A behavior-manipulating virus relative as a source of adaptive genes for Drosophila parasitoids

D Di Giovanni, D. Lepetit, B Guinet, B Bennetot, M. Boulesteix, Yohann Couté, O. Bouchez, M Ravallec, J. Varaldi

A Behavior-Manipulating Virus Relative as a Source of Adaptive Genes for *Drosophila* Parasitoids

Deborah Di Giovanni,[1] David Lepetit,[1] Benjamin Guinet,[1] Bastien Bennetot,[†,1] Matthieu Boulesteix,[1] Yohann Couté,[2] Olivier Bouchez,[3] Marc Ravallec,[4] and Julien Varaldi[*,1]

[1]Université de Lyon Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR 5558, Villeurbanne, France

[2]Université de Grenoble Alpes, CEA, Inserm, IRIG-BGE, Grenoble, France

[3]Institut National de la Recherche Agronomique (INRA), US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

[4]UMR 1333 INRAE – Université Montpellier "Diversité, Génomes et Interactions Microorganismes-Insectes" (DGIMI), Montpellier, France

[†]Present address: Ecologie Systématique & Evolution (UMR 8079), Université Paris Sud, Orsay, France

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

Associate editor: Harmit Malik

Accession numbers are RICB00000000, RJVV00000000 and WNHD00000000, respectively, for *Leptopilina heterotoma*, *Ganaspis brasiliensis* and *G.* sp genome drafts. The raw reads can be found at SRA with accessions PRJNA499032, PRJNA503097, and PRJNA587987.

Abstract

Some species of parasitic wasps have domesticated viral machineries to deliver immunosuppressive factors to their hosts. Up to now, all described cases fall into the Ichneumonoidea superfamily, which only represents around 10% of hymenoptera diversity, raising the question of whether such domestication occurred outside this clade. Furthermore, the biology of the ancestral donor viruses is completely unknown. Since the 1980s, we know that *Drosophila* parasitoids belonging to the *Leptopilina* genus, which diverged from the Ichneumonoidea superfamily 225 Ma, do produce immunosuppressive virus-like structure in their reproductive apparatus. However, the viral origin of these structures has been the subject of debate. In this article, we provide genomic and experimental evidence that those structures do derive from an ancestral virus endogenization event. Interestingly, its close relatives induce a behavior manipulation in present-day wasps. Thus, we conclude that virus domestication is more prevalent than previously thought and that behavior manipulation may have been instrumental in the birth of such associations.

Key words: **symbiosis, horizontal gene transfer, domestication, virus, parasitoid, wasps, superparasitism.**

Introduction

Genetic information is typically passed on from generation to generation through reproduction, that is, vertical transmission. However, at some point during the course of evolution, organisms may gain DNA from unrelated organisms, through horizontal gene transfer (HGT). Most horizontally acquired DNA is probably purged from the genomes of the population either because it did not reach the germinal cells in case of metazoan species and/or because no advantage is carried by the foreign sequence. However, natural selection may retain the foreign DNA leading ultimately to genetic innovation in the population/species (Husnik and McCutcheon 2018).

The high frequency and relevance of such phenomena has been recognized for decades for bacteria but was considered to have had a marginal impact on the evolution of metazoans (Keeling and Palmer 2008). However, this view has been recently challenged due to the discovery of numerous examples of HGT in metazoans with some of them leading to genetic innovation (Boto 2014). The most notorious example involves retroviral envelope genes that have been endogenized, domesticated, and multiply replaced in mammalian genomes (Lavialle et al. 2013). In this case, the fusogenic and immunosuppressive properties of these viral proteins (syncitins) have been repeatedly recruited to permit the evolution of placental structures during mammalian diversification. Interestingly, a similar case of syncitin domestication was recently described in a clade of viviparous Scincidae lizards that also rely on a placenta-like structure to feed their offspring (Cornelis et al. 2017). Other examples include phytophagous mites and Lepidoptera that deal with chemical defenses of their host plant thanks to the acquisition of a bacterial gene involved in detoxification (Wybouw et al. 2016), several phytophagous arthropods (aphids, mites, and gall midges) who independently acquired genes involved in carotenoid biosynthesis from fungal donors (Moran and Jarvik 2010; Grbić et al. 2011; Cobbs et al. 2013), parasitic wasps that co-opted microsporidia genes (Martinson et al. 2016), caterpillar that acquired from their parasitic wasp a gene protecting them from bacterial invasion (Di Lelio et al. 2019), or parasitic nematodes that domesticated plant cell-wall degrading enzymes from bacteria (Danchin et al. 2010).

Regarding the question of domestication of horizontally transferred DNA in eukaryotes, endoparasitic wasps are of particular interest because they have repeatedly

domesticated not only single genes but also entire viral machineries (review in Gauthier et al. [2018] and since then Burke et al. [2018]). Endoparasitic wasps lay their eggs inside the body of other arthropods, usually other insects, ultimately killing them. Their progeny are thus exposed to the host immune system. Notably, it has been found that the ancestor of at least three monophyletic groups of endoparasitic wasps has independently domesticated a battery of viral genes allowing them to deliver either DNA encoding immunosuppressive factors or immunosuppressive proteins themselves (Bezier et al. 2009; Volkoff et al. 2010; Burke et al. 2018). Strikingly, in the case DNA is delivered into the host (so-called polydnaviruses, PDVs), it integrates into the host hemocyte's DNA and gets expressed (Bitra et al. 2011; Chevignon et al. 2014), manipulating the host physiology and behavior, ultimately favoring the development of wasp offspring. In cases where proteins are delivered (so-called virus-like particles, VLPs), the viral machinery permits the delivery of these virulence proteins into host immune cells, thus inhibiting the host immune response (Reineke et al. 2006). In both cases, virally derived genes are used by the wasp to produce a vector toolset composed of capsids and/or envelopes. However, the virulence factors themselves (or the DNA encoding the virulence factors) are of eukaryotic origin, probably predating the domestication event (Espagne et al. 2004). Evolution has thus repeatedly favored the domestication of kits of viral genes allowing the production of virus-like structures in the reproductive apparatus of parasitic wasps with clear functional convergence.

One clear pattern emerging from the data is that all described cases documented so far involve wasps belonging to the Ichneumonoidea superfamily (Gauthier et al. 2018). Although this superfamily is very speciose (most likely around 100,000 species), it represents a modest fraction of parasitic Hymenoptera diversity (most likely around 1 million species; Forbes et al. 2018). Another feature of the current data is that the biology of the ancestral donor virus is unclear. For one such domestication event (in the Campopleginae subfamily, Ichneumonidae family), the ancestral virus has not been identified at all, whereas a betanudivirus has been identified as the donor virus for wasps belonging to the microgastroid complex of the Braconidae family (Bezier et al. 2009). In *Venturia canescens* (Campopleginae subfamily, Ichneumonidae family) where a VLP system replaced a PDV one, and in some wasp species from the genus *Fopius* (subfamily Opiinae, Braconidae family), it has been shown that an alpha-nudivirus was the donor (Pichon et al. 2015; Burke et al. 2018). However, close relatives of the donor viruses are not known to infect present-day wasps, nor to infect their hosts. One possible explanation is that the "donor" viral lineages and their relatives went extinct and/or have not been sampled yet. The exact nature of the association wasp/virus that permitted such massive domestication events is thus still unclear.

In this work, we identify a new independent case of virus domestication in the genus *Leptopilina* which belongs to a very distantly related wasp superfamily (Cynipoidea, Figitidae) compared with all previously described cases. Those wasps are parasitoids of *Drosophila* larvae. We provide strong evidence that the genes of viral origin permit all *Leptopilina* wasp species to produce so-called VLPs. VLPs have been known for decades in this genus (Rizki RM and Rizki TM 1990). They are produced in the venom gland of the wasp, are devoid of DNA but contain virulence proteins that are injected, together with the egg, into the *Drosophila* larva (Colinet et al. 2007). They protect wasp eggs from the *Drosophila* immune response (Rizki RM and Rizki TM 1990; Colinet et al. 2010). We show that a close relative of the ancestral donor virus is still segregating in the species *L. boulardi* and its biology has been extensively studied by our group (Martinez et al. 2012, 2015; Patot et al. 2012; Lepetit et al. 2016; Varaldi and Lepetit 2018; Varaldi et al. 2003). The virus, known as LbFV (*L. boulardi* filamentous virus), belongs to a possibly new double-stranded DNA (dsDNA) virus family related to Hytrosaviridae, and more distantly related to Nudiviridae and Baculoviridae (Lepetit et al. 2016). The virus is vertically transmitted and manipulates wasp behavior by causing infected females to lay their eggs into already parasitized larvae. This virus-induced "host-sharing" benefits the virus because it allows its horizontal transmission to new parasitoid lineages. On the contrary, this "superparasitism" behavior comes with a cost to wasp fitness, making it a nice example of behavior manipulation (Gandon et al. 2006). This result suggests that heritable viruses, such as LbFV, might have been instrumental in the birth of such association between wasps and viruses. In addition, it shows that virus domestication by parasitic wasps is not restricted to the Ichneumonoidea superfamily but may concern more diversity than previously thought.

