suggesting a long-standing specialization of these viruses to parasitic wasps. Finally, we propose a taxonomical revision of the class *Naldaviricetes* in which FVs related to the Leptopilina boulardi FV constitute a fifth family. We propose to name this new family, Filamentoviridae.

Keywords: dsDNA virus; filamentous virus; *Naldaviricetes*; *Lefavirales*; EVE; horizontal gene transfer; parasitoid wasp.

1. Introduction

All cellular life forms are associated with viruses (Kristensen et al. 2010; Koonin and Dolja 2013). However, a vast number of viruses remains to be discovered. As Arthropods are the most diverse taxonomic group of animals, a large fraction of this unknown viral diversity is expected to reside within these hosts. Assessment

of arthropod viral diversity has recently progressed thanks to the development of high-throughput sequencing technologies. Deep (meta-) genomic and transcriptomic analyses uncovered the potential breadth of this unknown diversity (Shi et al. 2018; Schulz et al. 2020; Wu et al. 2020; Roux and Emerson 2022). For RNA viruses, the discovery of new lineages is facilitated by the presence of the universal RNA-dependent RNA polymerase. By screening RNAseq datasets, it has been possible to identify thousands of new viruses in single large-scale bioinformatic analysis (Shi et al. 2018; Wu et al. 2020; Edgar et al. 2022). In contrast, DNA viruses do not share any universal gene and are prone to extensive gene loss and gain (Koonin et al. 2020). These features make the exploration of the 'DNA virosphere' more challenging (Schulz et al. 2020).

Among the large eukaryotic dsDNA viruses, the *Naldaviricetes* forms a monophyletic class of arthropod-infecting viruses with specific features (Van Oers et al. 2023). These viruses have circular genomes replicating in the nucleus and packaged into enveloped rod-shaped nucleocapsids. They belong to four families: the *Baculoviridae*, *Nudiviridae*, *Hytrosaviridae*, and *Nimaviridae* ((Walker et al. 2021); see https://ictv.global/taxonomy/taxondetails?taxnode_id=202209115). They share a group of genes that encode the so-called *per os* infectivity factors (PIFs). In *Baculoviridae*, PIFs are required for virus entry into midgut goblet cells during primary infection (Rohrmann 2019), while they may be involved in particle entry into other cell types for other viruses, i.e., salivary glands for hytrosaviruses (Abd-Alla et al. 2008; Garcia-Maruniak et al. 2008), gonad cells for Helicoverpa zea nudivirus 2 (Hamm, Carpenter, and Styer 1996; Burand et al. 2012), and virtually all cell types of parasitized Lepidoptera for bracoviruses (endogenous nudiviruses found in a parasitic wasp lineage of Braconidae) (Bézier et al. 2009; Muller et al. 2021). Among *Naldaviricetes*, the *Baculoviridae*, the *Nudiviridae*, and the *Hytrosaviridae* form the *Lefavirales* order. This name stems from the so-called *lef* (for late expression factor) genes encoding the subunits of a specific viral RNA polymerase, which ensures the transcription of viral genes in the late phase of infection (Rohrmann 2019).

Since the 1970s, investigations using electron microscopy uncovered a variety of viruses in the reproductive tract of parasitic Hymenoptera. In particular, several species of endoparasitoid wasps (i.e., wasps laying their eggs within the body of other insects) harbor long, flexible, filamentous-shaped enveloped particles replicating in their ovaries. These particles are often secreted in the genital fluid, which is injected during wasp oviposition into the parasitized host. Because of the peculiar shape of these particles, these viruses were referred to as 'Filamentous Viruses' (FVs), but their phenotypic effect and transmission means were largely unknown. Such FVs were described in taxonomically diverse wasp species: a Campopleginae (*Diadegma terebrans*) (Krell 1987), a series of Braconidae (*Cotesia congregata, C. hyphantriae, C. marginiventris, Microplitis croceipes, M. rufiventris*, and *M. mediator*) (Stoltz and Vinson 1979; Styer, Hamm, and Nordlund 1987; Tanaka 1987; Hamm, Styer, and Lewis 1990; de Buron and Beckage 1992; Hegazi et al. 2005) and a Figitidae (*Leptopilina boulardi*) (Varaldi et al. 2003, 2006). Of note, another type of 'filamentous virus' infects *Apis mellifera* (AmFV), but it is very distantly related to the FVs we describe in this paper (Gauthier et al. 2015; Yang et al. 2022).

More recently, a first FV detected in the *Drosophila* parasitoid *L. boulardi* was characterized at the phenotypic and molecular level (Varaldi et al. 2003, 2006; Lepetit et al. 2017). Phylogenetic analysis revealed that this virus, named Leptopilina boulardi FV (LbFV), was related to the Hytrosaviruses, known to induce salivary gland hypertrophy in dipteran adults and to impair the mass rearing of tsetse flies for use in the sterile insect technique (Abd-Alla et al. 2009, 2013). However, LbFV was suggested to belong to a new viral family based on its high-sequence divergence from *Hytrosaviridae* (Lepetit et al. 2017). LbFV is vertically transmitted through the maternal lineage and has a strong specific impact on the wasp egg-laying behavior as infected females readily accept to lay their eggs in already parasitized hosts, contrary to the uninfected ones (Varaldi et al. 2003, 2006). This induction of 'superparasitism' enables horizontal transmission of the virus between wasps sharing a same host, thus increasing the virus fitness at the expense of that of the wasp (Gandon, Rivero, and Varaldi 2006). This combination of vertical and horizontal transmission allows the virus to reach high prevalence (>90 per cent) in some wasp populations (Patot et al. 2010). Furthermore, an ancestral genomic integration of a related virus has been described in *Leptopilina* species (Di Giovanni et al. 2020). Endogenized viral genes have been retained by selection and are employed by female wasps as a means of delivering virulence factors (Di Giovanni et al. 2020), thereby protecting their eggs from the host immune system (Rizki and Rizki 1990). FVs may thus be a source of adaptive genes for parasitic wasps, as has been repeatedly observed in numerous viral domestication events (Bézier et al. 2009; Volkoff et al. 2010; Pichon et al. 2015; Burke et al. 2018).

In this study, we characterized six new genomes of LbFV-related viruses infecting wasps belonging to four Hymenoptera superfamilies (Cynipoidea, Chalcidoidea, Platygastroidea, and Ichneumonoidea) by in-depth comparative genomic and phylogenomic analyses, coupled with electron microscopic studies of particle morphogenesis. Our methodology highlights clear common features that distinguish these FVs from their closest relatives, the *Hytrosaviridae*, and shows that FVs likely constitute a monophyletic group within *Naldaviricetes*. As a result, we propose the creation of a new family within the *Lefavirales* order, which we tentatively name Filamentoviridae. Moreover, by mining public databases, we provide evidence that FVs preferentially infect Hymenoptera with parasitoid lifestyle and that their DNA has been repeatedly integrated into insect genomes.

2. Results

2.1 Genomes of FVs related to LbFV have similar structures

We analyzed the genomes of six novel FVs obtained from parasitoid wasps, *Leptopilina heterotoma* (Cynipoidea, Figitidae), *Encarsia formosa* (Chalcidoidea, Aphelinidae), *Platygaster orseoliae* (Platygastroidea, Platygastridae), *Psyttalia concolor* (Ichneumonoidea, Braconidae), and two *Cotesia congregata* incipient species (Ichneumonoidea, Braconidae) (Bredlau et al. 2019). These parasitic wasps belong to four out of the seven main superfamilies of Hymenoptera and develop from Diptera (*L. heterotoma, P. concolor* and *P. orseoliae*), Lepidoptera (*Cotesia sp.*), and Hemiptera (*E. formosa*) hosts. One genome was retrieved following viral purification and the other five were discovered while sequencing the parasitoid genomes (Supplementary Table S1). Hereafter, the viruses are designated by the initials of the species name of the wasp followed by FV for 'Filamentous Virus' (i.e., LhFV, EfFV, PoFV, PcFV, CcFV1, and CcFV2, respectively).

2.1.1 *Completeness of the six new FV genomes*

To confirm the exogenous nature of the five viral genomes obtained from parasitoid whole-genome sequencing projects, we compared the sequencing depths of each virus to those of the host. As genome coverage was significantly different between viral

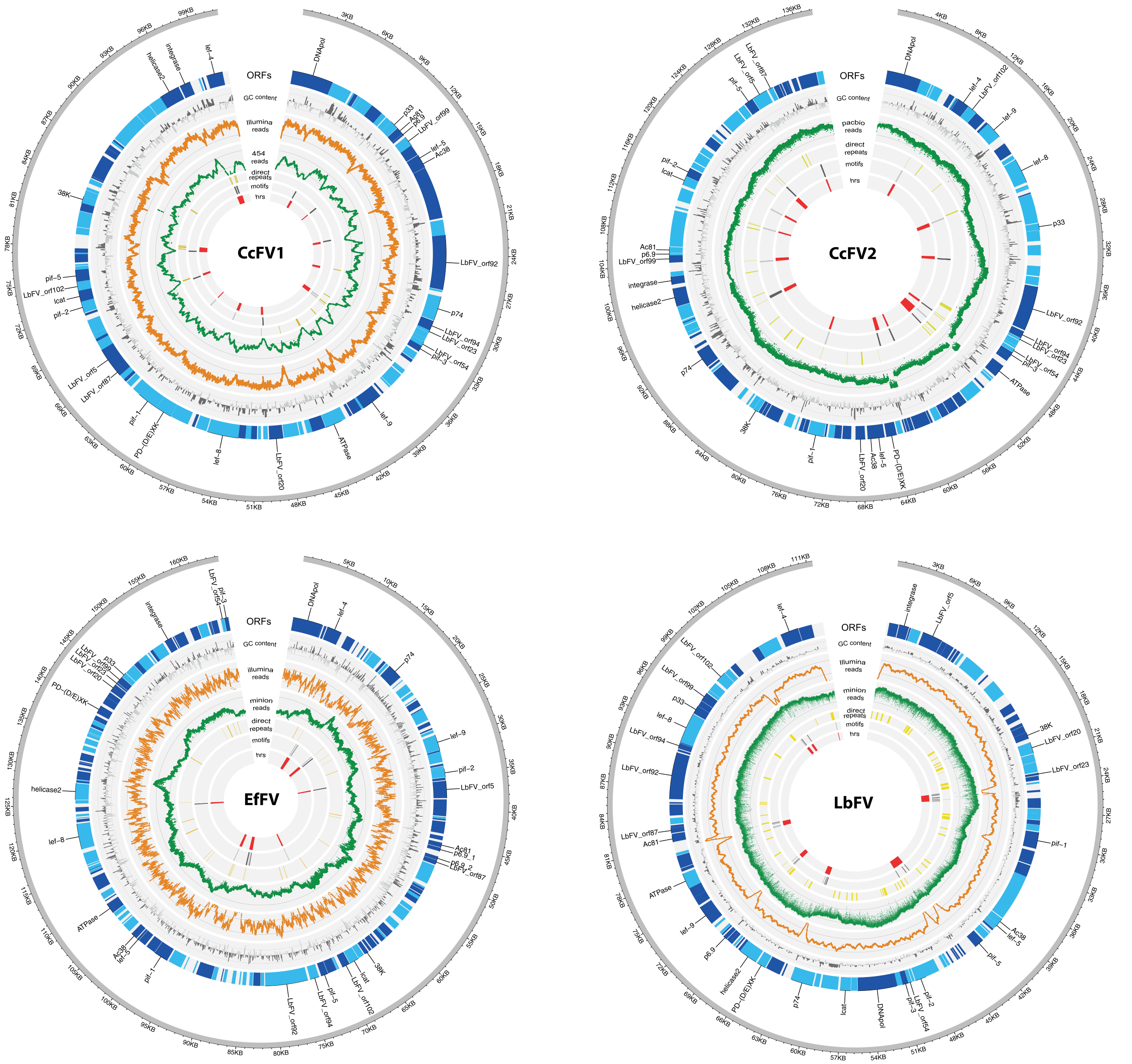

Figure 1. Diagrams of the four circularized FV genomes. CcFV1 = Cotesia congregata FV 1 from *Cotesia congregata* (laboratory line, MsT), CcFV2 = Cotesia congregata catalpae FV 2 from *Cotesia congregata catalpae* (line recently obtained from a wild population, CcC), EfFV = Encarsia formosa FV and the previously published LbFV. Predicted ORFs are displayed as blue rectangles, where dark blue and light blue represent positive and negative strands, respectively. The GC content along the genome is displayed with bars, where black bars represent GC > mean GC content and grey bars represent GC < mean GC content. Coverage corresponding to long read technologies (Nanopore, 454 or PacBio) is plotted in green, while coverage corresponding to short reads (Illumina) is displayed in orange. Repeated elements of interest are plotted in the center with the homologous repeat sequences (*hrs*) in red, the best motifs elements in grey (with light grey = sense strand and dark grey = antisense strand) and the direct repeats in yellow. First ORF was set as the *DNApol* for all three new genomes, as it is usually done for related viruses, except for the previously published LbFV where the *DNApol* corresponds to *orf*58.

and parasitoid contigs (Supplementary Fig. S1 and Supplementary Table S1), this strongly suggests that viral contigs do not correspond to endogenous viruses (integrated form of viruses), but genuine exogenous viruses associated with the wasps. Furthermore, various combinations of short and long-read sequencing approaches (Supplementary Table S1 and Section 4.3) allowed the assembly of single circular molecules for EfFV, CcFV1, and CcFV2, in addition to the previously obtained circular form of LbFV (Lepetit et al. 2017) (Fig. 1).

Even though no circular forms could be obtained for the other three genomes (LhFV, PoFV, and PcFV), we consider the assembled sequences to be nearly complete, since contig cumulative sizes ranged from ∼105 to 137 kb, comparable to the 101–164 kb of the FV circular genomes (Supplementary Table S1), and similar to other *Lefavirales* genomes (80–232 kb) (Supplementary Table S2). Additionally, *de novo* gene prediction showed a composition of 107–128 ORFs for these three genomes, which is comparable to the number of ORFs predicted for the four FV circular genomes (104–156; Supplementary Tables S1 and S3), and for other related viruses (*Naldaviricetes* = 90–241). Repetitive sequences (Section 2.1.3), as previously observed for LbFV (Lepetit et al. 2017), combined with the use of short (Illumina) reads only, most probably impeded the circular assembly of these three genomes, despite sufficient sequencing depth (Supplementary

Table S1). Finally, all six new virus genomes had high-coding densities (79.9–92 CD%) as expected for exogenous viruses belonging to *Naldaviricetes* (Fig. 1, Supplementary Tables S1 and S2).

2.1.2 *Global gene composition*

The predicted genes were uniformly distributed on both DNA strands for all the seven viruses studied (df = 1, all P-value > 0.05; Fig. 1 and Supplementary Table S3). The deduced proteins contained 49–2,145 amino acids with a median of 201 amino acids and 20–25 per cent had transmembrane domains (1–5, Supplementary Table S3). Putative function could be inferred for 31 per cent of the 840 predicted ORFs (Supplementary Table S3). Of note, 113 predicted ORFs belonged to multigene families (i.e., several variants found within the same genome; Supplementary Table S3 'multigenic family' column), some of them having predicted functions such as the *jmjd* or the well-known *bro* and *iap* (Section 2.3), while others have unknown function and are for the most part specific to a single virus

2.1.3 *Repeated regions as a common feature*

Homologous regions (*hrs*) are repeated sequences commonly encountered in dsDNA viruses. Mostly studied in *Baculoviridae*, *hrs* was found to enhance gene transcription and to serve as origins of replication (Theilmann and Stewart 1992; Leisy and Rohrmann 1993; Kool et al. 1995). By aligning each genome sequence to itself, we found *hrs* in all circularized FV genomes, but none in hytrosaviruses, revealing an important difference between the two virus clades (Supplementary Fig. S2).

FV genomes harbored from seven to fourteen *hrs*, composed of 147–300 nucleotide-long highly conserved motifs (Supplementary Fig. S2 and Supplementary Table S4) and from 19 to 53 direct repeats (Fig. 1). Interestingly, the *hrs* share a similar organization, as three–six palindromic sequences constitute a large part of the conserved motif in each virus (Fig. 1, Supplementary Fig. S2 and Supplementary Table S4). The sequences are highly conserved among *hrs* of the same virus (Supplementary Fig. S2A–D) but poorly conserved between viruses, although a 153-nucleotide long consensus motif comprising two palindromes was obtained (Supplementary Fig. S2E), which could reflect a common evolutionary origin of all these *hrs*. In summary, the presence of direct repeats and *hrs* is a feature shared by all FVs analyzed and these sequences are mainly virus-specific as in Baculoviruses.

