The domestication of a large DNA virus by the wasp Venturia canescens involves targeted genome reduction through pseudogenization

Matthieu Leobold, Annie Bézier, Apolline Pichon, Elisabeth A Herniou, Anne-Nathalie Volkoff, Jean-Michel Drezen

The Domestication of a Large DNA Virus by the Wasp *Venturia canescens* Involves Targeted Genome Reduction through Pseudogenization

Matthieu Leobold[1], Annie Bézier[1], Apolline Pichon[2], Elisabeth A. Herniou[1], Anne-Nathalie Volkoff[2], and Jean-Michel Drezen[1,*]

[1]Institut de Recherche sur la Biologie de l'Insecte, UMR 7261, CNRS - Université de Tours, UFR des Sciences et Techniques, Parc de Grandmont, Tours, France

[2]Diversity, Genomes and Interactions Microorganisms-Insect, UMR INRA 1333, Université de Montpellier 2, Montpellier, France

*Corresponding author: E-mail: drezen@univ-tours.fr.

Accepted: June 18, 2018

Data deposition: *Venturia canescens* annotated scaffolds have been deposited in GenBank/DDBJ/EMBL under accession numbers (KP972595–KP972600 and MH151087–MH151098). *Venturia canescens* genome is available at http://bipaa.genouest.org/sp/venturia_canescens (last accessed June 29, 2018).

Abstract

Polydnaviruses (PDVs) are compelling examples of viral domestication, in which wasps express a large set of genes originating from a chromosomally integrated virus to produce particles necessary for their reproductive success. Parasitoid wasps generally use PDVs as a virulence gene delivery system allowing the protection of their progeny in the body of parasitized host. However, in the wasp *Venturia canescens* an independent viral domestication process led to an alternative strategy as the wasp incorporates virulence proteins in viral liposomes named virus-like particles (VLPs), instead of DNA molecules. Proteomic analysis of purified VLPs and transcriptome sequencing revealed the loss of some viral functions. In particular, the genes coding for capsid components are no longer expressed, which explains why VLPs do not incorporate DNA. Here a thorough examination of *V. canescens* genome revealed the presence of the pseudogenes corresponding to most of the genes involved in lost functions. This strongly suggests that an accumulation of mutations that leads to gene specific pseudogenization precedes the loss of viral genes observed during virus domestication. No evidence was found for block loss of collinear genes, although extensive gene order reshuffling of the viral genome was identified from comparisons between endogenous and exogenous viruses. These results provide the first insights on the early stages of large DNA virus domestication implicating massive genome reduction through gene-specific pseudogenization, a process which differs from the large deletions described for bacterial endosymbionts.

Key words: polydnavirus, nudivirus, virus-like particles, endosymbiosis, parasitoid wasp, endogenous viral elements.

Introduction

Positive interactions between organisms are widespread. Among these relationships, endosymbiosis in which a microorganism lives inside host cells represents an ultimate stage of interactions that allows an organism to simultaneously co-opt a whole new set of genes and functions from its microbial partner. In insects numerous examples of endosymbiosis have been described. All known obligatory hematophagous insects, as well as a variety of other insects such as aphids (Munson et al. 1991; Shigenobu et al. 2000), weevils (Heddi et al. 1999), or cockroaches (Bandi et al. 1994) carry symbiotic microorganisms, which supplement their nutrition (Rio et al. 2016). These obligate partnerships between insects and symbionts have coevolved for millions of years resulting invariably in a dramatic reduction of the microbial symbiont genomes both in terms of size and gene content compared with related free living bacteria (Moran 2003; Moran and Bennett 2014).

It is less broadly known that viruses can also be involved in obligatory mutualistic relationships with their hosts. Viruses commonly interact in the long term with their hosts following genomic integrations in the germline allowing their vertical transmission. Such viruses fixed in host genome are known as

endogenous viral elements (EVEs) (Katzourakis and Gifford 2010; Holmes 2011; Feschotte and Gilbert 2012). Most EVEs are fragmented and nonfunctional, because they do not confer any adaptive advantage. They therefore evolve under neutral selection, accumulating a series of mutations after their integration. Worse, they may also be a costly burden to their hosts by vertically transmitting disease (Tarlinton et al. 2006; Arbuckle et al. 2010), and their pseudogenization is likely selected for. However, some EVEs may confer a selective advantage to their hosts by protecting against infections by closely related viruses (Frank and Feschotte 2017), or providing specific gene contributing to host development, such as the placental syncytins repeatedly acquired from retroviruses by different mammalian lineages (Mangeney et al. 2007; Lavialle et al. 2013) or as the Arc protein, which also derives from a retrovirus and is involved in synaptic plasticity in mammals and insects (Ashley et al. 2018; Pastuzyn et al. 2018). In all these examples, a single virus gene has been conserved and co-opted in the host genome. In contrast, the EVEs of large double-stranded DNA viruses integrated in the genomes of parasitic wasps (Drezen et al. 2017) have retained tens of genes encoding essential functions, including the production of infectious viral particle in the case of polydnaviruses (PDVs) (Bézier et al. 2009; Volkoff et al. 2010; Burke et al. 2013) or of virus-like particles (VLPs) from the wasp *Venturia canescens* (Pichon et al. 2015). These EVEs, provide their wasp hosts, the tools to deliver virulence molecules to the lepidopteran hosts they parasitize (Drezen et al. 2017). PDVs and VLPs are thus obvious models for studying how complex viruses encoding more than hundred genes evolve when they become beneficial to their host.

Our study focuses on EVEs derived from nudiviruses. The *Nudiviridae* is a family of large dsDNA viruses infecting arthropods, which evolved over 220 million years ago (Mya) and is the sister group of the *Baculoviridae* (Thézé et al. 2011). Exogenous nudiviruses have circular genomes typically 150 kb long and encoding 130 genes, of which 32 are conserved core genes (Bézier et al. 2015). Integration of nudiviruses in insect genomes appears relatively widespread. The exogenous nudivirus of the moth *Heliothis zea* (HzNV-1) is known to persist as a latent infection with both circular and integrated DNA present in cultured cells (Lin et al. 1999). Moreover, large nudivirus genome fragments (Nilaparvata lugens endogenous nudivirus, NlENV) were identified in the genome of the brown plant hopper *N. lugens* (Cheng et al. 2014). The first functional EVEs found to derive from a nudivirus are bracoviruses (Bézier et al. 2009), which integrated in the genome of the wasp ancestor of the microgastroid complex (Braconidae) about 100 Mya (Murphy et al. 2008; Thézé et al. 2011; Whitfield et al. 2018). *Bracovirus* is one of the two genera comprised in the family *Polydnaviridae.* The second PDV genus called *Ichnovirus* regroups functional EVEs associated with ichneumonid wasps (Gundersen-Rindal et al. 2013) that derived from the genomic integration of a virus from a new family, belonging probably to nucleocytoplasmic large DNA viruses (Volkoff et al. 2010; Beliveau et al. 2015; Drezen et al. 2017). In the ichneumonid wasp *V. canescens*, this ichnovirus is no longer active and an independent event of nudivirus genome acquisition is at the origin of the immune suppressive VLPs that enable wasp eggs to escape lepidopteran immune defences (Pichon et al. 2015). For clarity, we henceforth named the integrated nudiviral sequences, involved in VLP production, *V. canescens* endogenous nudivirus (VcENV). *Venturia canescens* nudiviral genes are most similar to those of Oryctes rhinoceros nudivirus (OrNV) (Pichon et al. 2015) and recently sequenced related viruses, which all belong to the genus *Alphanudivirus* (https://talk.ictvonline.org; last accessed June 29, 2018) whereas bracoviruses originate from a virus of the genus *Betanudivirus* (Thézé et al. 2011; Jehle et al. 2013).

Several empirical and theoretical lines of evidence suggest hypotheses on the early evolutionary steps leading to the domestication of endogenous nudiviruses by parasitoid wasps (reviewed in Drezen et al. 2017; Gauthier et al. 2017). Bracovirus or VLP particles are produced by EVE sequences, which independently derive from single nudiviral genomic integrations. Nudiviral genes are now found as single genes or as gene clusters within large genomic scaffolds of the wasp genome (Pichon et al. 2015). As they are scattered over several scaffolds, the initial nudiviral genomes were involved in recombination events of the wasp genomes after their integration. In both bracovirus and VLP cases, all nudiviral core genes present in the wasp genomes share the same phylogenetic history and therefore result from a single endogenization event. This event occurred approximately 100 Mya in the braconid ancestor of the microgastroid complex (Murphy et al. 2008), whereas *V. canescens* is to date the only wasp in its lineage known to produce VLPs of nudiviral origin, which suggests a more recent endogenization event. Accordingly, the nudiviral genes are not as widely dispersed in the genome of *V. canescens* as they are in wasps associated with bracoviruses (Pichon et al. 2015). In both types of models, several endogenous nudiviral genes are expressed in the wasp ovaries and involved in the production of virulence particles (Bézier et al. 2009; Pichon et al. 2015; Drezen et al. 2017). Instead of transmitting a nudivirus genome, bracoviruses and VcENV particles, respectively, enclose virulence genes (Bézier et al. 2013) and virulence wasp proteins (Pichon et al. 2015).

The study of *V. canescens* genome into which nudiviral integration is most likely a relatively recent event might offer a unique opportunity to get insights into some of the early steps leading to viral domestication. An interesting issue is how virus genes are lost, after virus integration. Conceptually, two major mechanisms might be involved in the loss of viral genes. First, genes could be lost through chromosomal rearrangements causing virus genome dispersion. In this case we could expect that large parts of the endogenous virus genome would be lost by large deletions spanning

multiple genes. Alternatively, genes could be specifically lost, one by one, though the accumulation of a series of mutations under a neutral selection regime. In particular, this could apply to the genes not contributing to *V. canescens* parasitism success. The two mechanisms are not mutually exclusive but they are likely to leave very different footprints in the wasp genome. Indeed in the first case we do not expect to observe remnants of lost genes, whereas in the second case it might be possible to detect pseudogenes in the wasp genome. To address this question, we compared the gene organization of available exogenous and endogenous alphanudiviruses in order to trace rearrangements that have occurred in VcENV after its integration in the wasp genome. Moreover, we comprehensively searched for remains of nudiviral genes by scanning the genome of *V. canescens*.

Materials and Methods

Genomic Data

Venturia canescens genomic and nudiviral gene expression data were obtained as previously described (Pichon et al. 2015). The 238-Mb long *V. canescens* genome (N50 = 115 kb), derives from the progeny (haploid males) of a single virgin female of a natural arrhenotokous strain collected in the area of Valence (France). It comprises 62,001 scaffolds and is available at the Bioinformatic Platform for Agroecosystem Arthropods (http://bipaa.genouest.org/sp/venturia_canescens/; last accessed June 29, 2018). VcENV is composed of 6 clusters containing 51 nudiviral genes. Annotations of scaffolds containing nudiviral pseudogenes have been released under accession numbers KP972595–KP972600 and MH151087–MH151098. The expression of these genes and of newly identified pseudogenes was studied by examining previously obtained transcriptome data from the calyx of a *V. canescens* thelytokous strain (parthenogenetic) collected in the same area. To measure pseudogene expression levels, the cDNA reads from transcriptome data were mapped onto the wasp genome scaffolds with the short-read aligner BOWTIE 1.1 (Pichon et al. 2015). The parameters chosen were –v 0 and –a. The number of reads was determined and RPKM was calculated for all validated peptides corresponding to pseudogenes.

VcENV sequences were compared with the genomic sequences of all exogenous and endogenous alphanudiviruses available at the time of analyses. OrNV is a highly pathogenic virus infecting the coleopteran *Oryctes rhinoceros* which has a circular 127-kb long dsDNA genome containing 138 genes (Wang et al. 2011). Drosophila innubia NV (DiNV), Kallithea virus (DmNV_kal) and Tomelloso Virus (DmNV_tom) are nudiviruses infecting *Drosophila* species respectively with 156-, 152-, and 112-kb circular genome encoding 106, 95, and 96 predicted proteins (Webster et al. 2015). Gryllus bimaculatus NV (GbNV) infects the cricket *Gryllus bimaculatus*, its 96-kb long genome contains 98 genes (Wang et al. 2007; Hill and Unckless 2018). Nudiviral genes in *N. lugens* (Cheng et al. 2014) are organized in 12 clusters encoding a total of 66 nudiviral genes over 98-kb. VcENV genome consists of six nudiviral clusters dispersed in the wasp genome. In these clusters of total length 117-kb, 51 genes have been annotated (Pichon et al. 2015).