Results

We analyzed the genomic sequences of *L. boulardi* (Varaldi and Lepetit 2018), *L. clavipes* (Kraaijeveld et al. 2016), *L. heterotoma* (this study), and two related species in the genus *Ganaspis* (*G. brasiliensis* and *Ganaspis* species [*G.* sp], this study). All *Leptopilina* species as well as *Ganaspis* species belong to the family Figitidae and are endoparasitoids developing from various species of *Drosophila* (Carton et al. 1986).

The basic statistics for the assemblies used in this study are presented in supplementary table S1, Supplementary Material online. With an N50 of, respectively, 2,080 and 5,595 bp the *G. brasiliensis* and *G.* sp assemblies appeared more fragmented than those from the *Leptopilina* species whose N50 ranges from 12,807 to 17,657 bp. This reflects their two to three times larger genome size likely due to their higher content of repetitive sequences (~40% vs. 24.02–28.82%). All five genomes were sequenced with coverage depth above 24× (between 24× and 85×), which is most likely sufficient to get the whole gene set (Malmstrøm et al. 2017). Accordingly, a BUSCO (Simão et al. 2015) analysis revealed that the vast majority of the 1,066 single-copy genes expected to be found in most arthropods are indeed present in all four assemblies (from 96.6% in *G. brasiliensis* to 99.1% in *L. boulardi*), making these assemblies suitable for HGT detection.

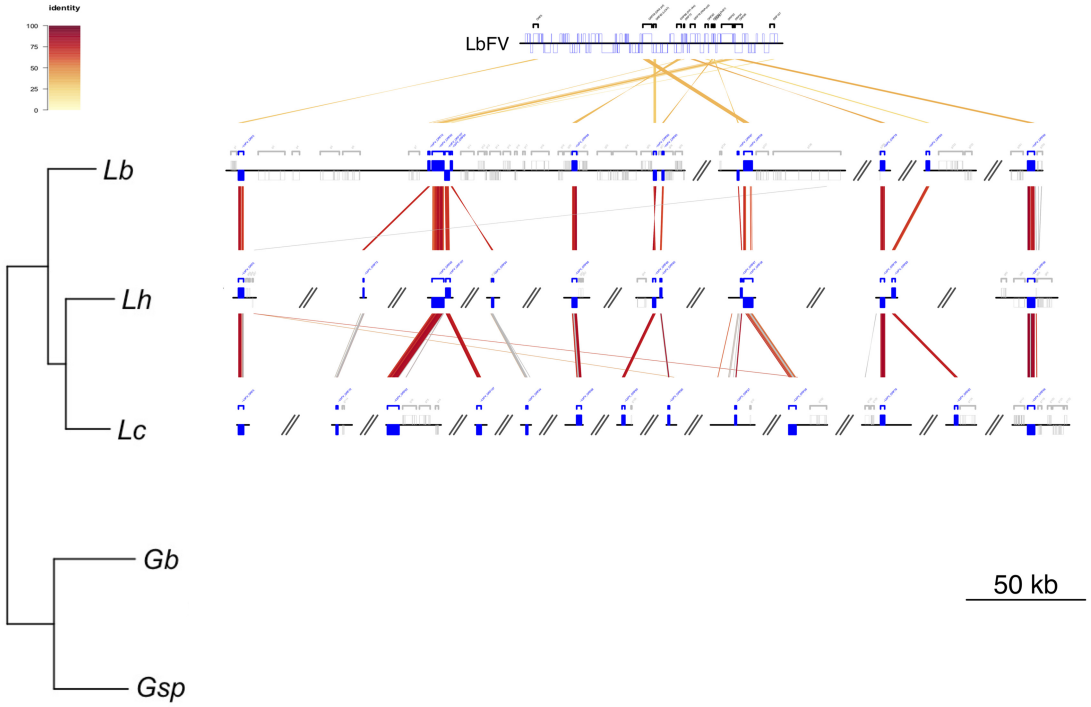

Fig. 1. Comparative genomics of wasp scaffolds sharing similarities with virus proteins. Lb, *Leptopilina boulardi*; Lh, *Leptopilina heterotoma*; Lc, *Leptopilina clavipes*; Gb, *Ganaspis brasiliensis*; Gsp, *Ganaspis* sp. The species tree on the left has been obtained using a concatenation of 541 universal arthropod genes. All nodes have an aLRT value of 1 (*Apis mellifera* was used as an outgroup). The red/yellow color code depicts the percentage of protein identity between homologous sequence pairs (viral or virally derived loci). Blue boxes identify the virally derived genes and their orientation (above: sense, below: antisense), whereas genes of eukaryotic origin are depicted in gray on the scaffolds. Gray connections indicate homology between nonvirally derived regions. The figure has been drawn using the genoPlotR package (Guy et al. 2010). The scaffolds are ordered from left to right in an arbitrary manner.

We inferred the relationships among the wasps under study using a set of 541 genes ubiquitous to all arthropods (see Materials and Methods). As expected, the three *Leptopilina* species form a monophyletic clade with *L. heterotoma* being more closely related to *L. clavipes* than to *L. boulardi* (fig. 1).

In order to identify putative horizontal transfers between an LbFV-like virus and the wasps, we blasted the 108 proteins encoded by the behavior-manipulating virus that infects *L. boulardi* (LbFV) against the *Leptopilina* and *Ganaspis* genomes (TBlastN). Interestingly, we found that 17 viral proteins had highly significant hits in wasp genomes ($1.3 \times 10^{-178} < e\text{-values} < 10^{-5}$). Among them, two classes should be distinguished. The first class is composed of four viral genes (open reading frames [ORFs] 11, 13, 27, and 66) that have strong similarities with both *Leptopilina* and *Ganaspis* genes (supplementary fig. S1, Supplementary Material online). We previously reported that these genes have probably been acquired horizontally by the virus from an ancestral insect before the *Leptopilina* diversification (Lepetit et al. 2016; supplementary figs. S1 and S2A, Supplementary Material online). Two of them (ORFs 27 and 66) are predicted to encode inhibitors of apoptosis, whereas ORFs 11 and 13 encode a putative demethylase (Lepetit et al. 2016). These two last genes may derive from a single horizontal transfer followed by a subsequent gene duplication (Lepetit et al. 2016). In the following section, we will focus on the second class of genes identified by this BLAST analysis.

Leptopilina Species Captured 13 Viral Genes from an LbFV-like Virus

More surprisingly, we found clear evidence that a massive integration of viral DNA into wasp genomes occurred before the diversification of the *Leptopilina* genus and most likely after the divergence between *Ganaspis* and *Leptopilina* (Table 1). This event led to the integration of 13 viral genes into the genome of the wasps (supplementary fig. S2B, Supplementary Material online). The corresponding 13 viral proteins have highly significant hits with all *Leptopilina* species ($4 \times 10^{-4} < e\text{-values} < 1.3 \times 10^{-178}$, median $= 10^{-33}$), but not with *G. brasiliensis* or *G.* sp. The percentages of identity between these 13 LbFV proteins and *Leptopilina* homologs ranged from 21.9 to 41.9 (supplementary tables S2–S4 and figs. S3–S15, Supplementary Material online). All 13 loci displayed complete ORF starting with a methionine and ending with a stop codon in the three wasp species, and their length was very similar to the corresponding ORF in LbFV genome (supplementary tables S2–S4, Supplementary Material online; the regression slopes of ORF length in the wasp vs. ORF length in LbFV were, respectively, 0.95, 1.02, and 0.894 for *L. boulardi*, *L. heterotoma*, and *L. clavipes*; all $R^2 >$ 0.95 and all *P*-values $< 10^{-9}$ on 11 df). This suggests that those genes do not contain introns.