2.2 Determining the FV core gene set

Gene composition and evolutionary conserved core genes are important criteria to identify viruses belonging to the same taxonomic group (Lefkowitz et al. 2018). The analysis of seven FV genomes resulted in the identification of a set of twenty-nine core genes (Fig. 2 and Supplementary Table S5), four-fifths of which are found in other *Lefavirales* families. Based on known functions in Baculoviruses, they are most likely involved in viral transcription, viral DNA packaging in the nucleocapsids, particle assembly and morphogenesis, as well as in virus entry into cells (infectivity). Eight core genes, including five specific FV genes, encode proteins

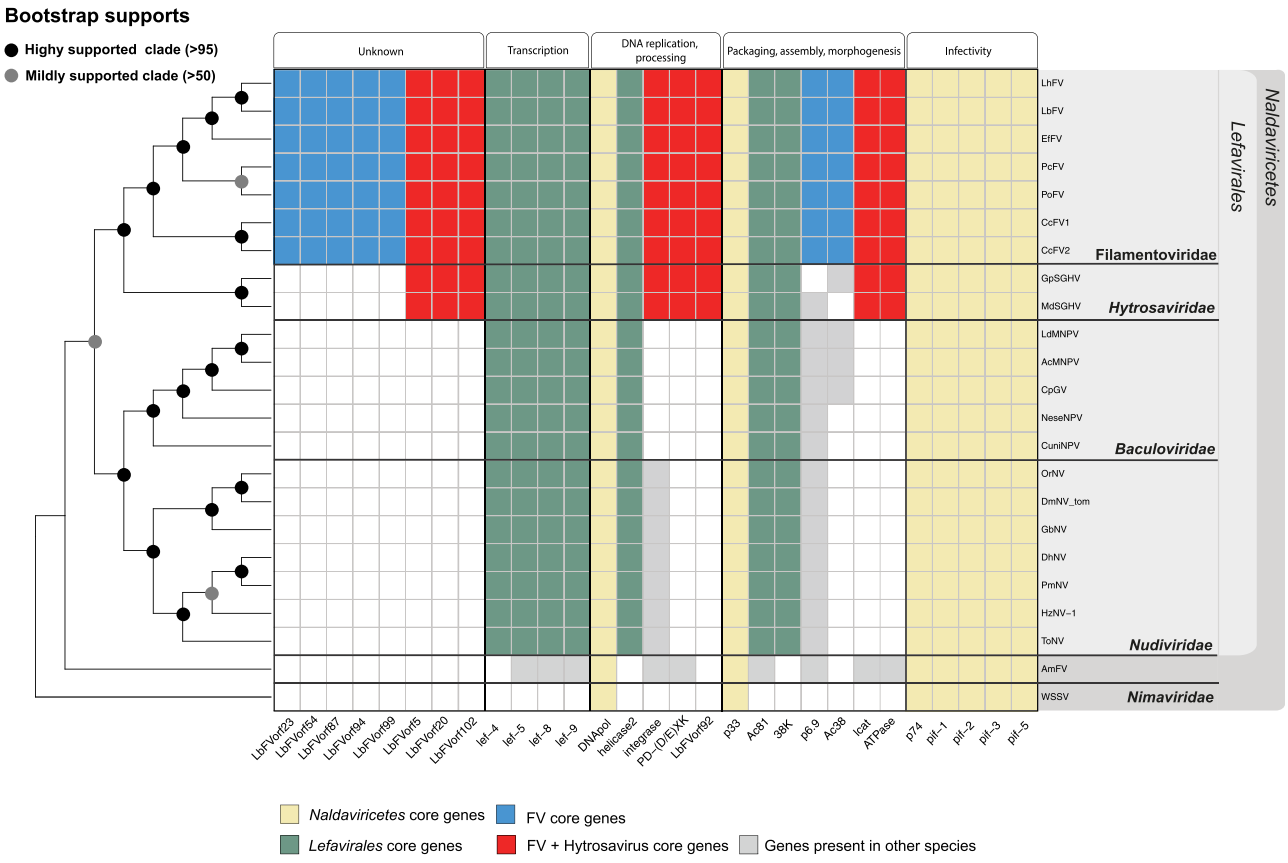

Figure 2. Heatmap representing the core gene content of FVs related to LbFV compared to virus representatives of *Naldaviricetes*. A cladogram phylogeny is reported on the left. Taxonomic affiliation of considered viruses is reported on the right. The rows represent the viral species, and the columns represent the genes grouped according to their potential functions. Colored cells indicate the presence of the gene in viral genomes. Of note, the homolog of Baculoviruses *helicase* is the *helicase 2* in Nudiviruses, Hytrosaviruses, and FVs.

of unknown function, as no sequence similarity or known domain was detected (Fig. 2, blue columns).

2.2.1 All the lefavirales core genes are conserved

All the seven core genes shared by the *Naldaviricetes* including the *per os* infectivity factors (*p74, pif-1, pif-2, pif-3*, and *pif-5*), the DNA polymerase (*DNA pol*), and the sulfhydryloxidase (*p33*) were detected in the analyzed FV genomes (Fig. 2, yellow columns).

The six specific *Lefavirales* core genes were also present (Fig. 2, green columns), i.e., three of the four subunits of the viral RNA polymerase (*lef-4, lef-8*, and *lef-9*), *lef-5*, *helicase* (nudiviral *helicase* 2 homolog) and *Ac81*. Of note, the *lef-5* ORF was readily identified in the genomes of CcFV1, CcFV2, EfFV, PcFV, and PoFV, but was predicted using an alternative start codon in LbFV and LhFV. However, RNAseq data from *L. boulardi* (Varaldi and Lepetit 2018) indicate that *lef-5* is fully transcribed (22,387 transcripts *per* million (TPM)) supporting the use of the alternative start codon and suggesting that the gene is functional, although proteomic data would be necessary to confirm this.

In addition, all FV genomes contained the *38 K* gene, encoding for a nucleocapsid protein also present in all baculoviruses and nudiviruses. To date, no homolog has been reported in hytrosaviruses. However, using an HMMER search, we found that MdSGHV073 from MdSGHV (YP_001883401.1, *e*-value = 0.000095) and SGHV044 from GpSGHV-Uga (YP_001686992.1) are indeed hytrosavirus 38 K homologs. Accordingly, the corresponding proteins were characterized as virion (Kariithi et al. 2010) and nucleocapsid components (Abd-Alla et al. 2016), as in Baculoviruses (Wu et al. 2008; Blissard and Theilmann 2018). Thus, *38 K* is an additional seventh core gene shared by all *Lefavirales* (Fig. 1, green columns).

2.2.2 Eight core genes are specifically shared by FVs and hytrosaviruses

FVs and *Hytrosaviridae* shared eight core genes (Fig. 2, red columns), seven of which encode virion, nucleocapsid, or envelope components in GpSGHV-Eth (Abd-Alla et al. 2016). Only four genes have predicted functions that we discuss below.

The *integrase* homologs are also core genes of nudiviruses and bracoviruses (Drezen et al. 2022; Petersen et al. 2022); in the latter, they are involved in the process of DNA excision/circularization producing the circles packaged in bracovirus particles and are viral particle components (Burke et al. 2013; Burke and Strand 2014).

PD-(D/E)XK nuclease (see IPR038726 from InterPro database) homologs could be involved in DNA replication or processing (Steczkiewicz et al. 2012). Two forms have been identified as virion-associated proteins in hytrosaviruses (Garcia-Maruniak et al. 2009; Abd-Alla et al. 2016). PD-(D/E)XK nuclease homologs are also found in prokaryotes, in giant viruses as well as in the betabaculovirus Diatraea saccharalis granulovirus (Ardisson-Araújo et al. 2016), and in AmFV (Gauthier et al. 2015).

Lecithin-cholesterol acyltransferases (LCAT) are catabolic enzymes, usually involved in the lipid metabolic process, which are activated by apolipoproteins (Saeedi, Li, and Frohlich 2015). Interestingly, in the protozoan parasite *Toxoplasma gondii*, a LCAT protein facilitates parasite cytolytic egress from infected cells (Pszenny et al. 2016). However, the function of the viral LCAT protein, which is not a component of viral particles in GpSGHV-Eth (Abd-Alla et al. 2016), remains unknown.

ATPases, from the AAA+ superfamily (IPR003593), participate not only in diverse cellular processes including membrane fusion, proteolysis, DNA recombination, replication and repair in cellular organisms, but also in viral replication (Ogura and Wilkinson 2001; Snider, Thibault, and Houry 2008; Khan et al. 2022). Hytrosavirus homologs further display similarity with cell division control protein 48 and/or vacuolar protein sorting-associated protein, which suggests these ATPases may also have a role in virus morphogenesis and egress (Kolesnikova et al. 2009; Hilbert et al. 2015; Ahmed et al. 2019; Huttunen et al. 2021).

The phylogenies obtained for LCAT and ATPase (Supplementary Fig. S3A and S3B, respectively) suggest that both genes have been acquired from bacteria through horizontal transfer. This acquisition most likely occurred following a single ancestral event for ATPase (in the common ancestor of FVs and *Hytrosaviridae*), while it is unclear whether a single or two independent events (one for each viral family) occurred for LCAT (Supplementary Fig. S3A).

Finally, among the four genes with no predicted functions, HHpred analyses (Gabler et al. 2020) suggest the *LbFV_orf92* gene might encode a viral primase-helicase (98.66 per cent probability from positions 1,358 to 1,503 from FV sequence alignment) or a ssDNA-binding protein (97.57 per cent probability from positions 1,487 to 1,613 from hytrosavirus sequence alignment). It was therefore classified here as having a predicted function in DNA replication/processing (Fig. 2), although it was reported as an envelope component in GpSGHV-Eth (Abd-Alla et al. 2016).

2.2.3 Seven core genes distinguish FVs from hytrosaviridae among which five are specific to FVs

Five core genes (*LbFV_orf23, LbFV_orf54, LbFV_orf87, LbFV_orf94*, and *LbFV_orf99*) are restricted to FVs (Fig. 2, blue columns). Although their function is unknown, their conservation suggests that they might play an important role in the specific biology of FVs. In addition to these five FV-specific genes, two FV core genes were also detected in one or the other hytrosavirus (*Ac38* and *p6.9*).

Firstly, FVs harbor the *Ac38* gene, which is found in all lepidopteran infecting baculoviruses and in GpSGHV but not in MdSGHV (Abd-Alla et al. 2008; Garcia-Maruniak et al. 2008). In Baculoviruses, Ac38 may be involved in budded virus production (Ge et al. 2007) and is associated with the envelope of budded virions (Wang et al. 2005). Ac38 proteins contain a conserved Nudix (nucleoside diphosphate X) motif (Mildvan et al. 2005), which is thought to negatively regulate viral gene expression by acting as a decapping enzyme in the vaccinia virus (Parrish and Moss 2007).

Secondly, all FVs contain *p6.9*, which is a core gene of *Baculoviridae* and *Nudiviridae*. In Baculoviruses, this gene encodes an abundantly expressed arginine/serine/threonine DNA-binding protein that is an essential nucleocapsid component (Hou et al. 2013; Irwin et al. 2021). This gene is often overlooked by automated computer analyses due to its small size and its low-sequence complexity. We were able to identify MdSGHV026 (YP_001883354.1) as a potential hytrosavirus P6.9 homolog, but, despite in-depth manual analyses, we could not find any homolog in GpSGHV genomes.

2.3 Focus on a few additional genes of interest

Besides core genes, viral genomes encode accessory genes that are specific to a clade or a particular virus, potentially conferring host adaptation. In the following, we will focus our attention on some genes that may play an important role in virus biology. To do so, we investigated their evolutionary history, whenever possible, knowing that viruses often coopt genes of various origins to

acquire new functions (Filée, Siguier, and Chandler 2007; Legendre et al. 2011). Among the 23 putative cases of horizontal gene transfer (HGT), we identified (Supplementary Fig. S3 A to W), eleven most likely involved eukaryotes as donors (such as *jmjd* and *iap* (Supplementary Fig. S3C and S3E, respectively), four (including the two previously described *lcat* and *ATPase* core genes (Supplementary Fig. S3A and S3B, respectively)) involved bacteria, and one involved virus (i.e., *bro* (Supplementary Fig. S3D)). Most HGT events were unique to a single FV species (16 out of 23), but some were shared by several, if not all, FVs suggesting ancient transfer events.

2.3.1 *The Jumonji C domain-containing gene family*

All FV genomes encode for at least one protein with a Jumonji C (JmjC) domain (IPR003347) or a related cupin domain (IPR041667). In cellular organisms, JmjC domain-containing proteins are predicted to be metalloenzymes potentially involved in chromatin regulation-related processes. So far, viral homologs have only been reported in two *Megaviricetes* (Colson et al. 2011; Legendre et al. 2011; Zhang et al. 2015) and a cyanophage (Supplementary Fig. S3C). Phylogenetic analyses show that FV JmjC domain-containing proteins are distant from known viral sequences and reveal a complex evolutionary history involving several horizontal transfer events, duplications as well as domain acquisitions or losses (Supplementary Fig. S3C). Because they do not clearly belong to a monophyletic group, we chose not to include *jmjd* in the set of core genes although they are present in all FVs analyzed.

2.3.2 *Genes interfering with cellular processes*

Many FV proteins contain interaction domains (Supplementary Table S3) commonly used by pathogens, including viruses, to facilitate infection processes (Correa et al. 2013; Matsushima et al. 2021; Matsushima and Kretsinger 2022). Several FV genomes encode one to three proteins with BRO-N domains, as well as other proteins with conserved domains frequently found in BRO (Baculovirus Repeated ORFs) proteins (Iyer, Koonin, and Aravind 2002). BRO proteins have DNA-binding ability and are supposedly involved in virus and/or host DNA replication and/or as transcriptional regulators (Zemskov, Kang, and Maeda 2000; Iyer, Koonin, and Aravind 2002). They commonly form multigene family in insect dsDNA viruses (*Baculoviridae*, *Ascoviridae*, *Poxviridae*, and *Iridoviridae*), as well as bacteriophage and bacterial transposons (Bideshi et al. 2003). Phylogenic analysis suggests that FV *bro* genes have been repeatedly acquired from other viruses following several horizontal transfers (Supplementary Fig. S3D).

Insect DNA viruses commonly interfere with the host immune system by preventing apoptosis (Marques and Imler 2016). Several FV genes encode for this function. In CcFV2, EfFV, LbFV, and PcFV, six genes in total encoded putative inhibitors of apoptosis (IAP) (Supplementary Table S3). FV IAPs display one or two type I BIR (baculovirus IAP repeat) domains (IPR001370) sometimes in combination with a C3H4 type (IPR013083) Zinc-/RING-finger domain, which is typical of insect and virus IAPs (Miller 1999; Verhagen, Coulson, and Vaux 2001). Phylogenetic trees showed FV *iap* genes probably derived from independent HGTs involving insect donors (Supplementary Fig. S3E), as observed in Baculoviruses (Hughes 2002). In addition to *iap* genes, a homolog of the anti-apoptotic protein *p35* gene (IPR036562) was detected in PoFV (Supplementary Table S3).

2.3.3. *The odv-e66 gene family within the lefavirales*

All the FV genomes, except EfFV, contain *odv-e66* (*Ac46*) genes (Supplementary Table S3). To our knowledge, *odv-e66* has exclusively been described in *Lefavirales* (IPR006934). Most FVs and hytrosaviruses (Garcia-Maruniak et al. 2009) harbor a single-gene copy while up to five copies are found in nudiviruses (Bateman et al. 2021) and up to two in baculoviruses (Rodrigues et al. 2020). Notably, *odv-e66* also forms a highly expanded gene family in bracoviruses with 36 genes (Bézier et al. 2009; Gauthier et al. 2021). They are structural components of baculovirus ODV envelopes (Hong, Braunagel, and Summers 1994; Boogaard, Van Oers, and Van Lent 2018) and bracovirus particles (Bézier et al. 2009; Burke et al. 2013). Likely originating from bacterial gene acquisition followed by several loss and gain events (Rodrigues et al. 2020; Bateman et al. 2021) combined with neofunctionalization or sub-functionalization processes (Francino 2005), *odv-e66* seems important for viral adaptation. FV ODV-E66 displays chondroitin AC/alginate lyase (IPR008929) and polysaccharide lyase 8 (IPR012970) overlapping domains, which may facilitate viral infection by digesting extracellular structures such as the insect peritrophic membrane as described in Baculoviruses (Sugiura et al. 2011; Hou et al. 2019).

2.3.4. *Candidate gene involved in virus morphogenesis processes*

Individual HHpred analyses (Gabler et al. 2020) combined with homologous sequence alignment led to the identification of a gene that may encode a nucleoporin (NUP) and that is conserved in all FV species (Supplementary Table S3; CcFV1_orf073, CcFV2_orf102, EfFV_orf124, LhFV_contig_22588_orf005, PcFV_scaffold_2882_orf006, PoFV_scaffold_7638_orf006), except LbFV. Viruses can hijack host nuclear pore complexes to promote some specific steps of viral morphogenesis, such as viral replication or viral particle nuclear import/export, or to interfere with the host immune response (Le Sage and Mouland 2013; Flatt and Greber 2017; Shen, Wang, and Palazzo 2021; Zhang et al. 2022). However, functional studies are required to determine the role of these genes in the FV infection process.

2.4. Position of FVs within the Naldaviricetes

The relationships between the available FVs and the *Naldaviricetes* (Supplementary Table S2) were inferred using two phylogenetic approaches. The first rather classical inference is based on the twenty-nine FV core genes (Section 2.2; Supplementary Table S5) using both Bayesian and maximum likelihood (ML) estimation models. The second 'all genes' method is an ML phylogenomic approach that takes into account seventy-two genes corresponding to all proteins encoded by at least four of the twenty-five viral species considered. The 'core genes' and 'all genes' datasets were composed of 7,031 and 17,114 amino acids, respectively.

All phylogenetic analyses revealed very similar tree topologies, showing the FVs as a strongly supported monophyletic clade, sister group to the *Hytrosaviridae* (Fig. 3 and Supplementary Fig. S4). On the other hand, the position of AmFV at the base or within *Lefavirales* was unclear and method-dependent (Fig. 3 and Supplementary Fig. S4). Within the FVs, the 'all genes' approach left some uncertainties, while the Bayesian and ML 'core genes'-based approaches were consistent and fully resolved the relationships within them (Supplementary Fig. S4). Evolutionary distances in the 'core gene' tree were calculated within and between virus families belonging to the *Lefavirales* order. As expected, all patristic distances within families were smaller (min = 0.36, max = 2.81)

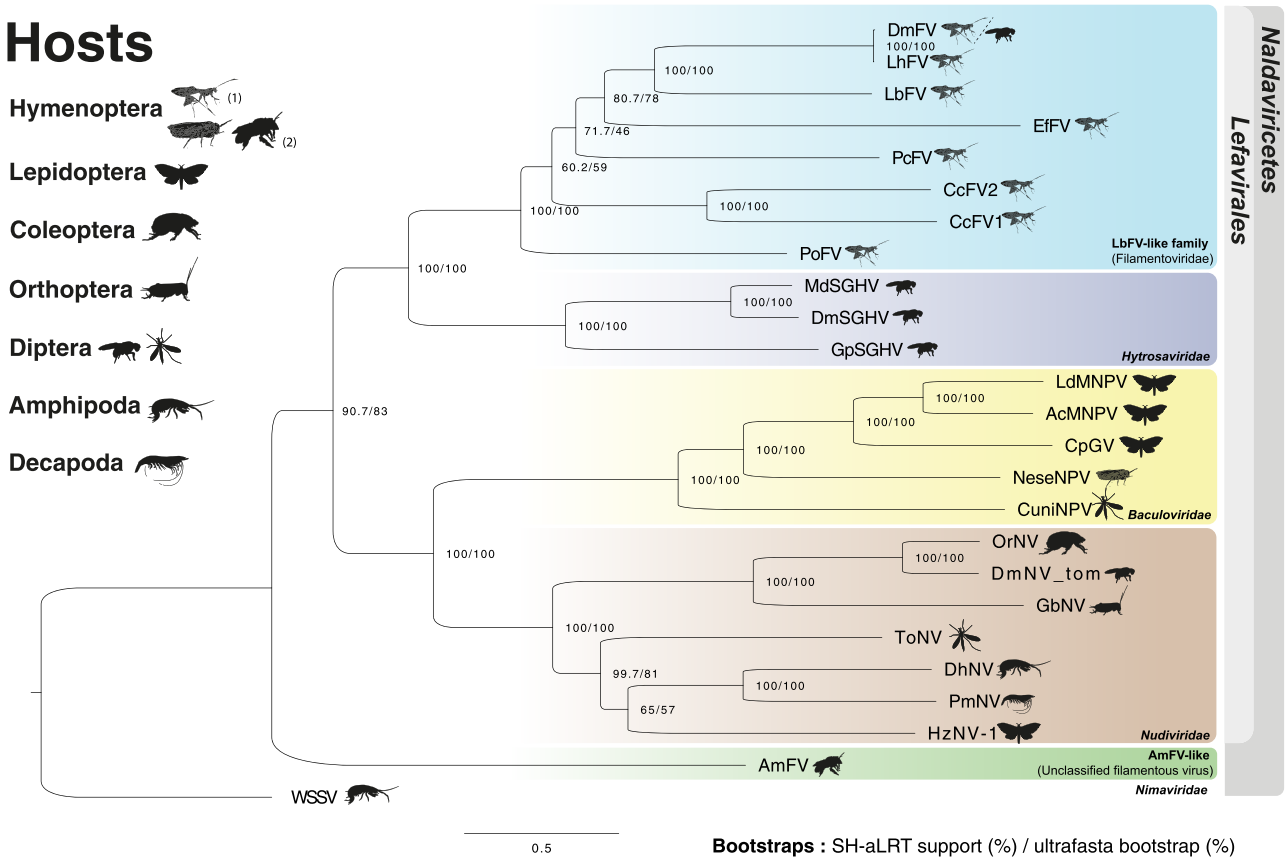

Figure 3. Phylogenetic tree of large dsDNA viruses from the *Naldaviricetes* class. Relationships were inferred using ML analysis in IQ-TREE v2 from twenty-five virus species with 17,114 sites with 16,650 distinct patterns at the amino-acid level. Bootstrap values are shown at each node (SH-aLRT support (%)/ultrafast bootstrap (%)). The scale bar indicates the average number of amino acid substitutions *per* site. Viral families are represented by the following colors: brown = *Nudiviridae*, yellow = *Baculoviridae*, purple = *Hytrosaviridae*, light blue = proposed Filamentoviridae and green = AmFV-like (Supplementary Tables S1 and S2). After each virus name a drawing represents the order of arthropod to which the infected host belongs. Hymenoptera have two lifestyle categories indicated: (1) endoparasitoids or (2) free-living species. DmFV and LhFV are likely the same virus infecting the *Drosophila* parasitoid *L. heterotoma*, although DmFV has been initially detected from a pool of wild-caught *Drosophila* (Section 2.5).

than those between families (min = 3.60, max = 4.57) (Supplementary Fig. S5). Moreover, the average distance within the FV clade was comparable to the average distance observed within recognized families and lower than any distance between families (Supplementary Fig. S5). Altogether, the phylogenetic analysis supports the hypothesis that FVs form a family distinct from *Hytrosaviridae* (Fig. 3).