Phylogenomic Analysis

Amino acid alignments were individually performed with default parameters using MAFFT alignment plugin v7.017 from Geneious software (Katoh et al. 2002; Kearse et al. 2012) on the 37 nudivirus-related predicted homologue set previously used (Pichon et al. 2015), to which *HzNVorf143* core gene homologues were added. Then the phylogenetic consensus tree was constructed using Geneious Tree builder on concatenated and refined multiple alignments with the following parameters: Genetic distance model =Jukes-Cantor, tree build method =Neighbor-Joining, resampling method =bootstrap, number of replicates =1000. Support threshold was set up at 40%. The sequences used for this analysis were obtained from three bracoviruses (Cotesia congregata bracovirus [CcBV], Microplitis demolitor bracovirus [MdBV], and Chelonus inanitus bracovirus [CiBV]), four pathogenic betanudiviruses (Tipula oleracea nudivirus [ToNV, KM610234], Penaeus monodon NV [PmNV, KJ184318], Heliothis zea nudivirus 1 [HzNV-1, AF451898], and Helicoverpa zea NV 2 [HzNV-2, NC_004156]), five pathogenic alphanudiviruses (Gryllus bimaculatus nudivirus [GbNV, EF203088], Drosophila innubia NV [DiNV, MF966379], Kallithea virus [DmNV_kal, NC_033829], Oryctes rhinoceros NV [OrNV, EU747721], and Tomelloso virus [DmNV_tom, KY457233]), two endogenous alphanudiviruses (Nilaparvata lugens endogenous nudivirus [NlENV] and *V. canescens* ENV [VcENV]), five baculoviruses (Autographa californica multiple nucleopolyhedrovirus [AcMNPV, NC_001623], Lymantria dispar MNPV [LdMNPV, NC_001973], Cydia pomonella granulovirus [CpGV, NC_002816], Culex nigripalpus nucleopolyhedrovirus [CuniNPV, NC_003084], and Neodiprion sertifer NPV [NeseNPV, NC_005905]), two hytrosaviruses (Musca domestica salivary gland hytrosavirus [MdSGHV, EU522111] and Glossina pallidipes SGHV [GpSGHV, NC_010356]) and a nimavirus (white spot syndrome virus [WSSV, AF369029]) used as outgroup based on gene content.

Synteny Analyses

Gene order comparison among exogenous and endogenous nudiviruses were performed based on homologous relationships reported in the annotations of the different virus genomes that were verified by reciprocal BlastP or TBlastN using amino acid sequences (Altschul et al. 1997). Duplicated genes of endogenous viruses were assigned to a single gene in OrNV. Gene positions were compiled in a correspondence table using Microsoft Office Excel as previously

described (Hu et al. 1998). Then gene parity plots were drawn from this table. For endogenous viruses, the different clusters were linked for the analysis following the order of their numbering; however, they were visualized separately on the gene parity plot using a specific color code.

Identification of Nudiviral Pseudogenes

The nudivirus set of core genes initially comprised 33 genes (Wang et al. 2012) based on the comparison of the few sequenced genomes. This list was recently updated to include P6.9, a protein involved in DNA condensation prior to packaging, as well as *OrNVorf41-like* (*11K*), *OrNVorf18-like* (*PmNVorf99-like*) and *OrNVorf61-like* (*PmNVorf62-like*) and to remove the genes *iap-3*, *ligase*, *rr1*, *rr2*, and *tk4* absent from ToNV (Bézier et al. 2015). We searched for remains of the nudiviral core genes not expressed in the calyx by a thorough examination of *V. canescens* genome. Protein sequences corresponding to OrNV nudiviral core genes were retrieved from database and used as queries in a TBlastN analysis of *V. canescens* genome. To identify seeds in order to start pseudogene identification BLAST analyses were performed locally using BioEdit and a personal database of *V. canescens* genome, using the matrix BLOSUM62 without filtering Expectation Value (Hall 1999). All retrieved sequences were then individually examined and peptide sequences of low complexity or containing repeated sequence were discarded. Then, scaffolds containing selected sequences were annotated using Geneious version R6 (Kearse et al. 2012). The second step consisted in generating small peptides using the three coding phases upstream and downstream of the seed in order to increase the length of the reconstructed protein. More precisely windows of 90 and 300 bp (the latter to cover long fragments or to jump over the gaps) were used to identify peptides. These peptides were then validated or discarded, based on their alignment to the homologous nudiviral protein using Geneious with BLOSUM65 matrix (gap open penalty of 12 and gap extension penalty of 5). If peptides were not validated in the first round of analysis, the presence of large gaps was assessed by using parameters gap open penalty of 6 and extension of 3. If it was the case, the gap was removed and the resulting fragment was realigned with the more stringent parameters. Once a fragment was validated, the analysis window was shifted and the annotation was further processed. Finally, a BlastP with BLOSUM62 matrix and a word size of 3 was performed on the nonredundant NCBI database (January 2018) using the block of contiguous fragments containing the seed peptide. Only pseudogenized amino acid sequences with a significant similarity (e-value threshold $<2e^{-1}$) to the nudiviral protein used initially to find the seed were validated. Other less conserved sequences were therefore discarded. It should be noted that this validation process is very restrictive because the available software do not take into account peptides separated from the seed by large gaps.

To compare our pseudogene annotation protocol with an automated approach, we tested Exonerate (Slater and Birney 2005) using the Galaxy (Afgan et al. 2016) server at BIPAA (https://bipaa.genouest.org/is/; last accessed June 29, 2018). Using 16 OrNV proteins, we obtained 457,278 hit of genes on *V. canescens* genome without filtering. Among the best hits we identified 7 pseudogenes (*OrNVorf130*, *OrNVorf128.2*, *OrNVorf99.2*, *OrNVorf2*, *pif-1*, *vlf-1*, and *integrase*) compared with 18 pseudogenes characterized with our protocol and only the seeds could be retrieved, not the peptides present in their vicinity, altogether this indicates that the protocol used in this study is more sensitive than Exonerate for the identification of nudiviral pseudogenes.

Quantitative Polymerase Chain Reaction Assay

A total of three pools of 24 newly emerged females were used for DNA extraction with the Wizard Genomic DNA purification kit (Promega). DNA was extracted either from head and thorax or from dissected calyx in 1X PBS solution. Gene copy number was analyzed for six nudiviral genes: *Ac81.1* (F: ACCGTTCATCGCAGGACTAC, R: AAATTTGTGCGTTCGACTGA), RNA polymerase *lef-8* (F: GGATTGTCGGAAAGACTGGA, R: GTATCGAGGCTTTGGTGGAA) and *p47* (F: TCGCATCATACCGCTTATCA, R: CGTCAGCACCGTAACTTCAA), *pif-3* (F: AGATTGCGATGAGGAAATGG, R: CGCTAGTGAATTGACGCAAG), *p74* (F: TTTACGGTGGAAGGCTGGAA, R: TGTTTAATGTCTGCGGCACG), and *GbNVorf19-like.2* (F: TCTCAATTTAATATCCCGGTTG, R: TGGACTCTTTAGTCTCGGGAA) as well as the wasp gene *neprylisin* (F: AGCGAACTCTTGGTGCTCGGTA, R: TTCACCAGTACCAGCAGCTACG). The wasp gene elongation factor 1-alpha (*elf1*) was used as reference gene (F: TCATTGACGCTCCTGGACAC, R: TTCACCAGTACCAGCAGCTACG). For each gene, a standard curve was prepared with plasmid DNA (Pichon et al. 2015). The DNA copy number was quantified using LightCycler 480 SYBR Green I Master Mix (Roche) in 384-well plates (Roche). Polymerase chain reactions (PCRs) were performed using 5 ng of DNA and 10 μM of each primer in a 10 μl volume having a 1× master mix concentration. We determined DNA levels with a LightCycler 480 System (Roche), by heating at 95°C for 10 min, followed by 45 cycles of 95°C for 10 s, 58°C for 10 s, and 72°C for 10 s. Each sample was evaluated in triplicate and the results were normalized with respect to *elf1*. The ratio between the copy number of each gene (normalized) in the calyx and the copy number of each gene in the head and thorax were calculated using the software Qbase+ (Biogazelle).

Protein Similarities Analyses

To examine the overall and differential levels of selection exerted on exogenous versus endogenous nudiviruses, similarity analyses were performed on nudiviral core genes by aligning amino acid sequences between OrNV and GbNV and

between OrNV and VcENV using Geneious with the BLOSUM65 matrix. To determine the baseline of similarity between all these genes, the average percentage was calculated without incorporating data corresponding to VcENV pseudogenes and duplicated genes, and nor their homologues in OrNV and GbNV. Average percentages of protein similarity were obtained for both OrNV/GbNV and OrNV/VcENV groups. Finally, the difference of similarity with the average of each protein was calculated in the two groups. The difference of protein similarity with the average (expressed as a percentage) is shown for all core genes from the two exogenous nudivirus OrNV/GbNV and from the endogenous versus exogenous nudivirus (OrNV/VcENV). These values are calculated to measure the increase or decrease of each protein divergence and then to observe change correlated with viral domestication.

Results

Synteny Analyses of Exogenous and Endogenous Nudiviruses

To identify the modifications that have occurred after the integration of VcENV ancestor, such as the loss of a block of collinear genes, the best approach would be to undertake comparisons with the genome of the virus originally integrated. Although this genome is not available, comparisons of *V. canescens* nudiviral clusters with sequenced nudiviruses might nonetheless provide some clues on the organization of the virus ancestrally captured. Thus in order to investigate in more details the relationships among alphanudiviruses and related EVEs we compared gene orders by gene parity-plot using OrNV as a common reference (figs. 1*A*–1*D*). Comparisons between sequenced alphanudiviruses and OrNV aimed to determine whether related exogenous nudiviruses shared a similar gene order. Comparison between endogenous nudiviruses (NlENV and VcENV) and OrNV (figs. 1*B*–1*D*) was intended to determine whether this gene order was conserved after endogenization and to assess whether OrNV may represent a good proxy of VcENV ancestor. In parallel, we have undertaken a phylogenetic analysis of exogenous and endogenous nudiviruses using the concatenated alignment of up to 38 homologous genes from related dsDNA viruses including newly sequenced nudiviruses. The phylogenetic tree obtained (fig. 1*E*) indicated that NlENV was the closest relative to VcENV, whereas OrNV formed a clade with recently sequenced viruses infecting Diptera (DiNV, DmNV_kal and DmNV_tom). GbNV was identified as basal to the group and the most distant to OrNV within alphanudiviruses.

Gene parity plots of alphanudiviruses versus OrNV showed that gene order within these viruses is generally conserved (fig. 1*A*), with more than 85% of the genes found in syntenic regions for GbNV, the most distant to OrNV. Thus all available exogenous alphanudiviruses share globally a similar organization. In contrast, gene parity plots of NlENV versus OrNV showed a lower conservation of gene order; with only 66% of genes found in syntenic regions (fig. 1*B*) indicating gene order is less conserved despite the closer relationship of NlENV with OrNV compared with GbNV, as indicated by the phylogenetic analysis. Strikingly the gene parity plots of OrNV versus VcENV allowed us to identify only four regions containing 3 or 4 genes (microsynteny) conserved between the two viruses, these regions containing in addition a few genes (one or two) which localization is not conserved (figs. 1–3). Thus in VcENV clusters most of the genes appear to have been reshuffled compared with the organization of exogenous alphanudiviruses, which was most probably shared by VcENV ancestor.

Because of this extensive reshuffling (fig. 1*D*) it is not possible to use an exogenous nudivirus as a proxy of VcENV ancestor and to identify, by comparing VcENV to this proxy, an eventual loss of several genes as a block. Thus to study how genes are lost we used an alternative approach consisting in searching for remnants of nudiviral genes by scanning individually for their presence in *V. canescens* genome.

Determination of the list of lost core genes to search for in *V. canescens* Genome

Because viruses are known to pick up genes from the cells they infect (Hughes and Friedman 2003; Thézé et al. 2015), not all the genes present in a virus genome are informative regarding the family to which the virus belongs. For an exogenous virus, the virus genes are simply defined as those enclosed in the genome encapsidated in the particles independently on whether they have a viral or cellular origin. However for an endogenous virus it can be almost impossible to determine whether a gene sharing similarities with cellular genes originated from the ancestral virus, in particular if the viral genome has been dispersed in the host genome. However some genes are truly typical of viral genomes, in particular among core genes that are defined as the set of genes that are essential for the virus cycle and present in all members of a virus family. We first focused our analysis on these nudiviral core genes. We hypothesized that a whole nudivirus genome was originally integrated and that nudiviral core genes not expressed in *V. canescens* calyx might still be present in the wasp genome as remnants.