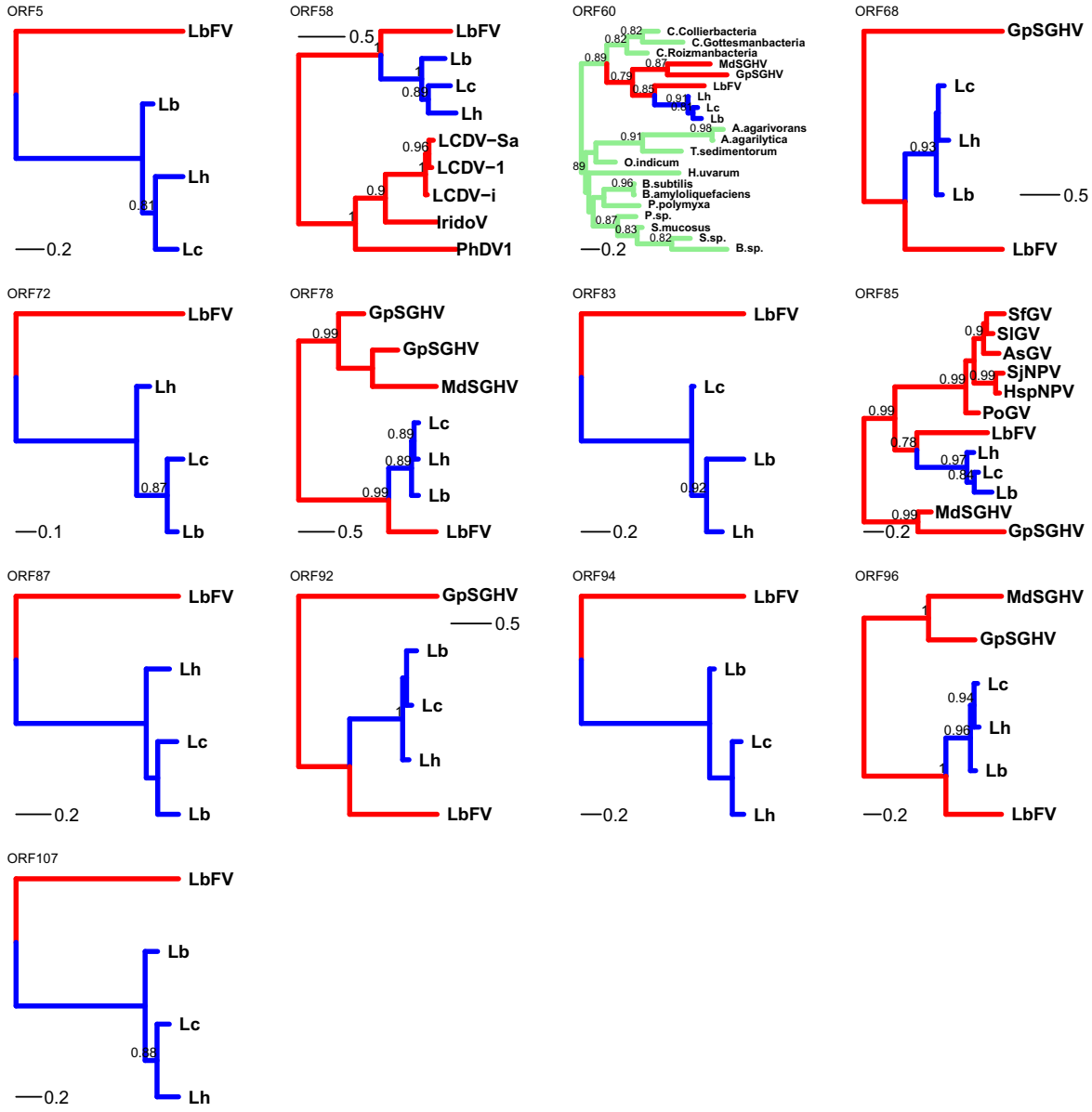

FIG. 2. Phylogenetic evidence for a massive horizontal transfer of 13 viral genes into the genome of *Leptopilina* wasps. The names of the ORFs refer to the ORF number in LbFV genome. Blue, red, and green colors represent, respectively (supposedly), eukaryotic, viral or bacterial branches. Only aLRT supports $\geqslant$ 0.7 are shown. The midpoint rooting method was used. Accession numbers of the corresponding sequences are available in supplementary table S5, Supplementary Material online.

To define a set of expected features for typical scaffolds belonging to wasp genomes, we calculated the GC content and sequencing depth for scaffolds containing single-copy arthropod-universal BUSCO genes (supplementary fig. S16, Supplementary Material online). As GC usually varies according to genomes, and coverage depth is directly related to the relative concentration of the DNA sequence under consideration, we used these expectations to ensure that the putatively endogenized sequenced were indeed part of the wasp genomes.

The general features (GC, sequencing depth) of wasp scaffolds sharing similarities with LbFV proteins were very similar to those calculated for the BUSCO-containing scaffolds (supplementary tables S2–S4 and fig. S16, Supplementary Material online). On the contrary, by analyzing these statistics (GC and coverage), we could easily detect the presence of some known extrachromosomal symbionts such as the virus LbFV in *L. boulardi* (supplementary fig. S16*A*, Supplementary Material online) or the bacteria *Wolbachia* in *L. heterotoma* (supplementary fig. S16*B*, Supplementary Material online). In addition, several typical intron-containing eukaryotic genes were predicted in the vicinity of these genes (depicted in gray in fig. 1). Note that apart from these 13 loci specifically found in *Leptopilina* genomes, most flanking *Leptopilina*-predicted proteins were also detected in both *G. brasiliensis* (66/72 for *L. boulardi*, 8/11 for *L. heterotoma*, and 10/15 for *L. clavipes*) and *G.* sp genomes (57/72 for *L. boulardi*, 8/11

for *L. heterotoma*, and 11/15 for *L. clavipes*) showing that the absence of homologs in *Ganaspsis* genomes was not the consequence of less reliable assemblies. Taken together, these observations demonstrate that the *Leptopilina* scaffolds containing viral-like genes are part of the wasp genomes.

The evolutionary history of the 13 genes is consistent with a horizontal transfer from an ancestor of the virus LbFV (or a virus closely related to this ancestor) to *Leptopilina* species (fig. 2). Indeed, when other sequences with homology to the proteins of interest were available in public databases, the three wasp genomes always formed a highly supported monophyletic clade with LbFV as a sister group of *Leptopilina* sequences (ORFs 58, 78, 92, 60, 68, 85, 96). In addition, for the six remaining phylogenies (for which no homologs were available in public databases), the midpoint rooting method always led to similar topologies with LbFV as the sister group of *Leptopilina* sequences. Furthermore, the divergence LbFV-*Leptopilina* relative to the divergence among *Leptopilina* species was identical for both types of loci (supplementary fig. S17, Supplementary Material online), further suggesting that both loci have the same evolutionary history. Interestingly, it appeared from the analysis of ORF60, that before being transferred to *Leptopilina* wasps, the gene has probably been acquired by the donor virus from an ancestral bacterium (fig. 2).

The clustering of most of these loci on the same scaffold in *L. boulardi* (8 out of 13 on scaffold 159, $N = 75{,}550$ scaffolds, see fig. 1) strongly suggests that a single event is at the origin of the phenomenon. In addition, for a few pairs of *L. boulardi* and *L. heterotoma* scaffolds, it was possible to test for the synteny of their virally derived genes (ORFs 92 and 107 in scaffolds 159 in Lb and IDBA_7081 in Lh, and ORFs 87 and 58 in scaffolds 2503 of Lb and IDBA_5653 in Lh). In all cases, the synteny appeared to be maintained between the two *Leptopilina* species (fig. 1). In addition, a few flanking non-virally derived sequences occurred around the same viral genes in different *Leptopilina* species (gray connections in fig.1, see supplementary fig. S18, Supplementary Material online, for details). The overall shared organization of these genes in the three *Leptopilina* species suggests that they have been vertically inherited since a single ancestral endogenization event.

To further assess the distribution of those virally derived genes in the diversity of *Leptopilina* wasps, we designed primers for ORF96 which is the most conserved gene. We successfully PCR amplified and sequenced the corresponding PCR product from DNA extracts obtained from all *Leptopilina* species tested (*L. guineaensis*, *L. freyae*, *L. victoriae* in addition to *L. boulardi*, *L. heterotoma*, and *L. clavipes*; supplementary fig. S19*A*, Supplementary Material online). The phylogeny obtained after the sequencing of the PCR products was congruent with the species-tree estimated from a phylogeny based on ITS2 (internal transcribed spacer 2) sequences (supplementary fig. S19*B*, Supplementary Material online). As expected, no PCR product was obtained from *Ganaspis* extracts.

Virally Derived Genes Are under Strong Purifying Selection in Wasp Genomes

In order to assess the way natural selection has acted on these virally derived genes since their endogenization, we calculated the dN/dS ratios using alignments involving the three *Leptopilina* species. We also calculated dN/dS ratios for a set of 958 genes found in the three *Leptopilina* species and that are also shared by at least 90% of all arthropods (Simão et al. 2015). Those genes are thus expected to be under strong purifying selection. Accordingly, the "universal" arthropod gene set had a very low dN/dS mean value (mean = 0.0733, median = 0.058), with a distribution skewed toward 0 (supplementary fig. S20, Supplementary Material online). Interestingly, the 13 virally derived genes had dN/dS below 1 (mean = 0.142, median = 0.136, min = 0.072, max = 0.236) and fell within the range of dN/dS values observed for BUSCO genes (min = 0, max = 0.452), suggesting that they are all essential for the survival and/or reproduction of *Leptopilina* wasps. On average, they showed, however, a slightly higher dN/dS values than the BUSCO genes (bootstrap sampling, *P*-value = 0.002), which could either indicate a slightly lower intensity of stabilizing selection or some history of adaptive evolution for some sites and/or some genes.