2.5 DmFV and LhFV are the same virus most likely infecting the parasitoid *Leptopilina heterotoma*

As can be observed in the phylogeny (Fig. 3), LhFV, which was obtained from the *Drosophila* parasitoid *L. heterotoma*, was almost identical (i.e., 97.98 per cent and 99.999 per cent at nucleotide and protein levels, respectively) to the recently published sequences of *Drosophila melanogaster*-associated FV (here called DmFV to follow the rule established for our own viruses). DmFV was detected in one of the 167 pool-seq libraries sequenced during a large sampling of wild *Drosophila* in forty-seven European locations (DrosEU consortium, (Walker et al. 2021)). We hypothesized that these reads were in fact derived from a virus injected by the *Drosophila* parasitoid *L. heterotoma*, which is known to be present in this geographic area (Fleury et al. 2009). Indeed, this situation may occur if the *Drosophila* survives parasitoid infestation, which is common in nature (Fellowes, Kraaijeveld, and Godfray 1998) and if viral DNA injected by the wasp persists in the adult fly.

To test this hypothesis, we screened SRA datasets for reads that have been assigned to its closest relative, LbFV, which was shown to be strictly specific to the parasitoid (Varaldi et al. 2003; Patot et al. 2009, 2012), using the SRA Sequence Taxonomic Analysis Tool (STAT, (Katz et al. 2021)). Seven *Drosophila* SRA entries displayed reads (from 5 to 986 reads) matching with LbFV sequences, all from DNA pool-seq libraries made of wild caught individuals, mostly generated by the DrosEU consortium, and including the library positive for LhFV/DmFV which thus contained both LbFV and LhFV/DmFV reads (Supplementary Table S6). This suggests that some of the *D. melanogaster* pooled in that sampling have survived after parasitoid attack by LbFV-infected wasps as suggested by the presence of a few reads confidently assigned to *L. boulardi* (Supplementary Table S6). This result suggests that similarly, some *Drosophila* pooled in the sample positive for DmFV may have survived infestation by LhFV/DmFV-infected wasps, although no trace of *L. heterotoma* DNA was detected in this particular sample (Supplementary Table S6). This possibility, acknowledged by Wallace and collaborators (Wallace et al. 2021), is also consistent with the fact that FV infection was never observed in 230 lab-established *D. melanogaster* isofemale lines tested even though both LbFV and LhFV/DmFV infect *Leptopilina* collected from the same collection sites (Varaldi et al. 2024). Alternatively, it is also

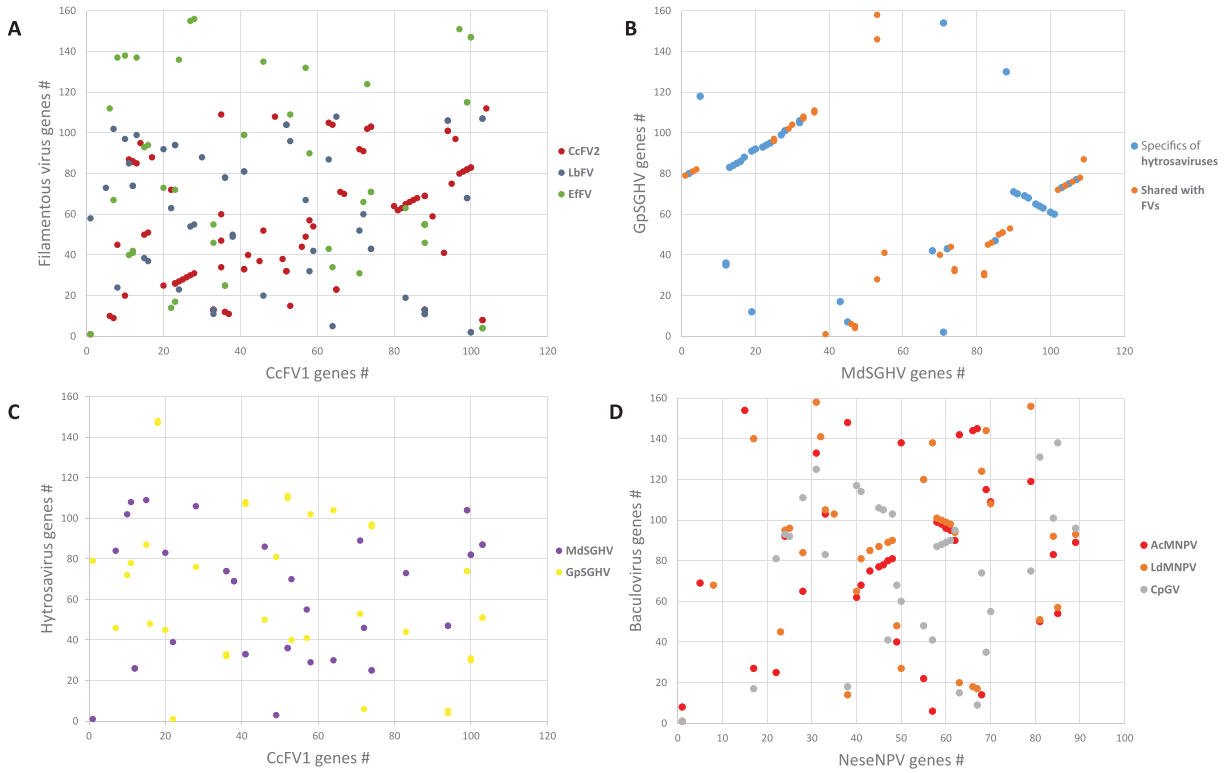

Figure 4. Gene-order conservation among *Naldaviricetes* assessed using gene parity plot comparisons. All genes are represented by dots following the order of the genes in the reference genome on the x-axis, and the positions of the homologs in the other genome on the y-axis. (A) Gene parity plot comparison of the circularized FV genomes. CcFV1 genome is set as reference and its gene order is compared to that of CcFV2 in red, LbFV in deep blue and EfFV in green. Orange boxes highlight microsyntenies present in all the FVs presented. (B) GpSGHV-Uga gene organization compared to that of MdSGHV. Genes specific to hytrosaviruses are represented in light blue and genes shared with FVs in orange. (C) Hytrosavirus gene order relative to CcFV1, with MdSGHV in purple and GpSGHV-Uga in yellow. (D) Gene-order comparison between a gammabaculovirus: NeseNPV (set as reference), two alphabaculoviruses: AcMNPV in red and LdMNPV in orange, and a betabaculovirus: CpGV in grey. Virus species abbreviations and number of genes (#) are those given in Supplementary Table S2.

possible that the FV reads detected in wild *Drosophila* are derived from a rare spillover from *L. heterotoma* to *D. melanogaster*, or from the presence of low-frequency endogenized versions of FV in the *Drosophila* genome (D. Obbard personal communication). In any case, it strongly suggests that the main driver of LhFV/DmFV infection is *L. heterotoma*, rather than *D. melanogaster*.

2.6 Synteny analysis of FVs and hytrosaviruses

Gene order is generally conserved among closely related viruses, whereas it breaks down with evolutionary distance and is limited to microsyntenies between viruses belonging to different genera (Herniou et al. 2003; Wang and Jehle 2009; Leobold et al. 2018). The synteny of the four circularized FV genomes was investigated using gene parity plot analysis (Goldbach et al. 1998). We found that gene order is weakly conserved among all FVs (Fig. 4 and Supplementary Table S7) apart from three microsyntenic regions, each consisting of two colinear FV core genes (*Ac38* and *lef-5*, *LbFV_orf92-like* and *LbFV_orf94-like*, and *LbFV_orf54* and *pif-3*). These microsyntenic regions are a specific feature of FVs (Fig. 4A and Fig. 4C) that differentiate them in particular from hytrosaviruses (in part, since some of these core genes are not present in hytrosaviruses).

Next, we compared the genome plasticity within the FVs to related viral clades with similar patristic distances (i.e., *Hytrosaviridae* and *Baculoviridae*). With three syntenic blocks of four to seven genes conserved a gene-order conservation between CcFV1 and CcFV2 (Fig. 4A, red dots for CcFV1/CcFV2) was very similar to that of the Baculoviruses, NeseNPV, and CpGV belonging to different genera (Fig. 4D, grey dots), while the two hytrosaviruses, GpSGHV-Uga and MdSGHV, showed higher synteny (Fig. 4B) (fragmentation score = 41.52, 38.21 and 17.09 for CcFV1/CcFV2, NeseNPV/CpGV, and GpSGHV/MdSGHV, respectively) suggesting a lower genome plasticity in hytrosaviruses. Lastly, the number of syntenic elements in FVs (Fig. 4A and Supplementary Table S6) is comparable to what is found between genera in Baculoviruses and Nudiviruses (Fig. 4D and Supplementary Table S7; (Leobold et al. 2018)) but far fewer than within the same genus of both viral families (Fig. 4B and Supplementary Table S7; (Herniou et al. 2003)). This suggests that the FV genomes, hereby described, could represent viral species from different genera.

2.7 FVs are preferentially associated with parasitoid wasps

All six species of FVs for which complete, or nearly complete genomes are available, do infect endoparasitoid wasps. However, determining the exact host range of FVs requires more extensive sampling. As most FV sequences available so far derive from host-genome sequencing, a data-mining approach was undertaken to search for FV-like sequences in all available genome assemblies (NCBI and BIPAA databases, November 2022) of hymenopteran ($n = 368$), dipteran ($n = 369$) and lepidopteran ($n = 911$) species. This sampling included free-living and parasitoid Hymenoptera, as

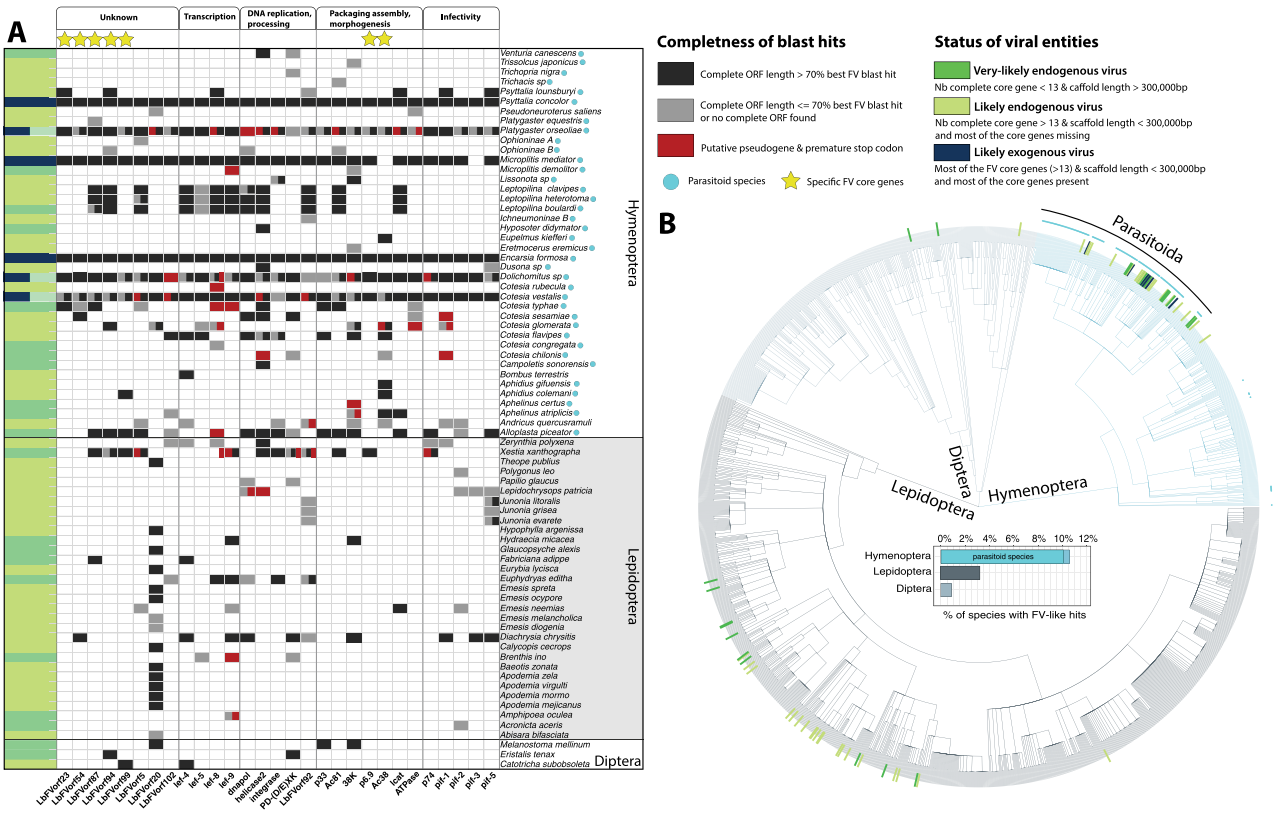

Figure 5. FV-like core sequences were detected in the genome assemblies of numerous Hymenoptera, Diptera, and Lepidoptera. (A) Table showing presence/absence of putative FV-like core genes in considered insect assemblies. A TBLASTN approach combined with alien index filtering and phylogenetic validation was used (note that a given species may be represented by several assemblies derived from different populations) (see also Supplementary Fig. S6 and S7 and Supplementary Table S8). The colors of the boxes represent the ORFs completeness status: black color represents a most probably complete ORF (that spans at least 70 per cent the size of the best FV BLAST hit), gray color represents a probably incomplete ORF (spanning less than 70 per cent the size of the best FV BLAST hit), red color represents a pseudogene or an ORF with a premature stop codon. More than one color in the same box indicates that the genome assembly of the species contains several copies with different status endogenous/exogenous as summarized in the left column. Several colors in one box indicate the probable presence of both endogenous and exogenous viral elements. The cyan circles along the taxon names indicate parasitoid species. (B) The circular phylogenetic tree shows the evolutionary relationships between the three insect orders: blue = Hymenoptera, light gray = Diptera, dark gray = Lepidoptera. Each colored dash along the leaves of the tree stands for the status of the FV elements (green = endogenous FV, dark blue = exogenous FV). The phylogenetic cladogram was reconstructed based on the taxonomical NCBI level of all genomes surveyed in this analysis using the NCBITaxa ete3 function in Python. The percentage of species with FV-like core sequences is displayed for each insect order in the inset.

well as Diptera and Lepidoptera, which are prey to numerous parasitoids and thus potentially exposed to FVs. We reasoned that mining genomic assemblies may enable us to identify new exogenous FVs, infecting the specimens used for the sequencing project, and/or putative endogenous viral elements (EVEs) derived from ancient FV infections.

To this end, *Naldaviricetes* homologs of the twenty-nine FV core genes (Supplementary Table S5) were first used to query the 2,815 genome assemblies representing 1,648 insect species. Each candidate sequence was then assessed using both an alien index (Supplementary Table S8, which includes only sequences that passed this first filter) and a phylogenetic validation (Supplementary Fig. S6 (1–29) and Supplementary Table S8). Finally, the endogenous or exogenous status of these sequences was determined mainly based not only on (1) the cumulative size of the identified scaffolds/contigs (which for an exogenous virus should not exceed the 300 kb size of FV genomes), but also on (2) the number of FV core genes detected (expected to be high in exogenous viruses due to their important function), and (3) the presence of viral pseudogenes and/or cellular genes within the identified scaffolds/contigs, which are both indicative of EVEs. The entire pipeline is schematized in Supplementary Fig. S7 and the validated candidate sequences are shown in Fig. 5A. In special cases, especially when the number of identified sequences exceeded a dozen, a detailed manual analysis was performed (Section 2.7.1).

2.7.1 *Detection of four additional exogenous FVs associated with parasitoid wasps*

In addition to the expected presence of the three exogenous FVs in the *E. formosa*, *P. concolor*, and *P. orseoliae* assemblies from which we obtained the EfFV, PcFV, and PoFV genomes, two FVs were identified in *Dolichomitus sp.* (Ichneumonidae: Pimplinae) and *Cotesia vestalis* (Braconidae: Microgastrinae; Section 2.7.2), and another one was strongly suspected in *Microplitis mediator* (Braconidae: Microgastrinae) a species in which a FV was previously described by microscopic analysis (Tanaka 1987) (Fig. 5A). From these three hymenopteran species, a complete or nearly complete set of FV core genes was detected and the contig properties were indeed consistent with the hypothesis that they are exogenous viruses. For *C. vestalis*, this result was also consistent with a previous analysis relying on a different pipeline (Burke, Hines, and Sharanowski

2021). In *C. flavipes* (Gauthier et al. 2021), 31 FV-like ORFs including ten core genes are clustered on five scaffolds with a cumulative size of ∼48 kb (Supplementary Fig. S8B). The high coding density (88.5 per cent), the absence of viral pseudogenes and of genes with eukaryotic architecture in these contigs altogether suggest that they belong to a fourth exogenous FV, which genome sequence is incomplete. In addition to these exogenous viruses, the presence of endogenous sequences was also clearly demonstrated for *Dolichomitus* sp., *P. orseoliae*, and *C. vestalis* (Fig. 5A), thus confirming previous reports (Burke, Hines, and Sharanowski 2021; Guinet et al. 2023).