The list of 32 nudiviral core genes (Bézier et al. 2017) was used as a reference in the present study. The functions of nudiviral genes are unknown because they have not been studied experimentally, however, some genes can be tentatively assigned based on their homology to baculovirus genes. Endogenous nudiviruses producing bracoviruses and VcENV have both undergone a reduction of core genes set expressed compared with regular nudiviruses with the loss of 5 and 8 nudiviral core genes, respectively (table 1). Two core genes involved in DNA replication (*dnapol* and *helicase-2*) are lacking in both *V. canescens* and in bracovirus-associated wasps; whereas the status of tyrosine kinase (*tk1*, *tk2*, and *tk3*) is difficult to determine because they are too close to cellular proteins. Six core genes have been lost specifically in VcENV:

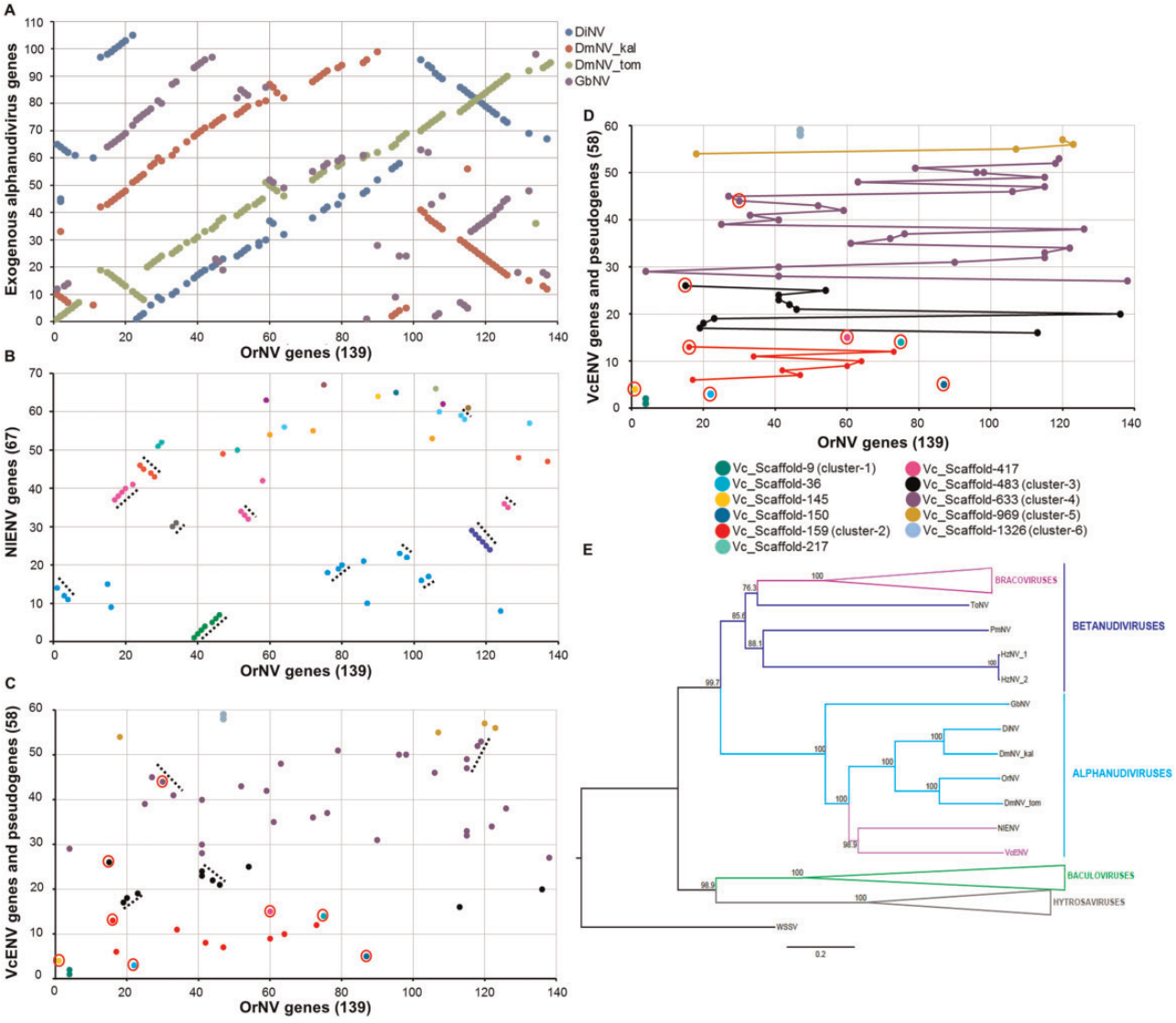

Fig. 1.—Gene parity plot comparisons of exogenous and endogenous alphanudiviruses. OrNV gene organization compared with (*A*) sequenced endogenous nudiviruses, (*B*) NlENV, and (*C*) VcENV. Genes are represented by dots following gene order within the OrNV genome on the *x* axis, the position of the homolog in the other genome being indicated on the *y* axis. For a better visualization of gene order conservation, NlENV and VcENV genes of the same cluster are shown using the same dot color. Syntenies are highlighted by black dotted lines. Dots corresponding to nudiviral pseudogenes identified are surrounded by a red circle. (*D*) Visualization of VcENV reshuffling compared with OrNV, using dots connected following VcENV clusters gene order. (*E*) Phylogenetic tree of dsDNA viruses including free and endogenous nudiviruses. Numbers on the node indicate bootstrap supports (1,000 replicates). Full names and accession numbers of the viruses included in the phylogeny are indicated in the materials and methods section. Clades corresponding to bracoviruses, baculoviruses, and hytrosaviruses were collapsed.

The integrase (*int-1*) and the flap-endonuclease (*fen-1*) involved in DNA replication, but also the four genes *vlf-1*, *vp39*, *38 K* encoding capsid components, and *p6.9* involved in viral DNA condensation, which is consistent with the lack of capsids and DNA in the VLPs (table 1).

Detection of Pseudogenes Corresponding to Lost Core Genes

We performed a thorough data mining of *V. canescens* genome to identify traces of lost genes. This search allowed the identification of remnants of nudiviral core genes that are not expressed in the calyx in which VLPs are produced (table 1 and fig. 4). Their nudiviral (versus wasp) origin was verified by reciprocal BLAST search. Once validated the peptides deduced from pseudogenes were aligned with homologous OrNV proteins (fig. 5). Strikingly we identified pseudogenes corresponding to most of the lost core genes (7 out of 11, table 1) including all those coding for the capsid components and DNA condensation (details of pseudogene annotation are reported in figs. 4 and 5). In particular we identified a

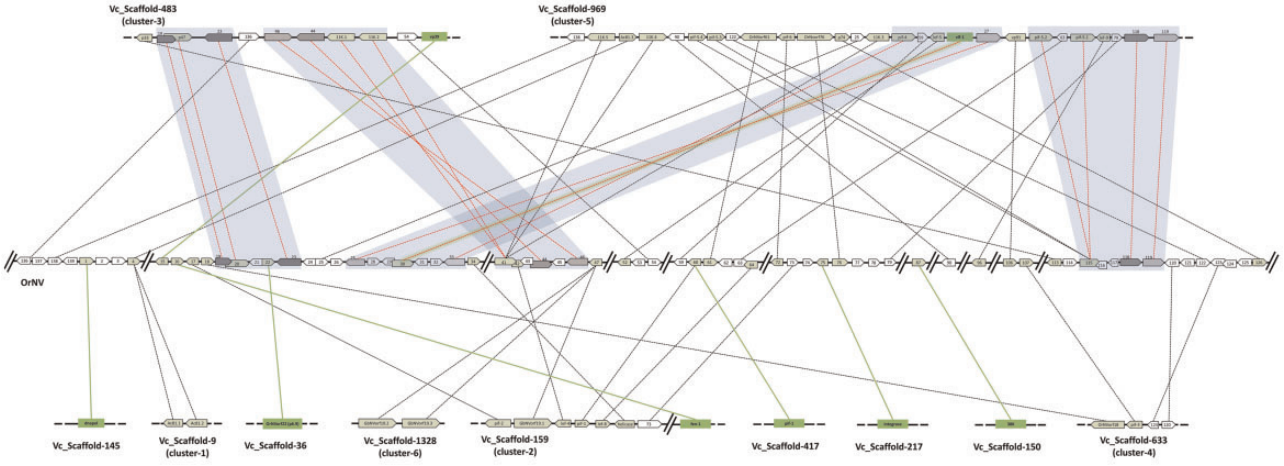

Fig. 2.—Overall comparison of VcENV and OrNV genomic organization. Boxes represent nudiviral genes and numbers correspond to their positions within the OrNV genome. Homologous genes are connected by dotted line. Syntenies are highlighted in light blue and conserved genes are visualized using gray boxes of different shades. Core genes are shown in light green and pseudogenes of lost core genes are shown in intense green.

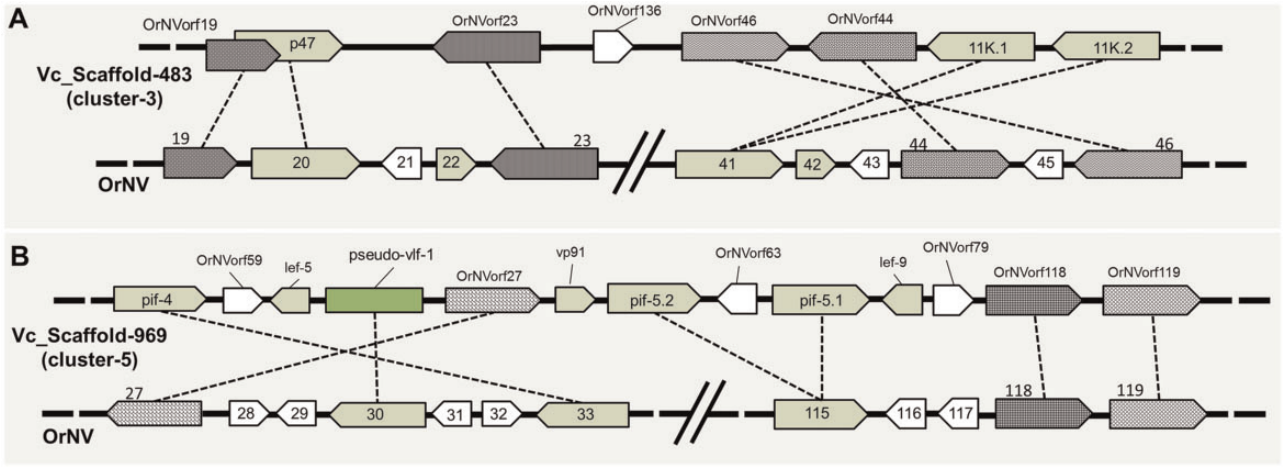

Fig. 3.—Details of microsyntenies between VcENV and OrNV. Homologous genes are displayed using the same gray shade and connected with dotted line as previous figure. Core genes are represented by light green boxes and pseudogenes by intense green. (*A*) and (*B*) display microsyntenies for nudiviral cluster 3 and nudiviral cluster 5, respectively.

pseudogene corresponding to the nudiviral DNA polymerase (*pseudo-dnapol*), which has been unsuccessfully searched for in the DNA of braconid wasps. Indeed DNA replication is the sole viral function lacking to bracoviruses, the amplification of viral sequences in calyx cells probably involving wasp replication genes (Bézier et al. 2009; Louis et al. 2013; Burke et al. 2015). Thus in *V. canescens*, and unlike in bracovirus-associated species, the genes functionally lost can still be detected as remnants in the genome (table 1). Pseudogenes could not be formally characterized for four core genes (*tk1*, *tk2*, *tk3*, and *helicase-2*) although several candidate sequences could be found, because they are too closely related to insect proteins to firmly conclude whether they have a viral or cellular origin. Indeed these genes are conserved in nudiviruses but not typical, which make them difficult to formally identify as viral within the context of the wasp genome.

The nudiviral pseudogene annotation revealed that they contained a large number of stop codons, and that coding sequences have been fragmented by numerous deletions and insertions (figs. 4 and 5). The alignments between peptides deduced from pseudogenes and their OrNV homologues show that the similarity and the size may vary from one fragment to another, the best match being found, as expected, with the sequence containing the seed (63–167 amino acids; 26–33% of similarity). The differences of similarities along the reconstructed protein could reflect an ancestral divergence

Table 1

Pseudogenes Identified within *V. canescens* Genome Corresponding to Lost Nudiviral Core Genes

Function	Gene Name	Bracovirus	VcENV	Transcript Abundance in *Venturia* (RPKM)
DNA replication	*dnapol*	–	*Pseudogene*	4
	helicase	+	+	27
	helicase 2	*Nd*	*nd*	*Nd*
	Integrase	+	*Pseudogene*	0
	fen-1	+	*Pseudogene*	>1
Transcription	*lef-4 (RNApol s.u.)*	+	+	488
	lef-5 (RNApol s.u.)	+	+	245
	lef-8 (RNApol s.u.)	+	+	125
	lef-9 (RNApol s.u.)	+	+	724
	p47 (initiation factor)	+	+	916
Envelope component	*p74*	+	+	12,900
	pif-1	+	+[a]	5,196
	pif-2	+	+	1,980
	pif-3	+	+	2,655
	pif-4	+	+	7,731
	pif-5 (4 copies[b])	+	+	527; 863; 1,048; 2,286
	pif-6	+	+	3,011
	p33	+	+	5,553
	vp91	+	+	2,966
	Ac81 (3 copies[b])	–	+	272; 1,777; 1,080
	11K (5 copies[b])	+	+	40,978; 41,939; 980; 1,831; 4,225
Capsid component	*38K (major component)*	+	*Pseudogene*	0
	vp39 (major component)	+	*Pseudogene*	4
	vlf-1	+	*Pseudogene*	17
	OrNVorf22 (p6.9)	+	*Pseudogene*	9
Unknown	*GbNVorf19 (3 copies)*	+	+	4,555; 127; 1,032
	OrNVorf61 (PmNVorf62)	–	+	4,044
	OrNVorf18 (PmNVorf99)	+	+	6,319
	OrNVorf76 (HzNVorf143)	–	+	4,012
Nucleotid metabolism	*tk1*	*Nd*	*nd*	*Nd*
	tk2	*Nd*	*nd*	*Nd*
	tk3	*Nd*	*nd*	*Nd*

NOTE. —Table cells are highlighted in light or intense green depending on the presence of a functional gene or a detected pseudogene. Functions are those determined for homologous baculovirus proteins. Core genes of nudiviral origin conserved in braconid wasps associated with bracoviruses and VcENV are listed for comparison. The different values indicated in transcript abundance column correspond to the expression of the different gene copies.

[a]Pseudogene identified in addition to the functional core gene.

[b]Gene family.

–, no pseudogene detected; nd, not determined; s.u., subunit.

between OrNV proteins and those of the captured alphanudivirus at the origin of VcENV (fig. 5). For example, essential domains for the protein function may have originally been more conserved between these two viruses. The smallest peptides tend to be more similar but this is most probably a bias, because peptides with the same size but less similar would not be validated. Pseudogene expression level (table 1) was assessed using the previously obtained calyx transcriptome (Pichon et al. 2015). At most, few reads corresponding to pseudogenes were detected indicating that they were not expressed (from 0 to 17 RPKM for *int-1* and *vlf-1*, respectively, the reads from the latter presenting an uneven coverage of the sequence). In contrast, functional nudiviral genes were generally expressed at a much higher level in the calyx (more than 100–12,000 RPKM) except for the *helicase* gene (27 RPKM) (Pichon et al. 2015).