Virally Derived Genes Are Only Expressed in Female Venom Glands at the Onset of VLP Production

All *Leptopilina* species studied so far (*L. heterotoma*, *L. boulardi*, and *L. victoriae*) produce VLPs in their venom gland (Rizki RM and Rizki TM 1990; Dupas et al. 1996; Morales et al. 2005). As expected, we found that *L. clavipes* also produce VLPs in their venom gland, very similar in shape to other *Leptopilina* VLPs, further suggesting that this is a common feature for all *Leptopilina* species (supplementary fig. S21, Supplementary Material online). Interestingly, transmission electron microscopy (TEM) analysis of the *G. xanthopoda* venom gland revealed the presence of uncharacterized membranes in the lumen. However, those structures were much less electron dense than the VLPs found in *Leptopilina* (supplementary fig. S22, Supplementary Material online). Because *Leptopilina* VLPs are known to protect their eggs from *Drosophila* immune response (Rizki RM and Rizki TM 1990; Labrosse et al. 2003; Morales et al. 2005), we wondered whether the 13 virally derived genes were in fact responsible for their production. Under this hypothesis, our prediction was that the 13 genes would be expressed only in the venom gland of females because VLPs are specifically produced in this tissue, and only when VLPs are being produced, that is, during pupation.

To test this idea, we first followed the expression of the 13 virally derived genes in whole individual females from early larval stages to adulthood. These data revealed that the virally derived genes were almost exclusively expressed during the pupal stage of the wasps (supplementary fig. S23, Supplementary Material online). During that period, the venom gland is being formed and is matured (supplementary fig. S24, Supplementary Material online). The venom gland produces the VLPs that are released in the lumen (fig. 3) and

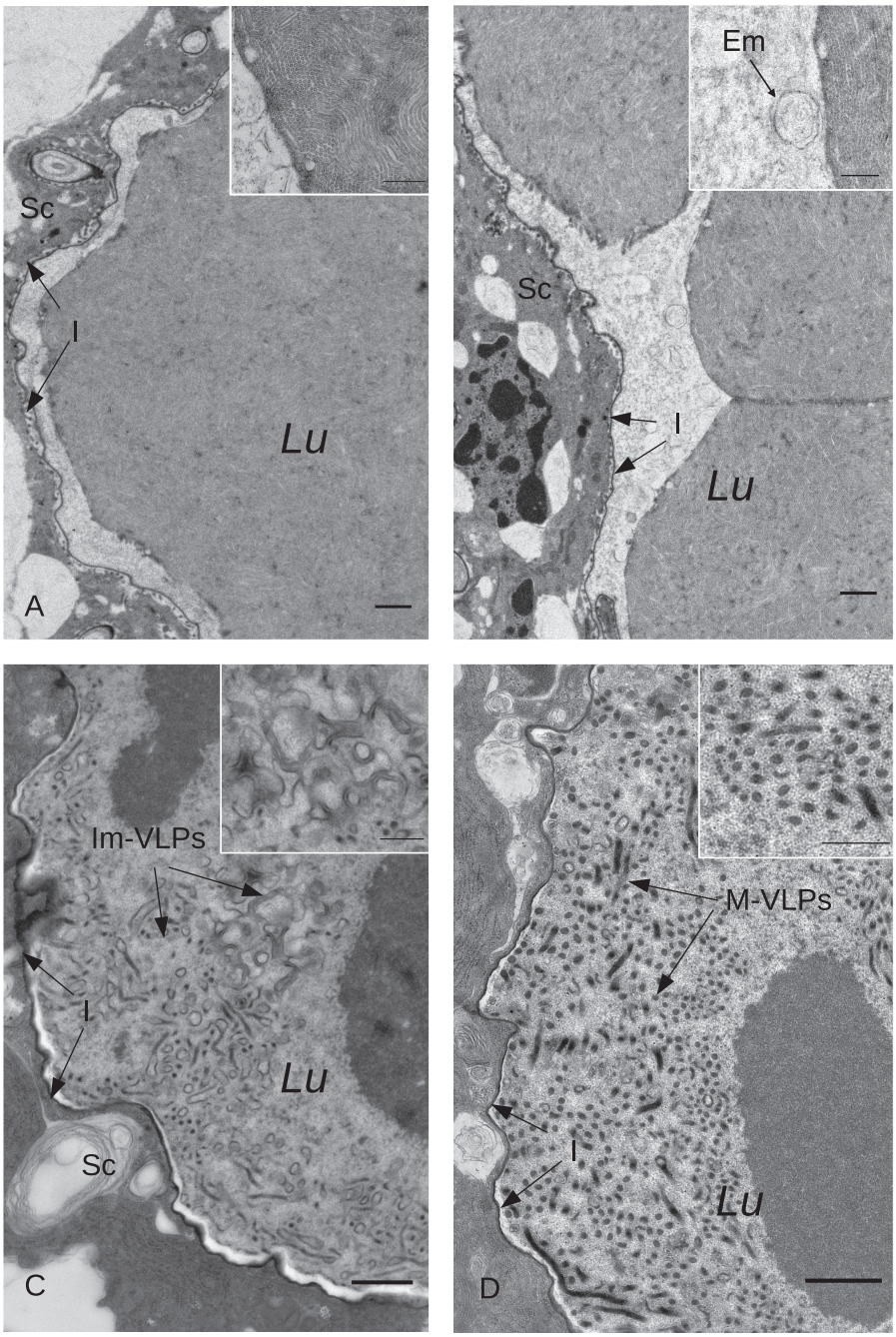

FIG. 3. Biogenesis of VLPs in the venom gland of *Leptopilina boulardi* during the pupal stage until adult emergence: (*A*) 14 days (pupae), (*B*) 16 days (pupae), (*C*) 18 days (pupae), (*D*) 21 days (adult). At days 14 and 16, secretory cells (SC) are releasing empty membranes (Em) into the Lumen (Lu) of the venom gland where they accumulate. Then at day 18, empty membranes start to be filled with electron-dense material (probably virulence proteins, such as LbGAP) to produce immature VLPs (im-VLPs). Finally at emergence (day 21), the venom gland lumen is filled with mature VLPs (m-VLPs) ready to be injected into the host. I: cuticular intima delineating the lumen. Insets show details of the reservoir. Bars represent 1 μM, except in insets where they represent 500 nM.

that finally reach the reservoir where they are stored until the emergence (see the size of the reservoir in supplementary fig. S24*E*, Supplementary Material online). Next, we followed their expression from the very beginning of the pupal stage (day 11) until the emergence from the host (day 21) in the venom glands (+ its reservoir), the ovaries, the rest of the body of *L. boulardi* females, and also in *L. boulardi* males. The patterns of expression of all 13 genes fit our prediction: They are all specifically expressed in the venom glands of females but not in other tissues, nor in males (fig. 4). Some virally derived genes were particularly expressed at the very beginning of venom gland morphogenesis (day 11), whereas the other genes had their peak of expression at day 14, when the reservoir of the gland starts to be filled with VLPs.