2.7.2 The case of Cotesia species

In addition to the *C. congregata* assemblies from which CcFV1 and CcFV2 were obtained, several assemblies from wasps of the *Cotesia* genus were further analyzed, including *C. vestalis* and *C. flavipes*. In particular, for *C. vestalis* (diamondback moth parasitoid), several assemblies have been reported originating from three locations, Hangzhou (China; (Burke, Hines, and Sharanowski 2021; Guinet et al. 2023)), Andong (South Korea; (Burke, Hines, and Sharanowski 2021)), and Wageningen (the Netherlands; (Gauthier et al. 2021)).

FV-like sequences were detected in 31 contigs (ranging from ∼0.3 to 12.1 kb) of the Hangzhou isolate for a cumulative size of 117.4 kb. These contigs encoded the full set of FV core genes ($n = 29$), as well as 60 additional FV genes. They displayed high coding density (87 per cent) and neither viral pseudogenes nor cellular genes were detected (Supplementary Fig. S8A). Although this genome is highly fragmented, we confirm the previous report (Burke, Hines, and Sharanowski 2021) that an exogenous FV probably infects the Chinese *C. vestalis* isolate, and we found that its closest relative is CcFV2 (based on sequence similarity, synteny, and gene phylogenies, Supplementary Fig. S6 (1 to 29)).

In contrast, for the Korean isolate, only 22 FV core genes were detected on four long contigs ranging from 50 to 97 kb for a cumulative length exceeding usual FV genome size (308.1 kb; Supplementary Table S1). Moreover, some FV genes were pseudogenized (e.g., *DNApol* or *LbFV_orf102-like*), the gene density was relatively low (64–77 per cent) and the contigs contained cellular genes and/or transposable elements (Burke, Hines, and Sharanowski 2021). Thus, we confirm that the Korean *C. vestalis* isolate clearly harbors an endogenous FV, which is not fixed in the species.

At last, only seven FV core gene sequences could be identified in the Wageningen *C.vestalis* isolate assembly and all of them were highly pseudogenized, except for the *helicase 2* (scaffold_4924). Interestingly, this *helicase 2* copy is present in the genome of all three *C. vestalis* isolates (Hangzhou from LQNH01159677.1 and Andong from JZSA01002659.1), as well as in the genomes of *C. sesamiae* (scaffold_915), *C. flavipes* (scaffold_1343), *C. typhae* (JAAOIC020000016.1), and *C. chilonis* (RJVT01000058.1) but not of *C. congregata, C. glomerata*, and *C. rubecula*, which belong to a different clade within the *Cotesia* genus (Gauthier et al. 2021). Thus, this *helicase 2* copy probably originates from an ancient endogenous FV or from the horizontal transfer of a single gene that occurred in the last common ancestor of the *C. flavipes* clade (Supplementary Fig. S6 10). Of note, the *helicase 2* gene identified in the FV from *M. mediator* (see above) groups with the same clade; this *helicase 2* copy could thus have been captured by the *C. flavipes* clade common ancestor from a FV related to the one infecting *M. mediator*.

Overall, this thorough analysis in *C. vestalis* highlights that FV sequences in the same endoparasitoid species can have different status (exogenous/endogenous virus) and that integrated genes can originate from different endogenization and/or horizontal transfer events.

2.7.3 The genomes of hymenoptera parasitoids are enriched for filamentous EVEs

In total, FV-like core gene sequences were identified in seventy-four species. Analysis of the distribution of these FV-like sequences among Hymenoptera, Diptera, and Lepidoptera revealed a striking pattern. Homologs of FV core genes were identified in 10.9 per cent of the hymenopteran species ($n = 40/368$), 3.4 per cent of the lepidopteran species ($n = 31/911$), and 0.8 per cent of the dipteran species ($n = 3/369$) (Fig. 5B). Considering the number of species available for each order, this indicates that Hymenoptera are enriched for FV-EVEs ($\chi^2 = 49.07$, $df = 2$, $P = 2.215e\text{-}11$).

More specifically, within the Hymenoptera, a Phylogenetic Generalized Least Squares (PGLS) model considering phylogenetic autocorrelation revealed that the presence of endogenized FV sequences was strongly associated with the parasitoid lifestyle: out of the forty positive species detected in the whole Hymenoptera order, thirty-seven have a parasitoid lifestyle (PGLS results: $\beta = 0.174$, $SE = 0.032$, $t = 5.42$, $P < 0.001$). This strongly suggests that FVs are specialized on Hymenoptera having a parasitoid way of life. As an example, we can cite *Alloplasta piceator* (Ichneumonidae), a widespread endoparasitoid of Lepidoptera larvae, for which we identified 17 FV-like core genes, most of which are located in a ∼132 kb region integrated into the wasp chromosome 8 (Supplementary Table S8).

2.7.4 FVs also integrated into the genomes of some lepidoptera and diptera

Although less frequent, FV endogenization events were also evident in some Lepidoptera and Diptera genomes (Fig. 5A and Supplementary Table S8). In particular, the genomes of Lepidoptera *Diachrysia chrysitis* (the burnished brass moth, Noctuidae) and *Xestia xanthographa* (the square-spot rustic, Noctuidae) contained homologous sequences for a dozen FV-like core genes. In both moth species, several assemblies were available and FV-like sequences were usually clustered in 20–270 kb long regions, comprising repetitions of several sets of two–twelve FV-like genes and pseudogenes (Supplementary Fig. S8C and S8D). All these sequences are invariably flanked by transposable elements. Phylogenetic analyses suggest these lepidopteran FV-like sequences derive from the endogenization of a virus related to CcFV1 for *D. chrysitis* and to PcFV for *X. xanthographa* (Supplementary Fig. S6 (1–29)). Interestingly, the gene *LbFV_orf20* has recurrently been identified as a complete ORF in several Lepidoptera (Fig. 5A). This opens the possibility that it has been retained by selection, as described for other horizontally transferred *Lefavirales* genes found in some Lepidoptera (Gasmi et al. 2015, 2021; Di Lelio et al. 2019).

In conclusion, this data mining not only revealed abundant FV-like sequences in the genomes of Hymenoptera but also their presence in some Diptera and Lepidoptera. However, FVs were much more prevalent both as exogenous and endogenous viruses in parasitic Hymenoptera.

2.8 FV morphogenesis

To determine common features of virus particle production in FVs, transmission electron microscopy (TEM) was used to investigate how CcFV2 replicates in the ovaries of *C. congregata* relative to LbFV

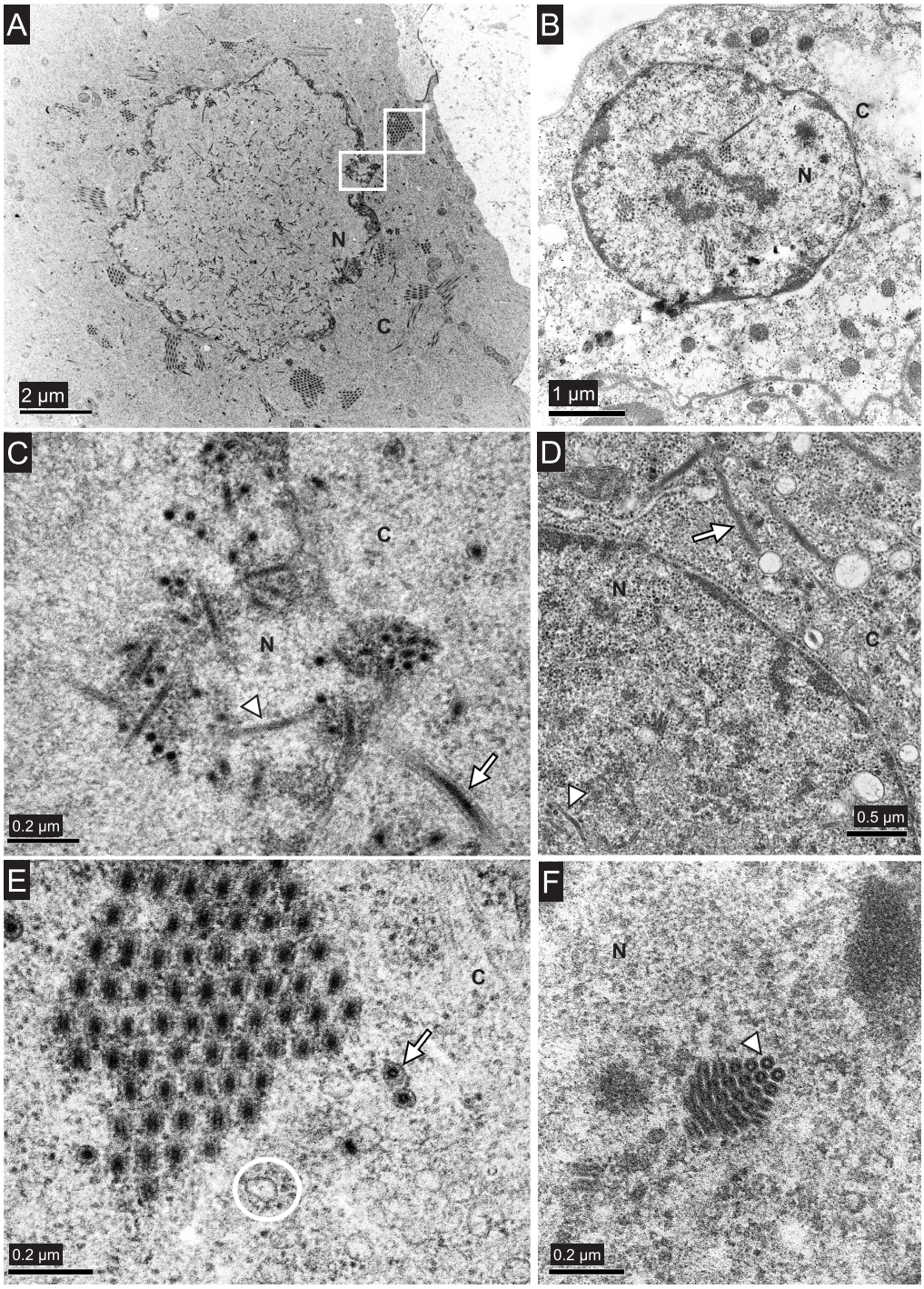

Figure 6. Typical structures observed by TEM in cells from two parasitoid wasps infected by FVs. (A, C, E) *Cotesia congregata* adult wasp calyx (CcC wasp line) and (B, D, F) *Leptopilina boulardi* adult wasp oviduct (B, D, F). (A, B) Details obtained using higher magnification of the views in the white frames TEM pictures of a calyx cell producing FVs, the white frames represent magnified sections in panels C and E. (C, D) Magnified TEM images of filamentous particles observed within *C. congregata* and *L. boulardi* cells, respectively. (E, F) TEM images of typical structures showing arrays filamentous particles in *C. congregata* and *L. boulardi*, respectively. N: nucleus and C: cytoplasm. The arrows designate enveloped particles, arrowheads nucleocapsids (in longitudinal section for C, D and cross-section for E, F), and white circle one of the abundant membrane vesicles observed in the cytoplasm.

(Fig. 6) (Varaldi et al. 2003, 2006) and was compared with the historical literature (Stoltz and Vinson 1979; Hamm, Styer, and Lewis 1990; de Buron and Beckage 1992).

Long virus-enveloped particles, similar to those of LbFV, were observed in abundance throughout the calyx of six *C. congregata* females (CcC wasp line, (Bredlau et al. 2019)) infected by CcFV2. CcFV2 producing cells were homogeneously distributed in the calyx sections and cohabited with bracovirus particle (CcBV) producing cells, unlike what was previously described by de Buron and Beckage 1992 for a FV from *C. congregata* (MsT wasp line), which replicates only in the upper calyx where no bracovirus is produced. Interestingly, no cells were observed producing FV and bracovirus particles simultaneously, suggesting a possible replication exclusion mechanism. However, as both viruses are

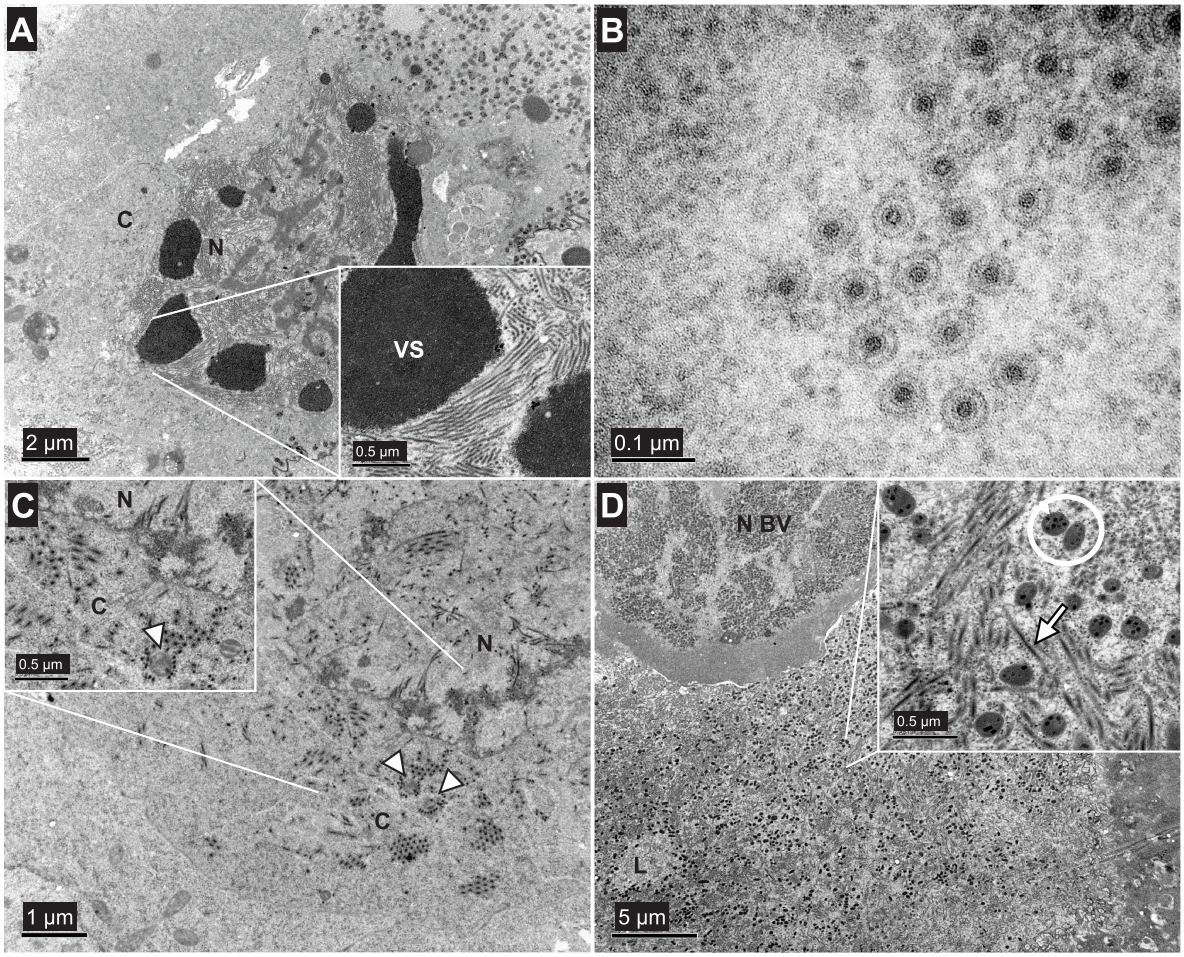

Figure 7. Electron microscope images showing filamentous particles in *C. congregata* adult wasp. (A) TEM photographs of a calyx cell producing large number of nucleocapsids, the inset at the bottom right represents a magnified section of the cell nucleus. (B) Image of enveloped filamentous particles obtained using high magnification. (C) Images of a calyx cell producing FVs, the inset at the top left represents a magnified section of the cell cytoplasm. (D) Images of wasp calyx lumen, the inset at the top right represents a magnified section of the lumen containing both FV and bracovirus particles. N: nucleus, C: cytoplasm, VS: virogenic stroma and N BV: nucleus containing bracovirus particles. White arrow: enveloped filamentous particle, white arrowheads: mitochondria lined with nucleocapsids, white circle: two bracovirus particles.

abundantly produced, the calyx lumen of *C. congregata* is filled with a mixture of both types of particles (Fig. 7D) suggesting both viruses are injected into the lepidopteran host during wasp oviposition. As described for LbFV (Varaldi et al. 2006), the replication of CcFV2 occurred in the nucleus where typical electron dense structure, the so-called virogenic stroma, was observed (Figs 6A–B and 7A). Long electron-dense, nonenveloped filament-shaped nucleocapsids were produced in the nuclei. Their structure is similar to that of LbFV nucleocapsids (Fig. 6A–D), although slightly shorter (up to 1 µm for LbFV and 0.8 µm long for CcFV2) and thinner (45 nm in diameter for LbFV and 20 nm for CcFV2).

The nucleocapsids are then released from the nucleus into the cytoplasm (Fig. 6C–D). The transport mechanism is unknown but is unlikely to involve budding like in Baculoviruses (Granados and Lawler 1981), as non-enveloped nucleocapsids (Fig. 6C) and abundant membrane vesicles are observed in the cytoplasm (Fig. 6E). Enveloped particles in the cytoplasm and calyx lumen have a diameter of 40–51 nm, compared to ~60 nm in LbFV (Figs 6 and 7). The particles of both CcFV2 and LbFV are often curved (Fig. 6C–D), and distinct from the rectilinear particles of hytrosaviruses (Abd-Alla et al. 2008). CcFV2 enveloped particles formed typical aggregates in the cytoplasm (Figs 6 and 7) in which they lined up with a distance of ~75 nm (Fig. 6E–F), as observed in previous electron microscopy studies (Krell 1987; Hamm, Styer, and Lewis 1990), whereas for LbFV nucleocapsid aggregates are observed in the nucleus (Fig. 6B and F) (Varaldi et al. 2006, 2024). As previously seen (de Buron and Beckage 1992; Kariithi et al. 2013), nucleocapsid lined the outer membranes of mitochondria near the nuclei in the cytoplasm (Fig. 7C), however, they do not appear to acquire their envelopes at these sites. In many cells, the presence of empty membrane vesicles near the CcFV2 nucleocapsid aggregates (Fig. 6E) suggests that the endoplasmic reticulum may be involved in the formation of the viral envelope. As extracellular viruses had no additional envelope (Fig. 7D), the passage of enveloped particles through the calyx lumen likely occurs by cell lysis (Fig. 7D). As CcFV2 enveloped particles spread through the lumen also contains wasp eggs, they are most likely transferred to the host Lepidoptera along with the eggs, which is also most likely the case for LbFV in *Drosophila*.