It is noteworthy that three pseudogenes (*vlf-1*, *fen-1*, and *vp39*) are present within nudiviral clusters (figs. 2 and 4), a localization that could correspond to their position in the captured ancestral virus. This interpretation is particularly supported in the case of *vlf-1*, which is located in one of the four blocks of microsynteny observed between OrNV and VcENV (figs. 2 and 3). Thus these genes have specifically undergone pseudogenization while adjacent genes have remained functional and expressed in the calyx. Four pseudogenes were identified as dispersed in the wasp genome (i.e. they are not located within nudiviral clusters), in regions that also contain wasp genes (figs. 2 and 4) confirming they are

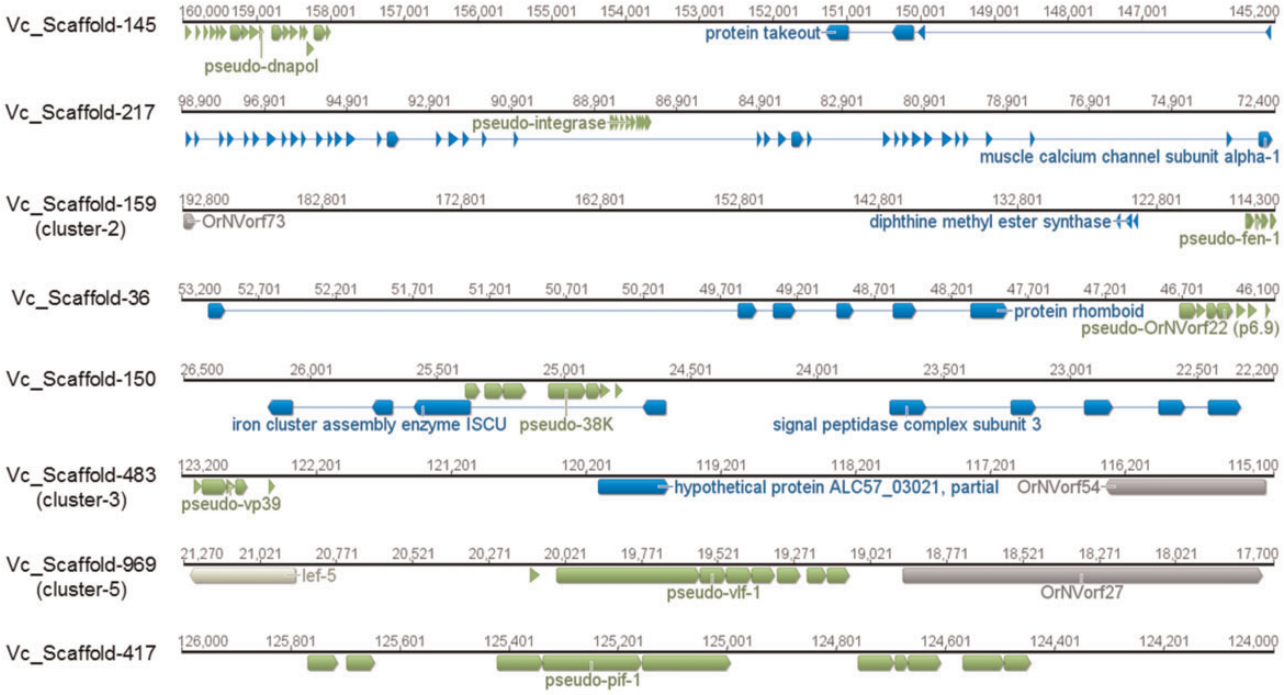

Fig. 4.—Schematic representation of pseudogenes identified within *V. canescens* genome (positions indicated refer to the whole scaffold). The pseudogenes corresponding to lost nudiviral core genes are shown in intense green, each stretch of coding sequences identified being represented by an arrow. Cellular genes are represented blue; the arrows correspond to the exons and the lines to the introns. Conserved nudiviral genes of VcENV are represented as gray and light green boxes for accessory and core genes respectively (they do not contain introns). The name of scaffolds and nudiviral clusters containing pseudogenes are indicated in the left column.

also integrated in the wasp genome. In addition a pseudogene of *pif-1* was identified but it represents a particular case because a functional copy is still present within the nudiviral cluster 2. This pseudogene was probably generated from a duplicated copy of *pif-1*. We did not find polyA tracts downstream of any of the pseudogenes identified, thus there is no evidence that they originated from the reintegration of a cDNA by a retrotransposase (retro-processed genes), although we cannot formally exclude that some of them might have been produced by this mechanism.

Detection of Pseudogenes Corresponding to Accessory Nudiviral Genes

Pseudogenes of most core genes have been retained although no longer providing a function for the host, which suggests that the time elapsed after their functional inactivation was not sufficient to allow their complete elimination from the wasp genome. We reasoned that it might also be the case for accessory genes. Thus we performed a comprehensive search of accessory genes that could remain as pseudogenes in *V. canescens* genome. We used as a reference a list of genes based on the accessory genes present in sequenced alphanudiviruses, with the exclusion of the genes with high similarity to cellular genes that have probably recently been acquired by the viruses (Hughes and Friedman 2003; Thézé et al. 2015). Using the same stringent validation method as for core genes we identified and validated 10 pseudogenes for accessory nudiviral genes (figs. 6 and 7). Three of them were located within nudiviral clusters (*pseudo-OrNVorf9-like*, *pseudo-OrNVorf62-like*, *pseudo-OrNVorf2-like*), six were isolated in new scaffolds containing at least one wasp gene (*pseudo-OrNVorf139-like*, *pseudo-OrNVorf99-like1*, *pseudo-OrNVorf99-like2*, *pseudo-OrNVorf28-like*, *pseudo-OrNVorf117-like*, *pseudo-OrNVorf130-like*), and one was identified in a very small scaffold (*pseudo-OrNVorf128-like*) of undetermined localization.

Measure of Active Gene Copy Number in the Calyx

Genes involved in VLP production are all located within the nudiviral clusters, although the genes from the smallest clusters (clusters 1 and 6) could be considered as dispersed because these clusters contain only two nudiviral genes. In contrast, among pseudogenes identified in large scaffolds only six are located in nudiviral clusters, whereas ten are apparently dispersed in the wasp genome. This higher dispersion of pseudogenes suggests that selection maintains active genes within nudiviral clusters. In wasps associated with bracoviruses a subset of nudivirus genes is maintained in a single cluster, which is amplified in the calyx (Louis et al. 2013; Burke et al. 2015) allowing high level of particles production. We

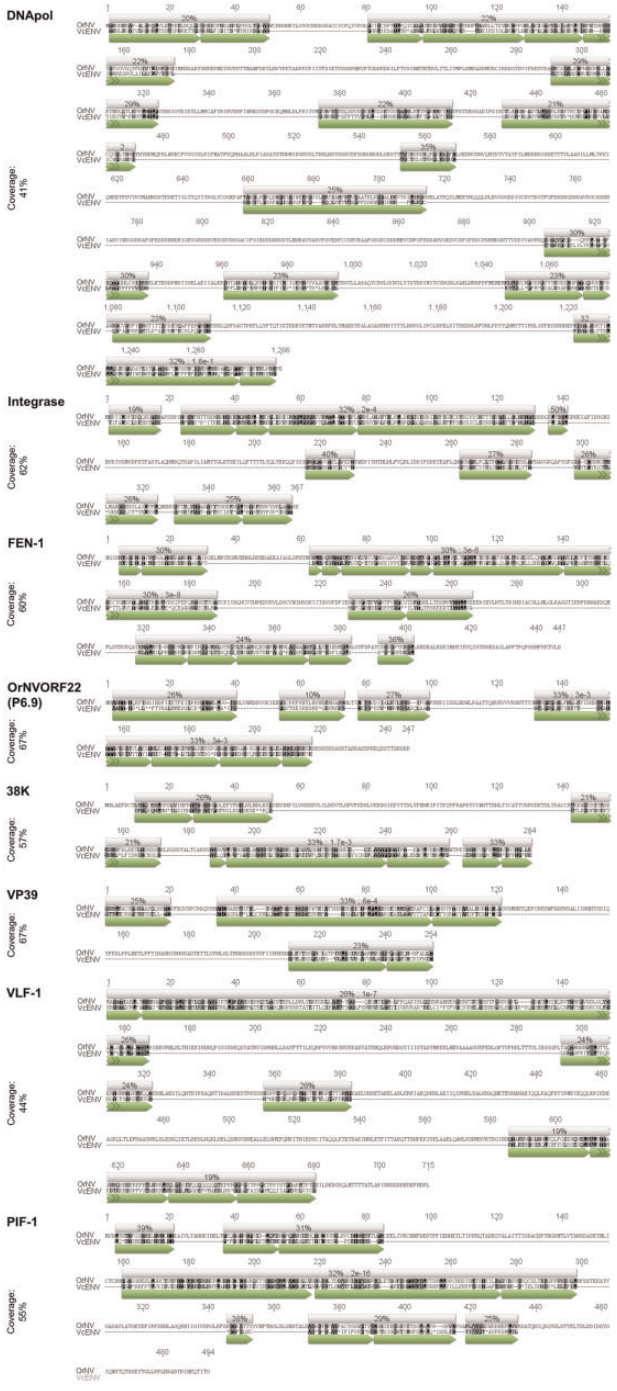

Fig. 5.—Alignment between peptides deduced from pseudogenes (highlighted by green boxes) and homologous OrNV proteins. Amino acid similarity is reported for each peptide in the box above the alignment. The e-value obtained by reciprocal BlastP is indicated in the box corresponding to the block of peptides used for validation. The query coverage of peptides deduced from a pseudogene with respect to OrNV protein homologue is given in the first column (protein coverage). Of note for *p6.9* the peptides identified correspond more precisely to *OrNVorf22-like* which product contain P6.9 in most *Alphanudiviridae*, the peptide precisely corresponding to P6.9 has been lost.

hypothesized this could be the case for genes involved in the abundant production of VLPs in the calyx. To assess whether nudiviral clusters were amplified, we performed a quantification of gene copy number using a quantitative PCR (qPCR) assay. We compared relative copy number of one gene representing each nudiviral cluster to that of a single copy gene encoding for the neprilysin (fig. 8), a metalloprotease packaged in VLPs. The genes from the two clusters containing two nudiviral genes (clusters 1 and 6) were not found to be significantly amplified compared with the *neprylisin* gene, whereas amplification was detected for the genes corresponding to the other clusters. The *lef-8* gene (from cluster 2) was amplified 7-fold and the *p47*, *pif-3*, and *p74* genes (from clusters 3, 4, and 5, respectively) were amplified at a similar level, considering standard deviation, of approximately 25-fold (fig. 8). This amplification probably has a functional role, which may explain that active genes are maintained in the clusters, whereas pseudogenes are relatively more dispersed in the wasp genome.

Evolution of Core Genes in the Context of Symbiogenesis

Following integration and domestication leading to the production of VLPs, nudiviral genes have evolved depending on the particular constraints imposed by the symbiotic association with the wasp, which are most likely very different from that acting on exogenous viruses. To evaluate which genes have been well conserved, or have particularly diverged, we compared homologous proteins divergence between OrNV and the exogenous nudivirus GbNV and between OrNV and the domesticated nudivirus VcENV (fig. 9). Because *V. canescens* is the only species in the Campopleginae known to be associated with VLPs, it is not possible to perform direct measurements of selection pressures by calculating ratios of synonymous versus/non synonymous mutation (dn/ds) by comparing homologues from closely related species. We measured the conservation of the different proteins relative to the average of core proteins corresponding to each comparison separately (VcENV/OrNV or GbNV/OrNV) because the average similarity between OrNV and VcENV is different from that between GbNV and OrNV proteins. With these data we could tentatively assess which VcENV core proteins were more conserved (higher similarity than the average of the core proteins) or more divergent (lower similarity) and examine whether their homologues in exogenous nudiviruses follow the same pattern. Unsurprisingly, amino acid sequences of pseudogenes have a remarkable lower than average protein similarity than their pathogenic homologues from OrNV and GbNV (fig. 9). This is probably because these genes have evolved under neutral selection after pseudogenization. The different copies of all duplicated genes (*pif-5-1* to *4*, *11K-1* to *5*, *GbNVorf19-like 1* to *3*) are also more divergent than the average suggesting duplication may have relaxed the selection pressure on individual copies. The *helicase* gene showed also a greater divergence, which was also reported as a general trend of nudiviruses (Hill