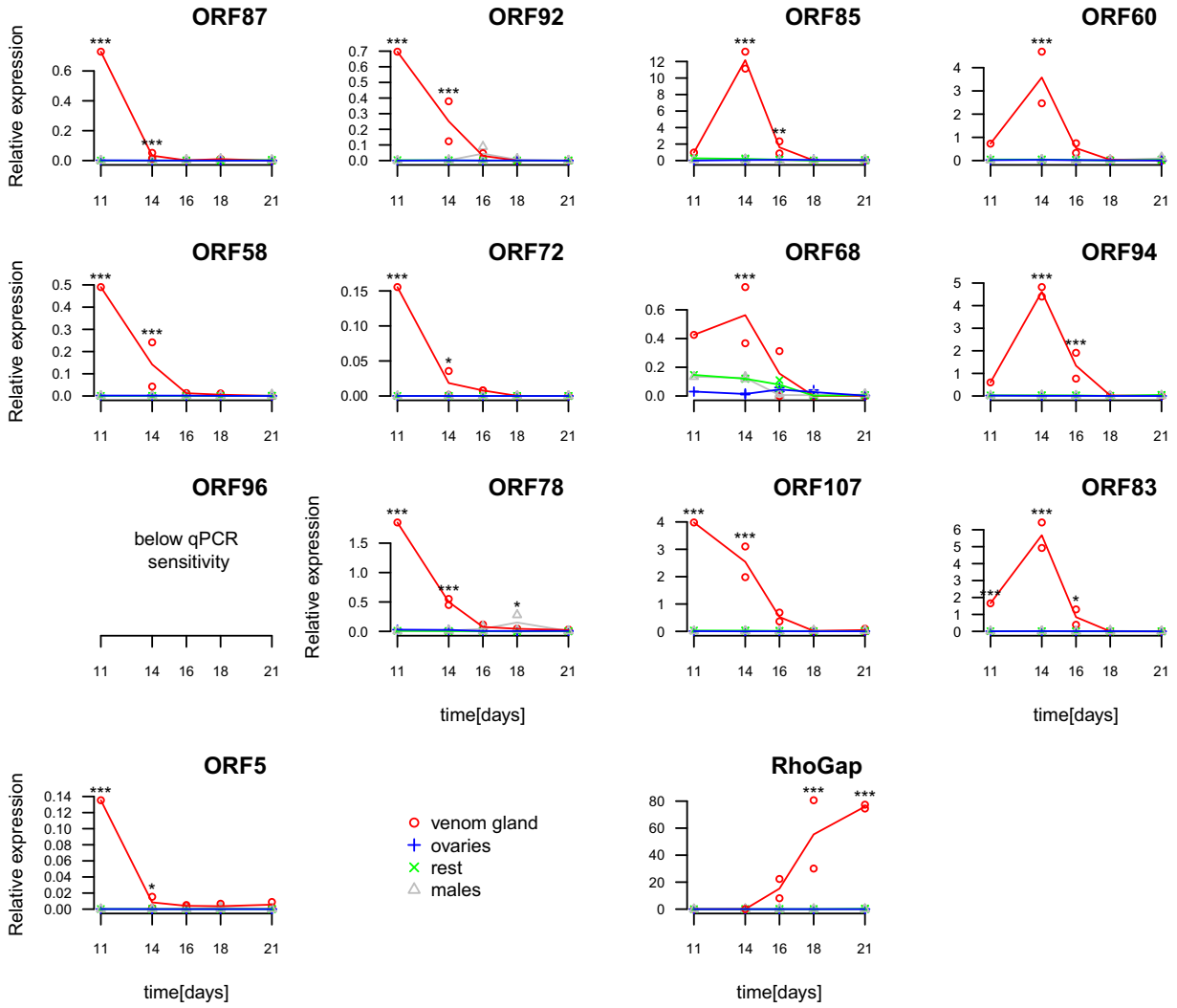

Fig. 4. Expression of the 13 virally derived genes and of the *Rho-GAP* in different tissues of *Leptopilina boulardi* from initial pupal stage to adult. *X*-axis represents days since egg-laying. Eleven days corresponds to the beginning of the pupal stage and 21 days to the emergence of adults from the *Drosophila* puparium. Stars correspond to the tissue effect tested at each time point (with holm correction for multiple tests): $* < 0.05$, $** < 0.01$, $*** < 0.001$.

Two sets of genes could also be identified based on their level of expression. One set of genes had an expression between 3 and 12 times that of the actin control gene (ORFs 94, 107, 60, 83, and 85), whereas the other genes had lower levels of expression, below 1.8 times that of the actin control (ORFs 5, 72, 68, 92, 87, 58, and 78). ORF96 was even below the detection threshold in our assay (Cqs above 35 cycles).

Under our hypothesis, viral genes have been co-opted by the wasps to deliver eukaryotic virulence proteins into *Drosophila* immune cells. Thus, to have a complete picture of the biogenesis of these supposedly "hybrid" structures, we also measured the expression of a wasp virulence protein. This protein of eukaryotic origin (Colinet et al. 2007) is known as a major component of wasp venom, most likely wrapped within the VLPs in *L. boulardi* (the RhoGAP LbGAP; Labrosse et al. 2005; Colinet et al. 2007; Goecks et al. 2013). As expected, this gene is also specifically expressed in the venom gland, and transcription starts just after the 14-day peak observed for most virally derived genes (fig. 4).

Most Virally Derived Genes but Not the Major Wasp Virulence Factor Are Amplified in the Venom Gland

Interestingly, among "early" virally derived genes, we identified a putative DNA polymerase (ORF58, see table 2). This opened the fascinating possibility that the DNA encoding those genes is amplified during this biological process. Using real-time PCR, we measured the relative DNA levels of each gene compared with an actin single-copy locus. As in the transcription assay, we measured it in the venom gland, ovaries, rest of the body, and in males of *L. boulardi*. We also included another single-copy gene (shake) as a control. As expected the relative copy number of shake did not show any trend in time, nor differences between tissues, thus validating our assay (fig. 5). We observed similar "flat" patterns for ORF87, ORF58, and ORF96 although a statistically significant effect was detected at day 11 for ORFs 87 and 96. On the contrary, all other virally derived genes were significantly amplified in the venom gland but not in other tissues. This amplification was highly significant for most genes at day 14,

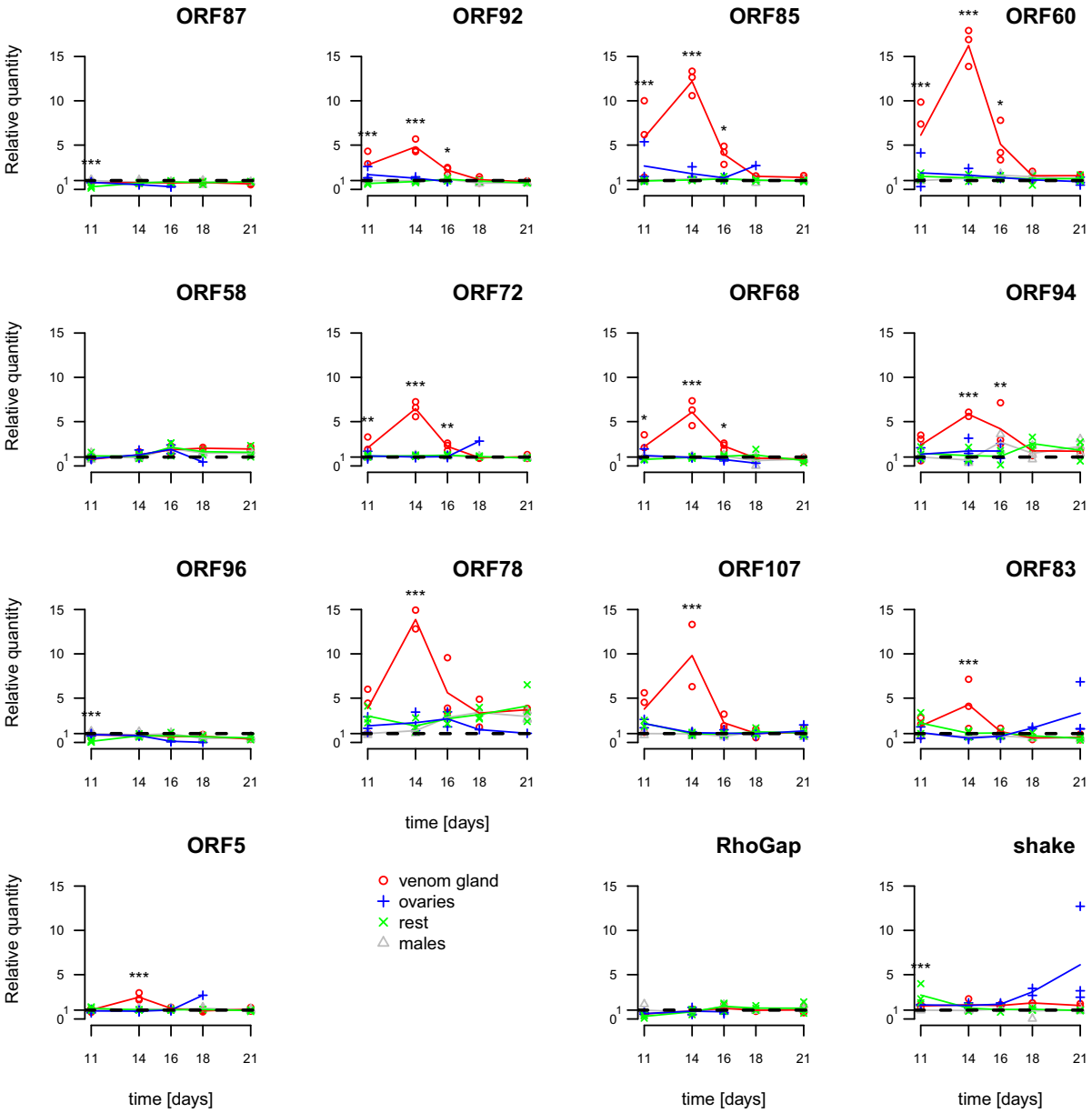

Fig. 5. Genomic amplification of virally derived genes measured by real time PCR in *Leptopilina boulardi*. The relative quantity of each target gene is represented relative to the actin control gene and normalized by the ratio observed in males at day 11. The expected value under no amplification (relative quantity = 1) is indicated as a dotted gray line. Stars correspond to the tissue effect tested at each time point (with holm correction for multiple tests): $* < 0.05$, $** < 0.01$, $*** < 0.001$.

where they all reached their peak of amplification. Interestingly, among the three genes that were not amplified is the putative DNA-polymerase (ORF58). This gene showed an early-transcription profile in the transcriptomic assay. The same "early-gene expression pattern" is also observed for the other nonamplified gene (ORF87). For most virally derived genes, we observed a striking correlation between the transcription and amplification profiles (compare figs. 4 and 5). Finally, our data set indicates that the gene encoding the major constituent of VLPs (LbGAP) is not amplified (fig. 5).