3. Discussion

3.1 Specific features of FVs

Filamentous-shaped viruses have been described in numerous species of parasitic wasps since the 1970s (Stoltz and Vinson 1977, 1979; Krell 1987; Styer, Hamm, and Nordlund 1987; Tanaka 1987;

Hamm, Styer, and Lewis 1990; de Buron and Beckage 1992; Hegazi et al. 2005). However, for lack of genomic data, it was neither possible to establish if they belonged to the same viral family, nor their relationships with other insect viruses. Moreover, their effects on the insects that host them remain largely unknown. Recently, a FV was discovered in the *Drosophila* parasitoid *L. boulardi* (LbFV). Phenotypically, this virus manipulates the superparasitism behavior of the wasp (Varaldi et al. 2003, 2005, 2006) and genome sequencing revealed this dsDNA virus belongs to the *Naldaviricetes* class. However, whether LbFV should be classified within the *Hytrosaviridae* or in a novel virus family was left as an open question (Lepetit et al. 2017). Here, the comparative analysis of six FV genomes provided conclusive insights into this question. Phylogenies showed that FVs from parasitoid wasps grouped as a strongly supported monophyletic clade within the *Lefavirales* and were distantly related to AmFV, the filamentous-shaped virus found in honeybees (Fig. 3 and Supplementary Fig. S4). Furthermore, our analysis revealed additional genomic (gene content, presence of *hrs*) as well as morphological features that distinguish FVs from *Hytrosaviridae*, supporting their classification in a new family, which we have provisionally named Filamentoviridae.

Thorough study of gene content showed that the Filamentoviridae share twenty-nine core genes, including all subunits of the PIF and transcriptional complexes that are the hallmarks of the order *Lefavirales* within the class *Naldaviricetes* (Fig. 2). In addition, the FVs contained five specific genes (homologs of *LbFV_orf23*, *LbFV_orf54*, *LbFV_orf87*, *LbFV_orf94*, and *LbFV_orf99*), as well as *Ac38* and *p6.9*, which are not found in all *Hytrosaviridae*. Similarly, to Baculoviruses (Herniou et al. 2003), FVs may comprise several genera, as suggested by phylogenetic distances and levels of collinearity between genomes. Additionally, electron microscopy analyses confirmed FVs replicate in cell nuclei which is a typical feature of the *Naldaviricetes* (Van Oers et al. 2023). Their large dsDNA genomes are enclosed within filamentous-shaped enveloped particles, which are large, flexible, and often curved, and thus distinguishable from the rectilinear particles of hytrosaviruses (Abd-Alla et al. 2008).

3.2 Parasitoid–FV association

Thus far, all Filamentoviridae, fully characterized at the genomic level, do infect endoparasitoid wasps. Extensive data mining of genomes assemblies from 1,648 species of Diptera, Lepidoptera, and Hymenoptera confirmed first that exogenous FV sequences are only found in endoparasitoid wasps (Fig. 5A), and secondly that endogenized FV genes were highly enriched in Hymenoptera genomes, and especially in endoparasitoids. These wasps, like most hymenopteran parasitoids, belong to the clade 'Parasitoida', which evolved around 230 million years ago in the Triassic era and diversified in the early Jurassic into seven hyperdiverse superfamilies (Peters et al. 2017; Tang et al. 2019). As FVs are as diverse and divergent as Baculoviruses and Nudiviruses (Fig. 3) and were found infecting Cynipoidea, Chalcidoidea, Platygastroidea, and Ichneumonoidea, they may also have diversified during the Jurassic era, like other *Naldaviricetes* (Thézé et al. 2011). Of note, the family of ascoviruses (*Ascoviridae*), large dsDNA viruses from the *Nucleocytoviricota* phylum (formerly known as NCLDV for Nucleocytoplasmic large DNA viruses), are also associated with endoparasitoids but unlike FVs they do not replicate in the wasp but use their stinging behavior for mechanical transmission of their particles to infect lepidopteran larvae, the hosts in which they replicate (Federici et al. 2009).

Parasitoid females inject eggs and fluids to protect them from the host immune response (Moreau and Asgari 2015). As FVs replicate in the same organs (oviduct cells and calyx cells in the ovaries) that produce these fluids, they are transferred together with the eggs and thus transmitted to subsequent generations of wasps, as has been found for LbFV (Varaldi et al. 2006) and for entomopoxviruses associated with endoparasitoids (Coffman, Hankinson, and Burke 2022). This transmission strategy might facilitate both vertical and horizontal transmission, as observed with LbFV, which manipulates the behavior of females to lay eggs in already parasitized hosts, thereby increasing horizontal transmission (Varaldi et al. 2003, 2006). This strategy has a high invasive power as LbFV may reach 90 per cent prevalence in some natural populations of *L. boulardi* (Patot et al. 2010).

The data we provided suggest that FVs can colonize parasitoids from four of the seven superfamilies in the Parasitoida clade (Ichneumonoidea, Cynipoidea, Chalcidoidea, and Platygastroidea) (Supplementary Table S8). This implies a broad distribution of FVs within the diverse world of parasitoid wasps (Forbes et al. 2018). We thus expect many more FVs to be described following sequencing efforts made on hymenopteran parasitoids. Accordingly, while we were writing this manuscript, a study describing a new FV in the endoparasitoid *Microctonus* (Braconidae) was released (Inwood et al. 2023). This wasp is used as a biocontrol agent in New Zealand to control the invasive pest weevil *Listronotus bonariensis*.The discovery of numerous EVEs within wasp genomes raises the question of their possible domestication, as viral exaptation is recurrent in endoparasitoids (Drezen et al. 2017; Gauthier, Drezen, and Herniou 2018; Guinet et al. 2023). Indeed, endoparasitoids acquire and domesticate genes deriving from dsDNA viruses more frequently than other hymenopterans (Guinet et al. 2023). Accordingly, all documented domestications have involved dsDNA viruses as donors, including nudiviruses (Bézier et al. 2009; Volkoff et al. 2010; Pichon et al. 2015; Burke et al. 2018). Recently, FV domestications have also been reported in *Leptopilina* parasitoids where thirteen EFVs (i.e., EVEs of FVs) allow the production of VLPs (Di Giovanni et al. 2020), and possibly also in *P. orseoliae* (Guinet et al. 2023), although further investigations are needed in this case.

EFVs have also been detected in the genomes of Diptera and Lepidoptera. Several hypotheses may explain the presence of EFVs in these orders. First, as many Diptera and Lepidoptera are attacked by parasitoids, FVs may have recurrently integrated their genomes through the host–parasite relationship without replication in the host. In the case of FV, we provide evidence that LbFV and LhFV viruses injected by the wasp can be detected in the adult host, which raises the possibility of horizontal transfer of FV genes that could reach the host germline. Although this idea would have been considered unrealistic a few years ago, several studies have now shown that host–parasitoid interactions may favor horizontal transmission (Schneider and Thomas 2014; Gasmi et al. 2015; Di Lelio et al. 2019, Muller et al. 2022; Heisserer et al. 2023). In many cases, these HGTs have been shown to be driven by the integration mechanism of these viruses into the caterpillar chromosomes during parasitism (Chevignon et al. 2018; Heisserer et al. 2023). Second, some FVs associated with endoparasitoid wasps can replicate within the parasitized host cells, as observed in certain lepidopteran species attacked by *C. marginiventris* infected by a virus-producing filamentous particles, where DNA sequence information is currently lacking (Styer, Hamm, and Nordlund 1987). If these hosts survive parasitoid infestation and viral infection, FV sequences may integrate into their germline, potentially transmitting to the next generation. Whatever the scenario, the presence of EFVs in Diptera and Lepidoptera prompts questions about their evolutionary fate following acquisition. Notably, many FV-derived

genes in these insects possess complete ORFs, suggesting potential functionality. Indeed, recent research in Lepidoptera (Di Lelio et al. 2019) has identified examples of gene exaptation, such as a viral-derived gene described as a 'parasitoid-killing factor' (Di Lelio et al. 2019). Therefore, future investigations combining molecular evolution and functional analyses are required to further elucidate the phenotypic impact of these FV-derived sequences in insect genomes.

3.3 Potential phenotypic effects of FVs

As parasitic wasps are extremely diverse, we believe that FVs are similarly diverse. However, so far there is very little data on their precise mode of transmission or their phenotypic effect, outside of the LbFV/*L.boulardi* interaction. In this system, viral transmission from mother to offspring is highly efficient (approximately 95 per cent) (Martinez, Fleury, and Varaldi 2015), and virus effects appear mostly limited to wasp oviposition behavior (Varaldi et al. 2006). This behavioral manipulation favors viral horizontal transmission between *L. boulardi* larvae sharing the same superparasitized host (Varaldi et al. 2003) and strongly reduces the competitiveness of infected wasp populations (Patot et al. 2012). Apart from that, the virus has a limited impact on other phenotypic traits (Varaldi et al. 2005) and confers a slight host-strain-dependent protective effect against the host immune response (Martinez et al. 2012).

Concerning CcFV1, it was sequenced from a *Cotesia* laboratory strain that had been originally established some 50 years ago (de Buron and Beckage 1992) and continuously maintained without refreshment from wild populations at least for the last 20 years. Interestingly, de Buron and Beckage already reported the presence of filamentous viral particles in female ovaries of this colony (de Buron and Beckage 1992). Moreover, PCRs assays on a series of tissues from the current *Cotesia* strain indicated that CcFV1 DNA was ubiquitous (see the GitHub, section Supplementary_data), suggesting that CcFV1 was maintained for numerous generations at no detectable cost to the wasp lab colony. As the common ancestor of LbFV and CcFV1 is at the root of FV clade, we can speculate that most FVs may similarly benefit from vertical transmission and impose low cost on their hosts, which would explain why, although common in parasitoid wasps, they often go unnoticed. Finally, an intriguing question is whether behavior manipulation is a general characteristic of all Filamentoviridae, as observed for LbFV and of note, the FV-specific core genes *LbFV_orf94* and *Ac38* are among the twenty LbFV genes that are potentially involved in the behavior manipulation induced by LbFV on *L. boulardi* (Varaldi and Lepetit 2018). Obviously, further phenotypic data and functional assays obtained on other wasps infected by FV will be necessary to evaluate this hypothesis.

3.4 Conclusion

Altogether, our results showed FVs share specific genomic, morphological, and evolutionary features and a probable tropism for hymenopteran host. These unique properties provide convincing evidence for assigning them as family rank (Simmonds et al. 2023), thereby supporting the creation of a fourth viral family in the *Lefavirales* order. Further research will be needed to fully explore the association between these viruses and their hosts to gain a better understanding of the impact of FVs in host/parasitoid interactions. This will be important both from a fundamental perspective, given the enormous diversity of Hymenoptera and their crucial ecological role, and from applied perspectives, as FVs may affect the efficacy of parasitic wasps used as biological control agents.

4. Materials and methods

4.1 Sampling

The six FVs characterized in this study are associated with five endoparasitoid wasp species belonging to four Hymenoptera families: *C. congregata* (Braconidae) which parasitizes Sphingidae, *Encarsia formosa* (Aphelinidae) which develops from whiteflies, *Psyttalia concolor* (Braconidae) that infests the olive fly (*Bactrocera oleae*), *Platygaster orseoliae* (Platygastridae) that parasitizes the cogongrass gall midge (*Orseolia javanica*) and *L. heterotoma* (Figitidae) which parasitizes *Drosophila* larvae. They were named as follows: *Cotesia congregata* FV 1 (CcFV1), *Cotesia congregata* FV 2 (CcFV2), *Encarsia formosa* FV (EfFV), *Psyttalia concolor* FV (PcFV), *Platygaster orseoliae* FV (PoFV), and *Leptopilina heterotoma* FV (LhFV). Among these viruses, only the LhFV partial genome was obtained after virus purification (protocol described in Varaldi et al. 2024 using lab-reared *L. heterotoma* iso-female lines originating from Igé, Burgundy, France). The others were obtained as by-products of wasp genomes sequencing projects (Supplementary Table S1).

Cotesia congregata FVs were identified from two *C. congregata* populations (Bredlau et al. 2019). CcFV1 infects a long-term-established laboratory strain (MsT wasps) reared on its natural host, *Manduca sexta* (Lepidoptera, Sphingidae), fed on artificial diet (de Buron and Beckage 1992). CcFV2 was obtained from a *C. congregata* strain collected in 2018 in Virginia (USA). This wasp strain, named CcC (Bredlau et al. 2019), develops specifically from the host *Ceratomia catalpae* (Lepidoptera, Sphingidae). MsT wasp DNA was extracted as described (Gauthier et al. 2021). *Ceratomia catalpae* caterpillars were collected from the field, and female wasps that emerged were used to parasitize a second set of caterpillars to produce all male broods. CcC wasp DNA was extracted from a single brood obtained from a virgin female using DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions. Several approaches were then used to sequence both wasp genomes: a 454/Illumina combined sequencing approach at the Genoscope platform (Evry, France) for MsT wasps (Gauthier et al. 2021) and a PacBio Sequel sequencing approach at University of Delaware sequencing laboratory (Newark, USA) for CcC wasps (Supplementary Table S1).

Psyttalia concolor DNA was extracted from a single female, while material for *P. orseoliae* and *E. formosa* was retrieved from a mix of dozens of individuals to obtain sufficient DNA, using the NucleoSpin Tissue extraction kit (Macherey-Nagel). The DNAs were used to construct Illumina TruSeq Nano DNA library sequenced as paired-end (2× 150 bp) at the GenoToul platform (Toulouse, France). In addition, long read sequencing was performed on *E. formosa* to obtain the complete circular EfFV genome. DNA extraction was performed on 100 individuals using the Blood and Cell Culture DNA mini kit (Qiagen). Sequencing was done using the MinION SQK-LSK109 protocol (Oxford Nanopore Technologies). A wash buffer was used after ligation of LFB adapters. The final elution was done in 30 µl with a final concentration of 90.4 ng/µl. Half of the solution was deposited on a nanopore FLO-MIN106 (R9.4.1) flow cell for sequencing.

4.2 Transmission electron microscopy

4.2.1 *Leptopilina boulardi*

Ovaries of *L. boulardi* superparaziting females (S strain described in (Varaldi et al. 2006) were fixed for 2 h in a 2 per cent glutaraldehyde solution (Agar Scientific) then postfixed for 1 h in a 2 per cent osmium tetroxide solution (Electron Microscopy Science), both prepared in cacodylate buffer (0.1 m sodium cacodylate-HCl, pH 7.4) (room temperature). Samples were further dehydrated in a

series of graded acetone solutions, embedded in Epon epoxy resin and incubated at 37°C for 24 h then at 60°C for 48 h for polymerization. Ultrathin sections (thickness = 70 nm) were cut using a LKB ultramicrotome (LKB-produkter AB) and stained with a 5 per cent uranyl acetate and Reynolds lead citrate. Finally, the sections were observed with an EM 10CR transmission microscope (Zeiss) at 80kV.

4.2.2 *Cotesia congregata*

Ovaries of *C. congregata* were fixed individually for 24 h in 2 per cent paraformaldehyde (Merck), 2 per cent glutaraldehyde (Agar Scientific) and 0.1 per cent sucrose in cacodylate buffer, washed 3 × 30 min in cacodylate buffer alone, then postfixed for 90 min in 2 per cent osmium tetroxide (Electron Microscopy Science) in cacodylate buffer and subjected to three final washes of 20 min (1× in cacodylate buffer alone and 2× in distillated water). Next, samples were dehydrated in a graded series of ethanol solutions (2× 50 per cent ethanol for 15 min; 2 × 70 per cent ethanol for 15 min and 1 × 70 per cent ethanol for an additional period of 14 h; 2 × 90 per cent ethanol for 20 min; and 3 × 100 per cent ethanol for 20 min). Final dehydration was carried out with 100 per cent propylene oxide (PrOx) (VWR) using three baths of 20 min each. Following, samples were incubated in PrOx/Epon epoxy resin (Fluka) mixtures in a 2:1 ratio for 2 h with closed caps, then in a 1:2 ratio for 2 h with closed caps and further 90 min with open caps, and finally in 100 per cent Epon for 16 h, at room temperature. Samples were finally transferred into a new 100 per cent Epon solution and incubated at 37°C for 24 h and at 60°C for 48 h for polymerization.

Both semi-thin (thickness = 800 nm) and ultrathin (thickness = 70 nm) sections were obtained using an Ultracut UCT ultramicrotome (Leica Microsystems GmbH). Semi-thin sections were first placed on glass, stained with Toluidine blue (Electron Microscopy Science) and embedded in Epon resin (Fluka), which was allowed to polymerize for 48 h at 60°C. They were then observed with an Eclipse 80i microscope connected to a DS-Vi1 camera driven by Nis-Element D 4.4 imaging software (Nikon). As for ultrathin sections, they were first placed on TEM one-slot Formvar® coated grids (Agar Scientific), stained 20 min with 5 per cent uranyl acetate (Electron Microscopy Science) then 5 min with Reynolds lead citrate. The sections were thereafter observed at 100 kV with a JEM-1011 TEM (Jeol) connected to a CMOS Gatan Rio 9 digital camera driven by DigitalMicrograph® software (Gatan).

4.3 Genome sequencing and assembly

The PcFV, PoFV, and EfFV paired-end reads were quality trimmed using fastq-mcf tool with the following parameters: -q15 -qual-mean 30 -D150 (https://github.com/ExpressionAnalysis/ea-utils) and assembled using IDBA-UD (Peng et al. 2012). Scaffolds were considered to belong to filamentous-like exogenous viruses if at least one homology to a LbFV ORF (min bit-score = 40) was detected and the scaffold coverage depth was significantly different from the coverage depth observed for BUSCO genes (analyses performed with BUSCO v3 based on the arthropod gene set) (Manni et al. 2021). This way, we were able to identify 14, 8, and 6 putative viral scaffolds from *P. orseoliae* (i.e., PoFV), *P. concolor* (i.e., PcFV), and *E. formosa* (i.e., EfFV) assemblies, respectively. The LhFV reads were assembled with Megahit v. 1.2.9 (Li et al. 2015).