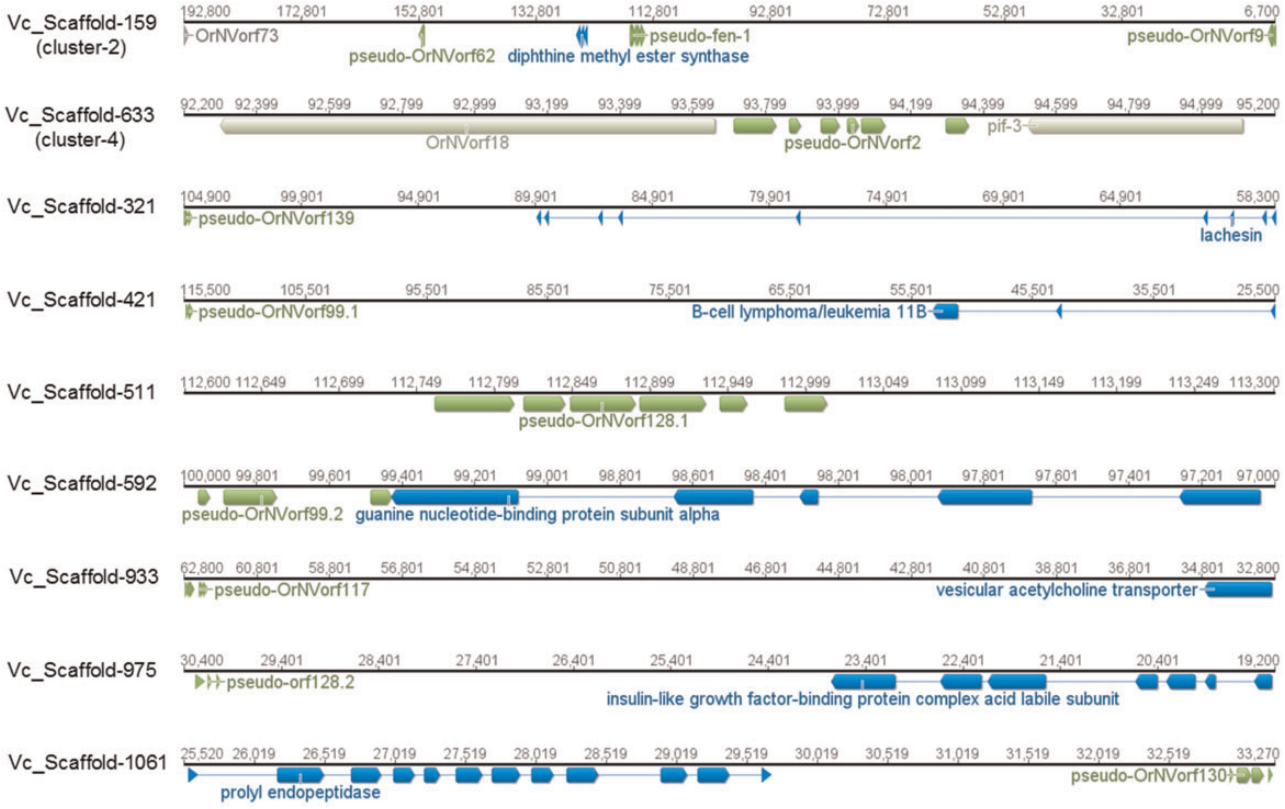

Fig. 6.—Schematic representation of nudiviral accessory pseudogenes identified within *V. canescens* genome. The pseudogenes corresponding to lost nudiviral genes are shown in green, each stretch of coding sequences identified being represented by an arrow. Cellular genes are represented in gray, the arrows correspond to the exons and the lines to the introns. Conserved nudiviral genes of VcENV are shown in blue (they do not contain introns). The numbers indicated in the left column are those of the corresponding scaffolds or clusters in *V. canescens* genome.

and Unckless 2018). The other prominent features of this analysis are the higher than average conservation of the proteins involved in viral transcription and of several PIF proteins (for *Per os* Infectivity Factors that are potentially involved in cell entry) in VcENV compared with exogenous nudiviruses. For proteins controlling viral transcription the effect is specific to VcENV, whereas the same PIF proteins are also conserved between GbNV and OrNV. These results suggest that a functional constraint is operating in the context of domestication to maintain the sequence of genes involved in viral transcription.

Discussion

Synteny between Endogenous and Exogenous Alphanudiviruses

Syntenies between VcENV and other endogenous and exogenous nudiviruses might provide a better understanding of the evolution and virus genome rearrangements after integration in the host genome. This could provide insights on how large dsDNA viruses evolve after endogenization, in the context of a mutualistic association with a eukaryote. By comparing OrNV with the other sequenced alphanudiviruses, we first determined whether gene order was conserved within the *Alphanudivirus* genus. We showed that 85% of the genes shared between OrNV and GbNV, the most distant alphanudiviruses from phylogenetic analysis, were conserved in syntenic blocks (fig. 1). This is comparable to the synteny observed between *Autographa californica multiple nucleopolyhedrovirus* (AcMNPV) and *Helicoverpa armigera single nucleopolyhedrosis virus* (HaSNPV), from the *Alphabaculovirus* genus (Chen et al. 2001). This result shows that despite the low conservation of gene order to the scale of the families *Baculoviridae* or *Nudiviridae* (Herniou et al. 2003; Wang and Jehle 2009), many genes can be found in synteny at the genus scale. Thus it should be possible to use any exogenous alphanudivirus as a proxy of VcENV ancestor. However, by comparing OrNV with the two reported endogenous alphanudiviruses we found that OrNV gene order was not well preserved in NlENV and even much less in VcENV, to the exception of four small blocks of microsynteny (figs. 1–3). Of note, the order of several genes present in microsyntenies between VcENV and OrNV, is also conserved in all alphanudiviruses. Indeed *OrNVorf19-like* and *p47* are contiguous (overlapping in VcENV), as well as *OrNVorf118-like* and *OrNVorf119-like* in both VcENV and OrNV and the microsynteny containing the genes from *OrNVorf41-like* (*11K*) to *OrNVorf44-like* are

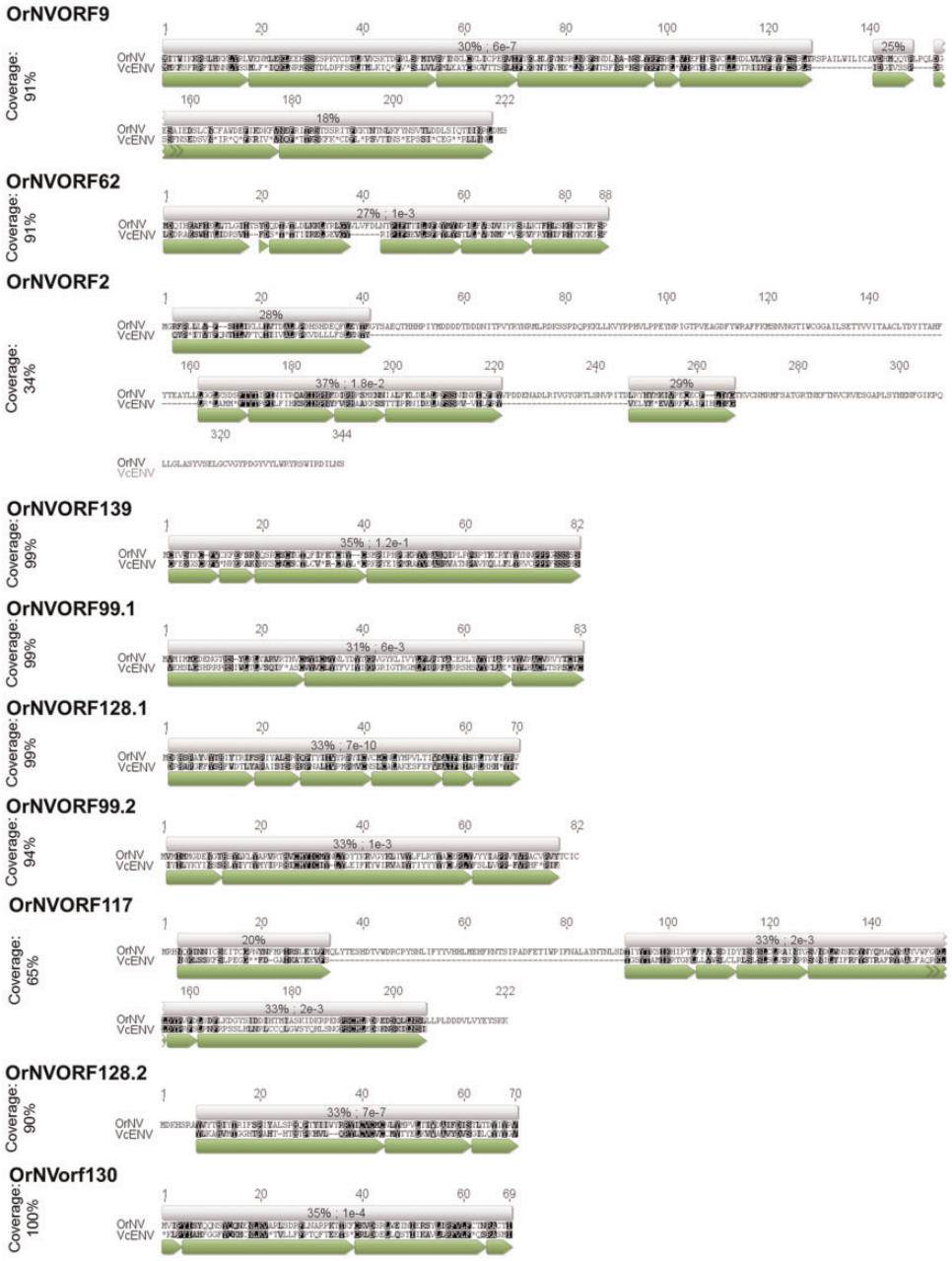

Fig. 7.—Alignment between peptides deduced from pseudogenes (highlighted by green boxes) and homologous OrNV proteins corresponding to nudiviral accessory gene products. The percentage of identity is reported for each peptide in the box above the alignment. The *e*-value obtained by reciprocal BlastP is indicated in the box corresponding to the block of peptides used for validation. The query coverage of peptides deduced from a pseudogene with respect to the protein homologue in OrNV is given in the title of each alignment.

observed in all known alphanudiviruses. This suggests that at least these regions have retained the organization of the ancestral virus. Overall we conclude that during viral domestication the ancestral nudiviral genome has been dispersed in clusters in the host genome and has undergone such reshuffling that the original gene order has only marginally been retained. The number of alphanudivirus sequences available is still limited. More extensive genome sequencing efforts may uncover an exogenous nudivirus more closely related to VcENV (i.e. belonging to the subclade of NlENV and VcENV). However, it is unlikely that it will share much more syntenies with VcENV than the alphanudiviruses already available. Genome dynamics might be involved in part in endogenous virus reshuffling, because comparison of 166 mosquito genomes have shown that small-scale rearrangements disrupt gene order within chromosomes arms over time, although at the chromosome level gene content is globally maintained (Neafsey et al. 2015).

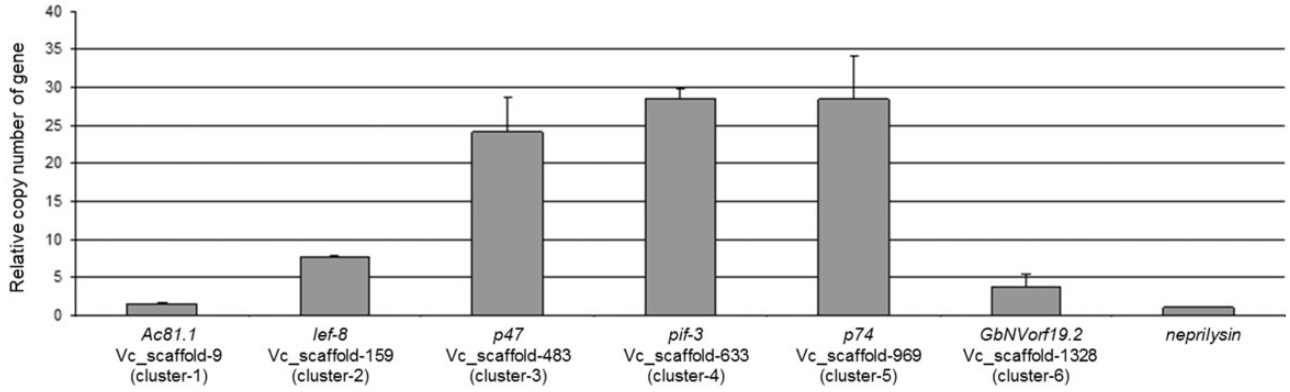

Fig. 8.—Relative quantification in adults wasp calyx of nudiviral genes DNA copies from nudiviral cluster compared with *neprilysin* gene, for which only one copy was detected in the *V. canescens* genome.