Annotation of Virally Derived Genes

Out of the 13 viral genes, 10 had similarities with known proteins from refseq or nr(table 2). However, conserved domains were only detected in five of them. First, five proteins had annotation suggesting involvement in DNA replication and or RNA transcription (ORFs 58, 68, 78, 92, and 96, yellow highlight in table 2). ORF60 bears a lecithine cholesterol acyl transferase (LCAT) domain, known to be involved in cholesterol metabolism (Saeedi et al. 2015). Three viral proteins had homologies with proteins of various putative functions (toll-like-receptor, FAD-containing oxidoreductase, *N*-acetyl-D-glucosamine kinase) but without conserved domains. Finally, ORF85 contains an Ac81 domain, a conserved domain found in all baculoviruses (van Oers and Vlak 2007). Interestingly, Ac81 is involved in virus envelopment in baculoviruses (Dong et al. 2016). They also showed that Ac81 contains a hydrophobic transmembrane domain that is necessary for this step. Interestingly, we also

Table 1. BLAST Hits for the 13 Viral Proteins against *Leptopilina* Genomes (TBlastN).

Query			*Leptopilina boulardi*			*Leptopilina heterotoma*			*Leptopilina clavipes*		
	ID	Length	Identity	Aln.length	*e*-Value	Identity	Aln.length	*e*-Value	Identity	Aln.length	*e*-Value
1	LbFV_ORF5	696	34.40	366	5.5e-41	29.70	370	3e-37	33.10	366	1.9e-40
2	LbFV_ORF72	106	31.80	107	5.2e-10	28.60	70	4e-04	32.70	107	8.8e-09
3	LbFV_ORF92	1,593	33.80	1,058	2.9e-151	38.10	501	5e-94	33.70	998	3.1e-136
4	LbFV_ORF107	625	29.80	322	1.3e-11	27.10	170	9e-09	28.30	378	5.3e-10
5	LbFV_ORF94	182	29.00	176	5.5e-14	27.60	174	1e-11	27.00	174	1.2e-12
6	LbFV_ORF68	645	34.10	646	6.7e-99	32.60	660	3e-92	34.00	674	3.5e-103
7	LbFV_ORF60	362	32.60	377	2.4e-36	26.00	381	7e-30	31.80	384	1.4e-33
8	LbFV_ORF85	215	36.40	225	3.0e-26	35.20	219	1e-23	33.00	218	1.3e-23
9	LbFV_ORF87	176	30.90	162	6.5e-12	29.00	162	1e-05	31.50	165	3.6e-11
10	LbFV_ORF58	1,308	36.70	932	1.3e-129	31.50	1,378	8e-158	31.50	1,042	1.8e-120
11	LbFV_ORF78	676	40.10	670	1.2e-134	41.00	646	2e-123	41.00	675	3.7e-135
12	LbFV_ORF83	433	24.80	435	1.6e-15	21.90	429	8e-15	24.50	436	1.8e-20
13	LbFV_ORF96	1,048	41.90	1,024	4.0e-169	36.60	1,043	2e-164	40.40	1,013	1.3e-178

NOTE.—The order is the same as in figure 1 for *L. boulardi*.

Table 2. Homology Search for the 13 LbFV Proteins Endogenized in *Leptopilina* Wasps.

ORF	Query Length	Representative Description (min *e*-value with this description)	*n* Sequences	Domains	Dom. Name	Dom. Pos	Dom. *e*-Value
ORF5	696	—	None				
ORF58	1,308	DNA polymerase (4e-72)	>500	DNA_pol_B	PF00136.20	639–870	1.4×10^{-24}
ORF60	362	Lecithin-cholesterol acyltransferase (LCAT) (10^{-55})	>500	LCAT	PF02450.14	76–172	6.7×10^{-11}
ORF68	645	ATP-dependent DNA helicase (2×10^{-35})	>500	PIF1-like helicase	PF05970.13	124–167; 248–320	3.3×10^{-8}; 3.5×10^{-6}
ORF72	106	Toll-like receptor (2×10^{-27})	57				
ORF78	676	DNA-directed RNA polymerase subunit (4×10^{-54})	>500	RNA_pol_Rpb1_2	PF00623.19	358–415	5.4×10^{-7}
ORF83	433	—	None				
ORF85	215	Ac81 (1×10^{-62})	41	Ac81	PF05820.10	56–201	1.3×10^{-25}
ORF87	176	FAD-containing oxidoreductase (1×10^{-29})	>500 (nr)				
ORF92	1,593	DNA primase/helicase (4×10^{-35})	206				
ORF94	182	*N*-acetyl-D-glucosamine kinase (5×10^{-51})	418				
ORF96	1,049	Lef-8 like protein (DNA-directed RNA polymerase subunit) (3×10^{-26})	3 (nr)				
ORF107	625	—	None				

NOTE.—A PSI-BLAST on refseq was first ran. When no homologs were found, a second PSI-BLAST against nr was performed (indicated in the "*n* Sequences Column"). Domain homology was detected using hmmer website. Only domain hits with *e*-values <0.15 are shown. *n* sequences indicate the number of sequences retrieved in the PSI-BLAST analysis. Light highlight indicates proteins putatively involved in DNA and RNA processing. ORF85 is also highlighted to indicate its presence as a protein in mature VLPs.

found the presence of a hydrophobic transmembrane domain for all three orthologs in *Leptopilina* sp. (supplementary fig. S25, Supplementary Material online).

A Virally Derived Protein Is Present in Mature VLPs of *Leptopilina* sp

In order to further test the hypothesis that the virally derived genes are involved in VLP formation, we purified VLPs from adult *L. boulardi* females. Mass spectrometry-based proteomics was then used to identify proteins present in two independent biological replicates (supplementary fig. S26, Supplementary Material online). This strategy allowed the identification of a total of 383 proteins, of which 236 were found in both replicates. Among these proteins, as expected, we were able to reproducibly identify typical virulence proteins known to be part of VLP content (such as the RhoGAP *LbGAP* [Colinet et al. 2007], superoxide dismutase [Colinet et al. 2011], serpin [Colinet et al. 2008], or calreticulin [Zhang et al. 2006]) confirming that we correctly purified the proteins (supplementary table S10, Supplementary Material online). We also reanalyzed a similar VLP proteomic data set obtained by others (Heavner et al. 2017) on the related species *L. heterotoma*. First, both *L. boulardi* and *L. heterotoma* data sets indicated enrichment for similar gene ontology terms, further suggesting that we correctly purified the VLP proteins (supplementary tables S8 and S9, Supplementary Material online). More importantly, in both *L. boulardi* biological replicates we found the presence of the endogenized version of LbFV ORF85 protein (three peptides in sample 1 and two in sample 2; supplementary table S10,

Supplementary Material online). We also detected the endogenized version of LbFV ORF85 protein, although with a single peptide, within the *L. heterotoma* data set (data not shown). Taken together, these data demonstrate that the virally derived protein ORF85, an Ac81 homolog encoded in the genome of *Leptopilina* species, is part of mature VLPs.

Discussion

In this article, we showed that all *Leptopilina* species tested so far contain a set of genes of viral origin deriving from either a direct ancestor of LbFV or a closely related one. We describe the genomic structure of those genes in detail in *L. boulardi*, *L. heterotoma*, and *L. clavipes*, for which the whole genome was obtained. In addition, we were able to detect the presence of one LbFV-derived gene (ORF96) in all *Leptopilina* DNA extracts tested so far, suggesting that those virally derived genes are shared by all *Leptopilina* species. Finally, one virally derived protein (ORF85) is detected in purified VLPs. From this analysis, we conclude that an ancestor of all *Leptopilina* species acquired a set of 13 viral genes deriving from a virus related to the behavior-manipulating virus LbFV, although we cannot rule out the possibility that the event occurred before the split between *Ganaspis* and *Leptopilina* with a subsequent loss in *Ganaspis*. These genes have been conserved in all *Leptopilina* species and allow them to produce immunosuppressive VLPs. This is very likely the consequence of a single event.

So far, all studied *Leptopilina* species are known to produce VLPs in their venom gland (Rizki RM and Rizki TM 1990; Morales et al. 2005; Gueguen et al. 2011). We confirmed this result in *L. boulardi* and found typical VLPs also in *L. clavipes*, suggesting that all *Leptopilina* species do produce VLPs. These particles are produced at the pupal stage and are stored in the reservoir of the venom gland. During oviposition, females inject not only their egg(s) but also some VLPs into their *Drosophila* hosts. VLPs are conceptually similar to liposomes that may contain virulence proteins. VLPs then permit the wasp to deliver these proteins to *Drosophila* immune cells (Colinet et al. 2007). The virulence proteins delivered to the target cells then induce important morphological changes in the lamellocytes, precluding them from initiating an efficient immune reaction against the parasitoid egg (Colinet et al. 2007). Thus, the VLPs are essential for the reproduction of the wasps. Because the proteins wrapped within the VLPs have a eukaryotic origin and because neither viral transcripts, viral proteins, nor viral DNA had been identified from venom gland analysis, it has been claimed that VLPs do not have a viral origin (Poirié et al. 2014; Heavner et al. 2017). In addition, the description of VLP proteins with eukaryotic microvesicular signature has been put forward as an evidence of a eukaryotic origin for these structures (Heavner et al. 2017). Following this argument, the authors proposed to change the denomination of VLPs for MSEV (mixed-strategy extracellular vesicle). On the contrary, our data strongly suggest that the VLPs found in *Leptopilina* do have a viral origin and derive from a massive endogenization event involving a virus related to an ancestor of the behavior-manipulating virus LbFV (supplementary fig. S2*B*, Supplementary Material online). Under this scenario, present-day VLPs are indeed eukaryotic structures but evolved thanks to the endogenization and domestication of ancient viral genes. These "hybrid" structures (relying on genes of both viral and eukaryotic origin) now allow the delivery of eukaryotic virulence proteins to *Drosophila* immune cells.