The EfFV genome was further circularized using both MinION long reads (Oxford Nanopore Technologies) and the previously obtained Illumina short reads. Nanopore adapters were removed using Porechop v0.2.4 (https://github.com/rrwick/Porechop/tree/v0.2.4), read quality was controlled using NanoPlot v1.33.0 (https://github.com/wdecoster/NanoPlot) and poor-quality reads trimming was done using NanoFilt v2.6.0 (https://github.com/wdecoster/NanoPlot) (-q 12 -headcrop 75). The assembly was first performed with Flye v2.9-b1774 algorithm (https://github.com/fenderglass/Flye) (-meta -scaffolds) then polished with Racon v1.4.3 (https://github.com/isovic/racon) (two rounds of polishing using Minimap2 v2.24 for mapping) using raw reads from Illumina TruSeq Nano DNA library. This allowed to reduce the number of scaffolds from six to two. Next, we isolated all R1 and R2 Illumina reads as well as the trimmed Nanopore reads that had a convincing homology with one of these two contigs (mmseqs2 search search-type = 3, e-value max = $9e^{-13}$, db = reads, query = the two contigs) then we submitted them to the Unicycler pipeline v0.5.0 -b2d57cb (Wick et al. 2017) which generated a single contig assembly graph showing circular string graphs, thus revealing the circular DNA structure of the full-length EfFV genome.

For CcFV1 genome, Illumina paired-end and 454 (single-end and 3-8-20kb Mate-pairs) reads (available on ERS4256209) were filtered for quality and trimmed using CutAdapt v.3.5 (Martin 2011) (-q 20,20 -e 0.10 -max-n 0.5 -minimum-length 30). Additionally, 454 reads were trimmed for homopolymers with a minimum length of eight nucleotides using NGS QC Toolkit v2.3 (Patel, Jain, and Liu 2012) (see more details about read processing in (Gauthier et al. 2021)). CcFV2 genome was obtained following PacBio Sequel sequencing of *C. congregata catalpae* strain.

From *de novo* assembly of the corresponding *C. congregata* subspecies genome, several contigs (seeds) were identified showing convincing similarities with the previously sequenced LbFV. Complete FV genome reconstructions were then achieved by the mapping of selected reads. In detail, alignments of processed reads on host genomes and 'seeds' were performed using BLASR v5.3.5 (Chaisson and Tesler 2012) for PacBio reads and bowtie v2 2.3.5.1 (with 'local' parameters) (Langmead et al. 2009) for Illumina and 454 reads. Mapped reads on respective 'seeds' were then *de novo* assembled using Trycycler v0.5.1 (Wick et al. 2021) and Unicycler v0.4.8 (Wick et al. 2017) to obtain CcFV2 and CcFV1 sequence assemblies, respectively. Assembled reads were finally realigned and genome quality and circularity were assessed using Qualimap v2.2.2 (Okonechnikov, Conesa, and García-Alcalde 2015) and IGV (Robinson et al. 2011).

4.4 Sequence analyses and genome annotation

4.4.1 *Gene prediction*

Several tools were combined to predict genes. In a first step, a *de novo* prediction of all potential ORFs was performed under Geneious Prime v2019.2.3 (Kearse et al. 2012) starting from the first detected ATG codon to the STOP codon. Only ORFs with a length greater than or equal to 150 bp were kept. The inner ORFs were not considered. In a second step, complementary ORF predictions were performed using both VGAS v1 (parameters: -l = 1 -n -p -l 150) (Zhang et al. 2019) and Prodigal v2.6.3 (default parameters) (Hyatt et al. 2010). Only ORFs confirmed by the Geneious prediction tool and at least one of the two other predictive methods (Prodigal and VGAS generally both confirmed Geneious proposal) were considered.

In a few cases, considering alternative START codons, like TTG or CTG, allowed identification of either new (*lef-5* gene) or longer ORFs whose sequence was more consistent with those predicted from other viral species (e.g., *lef-9* and *lef-4* genes). The use of alternative codons has been confirmed, in particular in hytrosaviruses (Abd-Alla et al. 2016) and is suggested for some FV genes by the Prodigal predictions. Additionally, by in-depth manual inspection,

we identified ORFs corresponding to *p6.9* homologs. Indeed, this gene is short and usually hardly detected by classical tools as it is often located in low complexity regions.

ORF similarities were identified using BLASTP, BLASTX, and/or TBLASTN (Altschul 1997; Altschul et al. 2005) against either the NCBI non-redundant protein database (default parameters, except for the expected *e*-value threshold set to 1), the NCBI non-redundant protein database restricted to virus taxa (taxid:10239) or against a local database composed of either all the predicted ORFs or all the genomic sequences from the retained *Naldaviricetes* (Supplementary Table S2) using the BLAST tool from Geneious Prime v2019.2.3. Further similarities and ORF assignments were inferred based on domain search using either the Interproscan tool plugin from Geneious (https://www.ebi.ac.uk/interpro/ (Blum et al. 2021)) or probabilistic methods like HHpred or HMMER from the MPI Bioinformatics Toolkit of the Max Planck Institute (https://toolkit.tuebingen.mpg.de). Potential transmembrane regions were identified using the transmembrane prediction tool from Geneious Prime v2019.2.3.

When a homolog was apparently missing in a viral genome, a consensus sequence was generated based on the protein alignment of all the available viral homologs. A TBLASTN was then performed against the target genome(s) to further check for the presence of the corresponding ORF. Such strategy was applied not only for the FVs but also for AmFV and the hytrosavirus missing homologs (see below). Some potential homologs for which the alignment was less obvious could also be conserved based on the synteny observed with the neighboring ORFs (see paragraph below). Finally, ORFs were named according to the similarities encountered either with genes of known function, or with conserved protein domains, or with genes of unknown function already identified in LbFV. For all other ORFs they were named according to their position within the circularized genomes (virus species name_position) or within the contig or scaffold for non-circularized genomes (virus species name_scaffold/contig number_position).

4.4.2 *Identification of repeated sequences*

We performed two sets of analyses to identify repeated sequences within FV genomes, one to highlight putative large homologous repeated regions that may be conserved at the genus and/or family level, and the other to bring out direct repeats. A search for large homologous repeated regions was also performed on the two available hytrosavirus genomes.

First, a reciprocal BLASTN approach was performed to identify repeated sequences that may correspond to homologous repeat regions (*hrs*) using default parameters. Alignments were then mapped to the corresponding viral genome under Geneious Prime v2019.2.3 to precisely delineate each identified candidate *hrs*. *De novo* motif search and their location were then performed in pre-defined genomic regions. In detail, the sequences from the identified *hrs* were first used to find the main repeated motif (gapped) using GLAM2 (Gapped Local Alignment of Motifs; (Frith et al. 2008)). The parameters used (–2 -z (number of *hrs* present in the considered viral genome) -a 100 -b 300 -w 100) were optimized to find the longest possible motif in the set of repeated genomic regions. The motif with the highest score was then reciprocally mapped onto the viral genome using FIMO (find individual motif occurrences; (Grant, Bailey, and Noble 2011)) (-thresh 1e-9) to confirm their location. Palindromic sequences were then searched in the *hrs* region containing the previously identified motif using MEME (Multiple EM for Motif Elicitation; (Bailey et al. 2006)) from the MEME Suite v5.4.1 (https://meme-suite.org/meme/) optimized to identify the largest possible palindromes (-dna -mod oops -nmotifs 10 -minw 30 -maxw 300 -evt 0.1 -revcomp -pal). Finally, the structure of each *hrs* was determined by plotting back these palindromic motifs and comparing them with each other within the same genome. Such a global approach was further conducted with all available *hrs* trying to identify specific patterns (including palindromic sequences) retained by all FVs.

Secondly, the eTandem program (https://www.bioinformatics.nl/cgi-bin/emboss/etandem (Rice, Longden, and Bleasby 2000) was run to retrieve simple direct repeats with default parameters modified to retrieve large direct repeats (min repeat = 20, max repeat = 150). Repeated sequences located within previously annotated genes were discarded.

4.4.3 *Graphical representation*

All the sequencing raw data, the positions, and names of the predicted ORFs and repeated regions were plotted along each considered genome using the shinyCircos v2.0 web application (https://venyao.xyz/shinyCircos/). The coverage data for LbFV were retrieved from Lepetit and collaborators (Lepetit et al. 2017).

4.5 Insect dsDNA virus phylogenomic analysis

4.5.1 *Insect dsDNA virus sequence dataset*

In order to update the phylogeny of *Naldaviricetes*, we retrieved all predicted ORFs from a selected set of twenty-five dsDNA viruses including five *Baculoviridae*, seven *Nudiviridae*, three *Hytrosaviridae*, nine FVs (the six FVs described in this study, the two previously published (LbFV and DmFV) to which we also added the distant AmFV), and a *Nimaviridae* used as an outgroup (Supplementary Table S2). The third used *Hytrosaviridae* was the partially sequenced Drosophila-associated salivary gland hypertrophy virus (here named DmSGHV) (Wallace et al. 2021) for which we tried to obtain the longest possible core gene sequences. To do so, we first retrieved the 18 contigs available for DmSGHV (MT469997 to MT470014) and the corresponding *D. melanogaster* SRA library (UA_Ode_16_47, SRR8494427). Then we used a TBLASTN strategy with the MdSGHV proteins as a query (ACD03460.1 to ACD03567.1) to improve and elongate step-by-step candidate contigs and genes of interest.

4.5.2 *Filamentoviridae core gene identification*

Whatever the considered scale (class, order, or family), virus core genes corresponded to genes shared by all the members of the same taxon and are therefore a decisive criterion for determining membership of this taxon. They generally have a common phylogenetic origin and encode essential functions of the virus cycle like replication, transcription or virus assembly, morphogenesis, and/or infectivity. Previously identified FVs were described as having the hytrosaviruses as their closest, however distant, relatives (Lepetit et al. 2017; Wallace et al. 2021; Yang et al. 2022). Core gene sets have already been established not only at the level of the *Naldaviricetes*, the *Lefavirales* but also at the level of each family composing them (*Baculoviridae, Nudiviridae*, and *Hytrosaviridae*). These data were thus used as a basis for identifying the specific core genes of our FVs.

In the first instance, based on the same sequence similarity strategy used to annotate genes (Section 4.4), we identified the genes that were homologous to the core gene set defined for the *Lefavirales* ((Walker et al. 2021); see https://ictv.global/taxonomy/taxondetails?taxnode_id=202209115). We further looked at conserved genes present in all the FV genomes whether they are restricted to the FV or shared with their closest relatives, the

hytrosaviruses. The absence of a gene in a partial genome was never considered to be a hindrance, given that it was present in all complete genomes and a majority of the partial genomes. On an ad hoc basis, when a core sequence appeared to be missing in one or more of the twenty-five dsDNA viruses, we conducted a thorough search using HMMER tools. Such analyses allowed the identification of previously unnoticed putative homologs not only in hytrosaviruses but also in AmFV, in particular 38 K, P6.9, LEF-5, LEF-8, and LEF-9 (Supplementary Table S5). For each of the core gene homologs, sequence alignments were performed and homology relevance was confirmed by expert review.

4.5.3 *Clustering of homologous sequences*

Several tools were combined (Supplementary Fig. S9) to cluster homologous sequences to generate a protein supermatrix for phylogenomic analyses. As a first step, clusters were established by gathering all selected *Naldaviricetes* (Supplementary Table S2) and FV protein sequences aligned with the MMseqs2 search (Steinegger and Söding 2017, 2018). Then, all sequences with a bit-score ≥ 40 were clustered together using a homemade python script available on GitHub (see 'Data availability' section). Following this, a HMMER profile v3.3.2 was created for each cluster based on amino acid alignments (Clustal-Omega v1.2.4, with default parameters; (Sievers et al. 2011)). The resulting profiles were then used as queries in a new round of HMMER search to detect possible connections between first-round clusters. When clusters displayed enough proximity (i.e., *e*-value $\geq$ 9e-05, value optimized to obtain the complete expected clusters with the twenty-nine FV core gene set previously defined), the corresponding clusters were pulled together. During this process, we ensured that each protein did not appear in several clusters and that, non-homologous proteins sharing small conserved motifs were not merged into the same cluster. In each case where misclustering was evident (e.g., PIF-3 and PIF-1), the analysis was repeated using more stringent clustering parameters (bit-score $\geq$ 100). When paralogs were identified in a single species, only the one with the lowest *e*-value compared to the LbFV homolog was kept for further analysis. The cluster composition and the distribution of the homologs in each virus strain are given as Supplementary material (Supplementary Table S9 and Supplementary Fig. S10).

4.5.4 *Tree reconstruction parameters*

Two distinct datasets were processed independently according to two approaches, the one designated as a 'core genes' approach, which might better reflect the evolutionary history of viruses (Simmonds et al. 2023) and the other as an 'all genes' approach. In the first one (i.e., 'core-genes'), only the twenty-nine previously defined core genes were used, while in the 'all genes' approach, all clusters with at least four out of the twenty-five considered viral species were retained. In both approaches, protein sequences were aligned with Clustal-Omega v1.2.4 (Sievers et al. 2011) and trimmed with Trimal v1.4.rev22 (option parameter: automated1) (Capella-Gutierrez, Silla-Martinez, and Gabaldon 2009) within each cluster. Alignments were then concatenated to generate a supermatrix with various partitions. A maximal symmetry test (Naser-Khdour et al. 2019) was used to eliminate heterogeneous partitions (-test-remove-bad option, P-value cutoff 0.05). Then phylogenetic trees were inferred using a maximum-likelihood framework (ML) and ModelFinder (Chernomor, von Haeseler, and Minh 2016; Kalyaanamoorthy et al. 2017) to identify models for each ORF partition (-MFP option), both implemented in IQ-TREE software v2.1.2 (Minh et al. 2020). The edge-linked partitioned model (-spp option), which allows each gene to have its own evolutionary rate, was chosen for tree reconstructions. Ultrafast bootstrap (Hoang et al. 2018) and SH-aLRT (options -bb 5,000 and -alrt 5,000) were computed to examine node supports for focal relationships using the ML method. Additionally, to reduce the risk of overestimating branch supports due to severe model violations, we used the command '-bnni' from IQ-TREE. In addition, a Mixed-model Bayesian phylogenomic analysis was run with MrBayes (Ronquist et al. 2012) on the twenty-nine 'core gene' partitions, using the best model schemes previously computed with IQ-TREE v2. For 1,000,000 generations, the Bayesian analysis was performed using four independent Monte Carlo Markov chains. Convergence of the chains was considered to have been achieved when the average standard deviation of the split frequencies was less than 0.01 and was validated by plotting the log-likelihood values against generation times in Tracer v1.7.2 (Rambaut et al. 2018). For all parameters, the effective sample size was greater than 100. The posterior probability consensus tree was obtained by integrating the results of the repeated analyses and discarding the first 25 per cent of sampled trees using relative burning. Following stabilization, every 1,000th tree was sampled to determine a majority rule consensus tree. FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize and root the consensus tree.

4.5.5 *Patristic distance calculation*

The patristic distances between the different taxa within and between viral families were performed from the phylogram of the 'core genes' phylogeny using the cophenetic.phylo function implemented in the APE (Analyses of Phylogenetics and Evolution) R package (Paradis, Claude, and Strimmer 2004). Graphics were then generated using the ggplot2 package (Wickham 2016).

4.6 Synteny analyses

Homologous relationships between genes were either obtained from previously published viral genomes (i.e., hytrosaviruses and LbFV) or inferred as described above. Duplicated genes were assigned to a single gene in the reference genome. Genes were compiled in a correspondence table using Microsoft Office Excel in order to compare positions between viruses. Then gene parity plots were drawn from this table. For fragmented viral genomes, contigs were arbitrarily ordered for the analysis.

From the gene plots, we designed scores to estimate a gene order conservation/fragmentation level. The Synteny score indicates the level of order conservation (Sc, interval:]0,1]) and reciprocally the Fragmented score (Fc, interval: [1, $+\infty$[) indicates the fragmentation level. Such scores take into consideration the number of homologous genes (Hn), the number of genes found in synteny (NgS) and the average number of genes found in syntenic blocks (AgSB). The syntenic blocks were defined from the tables of gene plots by allowing jump and/or inversion of two genes for the different series. The formulas for calculating the scores are shown below:

$$Sc = \frac{NgS \cdot AgSB}{(Hn)^2} \qquad Fc = \frac{1}{Sc}$$

4.7 Mining databases for FV-like sequence identification in genome assemblies

A search was conducted to identify exogenous or endogenous FV-like sequences in publicly available insect genome assemblies. We adopted a TBLASTN approach implemented in the MMseqs2

search program (Steinegger and Söding 2017) combined with systematic phylogenetic clade validations (Supplementary Figs. S6 and S11 (1–29)). The database was composed of 2,815 genome assemblies obtained from 368 Hymenoptera, 911 Lepidoptera and 369 Diptera species (downloaded from NCBI and BIPAA databases on 5 November 2022), complemented with parasitoid genomes published in a previous recent study (Guinet et al. 2023). The query corresponded to the set of twenty-nine core proteins detected in the genomes of the twenty-five *Naldaviricetes* used in the main phylogeny (Section 4.5). Only hits with a bit-score ≥ 50 and with more than 25 per cent sequence alignment coverage (to avoid identification of small domain homologies) were retained.

We characterized candidate sequences by estimating their 'filamentous nature' using an alien index. This index quantifies the difference in BLAST *e*-values between the best filamentous and non-filamentous results. The index formula is: Index = log10(Best *e*-value for non-filamentous)—log10(Best *e*-value for FV). A positive index suggests a stronger genetic proximity to FV species, indicating likely FV species affiliation. To establish the index threshold, we computed it for thirteen genes endogenized in three *Leptopilina* species (Di Giovanni et al. 2020) and used a minimum Index value from this distribution as a threshold to identify candidate sequences associated with Filamentoviridae. Details about these sequences can be found in Supplementary Table S8. To validate the sequence origin, phylogenetic analysis was performed, including insect loci candidates and the *Naldaviricetes* sequences used as initial baits (Supplementary Table S5). Sequences were assigned to FV as soon as they branched within a FV clade. A small number of sequences (<3 per cent) for which the phylogenetic signal was unclear were however assigned to FV when a majority of FV-assigned sequences were detected in their vicinity. The phylogenies were conducted as described previously (Section 4.5). Supporting gene phylogenies are available in Supplementary Fig. S6 (1–29).

Several metrics were chosen to further determine the endogenous or exogenous nature of the sequences identified in an insect assembly including the identified number of complete FV-like core ORFs (defined as a sequence of at least 150 nucleotides that includes a start and a stop codon) and the cumulative size of the scaffolds carrying these ORFs. Indeed, a low number of complete core ORFs (compared to the full set of core genes defining the family), the presence of degraded ORFs (having already undergone pseudogenization), or a cumulative size exceeding the usual size of virus genomes, would tend to favor the endogenous virus hypothesis. On the contrary, the sequences were assigned to a probable exogenous FV when at least 14 complete ORFs (approximately half the number of FV core genes) with complete sequences (at least 70 per cent of the size of the most similar FV ORF protein identified by blast analysis) without premature stop codons were identified on scaffolds whose cumulative size did not exceed 300,000 bp, which is the maximum size so far expected for an exogenous FV genome (see Supplementary Table S1 for details). Genomic sequence homologs of FV genes were considered as pseudogenized when the translation obtained from the locus identified by TBLASTN analysis contained at least one stop codons as deduced from translation of the nucleotide sequence.