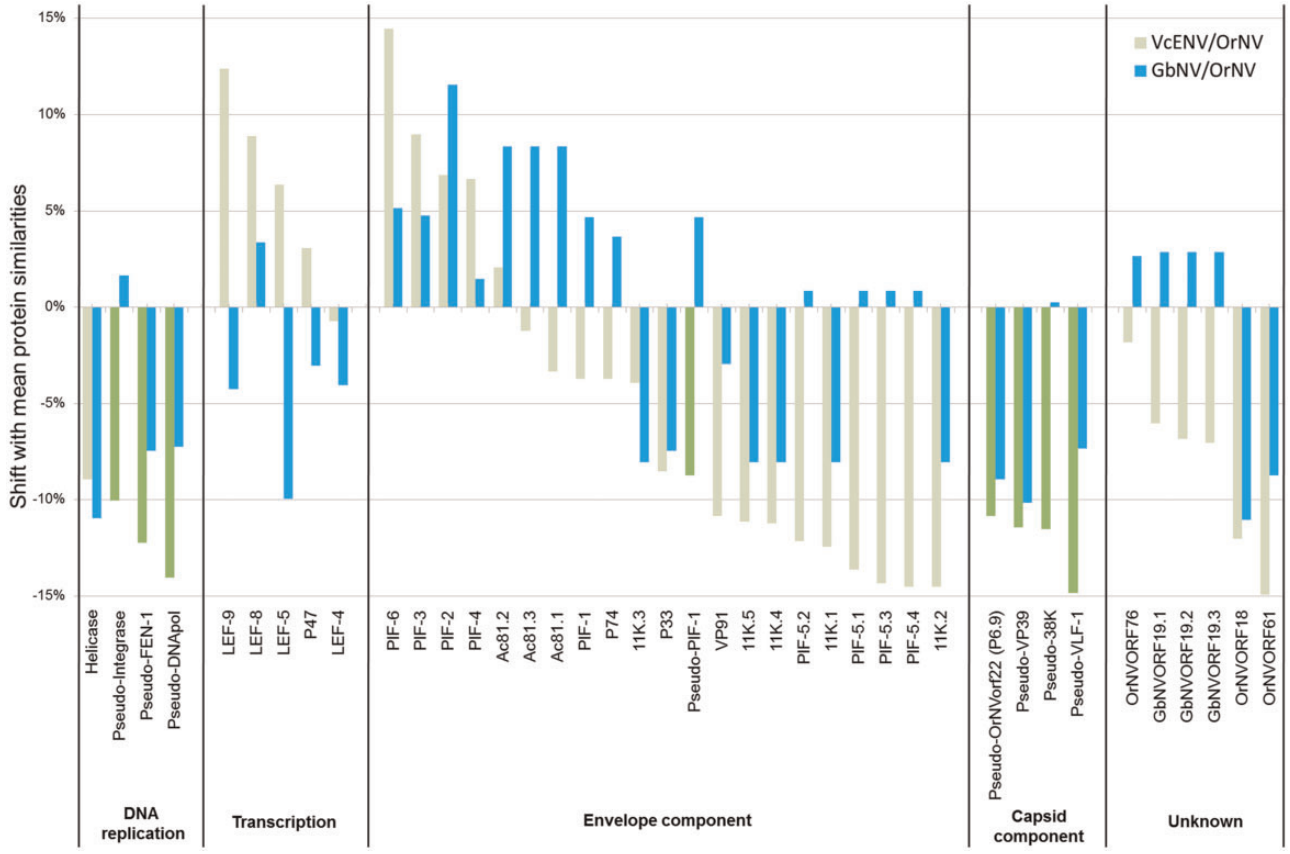

Fig. 9.—Divergence of the different nudiviral proteins relative to the average similarity of core proteins for the two comparisons VcENV/OrNV (in light green) and GbNV/OrNV (in blue). The percentages indicate higher (bottom of the panel) or lower (top of the panel) divergence of particular nudiviral proteins with respect to the mean similarity of the whole set of core proteins that are common to the two viruses used in each comparison. The genes are functionally classified according to table 1. The results for proteins reconstituted from pseudogenes (not used to calculate mean divergence) are showed in intense green.

Lost and Conserved Genes

Although VLPs and bracoviruses, respectively, originated from an alphanudivirus and a betanudivirus, they contribute to different parasitic strategies both allowing success of wasp larvae development within the parasitized lepidopteran host. Viral particles of braconid wasps constitute a gene transfer system in which virulence genes are expressed by infected host cells (Gundersen-Rindal et al. 2013), whereas in *V. canescens* virulence factors are produced by the wasp and packaged into VLPs, which function as a protein delivery system (Pichon et al. 2015). To compare both models, it might be useful to resume first what we know about well-studied

bracoviruses. Whereas during a regular cycle, virus particles infect target cells and replicates producing a viral progeny, the bracovirus life cycle is split between the wasp and its parasitized host. Virus particles production (corresponding to virus replication for a pathogenic virus) occurs in the calyx of the wasp and infection occurs in the lepidopteran host. In the calyx, wasp genes products involved in specific viral transcription during an infectious process, such as nudiviral RNA polymerase subunits, control the expression of some other viral genes, such as those coding for nucleocapsid components and viral envelope proteins (Drezen et al. 2012; Burke et al. 2013). Endogenous viral sequences are amplified in calyx cells (Pasquier-Barre et al. 2002; Burke et al. 2015) and are processed to produce the dsDNA circular molecules that are packaged into bracovirus particles. In the parasitized host the particles infect cells using a conserved set of viral envelope proteins (Burke et al. 2013) known as PIF proteins shown to be essential for baculoviruses primary infection in midgut cells (Peng et al. 2010). The circles present in the particles are released in the nuclei of host cells. Then these cells express circle genes (Chevignon et al. 2014) and the virulence proteins thus produced protect wasp larvae from host immune defences (Strand 2012). They are also involved in altering host physiology and developmental timing for parasite benefit (Dushay and Beckage 1993).

VcENV originated from a nudivirus more recently than bracoviruses (probably several Mya versus 100 Mya) but has lost one more viral function, which is the ability to package DNA into nucleocapsids. This loss of function correlates with the loss of many nudiviral core genes. Like bracoviruses, VcENV has lost functional genes involved in DNA replication, in particular the nudiviral *DNA polymerase* gene (table 1). There is apparently no need for DNA polymerase in *V. canescens* because VLPs do not encapsidate DNA and that VcENV is exclusively transmitted vertically as a part of the wasp genome. Nudivirus clusters might be amplified by wasp cellular machinery, as suggested for that of the bracovirus (Bézier et al. 2009; Burke et al. 2015). VcENV has also specifically lost other proteins (table 1) potentially involved in viral DNA processing such as P6.9 ensuring condensation of the DNA molecules required for their encapsidation in the particles (Wang et al. 2010). The integrase INT-1, the tyrosine recombinase VLF-1 (Vanarsdall et al. 2006) and the flap-endonuclease FEN-1 (Lieber 1997) have also been lost. These proteins are conserved in bracoviruses and probably contribute to produce the dsDNA circles packaged in the particles and/or to integrate these molecules into host cell DNA. VcENV has also lost nucleocapsid components that constitute the shells protecting the viral DNAs in the particles, in particular VP39 and 38K, which are major component of baculovirus capsids (Wu et al. 2008; Danquah et al. 2012), nudivirus capsids (Bézier et al. 2017) and bracovirus particles (Bézier et al. 2009; Wetterwald et al. 2010; Burke et al. 2013).

Loss of function is thought to occur through point mutations or small insertion or deletions generating a pseudogene that is eroded by subsequent rearrangements (Ding et al. 2006; Xiao et al. 2016). After inactivation, a gene accumulates a large number of substitutions, insertions, deletions, frame-shifts and stop codons to such an extent that the history of events is impossible to retrace (figs. 4–7). This is the case for pseudogenes of nudiviral core genes that have accumulated many mutations (figs. 4 and 5) but are still detectable in *V. canescens*, whereas with the same approach we did not detect any remnant of the nudiviral *DNA polymerase* in the genome of *Microplitis demolitor* producing MdBV (Burke et al. 2014). This is consistent with the more recent origin of VcENV-wasp symbiotic association.

In contrast, all the genes involved in viral transcription encoding the four RNA polymerase subunits LEF-8, LEF-9, LEF-4, and P47 (Guarino et al. 1998) have been conserved, suggesting that the viral RNA polymerase still controls the expression of other nudiviral genes (or at least a subset of them) like in bracoviruses (Burke et al. 2013). Moreover, these genes show lower than average divergence suggesting that a constraint is acting in the context of virus domestication to maintain their function (fig. 9). This might be related to the observation that intron acquisition by nudiviral genes seems to be very rare in VcENV such as in bracoviruses. In fact, only one nudiviral gene of VcENV, *lef-9* coding for RNA polymerase subunit, was found to contain introns, which was specifically verified by sequencing (Pichon et al. 2015). Furthermore, as far as we know all nudiviral genes from bracoviruses have retained a viral structure. Altogether this suggests that a specific constraint is operating on nudiviral genes transcription to preserve their viral structure. Transcription of nudiviral genes could occur in a specialized ultrastructure such as described for intronless histone genes, which are transcribed in histone-like bodies (Hentschel and Birnstiel 1981; Nizami et al. 2010).

The whole set of PIF proteins, which are involved in cell entry and thus infectivity, is also conserved suggesting that they are used to target host cells. In particular two components that form a molecular complex (PIF-2 and PIF-3) (Peng et al. 2012) show higher than average protein similarity to their OrNV and GbNV homologues (fig. 9), suggesting they might be under conservative selection. The conservation of the whole set of genes involved in infectivity, together with the nature of the virulence factors corresponding to molecules present in venoms of other parasitoid wasps (Parkinson et al. 2002; Labrosse et al. 2005; Colinet et al. 2007)—including a metalloprotease (Asgari et al. 2002) and a RhoGTPase activating protein (RhoGAP) (Reineke et al. 2002)—strongly suggest that VLPs do not confer a passive protection toward encapsulation by coating the parasitoid eggs as originally proposed (Feddersen et al. 1986), but rather deliver virulence proteins directly within host haemocytes inhibiting host defences by a local but active mechanism (Pichon et al. 2015).

Surprisingly, although VcENV has lost many nudiviral core genes, it has conserved a few genes lost by bracoviruses (table 1). Thus *Ac81*, which in baculoviruses encodes an

envelope protein involved in wrapping the particles in the viral envelope (Dong et al. 2016), is present in 3 copies in *V. canescens*. This is also the case for homologues of the core genes *OrNVorf76-like* (*HzNVorf143-like*) and *OrNVorf61-like* (*PmNVorf62-like*) for which no precise function has been identified to date, except *OrNVorf61-like* product is a VLP component and a major component of ToNV particles (Bézier et al. 2017). VLPs are secreted by budding successively through the nuclear and the cytoplasmic membrane (Pichon et al. 2015), as nudiviruses (Crawford and Sheehan 1985), whereas bracoviruses are released by calyx cell lysis (Wyler and Lanzrein 2003). So, it is possible that some genes among those specifically conserved in VcENV could be involved in this particles secretion mechanism.

Given the success in identifying core genes we reasoned that remains of accessory genes could also be found within *V. canescens* genome. Accordingly, we identified additional pseudogenes corresponding to 10 nudiviral proteins. Because they have no homologues in baculovirus there is no clue on their function. Only three of them were found in the nudiviral clusters confirming the trend for the dispersion of pseudogenes compared with active genes in the wasp genome. In the human genome, active L1 retrotransposons are thought to provide the retrotransposase allowing genomic integration of cDNA from genes expressed in germline cells (Zhang et al. 2003). Similarly a few virulence genes encoded by bracovirus circles are suspected to derive from cDNA retrotransposition in the proviral form of the bracovirus genome (Serbielle et al. 2008, 2012). However, we did not identify clues supporting such an origin of VcENV pseudogenes (in particular polyA tracts downstream of the pseudogenes), although retrotransposition would result in the production of dispersed genes.

Duplicated Genes

Another characteristic feature of VcENV is the presence of multicopies of some nudiviral core genes. This is the case for *pif-5* (four copies), *GbNVorf19-like* (three copies), and *11K* (five copies). Gene families can be found in large dsDNA virus genomes but these particular genes are generally present as a single copy in nudiviruses, except *pif-5*, which is present in three copies in ToNV genome (Bézier et al. 2015), and two copies in DmNV_kal and DiNV genomes (Hill and Unckless 2018). For most duplicated genes a strong decrease in similarity was observed compared with other core genes (fig. 9). This increase in protein divergence for duplicated genes could reflect a functional diversification of proteins after duplication (Francino 2005). Among these three genes, only the PIF-5 function has been studied in baculoviruses, but its precise role in infection is still unknown. PIF-5 is one of the factors required for baculovirus primary infection (*pif-5* gene deletion results in viruses unable to infect by the feeding route), and it thus formally encode a "*per os* infectivity factor"; however, it was shown that PIF-5 does not participate in binding and fusion to targeted cells (Sparks et al. 2011) suggesting it might be involved in infectivity prior to binding to midgut cells or after fusion of the particles with the cells. Because PIF-5 is not a component of the PIF complex (constituted by PIF-1, PIF-2, PIF-3, and PIF-4 [Peng et al. 2012], modified PIF-5 will not inhibit the production of this complex, which would favor *pif-5* copy divergence. Accordingly, a relaxed selective constraint was reported recently for a DiNV duplicated *pif-5* (Hill and Unckless 2018). More broadly, the chromosomal transmission of VcENV genes might favor the conservation of duplicated genes because the constraint operating on the size of the viral genome does not operate on the endogenous virus. Indeed, in contrast to conventional viruses, for which all the genes necessary for the infectious cycle need to be packaged in the particles, the nudiviral machinery is chromosomally integrated, and therefore does not rely on virus particles for its transmission, which occurs exclusively vertically.

Maintenance of Core Genes in the Integrated Wasp Genome

Altogether the identification of pseudogenes for most of the core genes lost in *V. canescens* genome and the common origin of both genes and pseudogenes (indicated by BLAST analyses) suggest that a single integration event of a nudivirus occurred in *V. canescens* lineage (fig. 10). Moreover, the genome of the virus ancestrally captured was most probably complete, because core genes are generally dispersed in nudivirus genomes, implying that a truncated genome would lack at least some of these core genes. Integration into host DNA may have constituted a property of the ancestor nudivirus, as suggested for Heliothis zea nudivirus 1 (HzNV-1) (Lin et al. 1999). Pseudogenes are distributed in various localizations in *V. canescens* genome, they can be found alone or in nudiviral cluster close to functional viral genes (figs. 2, 4, and 6). Moreover, one pseudogene (*pseudo-vlf-1*) is present in a microsyntenic region conserved in alphanudiviruses, which most likely correspond to the original gene position within VcENV ancestor genome (figs. 1 and 3). Concerning the evolution of the endogenous virus in the context of symbiotic relationship with the host, an important conclusion is that the lack of conservative selection appears as the driving force leading to specific gene loss by pseudogenization. Pseudogenes would have been produced by accumulating deleterious mutations leading to the functional inactivation of genes not contributing to VLP production. Although virus genome reshuffling makes it difficult to trace rearrangements (figs. 1 and 2), there is no evidence that genes may be lost as large chunks of DNA containing several genes during dispersion of viral genome in the wasp genome, because in this case many core genes would have been lost, with no trace remaining in the wasp genome (fig. 10).