As expected from this hypothesis, we found that the virally derived genes are specifically expressed in the venom gland, during the first part of the pupal stage, the time when VLPs are beginning to be produced. In addition, those genes are under strong purifying selection, as would be expected for genes involved in the production of fitness-related structures such as VLPs. Analyzing the putative biological function of the genes brings additional support in favor of this hypothesis. In particular, two of them have functions suggesting that they could be involved in membrane formation.

The first one is ORF60 which contains an LCAT domain. In humans, LCAT is involved in extracellular metabolism of plasma lipoproteins, including cholesterol. LCAT esterifies the majority of free cholesterol, catalyzing translocation of the fatty acid moiety of lecithin (phosphatidyl choline) to the free 3-OH group of cholesterol. It thus plays a major role in the maturation of HDL (high-density lipoprotein cholesterol) (Saeedi et al. 2015). This putative biological property makes sense under our hypothesis because VLPs resemble liposomes that may be composed of highly hydrophobic compounds such as cholesterol. We may thus speculate that ORF60 plays a crucial role in the early formation of the "empty" membranes observed in the lumen of the venom gland under TEM (fig. 3*A* and *B*). Interestingly, the phylogenetic reconstruction of this gene suggests that LbFV itself acquired the LCAT gene from a bacterial donor species.

The second relevant gene is ORF85. ORF85 is an homolog of Ac81, a conserved protein found in all baculoviruses (van Oers and Vlak 2007). Its role has been recently deciphered in *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV; Dong et al. 2016). Baculoviruses first produce budded virions and, late in infection, occlusion-derived virions. After the initial infection, budded virions are responsible for the spread of the infection from cell to cell within the infected insect, whereas occlusion-derived virions are produced at the final stage of the infection. At that point, nucleocapsids are retained in the nucleus where they acquire an envelope from microvesicles. They are then exported into the cytoplasm and are embedded into proteinaceous crystal matrix, thus forming occlusion bodies (OBs). The OBs are then released in the environment. OBs are absolutely necessary to initiate new insect infections through horizontal transmission. By a mutant analysis, Dong et al. (2016) showed that Ac81 is necessary for the capsid envelopment and embedding within the OBs. They also showed that Ac81 contains a hydrophobic transmembrane domain that is necessary for this step. Interestingly, all three orthologs in *Leptopilina* sp. also contain a TM domain (supplementary fig. S25, Supplementary Material online). Our hypothesis is that the homolog of Ac81 in *Leptopilina* species is involved in the wrapping of

virulence proteins into the VLPs, which is observed at day 18 under electron microscopy (fig. 3C). Interestingly, it has been found that the closest viral homolog of this protein (apart from LbFV) is a structural protein of the Hytrosaviridae GpSGHV (Kariithi et al. 2010). In line with this, we found that protein ORF85 is indeed part of mature VLPs in *L. boulardi* and *L. heterotoma* and very likely in all *Leptopilina* species. This protein thus probably plays a crucial role in wrapping virulence proteins into VLP membranes and/or in the fusion with the target *Drosophila* immune cells. Interestingly, a nudiviral homolog of Ac81 has also been domesticated by *V. canescens* where three paralogs are found (Leobold et al. 2018) and also by *Fopius* wasps (Burke et al. 2018).

Apart from these two genes, the other genes with some annotation reveal functions related to DNA replication and transcription (in yellow in table 2). The presence of a putative DNA polymerase (ORF58) and a helicase (ORF68) may sound surprising if one considers that VLPs do not contain DNA, contrary to PDVs. However, we observed that after the early transcription activation of the DNA polymerase (at day 11), 10 out of the 13 virally derived genes were subsequently amplified (at day 14). This genomic amplification correlates very well with their respective expression profile which suggests that the transcriptomic regulation of these virally derived genes is governed, at least partly, by the gene copy number in the cell. Interestingly, the DNA polymerase itself and the nearby virally derived gene (ORF87) are not amplified, suggesting that the amplification depends on the location of the loci in wasp chromosome. It is unclear at that point whether the genomic amplification involves the production of circular or linear amplicons or concatemers, and where are located the boundaries of the amplified loci. On the contrary, the gene encoding the major constituent of the VLPs (LbGAP), which does not have a viral origin, is not amplified, although it is highly transcribed from day 14 until the emergence of the wasp and finally detected in mature VLPs as a protein. This suggests that the virally derived DNA polymerase targets some specific sequences flanking the amplified loci.

Altogether, our data strongly suggest that VLP production is possible thanks to the domestication of 13 virally derived genes, captured from an ancestor of LbFV. Most of these virally derived genes may be involved in the amplification and transcription of other viral genes, ultimately leading to the incorporation of one virally derived protein in mature VLPs (ORF85). This is in contrast to other VLPs systems described so far, where usually several envelope proteins are incorporated into mature VLPs (Pichon et al. 2015; Burke et al. 2018). This may relate to some specificity of the *Drosophila*–wasp interaction and deserves further investigation.

Based on the gene clustering and synteny conservation in our *L. boulardi* assembly, we speculate that a single event led to the acquisition of the whole gene set. We can even hypothesize that a whole virus genome integrated into the chromosome of the *Leptopilina* ancestor. Several recent publications suggest that large, possibly full-genome insertions of symbiont into their host DNA do occur in the course of evolution, including from dsDNA viruses. For instance, whole-genome sequencing of the brown planthopper revealed a total of 66 putative ORFs (74,730 bp in total) deriving from a nudivirus genome, including 32 out of the 33 core nudiviral genes (Cheng et al. 2014). Also, it has been recently shown that an almost complete *Wolbachia* genome has been integrated into the chromosome of its host the common pillbug *Armadillidium vulgare*, with dramatic consequences on its sex determination system (Leclercq et al. 2016). After this suspected full-genome insertion of an ancestor of LbFV, we speculate that subsequent rearrangements have eliminated unnecessary genes and finally scattered, to a certain degree, the 13 remaining genes. Better genome assemblies are now necessary to gain insights on this aspect of the domestication process in the different *Leptopilina* lineages.

From the literature it is clear that the domestication of whole sets of viral genes has repeatedly occurred in endoparasitoid wasps belonging to the superfamily Ichneumonoidea, with at least two events leading to PDV systems (that deliver DNA circles encoding virulence factors to the host) in some Braconidae and Ichneumonidae and two events leading to the evolution of a VLP system (that deliver virulence proteins wrapped into a liposome-like structure to the host) in *Fopius* (Opiinae) (Burke et al. 2018) and in *V. canescens* (Campopleginae) (Herniou et al. 2013; Pichon et al. 2015). Actually, this last VLP domestication in *V. canescens* better corresponds to a replacement of a PDV system by a VLP system (Pichon et al. 2015), showing that domestication events have been frequent in this superfamily. With our results obtained on species belonging to the family Figitidae, which diverged from Ichneumonoidea 225 Ma (Peters et al. 2017), it is tempting to extend this conclusion to other clades of Hymenoptera endoparasitoids.

If this idea is confirmed, then a striking parallel comes up between virus domestication in Hymenoptera and syncytin domestication in mammals (Lavialle et al. 2013). In both cases, viral proteins have been repeatedly co-opted to permit cell–cell fusion, although in one case this is for materno-fetal communication and in the second case it is for virulence factor delivery. Future investigations should test more thoroughly this hypothesis.

One remaining open question for all those events is the type of interaction the ancestral virus and its wasp did have before the domestication happened. Regarding this question, very little data are available up to now. For PDV found in Campopleginae such as *H. dydimator* and in Banchinae such as *Glypta fumiferanae*, the ancestral virus has not been clearly identified (Volkoff et al. 2010; Béliveau et al. 2015). On the contrary, the putative virus donors have been identified as a beta-nudivirus for PDVs in Braconidae (Bezier et al. 2009) and as an alpha-nudivirus for VLPs found in *V. canescens* (Pichon et al. 2015) and in *Fopius* species (Burke et al. 2018). However, their closest viral relatives do not infect hymenoptera but rather other arthropods (Theze et al. 2011). In addition, the endogenization event is ancient, at least for Bracoviruses, which is the only case for which an estimate exists (103 My; Murphy et al. 2008), rendering difficult the inferences

on the type of association that existed upon emergence of the association.