To test the association between parasitoid lifestyle and endogenization of FVs, we performed a PGLS analysis within the Hymenoptera order. To conduct this analysis, we employed the 'geiger' R package under a ML framework (Pennell et al. 2014). The presence or absence of FV sequences within assemblies ('Status' variable) was studied this way with the variable called 'lifestyle' categorized as either parasitoid or free-living as the explanatory variable. The analysis was conducted assuming that the error structure follows a Brownian model.

4.8. SRA analysis

Taxonomic assignment of *Drosophila* short-read sequences to LbFV genome were queried from the full SRA-STAT database provided by SRA (November 2021 release, https://www.ncbi.nlm.nih.gov/sra/docs/sra-cloud-based-examples/ (Katz et al. 2021)) using python pandas and pyarrow in memory data analytics libraries.

4.9. Horizontal gene transfer in FVs

To find filamentous viral ORFs acquired from non-viral organism through HGT, we searched for sequence similarities between FV predicted ORFs and the NCBI non-redundant (nr) database (mmseqs2 search: query: all the ORFs, db: NR, using a threshold a bit-score minimum of 50). This BLASTP allowed us to gather multiple information for each ORF, such as the proportion of hits from viral origin compared to the proportion of eukaryotic, archaea, or bacterial origins. This allowed us to assign each ORF to a category: either likely 'viral' (if nb viral hits/nb non-viral hits > 1), 'uncertain' (if nb viral hits/nb non-viral hits > 1 but more than 5 non-viral hits where found), or likely 'non-viral' (if nb viral hits/nb non-viral hits < 1). From this, in order to find typical ORFs transferred into viral genomes, we assigned each cluster to a category depending on the composition of the cluster: viral_cluster (if nb 'viral' ORF/nb 'non-viral' ORFs > 1), uncertain_cluster (if nb 'viral' ORFs/nb 'non-viral' ORFs > 1 but more than five non-viral ORFs where found), or non-viral_cluster (if nb 'viral' ORFs/nb 'non-viral' ORFs < 1). We then kept cluster annotated as non-viral cluster for further phylogenetic analyses. Phylogenetic analysis was done after the protein alignment using Clustal-Omega v 1.2.4 (Sievers et al. 2011) (default parameters) followed by the inference of the trees using IQ-TREE (v 2.1.2) (Minh et al. 2020) (-m MFP -alrt 1,000). Finally, alignments and phylogenies were conducted as previously described (Section 4.5). Each phylogeny was analyzed by eye to infer the direction of the transfer. All supporting gene phylogenies are available in Supplementary Fig. S3 (A to W).

Data availability

All scripts used in this paper can be found in the GitHub page: https://github.com/BenjaminGuinet/Filamentous_viral_family_project. All viral genome assemblies as well as raw data and genome annotations can be found under the NCBI Bioproject PRJNA964713 and the ENA Bioproject PRJEB62786 (all informations are found in the Supplementary Table S1).

Supplementary data

Supplementary data is available at *Virus Evolution Journal* online.

Acknowledgements

We thank Damien de Vienne for his help in building the phylogeny presented in Fig. 5B, Darren Obbard for sharing data and thoughts on DmFV, Mohamed Amine Chebbi for his work performed at Viroscan to complete CcV1 and CcV2 genome assemblies, and the reviewers for helpful comments. We thank the DTAMB platform of the FR-BioEnvis (Lyon 1) for sequencing EfFV and Monique Van Oers for suggesting the name for this new viral family. This work was in part performed using the computing facilities of the CC LBBE/PRABI and the PST ASB.

Funding

The sequencing of EfFV, PcFV, PoFV, and LhFV was funded by the French National Research Agency (ANR 11-JSV7-0011 Viromics, ANR 17-CE02-0021 Horizon) and the FR-BioEnvis (LOWINSEQ) to J.Varaldi. *Cotesia congregata* MsT genomes sequencing was funded by French National Research Agency ANR (ABC Papogen project ANR-12-ADAP-0001 and CoteBio ANR17-CE32-0015-02 to L.Kaiser). The *C. congregata* CcC genome sequencing was funded by USDA-ARS Cooperative Agreement (Project No. 8042–22000-291-04-S to K.Kester). The electron microscopy approach in *C. congregata* CcC was funded by the Fédération de Recherche en Infectiologie (FéRI to M.Leobold).

Conflict of interest: None declared.

References

Abd-Alla, A. M. M. et al. (2008) 'Genome Analysis of a Glossina *Pallidipes* Salivary Gland Hypertrophy Virus Reveals a Novel, Large, Double-Stranded Circular DNA Virus', *Journal of Virology*, 82: 4595–611.

Abd-Alla, A. M. M. et al. (2009) '*Hytrosaviridae*: A Proposal for Classification and Nomenclature of a New Insect Virus Family', *Archives of Virology*, 154: 909–18.

Abd-Alla, A. M. M. et al. (2013) 'Improving Sterile Insect Technique (SIT) for Tsetse Flies through Research on Their Symbionts and Pathogens', *Journal of Invertebrate Pathology*, 112: S2–S10.

Abd-Alla, A. M. M. et al. (2016) 'Comprehensive Annotation of Glossina *Pallidipes* Salivary Gland Hypertrophy Virus from Ethiopian Tsetse Flies: A Proteogenomics Approach', *Journal of General Virology*, 97: 1010–31.

Ahmed, Y. L. et al. (2019) 'Crystal Structures of Rea1-MIDAS Bound to Its Ribosome Assembly Factor Ligands Resembling Integrin–ligand-type Complexes', *Nature Communications*, 10: 3050.

Altschul, S. (1997) 'Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs', *Nucleic Acids Research*, 25: 3389–402.

Altschul, S. F. et al. (2005) 'Protein Database Searches Using Compositionally Adjusted Substitution Matrices', *The FEBS Journal*, 272: 5101–9.

Ardisson-Araújo, D. M. P. et al. (2016) 'A Betabaculovirus Encoding A gp64 Homolog', *BMC Genomics*, 17: 94.

Bailey, T. L. et al. (2006) 'MEME: Discovering and Analyzing DNA and Protein Sequence Motifs', *Nucleic Acids Research*, 34: W369–W373.

Bateman, K. S. et al. (2021) 'Identification and Full Characterisation of Two Novel Crustacean Infecting Members of the Family *Nudiviridae* Provides Support for Two Subfamilies', *Viruses*, 13: 1694.

Bézier, A. et al. (2009) 'Polydnaviruses of Braconid Wasps Derive from an Ancestral Nudivirus', *Science*, 323: 926–30.

Bézier, A. et al. (2009) 'Polydnavirus Hidden Face: The Genes Producing Virus Particles of Parasitic Wasps', *Journal of Invertebrate Pathology*, 101: 194–203.

Bideshi, D. K. et al. (2003) 'Phylogenetic Analysis and Possible Function of Bro-like Genes, a Multigene Family Widespread among Large Double-stranded DNA Viruses of Invertebrates and Bacteria', *Journal of General Virology*, 84: 2531–44.

Blissard, G. W., and Theilmann, D. A. (2018) 'Baculovirus Entry and Egress from Insect Cells', *Annual Review of Virology*, 5: 113–39.

Blum, M. et al. (2021) 'The InterPro Protein Families and Domains Database: 20 Years On', *Nucleic Acids Research*, 49: D344–D354.

Boogaard, B., Van Oers, M., and Van Lent, J. (2018) 'An Advanced View on Baculovirus *per Os* Infectivity Factors', *Insects*, 9: 84.

Bredlau, J. P. et al. (2019) 'The Parasitic Wasp, *Cotesia congregata* (Say), Consists of Two Incipient Species Isolated by Asymmetric Reproductive Incompatibility and Hybrid Inability to Overcome Host Defenses', *Frontiers in Ecology and Evolution*, 7: 187.

Burand, J. P. et al. (2012) 'Analysis of the Genome of the Sexually Transmitted Insect Virus Helicoverpa zea Nudivirus 2', *Viruses*, 4: 28–61.

Burke, G. R. et al. (2013) 'Mutualistic Polydnaviruses Share Essential Replication Gene Functions with Pathogenic Ancestors', *PLoS Pathogens*, 9: e1003348.

Burke, G. R. et al. (2018) 'Rapid Viral Symbiogenesis via Changes in Parasitoid Wasp Genome Architecture', *Molecular Biology and Evolution*, 35: 2463–74.

Burke, G. R., Hines, H. M., and Sharanowski, B. J. (2021) 'The Presence of Ancient Core Genes Reveals Endogenization from Diverse Viral Ancestors in Parasitoid Wasps', *Genome Biology and Evolution*, 13: evab105.

Burke, G. R., and Strand, M. R. (2014) 'Systematic Analysis of a Wasp Parasitism Arsenal', *Molecular Ecology*, 23: 890–901.

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009) 'trimAl: A Tool for Automated Alignment Trimming in Large-scale Phylogenetic Analyses', *Bioinformatics*, 25: 1972–3.

Chaisson, M. J., and Tesler, G. (2012) 'Mapping Single Molecule Sequencing Reads Using Basic Local Alignment with Successive Refinement (BLASR): Application and Theory', *BMC Bioinformatics*, 13: 238.

Chernomor, O., von Haeseler, A., and Minh, B. Q. (2016) 'Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices', *Systematic Biology*, 65: 997–1008.

Chevignon, G. et al. (2018) 'Cotesia congregata Bracovirus Circles Encoding PTP and Ankyrin Genes Integrate into the DNA of Parasitized *Manduca sexta* Hemocytes', *Journal of Virology*, 92: e00438–18.

Coffman, K. A., Hankinson, Q. M., and Burke, G. R. (2022) 'A Viral Mutualist Employs Posthatch Transmission for Vertical and Horizontal Spread among Parasitoid Wasps', *Proceedings of the National Academy of Sciences*, 119: e2120048119.

Colson, P. et al. (2011) 'Viruses with More than 1,000 Genes: Mamavirus, a New Acanthamoeba polyphagamimivirus Strain, and Reannotation of Mimivirus Genes', *Genome Biology and Evolution*, 3: 737–42.

Correa, R. et al. (2013) 'The Role of F-Box Proteins during Viral Infection', *International Journal of Molecular Sciences*, 14: 4030–49.

de Buron, I., and Beckage, N. E. (1992) 'Characterization of a Polydnavirus (PDV) and Virus-like Filamentous Particle (VLFP) in the Braconid Wasp *Cotesia congregata* (Hymenoptera: Braconidae)', *Journal of Invertebrate Pathology*, 59: 315–27.

Di Giovanni, D. et al. (2020) 'A Behavior-manipulating Virus Relative as A Source of Adaptive Genes for *Drosophila* Parasitoids', *Molecular Biology and Evolution*, 37: 2791–807.

Di Lelio, I. et al. (2019) 'Evolution of an Insect Immune Barrier through Horizontal Gene Transfer Mediated by a Parasitic Wasp', *PLoS Genetics*, 15: e1007998.

Drezen, J. M. et al. (2017) 'Impact of Lateral Transfers on the Genomes of Lepidoptera', *Genes*, 8: 315.

Drezen, J. M. et al. (2022) 'Bracoviruses, Ichnoviruses, and Virus-like Particles from Parasitoid Wasps Retain Many Features of Their Virus Ancestors', *Current Opinion in Insect Science*, 49: 93–100.

Edgar, R. C. et al. (2022) 'Petabase-scale Sequence Alignment Catalyses Viral Discovery', *Nature*, 602: 142–7.

Federici, B. A. et al. (2009) 'Ascoviruses: Superb Manipulators of Apoptosis for Viral Replication and Transmission'. In: Compans, R. W. et al. (eds) *Lesser Known Large dsDNA Viruses*, Vol. 328, pp.

171–96. Springer: Berlin, Heidelberg. Series Title: Current Topics in Microbiology and Immunology.

Fellowes, M. D. E., Kraaijeveld, A. R., and Godfray, H. C. J. (1998) 'Trade–off Associated with Selection for Increased Ability to Resist Parasitoid Attack in *Drosophila melanogaster*', *Proceedings of the Royal Society of London B*, 265: 1553–8.

Filée, J., Siguier, P., and Chandler, M. (2007) 'I Am What I Eat and I Eat What I Am: Acquisition of Bacterial Genes by Giant Viruses', *Trends in Genetics*, 23: 10–5.

Flatt, J. W., and Greber, U. F. (2017) 'Viral Mechanisms for Docking and Delivering at Nuclear Pore Complexes', *Seminars in Cell and Developmental Biology*, 68: 59–71.

Fleury, F. et al. (2009) 'Chapter 1 Ecology and Life History Evolution of Frugivorous *Drosophila* Parasitoids Genevieve Prevost'. In: *Advances in Parasitology*, Vol. 70, pp. 3–44. Elsevier.

Forbes, A. A. et al. (2018) 'Quantifying the Unquantifiable: Why Hymenoptera, Not Coleoptera, Is the Most Speciose Animal Order', *BMC Ecology*, 18: 21.

Francino, M. P. (2005) 'An Adaptive Radiation Model for the Origin of New Gene Functions', *Nature Genetics*, 37: 573–8.

Frith, M. C. et al. (2008) 'Discovering Sequence Motifs with Arbitrary Insertions and Deletions', *PLoS Computational Biology*, 4: e1000071.

Gabler, F. et al. (2020) 'Protein Sequence Analysis Using the MPI Bioinformatics Toolkit', *Current Protocols in Bioinformatics*, 72: e108.

Gandon, S., Rivero, A., and Varaldi, J. (2006) 'Superparasitism Evolution: Adaptation or Manipulation?', *The American Naturalist*, 167: E1–E22.

Garcia-Maruniak, A. et al. (2008) 'Sequence Analysis of a Non-classified, Non-occluded DNA Virus that Causes Salivary Gland Hypertrophy of *Musca domestica*', *MdSGHV. Virology*, 377: 184–96.

Garcia-Maruniak, A. et al. (2009) 'Two Viruses that Cause Salivary Gland Hypertrophy in *Glossina pallidipes* and *Musca domestica* are Related and Form a Distinct Phylogenetic Clade', *Journal of General Virology*, 90: 334–46.

Gasmi, L. et al. (2015) 'Recurrent Domestication by Lepidoptera of Genes from Their Parasites Mediated by Bracoviruses', *PLoS Genetics*, 11: e1005470.

Gasmi, L. et al. (2021) 'Horizontally Transmitted Parasitoid Killing Factor Shapes Insect Defense to Parasitoids', *Science*, 373: 535–41.

Gauthier, L. et al. (2015) 'The Apis mellifera Filamentous Virus Genome', *Viruses*, 7: 3798–815.

Gauthier, J. et al. (2021) 'Chromosomal Scale Assembly of Parasitic Wasp Genome Reveals Symbiotic Virus Colonization', *Communications Biology*, 4: 104.

Gauthier, J., Drezen, J. M., and Herniou, E. A. (2018) 'The Recurrent Domestication of Viruses: Major Evolutionary Transitions in Parasitic Wasps', *Parasitology*, 145: 713–23.

Ge, J. et al. (2007) 'AcMNPV ORF38 Protein Has the Activity of ADP-ribose Pyrophosphatase and Is Important for Virus Replication', *Virology*, 361: 204–11.

Goldbach, R. W. et al. (1998) 'Distinct Gene Arrangement in the Buzura Suppressaria Single-nucleocapsid Nucleopolyhedrovirus Genome', *Journal of General Virology*, 79: 2841–51.

Granados, R. R., and Lawler, K. A. (1981) 'In Vivo Pathway of Autographa californica Baculovirus Invasion and Infection', *Virology*, 108: 297–308.

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011) 'FIMO: Scanning for Occurrences of a Given Motif', *Bioinformatics*, 27: 1017–8.

Guinet, B. et al. (2023) 'Endoparasitoid Lifestyle Promotes Endogenization and Domestication of dsDNA Viruses', *eLife*, 12: e85993.

Hamm, J. J., Carpenter, J. E., and Styer, E. L. (1996) 'Oviposition Day Effect on Incidence of Agonadal Progeny of *Helicoverpa zea* (Lepidoptera: Noctuidae) Infected with a Virus', *Annals of the Entomological Society of America*, 89: 266–75.

Hamm, J. J., Styer, E. L., and Lewis, W. J. (1990) 'Comparative Virogenesis of Filamentous Virus and Polydnavirus in the Female Reproductive Tract of *Cotesia marginiventris* (Hymenoptera: Braconidae)', *Journal of Invertebrate Pathology*, 55: 357–74.

Hegazi, E. M. et al. (2005) 'The Calyx Fluid of *Microplitis rufiventris* Parasitoid and Growth of Its Host *Spodoptera Littoralis* Larvae', *Journal of Insect Physiology*, 51: 777–87.

Heisserer, C. et al. (2023) 'Massive Somatic and Germline Chromosomal Integrations of Polydnaviruses in Lepidopterans', *Molecular Biology and Evolution*, 40: msad050.

Herniou, E. A. et al. (2003) 'The Genome Sequence and Evolution of Baculoviruses', *Annual Review of Entomology*, 48: 211–34.

Hilbert, B. J. et al. (2015) 'Structure and Mechanism of the ATPase that Powers Viral Genome Packaging', *Proceedings of the National Academy of Sciences*, 112: E3792–9.

Hoang, D. T. et al. (2018) 'UFBoot2: Improving the Ultrafast Bootstrap Approximation', *Molecular Biology and Evolution*, 35: 518–22.

Hong, T., Braunagel, S. C., and Summers, M. D. (1994) 'Transcription, Translation, and Cellular Localization of PDV-E66: A Structural Protein of the PDV Envelope of Autographa californica Nuclear Polyhedrosis Virus', *Virology*, 204: 210–22.

Hou, D. et al. (2013) 'Comparative Proteomics Reveal Fundamental Structural and Functional Differences between the Two Progeny Phenotypes of a Baculovirus', *Journal of Virology*, 87: 829–39.

Hou, D. et al. (2019) 'Baculovirus ODV-E66 Degrades Larval Peritrophic Membrane to Facilitate Baculovirus Oral Infection', *Virology*, 537: 157–64.

Hughes, A. L. (2002) 'Evolution of Inhibitors of Apoptosis in Baculoviruses and Their Insect Hosts', *Infection Genetics & Evolution*, 2: 3–10.

Huttunen, M. et al. (2021) 'Vaccinia Virus Hijacks ESCRT-mediated Multivesicular Body Formation for Virus Egress', *Life Science Alliance*, 4: e202000910.

Hyatt, D. et al. (2010) 'Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification', *BMC Bioinformatics*, 11: 119.

Inwood, S. N. et al. (2023) 'Chromosome-level Genome Assemblies of Two Parasitoid Biocontrol Wasps Reveal the Parthenogenesis Mechanism and an Associated Novel Virus', *BMC Genomics*, 24: 440.

Irwin, N. A. T. et al. (2021) 'Systematic Evaluation of Horizontal Gene Transfer between Eukaryotes and Viruses', *Nature Microbiology*, 7: 327–36.