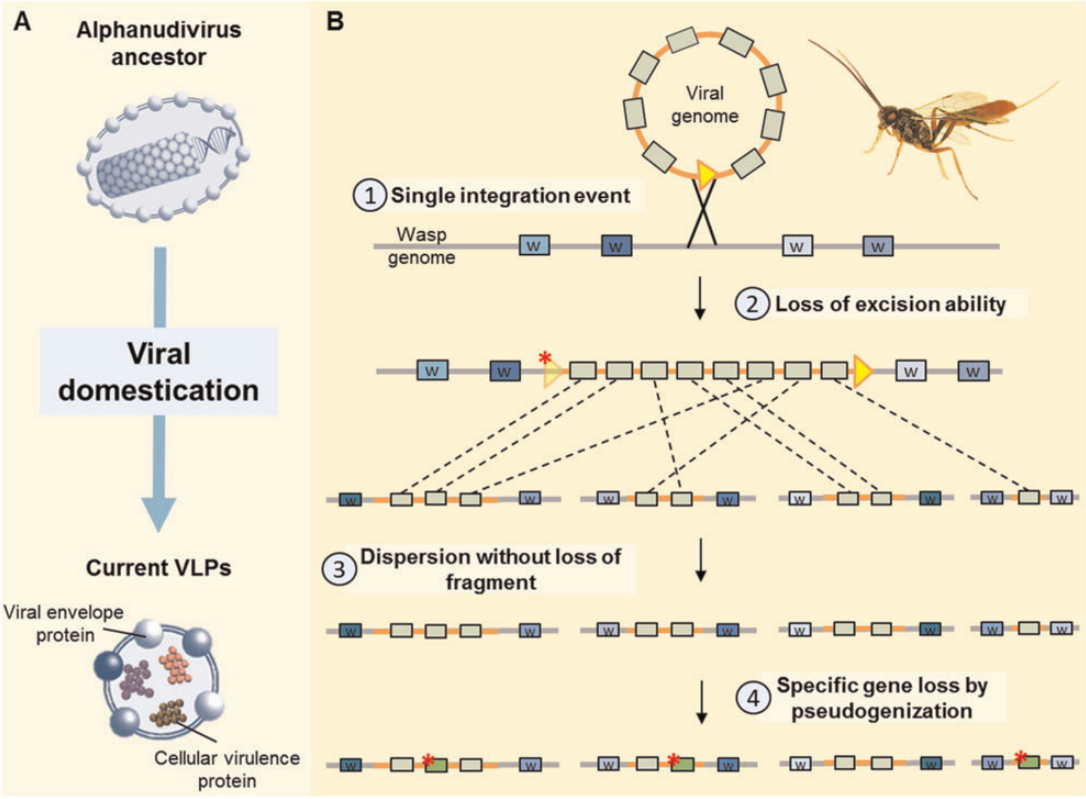

Fig. 10.—Hypothetical scenarios of early evolution of a large dsDNA virus integrated within *V. canescens* genome that provides a beneficial function to its host. (*A*) Production of *V. canescens* VLPs from an endogenous alphanudivirus, the VLPs correspond to viral envelopes produced by VcENV used to package virulence proteins of wasp origin. (*B*) Schematic representation of viral genome evolution within the host genome. (1) A complete nudivirus genome was ancestrally integrated possibly via a specific mechanism involved in latent infection. (2) Loss of excision ability might have resulted from a deletion of a regulatory sequence involved in integration and excision. (3) The viral genome was later dispersed in clusters and the gene order was extensively modified, however no loss of large fragments containing collinear genes has occurred, (4) instead the genes not involved in VLP production have been lost by pseudogenization following the accumulation of mutations authorized by neutral selection. Blue boxes represents wasp genes, light green boxes represents nudiviral genes, intense green boxes represents nudiviral pseudogenes, yellow triangles represent putative binding sites of an integrase/recombinase involved in the insertion/excision of nudivirus genome in host DNA and stars indicate mutations.

Amplification of the Nudiviral Clusters

Nudiviral clusters contain all the genes involved in VLP production, but less than half of the pseudogenes identified. This suggests that selection might maintain active genes within nudiviral clusters. This constraint could be relaxed for no longer useful genes, allowing their dispersion in the wasp genome. In the genomes of wasps associated with bracoviruses a subset of nudiviral genes are maintained in a single cluster, whereas most genes are dispersed in the genome. The nudiviral cluster of bracovirus associated wasps, containing in particular genes encoding major capsid components (VP39 and 38K), is amplified during virus particles production in the calyx (Louis et al. 2013; Burke et al. 2015). It is thought that high copy number of the genes present in the cluster is necessary to achieve high level bracovirus particles production from a relatively small population of virus particles producing cells. We reasoned the same constraint could operate for genes involved in VLPs production in the calyx, because specialized cells produce them in abundance. Accordingly, we found that genes from 4 different nudiviral clusters (fig. 8), encoding most of the nudiviral genes, were amplified (from 5- to 25-fold). This suggests that the localization within a nudiviral cluster may have functional consequences for gene expression. It is not simply that this confers high-level expression. Indeed expression levels of genes encoded by the clusters are heterogeneous (Pichon et al. 2015). However it is likely that the dispersion of a gene would result in a dramatic drop of its expression resulting in a disruption of the relative ratio of nudiviral genes expression and stoichiometry of their products, which could be counter selected.

Comparison of Viral and Bacterial Symbionts

To our knowledge PDVs and VcENV are the only endogenous virus symbionts with complex virus machinery providing a benefit to their host. In contrast, a cohort of bacteria

obligatorily associated with insects for tens of millions years have been studied during the last decades, in particular among Gammaproteobacteria (Moran and Bennett 2014). This group comprises free-living *Escherichia coli*—on which many data on gene functions are available—as well as *Buchnera*, *Wigglessworthia*, *Blochmania*, and *Baumania* endosymbionts associated with the pea aphid, the tse–tse fly, ants and the leafhopper, respectively. These symbionts provide to the insects the nutrients that they are not able to synthetize or uptake from their food source. Synteny comparisons between sequenced genomes of symbiotic and free-living bacteria have shown that a drastic genome size reduction had invariably occurred during symbiont evolution accompanied by massive gene loss (Moran 2003; McCutcheon and Moran 2010) resulting ultimately in the sole conservation of the genes useful for the insect, for example, in the smallest 150 kb genome of *Candidatus Nasuia deltocephalinicola* an obligate symbiont of the leafhopper *Macrosteles quadrilineatus* (Moran and Bennett 2014). This loss is attributed to an inherent deletion bias of bacterial mutation (Mira et al. 2001), which is generally involved in the compaction resulting in high gene density of bacterial genomes. Study of *Buchnera aphidicola*, the obligatory symbiont of aphid, showed that in obligate endosymbionts genome reduction has two phases. Early in the symbiosis large sequence blocks corresponding to many unrelated genes can be lost (Moran and Mira 2001) by the fixation of large scale deletions associated with rearrangements (Wernegreen 2002). This is later followed by DNA loss in smaller blocks (Moran 2003). Thorough analysis of syntenic blocks further showed that a smaller proportion of lost genes are also eliminated by mutation accumulations and pseudogenization in their genomic context in addition to those lost by large deletions (Moran and Mira 2001).

Genome sequencing of *Serratia symbiotica*, a symbiont recently acquired by aphids and undergoing the early stages of genomic change affecting bacteria that transition to a symbiotic lifestyle, confirmed that most genomic changes occurs during the early phase of symbiosis. Comparison with free-living *Serratia* species revealed a decaying genome dramatically reduced in size, with many pseudogenes and rearrangements, and an increase of mobile DNA proportion (Burke and Moran 2011). Analyses performed on a recently acquired symbiont of the weevil *Sitophilus oryzae* primary endosymbiont further confirmed that early during symbiosis extensive genome degeneration occurs characterized by gene inactivation, deletion, and mobile element-mediated genome rearrangements (Oakeson et al. 2014). Interestingly, some of the genes that appear as nonessential in the symbiotic context and will be ultimately lost, as predicted by the study of older associations, have not yet been pseudogenized in *S. symbiotica* (Burke and Moran 2011).

Whereas bacterial endosymbionts evolution has been thoroughly studied, little is known on the fate of viral symbionts integrated into host genomes. A major difference is that vertical transmission experienced by viral genes is similar to wasp genes, whereas for bacterial symbionts, there is relaxed selection due to the bottlenecks during vertical transmission. If one might speculate on the evolutionary scenario that led to VLPs, a unique event such as a mutation in regulatory regions involved in excision of viral DNA could have been sufficient to impair the ability to incorporate DNA in the particles (fig. 10). This initial event might have released the conservative selection pressure on all the genes involved in nucleocapsid production resulting in their subsequent pseudogenization, whereas the genes contributing to parasitism success would have remained functional (fig. 10). Studies of symbiotic associations with bacteria suggest that major genomic changes appear early after the symbiosis is established (Moran and Plague 2004) followed by a period of stasis characterized by remarkable symbiont genome stability. Dating of ichneumonid wasps phylogeny and search for the presence of VLPs in closely related species to *V. canescens* would enable estimating more precisely how long VcENV has evolved in the context of a symbiotic relationship with the wasp. In any case, it is much more recent that the betanudivirus capture leading to bracoviruses, but strikingly there is more functional nudiviral core genes missing in the case of VcENV than for bracoviruses. Moreover there is a perfect correlation between whether genes are conserved or pseudogenized and the function of VLPs. This suggests that the VcENV symbiosis has already reached such a stasis state (fig. 10). However, whereas most lost genes are deleted in bacterial symbionts we could detect pseudogenes corresponding to lost functions in the wasp genome, this suggests that the DNA of the symbiont is more likely to be conserved for a longer time in the context of a chromosomal integration than as part of a bacterial genome, prone to undergo deletions.

Conclusion

Virus domestication is probably recurrent in the evolution of metazoans, although to date the only examples described concern endoparasitoid wasps. VLPs have been described specifically in the species *V. canescens* belonging to Campopleginae, a subfamily of wasps associated with ichnoviruses, indicating that a symbiotic virus replacement occurred in this lineage involving the integration of a nudivirus. Analysis of gene order conservation in VcENV indicates that the viral genome has been reshuffled, marginally retaining ancestral gene order. However, pseudogenes corresponding to most of the lost nudiviral core genes were identified, some still in their ancestral context within nudiviral clusters. This suggests that the loss of virus genes during large dsDNA virus domestication result from mutations of individual genes rather than by deletion of large chunks of viral genome. Further comparative genomic studies of closely related species to *V. canescens* or harboring other endogenous nudiviruses should provide more

information on the evolution of virus sequences within host genome depending on whether they are beneficial neutral or detrimental for their host.

Acknowledgments

This study was supported by core funding from CNRS and Université de Tours. Grants were obtained by A.-N. Volkoff from INRA « Santé des Plantes et Environnement » and by M.L. from the Fédération de Recherche en Infectiologie (FéRI). *Venturia canescens* genome sequence was obtained in the frame of the Genoscope project to J.-M. Drezen "The role of viruses in parasitoid wasp evolution." M.L. salary was funded by a PhD grant from region Centre-Val de Loire.

Authors' Contribution

M.L. has set up the approach to identify pseudogenes, performed analyses, studied nudivirus syntenies, and wrote the manuscript. A.B. helped with nudiviral sequences and virus phylogeny analyses, and contributed to the writing; A.P. performed qPCR assay; Elisabeth Herniou provided expertise on the analysis of virus genes divergence and contributed to the writing. A.-N.V. analyzed the expression of pseudogenes and contributed to the writing. J.-M.D. directed the study and wrote the manuscript.

Literature Cited

Afgan E, et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44(W1):W3–W10.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389–3402.

Arbuckle JH, et al. 2010. The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes *in vivo* and *in vitro*. Proc Natl Acad Sci USA. 107(12):5563–5568.

Asgari S, Reineke A, Beck M, Schmidt O. 2002. Isolation and characterization of a neprilysin-like protein from *Venturia canescens* virus-like particles. Insect Mol Biol. 11(5):477–485.

Ashley J, et al. 2018. Retrovirus-like gag protein Arc1 binds RNA and traffics across synaptic boutons. Cell 172(1–2):262–274.e211.

Bandi C, et al. 1994. Flavobacteria as intracellular symbionts in cockroaches. Proc Biol Sci. 257(1348):43–48.

Beliveau C, et al. 2015. Genomic and proteomic analyses indicate that banchine and campoplegine polydnaviruses have similar, if not identical, viral ancestors. J Virol. 89(17):8909–8921.

Bézier A, Herbinière J, Lanzrein B, Drezen JM. 2009. Polydnavirus hidden face: the genes producing virus particles of parasitic wasps. J Invertebr Pathol. 101(3):194–203.

Bézier A, Harichaux G, Musset K, Labas V, Herniou EA. 2017. Qualitative proteomic analysis of Tipula oleracea nudivirus occlusion bodies. J Gen Virol. 98(2):284–295.

Bézier A, et al. 2015. The genome of the nucleopolyhedrosis-causing virus from *Tipula oleracea* sheds new light on the *Nudiviridae* family. J Virol. 89(6):3008–3025.