Iyer, L. M., Koonin, E. V., and Aravind, L. (2002) 'Extensive Domain Shuffling in Transcription Regulators of DNA Viruses and Implications for the Origin of Fungal APSES Transcription Factors', *Genome Biology*, 3: research0012.

Kalyaanamoorthy, S. et al. (2017) 'ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates', *Nature Methods*, 14: 587–9.

Kariithi, H. M. et al. (2010) 'Proteomic Analysis of Glossina pallidipes Salivary Gland Hypertrophy Virus Virions for Immune Intervention in Tsetse Fly Colonies', *Journal of General Virology*, 91: 3065–74.

Kariithi, H. M. et al. (2013) 'Correlation between Structure, Protein Composition, Morphogenesis and Cytopathology of Glossina pallidipes Salivary Gland Hypertrophy Virus', *Journal of General Virology*, 94: 193–208.

Katz, K. S. et al. (2021) 'STAT: A Fast, Scalable, MinHash-based K-mer Tool to Assess Sequence Read Archive Next-generation Sequence Submissions', *Genome Biology*, 22: 270.

Kearse, M. et al. (2012) 'Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data', *Bioinformatics*, 28: 1647–9.

Khan, M. M. et al. (2022) 'Oxidative Stress Protein Oxr1 Promotes V-ATPase Holoenzyme Disassembly in Catalytic Activity-independent Manner', *The EMBO Journal*, 41.

Kolesnikova, L. et al. (2009) 'Vacuolar Protein Sorting Pathway Contributes to the Release of Marburg Virus', *Journal of Virology*, 83: 2327–37.

Kool, M. et al. (1995) 'Replication of Baculovirus DNA', *Journal of General Virology*, 76: 2103–18.

Koonin, E. V. et al. (2020) 'Global Organization and Proposed Megataxonomy of the Virus World', *Microbiology and Molecular Biology Reviews*, 84: e00061–19.

Koonin, E. V., and Dolja, V. V. (2013) 'A Virocentric Perspective on the Evolution of Life', *Current Opinion in Virology*, 3: 546–57.

Krell, P. J. (1987) 'Replication of Long Virus-like Particles in the Reproductive Tract of the Ichneumonid Wasp *Diadegma terebrans*', *Journal of General Virology*, 68: 1477–83.

Kristensen, D. M. et al. (2010) 'New Dimensions of the Virus World Discovered through Metagenomics', *Trends in Microbiology*, 18: 11–9.

Langmead, B. et al. (2009) 'Ultrafast and Memory-efficient Alignment of Short DNA Sequences to the Human Genome', *Genome Biology*, 10: R25.

Lefkowitz, E. J. et al. (2018) 'Virus Taxonomy: The Database of the International Committee on Taxonomy of Viruses (ICTV)', *Nucleic Acids Research*, 46: D708–D717.

Legendre, M. et al. (2011) 'Breaking the 1000-gene Barrier for Mimivirus Using Ultra-deep Genome and Transcriptome Sequencing', *Virology Journal*, 8: 99.

Leisy, D. J., and Rohrmann, G. F. (1993) 'Characterization of the Replication of Plasmids Containing *Hr* Sequences in Baculovirus-Infected *Spodoptera frugiperda* Cells', *Virology*, 196: 722–30.

Leobold, M. et al. (2018) 'The Domestication of a Large DNA Virus by the Wasp *Venturia canescens* Involves Targeted Genome Reduction through Pseudogenization', *Genome Biology and Evolution*, 10: 1745–64.

Lepetit, D. et al. (2017) 'Genome Sequencing of the Behaviour Manipulating Virus LbFV Reveals a Possible New Virus Family', *Genome Biology and Evolution*, 8, 3718–39.

Le Sage, V., and Mouland, A. (2013) 'Viral Subversion of the Nuclear Pore Complex', *Viruses*, 5: 2019–42.

Li, D. et al. (2015) 'MEGAHIT: An Ultra-fast Single-node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph', *Bioinformatics*, 31: 1674–6.

Manni, M. et al. (2021) 'BUSCO: Assessing Genomic Data Quality and Beyond', *Current Protocols*, 1: e323.

Marques, J. T., and Imler, J. L. (2016) 'The Diversity of Insect Antiviral Immunity: Insights from Viruses', *Current Opinion in Microbiology*, 32: 71–6.

Martin, M. (2011) 'Cutadapt Removes Adapter Sequences from High-throughput Sequencing Reads', *EMBnet.journal*, 17: 10–12.

Martinez, J. et al. (2012) 'Influence of the Virus LbFV and of Wolbachia in a Host-parasitoid Interaction', *PLoS One*, 7: e35081.

Martinez, J., Fleury, F., and Varaldi, J. (2015) 'Competitive Outcome of Multiple Infections in a Behavior-manipulating Virus/wasp Interaction', *Ecology and Evolution*, 5: 5934–45.

Matsushima, N. et al. (2021) 'Shrinking of Repeating Unit Length in Leucine-rich Repeats from Double-stranded DNA Viruses', *Archives of Virology*, 166: 43–64.

Matsushima, N., and Kretsinger, R. H. (2022) 'Numerous Variants of Leucine Rich Repeats in Proteins from Nucleo-cytoplasmic Large DNA Viruses', *Gene*, 817: 146156.

Mildvan, A. S. et al. (2005) 'Structures and Mechanisms of Nudix Hydrolases', *Archives of Biochemistry and Biophysics*, 433: 129–43.

Miller, L. K. (1999) 'An Exegesis of IAPs: Salvation and Surprises from BIR Motifs', *Trends in Cell Biology*, 9: 323–8.

Minh, B. Q. et al. (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37: 1530–4.

Moreau, S., and Asgari, S. (2015) 'Venom Proteins from Parasitoid Wasps and Their Biological Functions', *Toxins*, 7: 2385–412.

Muller, H. et al. (2021) 'Genome-Wide Patterns of Bracovirus Chromosomal Integration into Multiple Host Tissues during Parasitism', *Journal of Virology*, 95: e00684–21.

Muller, H. et al. (2022) 'Investigating Bracovirus Chromosomal Integration and Inheritance in Lepidopteran Host and Nontarget Species', *Molecular Ecology*, 31: 5538–51.

Naser-Khdour, S. et al. (2019) 'The Prevalence and Impact of Model Violations in Phylogenetic Analysis', *Genome Biology and Evolution*, 11: 3341–52.

Ogura, T., and Wilkinson, A. J. (2001) 'AAA Superfamily ATPases: Common Structure-diverse Function: AAA Superfamily ATPases', *Genes to Cells*, 6: 575–97.

Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2015) 'Qualimap 2: Advanced Multi-sample Quality Control for High-throughput Sequencing Data', *Bioinformatics* 32, 292–294.

Paradis, E., Claude, J., and Strimmer, K. (2004) 'Ape: Analyses of Phylogenetics and Evolution in R Language', *Bioinformatics*, 20: 289–90.

Parrish, S., and Moss, B. (2007) 'Characterization of a Second Vaccinia Virus mRNA-Decapping Enzyme Conserved in Poxviruses', *Journal of Virology*, 81: 12973–8.

Patel, R. K., Jain, M., and Liu, Z. (2012) 'NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data', *PLoS One*, 7: e30619.

Patot, S. et al. (2009) 'Molecular Detection, Penetrance, and Transmission of an Inherited Virus Responsible for Behavioral Manipulation of an Insect Parasitoid', *ApplEnvironMicrobiol*, 75: 703–10.

Patot, S. et al. (2010) 'Prevalence of a Virus Inducing Behavioural Manipulation near Species Range Border: Virus Responsible For Behavioural Manipulation', *Molecular Ecology*, 19: 2995–3007.

Patot, S. et al. (2012) 'An Inherited Virus Influences the Coexistence of Parasitoid Species through Behaviour Manipulation: A Symbiont Mediates Interspecific Competition', *Ecology Letters*, 15: 603–10.

Peng, Y. et al. (2012) 'IDBA-UD: A *de Novo* Assembler for Single-cell and Metagenomic Sequencing Data with Highly Uneven Depth', *Bioinformatics*, 28: 1420–8.

Pennell, M. W. et al. (2014) 'Geiger V2.0: An Expanded Suite of Methods for Fitting Macroevolutionary Models to Phylogenetic Trees', *Bioinformatics*, 30: 2216–8.

Peters, R. S. et al. (2017) 'Evolutionary History of the Hymenoptera', *Current Biology*, 27: 1013–8.

Petersen, J. M. et al. (2022) 'The Naked Truth: An Updated Review on Nudiviruses and Their Relationship to Bracoviruses and Baculoviruses', *Journal of Invertebrate Pathology*, 189: 107718.

Pichon, A. et al. (2015) 'Recurrent DNA Virus Domestication Leading to Different Parasite Virulence Strategies', *Science Advances*, 1: e1501150.

Pszenny, V. et al. (2016) 'A Lipolytic Lecithin:Cholesterol Acyltransferase Secreted by Toxoplasma Facilitates Parasite Replication and Egress', *Journal of Biological Chemistry*, 291: 3725–46.

Rambaut, A. et al. (2018) 'Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7', *Systematic Biology*, 67: 901–4.

Rice, P., Longden, I., and Bleasby, A. (2000) 'EMBOSS: The European Molecular Biology Open Software Suite', *Trends in Genetics*, 16: 276–7.

Rizki, R. M., and Rizki, T. M. (1990) 'Parasitoid Virus-Like Particles Destroy *Drosophila* Cellular Immunity', *Proceedings of the National Academy of Sciences of the United States of America*, 87: 8388–92.

Robinson, J. T. et al. (2011) 'Integrative Genomics Viewer', *Nature Biotechnology*, 29: 24–6.

Rodrigues, D. T. et al. (2020) 'Characterization of a Novel Alphabaculovirus Isolated from the Southern Armyworm, *Spodoptera eridania* (Cramer, 1782) (Lepidoptera: Noctuidae) and the Evolution of *Odv-e66*, a Bacterium-acquired Baculoviral Chondroitinase Gene', *Genomics*, 112: 3903–14.

Rohrmann, G. (2019) *Cover of Baculovirus Molecular Biology Baculovirus Molecular Biology*., 4th edition. National Center for Biotechnology Information: US.

Ronquist, F. et al. (2012) 'MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space', *Systematic Biology*, 61: 539–42.

Roux, S., and Emerson, J. B. (2022) 'Diversity in the Soil Virosphere: To Infinity and Beyond?', *Trends in Microbiology*, 30: 1025–35.

Saeedi, R., Li, M., and Frohlich, J. (2015) 'A Review on Lecithin: cholesterol Acyltransferase Deficiency', *Clinical Biochemistry*, 48: 472–5.

Schneider, S. E., and Thomas, J. H. (2014) 'Accidental Genetic Engineers: Horizontal Sequence Transfer from Parasitoid Wasps to Their Lepidopteran Hosts', *PLoS One*, 9: e109446.

Schulz, F. et al. (2020) 'Giant Virus Diversity and Host Interactions through Global Metagenomics', *Nature*, 578: 432–6.

Shen, Q., Wang, Y. E., and Palazzo, A. F. (2021) 'Crosstalk between Nucleocytoplasmic Trafficking and the Innate Immune Response to Viral Infection', *Journal of Biological Chemistry*, 297: 100856.

Shi, M. et al. (2018) 'The Evolutionary History of Vertebrate RNA Viruses', *Nature*, 556: 197–202.

Shi, M. et al. (2019) 'The Genomes of Two Parasitic Wasps that Parasitize the Diamondback Moth', *BMC Genomics*, 20: 893.

Sievers, F. et al. (2011) 'Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega', *Molecular Systems Biology*, 7: 539.

Simmonds, P. et al. (2023) 'Four Principles to Establish a Universal Virus Taxonomy', *PLoS Biology*, 21: e3001922.

Snider, J., Thibault, G., and Houry, W. A. (2008) 'The AAA+ Superfamily of Functionally Diverse Proteins', *Genome Biology*, 9: 216.

Steczkiewicz, K. et al. (2012) 'Sequence, Structure and Functional Diversity of PD-(D/E)XK Phosphodiesterase Superfamily', *Nucleic Acids Research*, 40: 7016–45.

Steinegger, M. (2018) 'Clustering Huge Protein Sequence Sets in Linear Time', *Nature Communications*, 9: 2542.

Steinegger, M., and Söding, J. (2017) 'MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets', *Nature Biotechnology*, 35: 1026–8.

Stoltz, D. B. (1979) 'Viruses and Parasitism in Insects Lauffer Max A, Bang Frederik B, Maramorosch Karl and Kenneth M. Smith'. In: *Advances in Virus Research*, Vol. 24, pp. 125–71. Elsevier.

Stoltz, D. B., and Vinson, S. B. (1977) 'Baculovirus-like Particles in the Reproductive Tracts of Female Parasitoid Wasps II: The Genus *Apanteles*', *Canadian Journal of Microbiology*, 23: 28–37.

Styer, E. L., Hamm, J. J., and Nordlund, D. A. (1987) 'A New Virus Associated with the Parasitoid *Cotesia marginiventris* (Hymenoptera: Braconidae): Replication in Noctuid Host Larvae', *Journal of Invertebrate Pathology*, 50: 302–9.

Sugiura, N. et al. (2011) 'Baculovirus Envelope Protein ODV-E66 Is a Novel Chondroitinase with Distinct Substrate Specificity', *Journal of Biological Chemistry*, 286: 29026–34.

Tanaka, T. (1987) 'Effect of the Venom of the Endoparasitoid, *Apanteles kariyai* Watanabe, on the Cellular Defence Reaction of the Host, *Pseudaletia separata* Walker', *Journal of Insect Physiology*, 33: 413–20.

Tang, P. et al. (2019) 'Mitochondrial Phylogenomics of the Hymenoptera', *Molecular Phylogenetics & Evolution*, 131: 8–18.

Theilmann, D. A., and Stewart, S. (1992) 'Tandemly Repeated Sequence at the 3 End of the IE-2 Gene of the Baculovirus *Orgyia pseudotsugata* Multicapsid Nuclear Polyhedrosis Virus Is an Enhancer Element', *Virology*, 187: 97–106.

Thézé, J. et al. (2011) 'Paleozoic Origin of Insect Large dsDNA Viruses', *Proceedings of the National Academy of Sciences*, 108: 15931–5.

Thézé, J. et al. (2015) 'Gene Acquisition Convergence between Entomopoxviruses and Baculoviruses', *Viruses*, 7: 1960–74.

Van Oers, M. M. et al. (2023) 'Developments in the Classification and Nomenclature of Arthropod-infecting Large DNA Viruses that Contain Pif Genes', *Archives of Virology*, 168: 182.

Varaldi, J. et al. (2003) 'Infectious Behavior in a Parasitoid', *Science*, 302: 1930–1930.

Varaldi, J. et al. (2005) 'Superparasitism Acceptance and Patch-leaving Mechanisms in Parasitoids: A Comparison between Two Sympatric Wasps', *Animal Behaviour*, 69: 1227–34.

Varaldi, J. et al. (2006) 'Artifical Transfer and Morphological Description of Virus Particles Associated with Superparasitism Behaviour in a Parasitoid Wasp', *Journal of Insect Physiology*, 52: 1202–12.

Varaldi, J. et al. (2024) 'Community Structure of Heritable Viruses in a *Drosophila*-parasitoids Complex', *Peer Community Journal*, 4: e16.

Varaldi, J., and Lepetit, D. (2018) 'Deciphering the Behaviour Manipulation Imposed by a Virus on Its Parasitoid Host: Insights from a Dual Transcriptomic Approach', *Parasitology*, 145: 1979–89.

Verhagen, A. M., Coulson, E. J., and Vaux, D. L. (2001) 'Inhibitor of Apoptosis Proteins and Their Relatives: IAPs and Other BIRPs', *Genome Biology*, 2: reviews3009.

Volkoff, A. N. et al. (2010) 'Analysis of Virion Structural Components Reveals Vestiges of the Ancestral Ichnovirus Genome', *PLoS Pathogens*, 6: e1000923.

Walker, P. J. et al. (2021) 'Changes to Virus Taxonomy and to the International Code of Virus Classification and Nomenclature Ratified by the International Committee on Taxonomy of Viruses (2021)', *Archives of Virology*, 166: 2633–48.

Wallace, M. A. et al. (2021) 'The Discovery, Distribution, and Diversity of DNA Viruses Associated with *Drosophila melanogaster* in Europe', *Virus Evolution*, 7: veab031.

Wang, D. et al. (2005) 'Characterization of Helicoverpa armigera Nucleopolyhedrovirus ORF33 that Encodes a Novel Budded Virion Derived Protein, BV-e31', *Archives of Virology*, 150: 1505–15.

Wang, Y., and Jehle, J. A. (2009) 'Nudiviruses and Other Large, Double-stranded Circular DNA Viruses of Invertebrates: New Insights on an Old Topic', *Journal of Invertebrate Pathology*, 101: 187–93.

Wick, R. R. et al. (2017) 'Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads', *PLoS Computational Biology*, 13: e1005595.

Wick, R. R. et al. (2021) 'Trycycler: Consensus Long-read Assemblies for Bacterial Genomes', *Genome Biology*, 22: 266.

Wickham, H. (2016) *Ggplot2: Elegant Graphics for Data Analysis*, 2nd. Use R! Springer International Publishing: Imprint: Springer: Cham.

Wu, W. et al. (2008) 'Autographa californica Multiple Nucleopolyhedrovirus 38K Is a Novel Nucleocapsid Protein that Interacts with VP1054, VP39, VP80, and Itself', *Journal of Virology*, 82: 12356–64.

Wu, H. et al. (2020) 'Abundant and Diverse RNA Viruses in Insects Revealed by RNA-Seq Analysis', *Ecological and Evolutionary Implications. mSystems*, 5: e00039–20.

Yang, D. et al. (2022) 'Genomics and Proteomics of Apis mellifera Filamentous Virus Isolated from Honeybees in China', *Virologica Sinica* 37, 483–490.

Zemskov, E. A., Kang, W., and Maeda, S. (2000) 'Evidence for Nucleic Acid Binding Ability and Nucleosome Association of Bombyx mori Nucleopolyhedrovirus BRO Proteins', *Journal of Virology*, 74: 6784–9.

Zhang, W. et al. (2015) 'Four Novel Algal Virus Genomes Discovered from Yellowstone Lake Metagenomes', *Scientific Reports*, 5: 15131.

Zhang, K. Y. et al. (2019) 'Vgas: A Viral Genome Annotation System', *Frontiers in Microbiology*, 10: 184.

Zhang, H. et al. (2022) 'Research Progress in Porcine Reproductive and Respiratory Syndrome Virus–Host Protein Interactions', *Animals*, 12: 1381.

Virus Evolution, 2024, **10(1)**, veae022
DOI: https://doi.org/10.1093/ve/veae022
Advance Access Publication 2 March 2024
Research Article