Bézier A, et al. 2013. Functional endogenous viral elements in the genome of the parasitoid wasp *Cotesia congregata*: insights into the evolutionary dynamics of bracoviruses. Philos Trans R Soc Lond B Biol Sci. 368(1626):20130047.

Bézier A, et al. 2009. Polydnaviruses of braconid wasps derive from an ancestral nudivirus. Science 323(5916):926–930.

Burke GR, Moran NA. 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. Genome Biol Evol. 3:195–208.

Burke GR, Thomas SA, Eum JH, Strand MR. 2013. Mutualistic polydnaviruses share essential replication gene functions with pathogenic ancestors. PLoS Pathog. 9(5):e1003348.

Burke GR, Simmonds TJ, Thomas SA, Strand MR. 2015. Microplitis demolitor bracovirus proviral loci and clustered replication genes exhibit distinct DNA amplification patterns during replication. J Virol. 89(18):9511–9523.

Burke GR, Walden KK, Whitfield JB, Robertson HM, Strand MR. 2014. Widespread genome reorganization of an obligate virus mutualist. PLoS Genet. 10(9):e1004660.

Chen X, et al. 2001. The sequence of the Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus genome. J Gen Virol. 82(1):241–257.

Cheng RL, et al. 2014. Brown planthopper nudivirus DNA integrated in its host genome. J Virol. 88(10):5310–5318.

Chevignon G, et al. 2014. Functional annotation of Cotesia congregata bracovirus: identification of the viral genes expressed in parasitized host immune tissues. J Virol. 88(16):8795–8812.

Colinet D, Schmitz A, Depoix D, Crochard D, Poirié M. 2007. Convergent use of RhoGAP toxins by eukaryotic parasites and bacterial pathogens. PLoS Pathog. 3(12):e203.

Crawford AM, Sheehan C. 1985. Replication of oryctes baculovirus in cell culture: viral morphogenesis, infectivity and protein synthesis. J Gen Virol. 66(3):529–539.

Danquah JO, Botchway S, Jeshtadi A, King LA. 2012. Direct interaction of baculovirus capsid proteins VP39 and EXON0 with kinesin-1 in insect cells determined by fluorescence resonance energy transfer-fluorescence lifetime imaging microscopy. J Virol. 86(2):844–853.

Ding W, Lin L, Chen B, Dai J. 2006. L1 elements, processed pseudogenes and retrogenes in mammalian genomes. IUBMB Life 58(12):677–685.

Dong F, Wang J, Deng R, Wang X. 2016. Autographa californica multiple nucleopolyhedrovirus gene *ac81* is required for nucleocapsid envelopment. Virus Res. 221:47–57.

Drezen JM, Herniou EA, Bézier A. 2012. Evolutionary progenitors of bracoviruses. In: Beckage NE, Drezen J-M, editors. Parasitoid viruses symbionts and pathogens. San Diego: Elsevier. p. 15–31.

Drezen JM, et al. 2017. Endogenous viruses of parasitic wasps: variations on a common theme. Curr Opin Virol. 25:41–48.

Dushay MS, Beckage NE. 1993. Dose-dependent separation of *Cotesia congregata* associated polydnavirus effects on *Manduca sexta* larval development and immunity. J. Insect Physiol. 39(12):1029–1040.

Feddersen I, Sanders K, Schmidt O. 1986. Virus-like particles with host protein-like antigenic determinants protect an insect parasitoid from encapsulation. Experientia 42(11–12):1278–1281.

Feschotte C, Gilbert C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 13(4):283–296.

Francino MP. 2005. An adaptive radiation model for the origin of new gene functions. Nat Genet. 37(6):573–577.

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

Gauthier J, Drezen JM, Herniou EA. Forthcoming 2017. The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps. Parasitology 1–13; doi:10.1017/S0031182017000725

Guarino LA, Xu B, Jin J, Dong W. 1998. A virus-encoded RNA polymerase purified from baculovirus-infected cells. J Virol. 72(10):7985–7991.

Gundersen-Rindal D, Dupuy C, Huguet E, Drezen JM. 2013. Parasitoid Polydnaviruses: evolution, pathology and applications. Biocontrol Sci Technol. 23(1):1–61.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 41:95–98.
Heddi A, Grenier AM, Khatchadourian C, Charles H, Nardon P. 1999. Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. Proc Natl Acad Sci USA. 96(12):6814–6819.
Hentschel CC, Birnstiel ML. 1981. The organization and expression of histone gene families. Cell 25(2):301–313.
Herniou EA, Olszewski JA, Cory JS, O'Reilly DR. 2003. The genome sequence and evolution of baculoviruses. Annu Rev Entomol. 48:211–234.
Hill T, Unckless RL. 2018. The dynamic evolution of Drosophila innubila nudivirus. Infect Genet Evol. 57:151–157.
Holmes EC. 2011. The evolution of endogenous viral elements. Cell Host Microbe 10(4):368–377.
Hu ZH, et al. 1998. Distinct gene arrangement in the Buzura suppressaria single-nucleocapsid nucleopolyhedrovirus genome. J Gen Virol. 79(11):2841–2851.
Hughes AL, Friedman R. 2003. Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses. Mol Biol Evol. 20(6):979–987.
Jehle JA, et al. 2013. Creation of a new Family *Nudiviridae* including two new genera and three species. Taxonomy Poposal. https://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/invertebrate-official/4819 (last accessed June 29, 2018)
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14):3059–3066.
Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. PLoS Genet. 6(11):e1001191.
Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649.
Labrosse C, et al. 2005. A RhoGAP protein as a main immune suppressive factor in the *Leptopilina boulardi* (Hymenoptera, Figitidae)-*Drosophila melanogaster* interaction. Insect Biochem Mol Biol. 35(2):93–103.
Lavialle C, et al. 2013. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. Philos Trans R Soc Lond B Biol Sci. 368(1626):20120507.
Lieber MR. 1997. The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. Bioessays 19(3):233–240.
Lin CL, et al. 1999. Persistent Hz-1 virus infection in insect cells: evidence for insertion of viral DNA into host chromosomes and viral infection in a latent status. J Virol. 73(1):128–139.
Louis F, et al. 2013. The bracovirus genome of the parasitoid wasp *Cotesia congregata* is amplified within 13 replication units, including sequences not packaged in the particles. J Virol. 87(17):9649–9660.
Mangeney M, et al. 2007. Placental syncytins: genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. Proc Natl Acad Sci USA. 104(51):20534–20539.
McCutcheon JP, Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol Evol. 2:708–718.
Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. Trends Genet. 17(10):589–596.
Moran NA. 2003. Tracing the evolution of gene loss in obligate bacterial symbionts. Curr Opin Microbiol. 6(5):512–518.
Moran NA, Mira A. 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. Genome Biol. 2:1–12.
Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. Curr Opin Genet Dev. 14(6):627–633.
Moran NA, Bennett GM. 2014. The tiniest tiny genomes. Annu Rev Microbiol. 68(1):195–215.
Munson MA, et al. 1991. Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. J Bacteriol. 173(20):6321–6324.
Murphy N, Banks JC, Whitfield JB, Austin AD. 2008. Phylogeny of the parasitic microgastroid subfamilies (Hymenoptera: braconidae) based on sequence data from seven genes, with an improved time estimate of the origin of the lineage. Mol Phylogenet Evol. 47(1):378–395.
Neafsey DE, et al. 2015. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles mosquitoes*. Science 347(6217):1258522.
Nizami Z, Deryusheva S, Gall JG. 2010. The Cajal body and histone locus body. Cold Spring Harb Perspect Biol. 2(7):a000653.
Oakeson KF, et al. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. Genome Biol Evol. 6(1):76–93.
Parkinson N, Conyers C, Smith I. 2002. A venom protein from the endoparasitoid wasp *Pimpla hypochondriaca* is similar to snake venom reprolysin-type metalloproteases. J Invertebr Pathol. 79(2):129–131.
Pasquier-Barre F, et al. 2002. Polydnavirus replication: the EP1 segment of the parasitoid wasp *Cotesia congregata* is amplified within a larger precursor molecule. J Gen Virol. 83(Pt 8):2035–2045.
Pastuzyn ED, et al. 2018. The neuronal gene *Arc* encodes a repurposed retrotransposon Gag protein that mediates intercellular RNA transfer. Cell 172(1–2):275–288.e218.
Peng K, van Oers MM, Hu Z, van Lent JW, Vlak JM. 2010. Baculovirus *per os* infectivity factors form a complex on the surface of occlusion-derived virus. J Virol. 84(18):9497–9504.
Peng K, et al. 2012. Characterization of novel components of the baculovirus *per os* infectivity factor complex. J Virol. 86(9):4981–4988.
Pichon A, et al. 2015. Recurrent DNA virus domestication leading to different parasite virulence strategies. Sci Adv. 1(10):e1501150.
Reineke A, Asgari S, Ma G, Beck M, Schmidt O. 2002. Sequence analysis and expression of a virus-like particle protein, VLP2, from the parasitic wasp *Venturia canescens*. Insect Mol Biol. 11(3):233–239.
Rio RV, Attardo GM, Weiss BL. 2016. Grandeur alliances: symbiont metabolic integration and obligate arthropod hematophagy. Trends Parasitol. 32(9):739–749.
Serbielle C, et al. 2012. Evolutionary mechanisms driving the evolution of a large polydnavirus gene family coding for protein tyrosine phosphatases. BMC Evol Biol. 12:253.
Serbielle C, et al. 2008. Viral cystatin evolution and three-dimensional structure modelling: a case of directional selection acting on a viral protein involved in a host-parasitoid interaction. BMC Biol. 6(1):38.
Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407:81–86.
Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6(1):31.
Sparks WO, Harrison RL, Bonning BC. 2011. Autographa californica multiple nucleopolyhedrovirus ODV-E56 is a *per os* infectivity factor, but is not essential for binding and fusion of occlusion-derived virus to the host midgut. Virology 409(1):69–76.
Strand MR. 2012. Polydnavirus gene products that interact with the host immune system. In: Beckage NE, Drezen J-M, editors. Parasitoid viruses symbionts and pathogens. San Diego: Elsevier. p. 149–161.
Tarlinton RE, Meers J, Young PR. 2006. Retroviral invasion of the koala genome. Nature 442(7098):79–81.
Thézé J, Bézier A, Periquet G, Drezen JM, Herniou EA. 2011. Paleozoic origin of insect large dsDNA viruses. Proc Natl Acad Sci USA. 108(38):15931–15935.

Thézé J, Takatsuka J, Nakai M, Arif B, Herniou EA. 2015. Gene acquisition convergence between entomopoxviruses and baculoviruses. Viruses 7(4):1960–1974.
Vanarsdall AL, Okano K, Rohrmann GF. 2006. Characterization of the role of very late expression factor 1 in baculovirus capsid structure and DNA processing. J Virol. 80(4):1724–1733.
Volkoff AN, et al. 2010. Analysis of virion structural components reveals vestiges of the ancestral ichnovirus genome. PLoS Pathog. 6(5):e1000923.
Wang M, et al. 2010. Specificity of baculovirus P6.9 basic DNA-binding proteins and critical role of the C terminus in virion formation. J Virol. 84(17):8821–8828.
Wang Y, Jehle JA. 2009. Nudiviruses and other large, double-stranded circular DNA viruses of invertebrates: new insights on an old topic. J Invert Pathol. 101(3):187–193.
Wang Y, Bininda-Emonds O, Jehle JA. 2012. Nudivirus genomics and phylogeny. In: Garcia ML, Romanowski V, editors. Viral genomes-molecular structure, diversity, gene expression mechanisms and host-virus interactions. Rijeka: Intech.
Wang Y, Kleespies RG, Huger AM, Jehle JA. 2007. The genome of Gryllus bimaculatus nudivirus indicates an ancient diversification of baculovirus-related nonoccluded nudiviruses of insects. J Virol. 81(10):5395–5406.
Wang Y, Bininda-Emonds OR, van Oers MM, Vlak JM, Jehle JA. 2011. The genome of Oryctes rhinoceros nudivirus provides novel insight into the evolution of nuclear arthropod-specific large circular double-stranded DNA viruses. Virus Genes 42(3):444–456.
Webster CL, et al. 2015. The discovery, distribution, and evolution of viruses associated with *Drosophila melanogaster*. PLoS Biol 13(7):e1002210.
Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet. 3(11):850–861.
Wetterwald C, et al. 2010. Identification of bracovirus particle proteins and analysis of their transcript levels at the stage of virion formation. J Gen Virol. 91(Pt 10):2610–2619.
Whitfield JB, Austin AD, Fernandez-Triana JL. 2018. Systematics, biology, and evolution of Microgastrine Parasitoid Wasps. Annu Rev Entomol. 63(1):389–406.
Wu W, et al. 2008. Autographa californica multiple nucleopolyhedrovirus 38K is a novel nucleocapsid protein that interacts with VP1054, VP39, VP80, and itself. J Virol. 82(24):12356–12364.
Wyler T, Lanzrein B. 2003. Ovary development and polydnavirus morphogenesis in the parasitic wasp *Chelonus inanitus*. II. Ultrastructural analysis of calyx cell development, virion formation and release. J Gen Virol. 84(Pt 5):1151–1163.
Xiao J, et al. 2016. Pseudogenes and their genome-wide prediction in plants. Int J Mol Sci. 17(12):1991.
Zhang Z, Harrison PM, Liu Y, Gerstein M. 2003. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res. 13(12):2541–2558.

Associate editor: Chantal Abergel