In *Leptopilina*, we unequivocally identified an ancestor (or a close relative) of the behavior-manipulating virus LbFV as the donor virus. First, it should be noted that in previous cases for which the ancestor has been identified the donor virus has a large circular genome composed of a dsDNA. Our results again show the same pattern. Second, the previous studies repeatedly identified nudiviruses as the donor family. Here we identify a virus belonging to another, possibly new, virus family (Lepetit et al. 2016). This virus is related to nudiviruses and baculoviruses, but is more closely related to the hytrosaviruses (Abd-Alla et al. 2009), which is known to induce salivary gland hypertrophy in tsetse flies and house flies, although it can also remain symptom-less (Abd-Alla et al. 2008).

Finally, this is the first time that the identified virus ancestor still has extant relatives infecting one of the wasp species. From our previous work on the interaction between LbFV and its host *L. boulardi*, we know that LbFV is vertically transmitted and replicate in cells of the oviduct (Varaldi et al. 2006). This result suggests that physical proximity with the germ line may have facilitated the initial endogenization event, thus allowing the initiation of the domestication process. The identification of a contemporary virus still infecting the wasp also opens the way for addressing experimentally the mechanisms by which the virus could integrate into wasp chromosomes. Finally, LbFV is responsible for a behavior manipulation in *L. boulardi*: It forces females to superparasitize, which allows its horizontal transmission to other wasps (Varaldi et al. 2003). This raises the fascinating possibility that the ancestral donor virus also manipulated the behavior of the wasp. The sampling of relatives of LbFV will be essential to be able to reconstruct the ancestral state for the lineage that actually gave rise to such genetic innovation in wasp genomes.

Materials and Methods

Wasp Biology and Rearing Conditions

All *Leptopilina* species are endoparasitoids of *Drosophila* species, and their biology has been the subject of intense investigations on behavior, physiology and ecology (Carton et al. 1986). Little is known, however, on *Ganaspis* biology, apart from the fact that at least some *Ganaspis* species parasitize *Drosophila* larvae (such as *G. brasiliensis* and the uncharacterized *G.* sp used in this study). Some *Ganaspsis* species have been investigated in terms of immune interaction with the *Drosophila* (Ferrarese et al. 2009; Mortimer et al. 2013), revealing important differences with *Leptopilina* species (Mortimer 2013).

Leptopilina boulardi, *L. heterotoma*, and *G. brasiliensis* were reared on *D. melanogaster* as host (StFoy strain) in a climatic chamber (25 °C, 60% humidity, 12/12 LD). The *G. brasiliensis* strain was kindly provided by Dr Shubha Govind, *L. clavipes* by Dr Elzemiek Geuverink, and *L. boulardi* and *L. heterotoma* strains were collected and identified by our group. *Drosophila* were fed with a standard medium (David 1962). All experiments on *L. boulardi* were performed on a strain uninfected with the behavior-manipulating virus (NSref).

Wasp Genome Sequences and Annotation

We previously reported the genome of *L. boulardi*, strain Sienna (accession number: PQAT00000000) which has been obtained from the sequencing of a single female (Varaldi and Lepetit 2018). Although this female was infected by LbFV, the draft genome does not contain contigs belonging to the virus genome as we removed them by comparison to the published virus genome sequence (Lepetit et al. 2016). The assembly was performed using IDBA_ud (Peng et al. 2012) followed by a scaffolding step with assembled RNAseq reads using the software L_RNA_scaffolder (Xue et al. 2013).

We sequenced the genomes of the related *L. heterotoma* (Gotheron strain, accession number RICB00000000), the more distantly related *G. brasiliensis* (Va strain, accession number RJVV00000000), and an uncharacterized *G.* sp emerged from *Drosophila* collected in Brazil (Cabo Frio, RJ, accession number WNHD00000000). *Leptopilina heterotoma* is refractory to infection by LbFV (Patot et al. 2012), and no reads mapping to LbFV genome have been found neither in *L. heterotoma* nor in *G. brasiliensis* or *G.* sp data sets. We extracted the DNA of a single female abdomen using Macherey-Nagel columns, similarly to what was performed for *L. boulardi* (Varaldi and Lepetit 2018). The DNAs were then used to prepare paired-end Illumina libraries using standard protocols (TruSeq PE Cluster v3, TruSeq SBS 200 cycles v3, TruSeq Multiplex Primer). The libraries were then sequenced on a Hiseq2500 (for Lh, 2 × 100 bp, insert size = 418 bp) or Hiseq3000 (2 × 150 bp; for Gb: insert size = 427 bp, for *G.* sp: insert size = 438 bp) machine on the Genotoul sequencing platform.

Similarly to what was done for *L. boulardi*, the drafts of *L. heterotoma*, *G. brasiliensis*, and *G.* sp were obtained after assembling genomic DNA reads with IDBA_ud (Peng et al. 2012). For *L. heterotoma* assembly, this was followed by scaffolding using publicly available assembled RNAseq reads (Goecks et al. 2013) by running the software L_RNA_scaffolder (Xue et al. 2013). This RNAseq scaffolding step was not performed for *Ganaspis* species because no RNAseq reads were available for these species in public databases.

The genome of an asexual strain of *L. clavipes* (strain GBW) which is not infected by LbFV was obtained and is described in Kraaijeveld et al. (2016) (accession number JUFY00000000). To have comparable assembly strategies, we included an additional RNA scaffolding step using publicly available sequences (Misof et al. 2014).

In order to test the completeness of the drafts generated, we ran the BUSCO pipeline (version 2.0) that looks for the presence of 1,066 ubiquitous genes shared by at least 90% of all arthropods (Simão et al. 2015).

The genome sizes were estimated using several methods. First of all, we simply divided the total number of bases mapped to the draft by the mean coverage observed on scaffolds containing complete BUSCO genes. Those scaffolds are expected to contain nonrepeated nuclear DNA, and their

coverage is a valuable estimate of the coverage for any nuclear locus. Second, after filtering out adapters containing reads with Skewer version 0.2.2 (Jiang et al. 2014), removing reads duplicates with FastUniq version 1.1 (Xu et al. 2012), filtering out reads mapping to mitochondrial contigs with Bowtie 2 version 2.3.4.1 (Langmead and Salzberg 2012) and samtools version 1.8 (Li et al. 2009), removing contaminant reads (from viruses, prokaryotes, and microbial eukaryotes) with Kaiju 1.6.2 used with the NR+euk 2018-02-23 database (Menzel et al. 2016), k-mer frequencies were established from the remaining reads for each species using Jellyfish 2.2.9 (Marçais and Kingsford 2011) and $k = 21$ (default value). From these 21-mers distributions, genome size was estimated with findGSE (Sun et al. 2018) used with default parameters. These estimates were then used to run DNAPipeTE version 1.3 (Goubert et al. 2015) (two samples per run, 0.1× coverage per sample) in order to assess the repetitive fraction of the genomes. Finally, independent estimates from flow cytometry experiments were obtained for *L. boulardi*, *L. heterotoma*, and *G. brasiliensis* from Gokhman et al. (2011) and for *L. clavipes* from Kraaijeveld et al. (2016).

We predicted genes in wasp sequences using the software AUGUSTUS 3.2.3 (Hoff and Stanke 2013), with training parameters obtained from the BUSCO outputs.

Homology Search

In order to identify homologies between viral proteins and wasp DNA, we used a simple TBlastN (v. 2.6.0) approach with viral proteins as query and each wasp genome as database. Default parameters were used except that an *e*-value threshold of 0.01 and a minimum alignment length of 70 amino acids were chosen.

Phylogenies

Species Tree

Based on 541 "universal arthropod" genes identified by the BUSCO pipeline (Simão et al. 2015), a species tree was constructed for *L. heterotoma*, *L. boulardi*, *L. clavipes*, *G. brasiliensis*, and *G.* sp, using *Apis mellifera* as outgroup. The protein sequences were aligned using the bioconductor msa package (Bodenhofer et al. 2015). Individual alignments were concatenated and a phylogenetic reconstruction was then performed using PhyML (parameters: -d aa -m LG -b -4 -v e -c 4 -a e -f m) (Guindon et al. 2010). In total, 248,282 variable sites were found and the branch supports were computed using approximate likelihood ratio test (aLRT). We also constructed a tree for ten *Leptopilina* species and *G. brasiliensis* using publicly available sequences of ITS2. Alignment was performed with muscle and a phylogeny was obtained with PhyML (parameters: -d nt -m GTR -b -4 -v 0.0 -c 4 -a e -f e). In total, 399 variable sites were used and the tree was rooted using midpoint rooting method.

Gene Tree

We searched orthologs of viral proteins of interest in other organisms by blasting (BlastP) them against nr (downloaded on October 2017) with an *e*-value threshold of 0.01. After retrieving the sequences, we selected one sequence per species and added them to the proteins identified in *Leptopilina* genomes. The sequences were then aligned using muscle algorithm v3.8.31. Because the proteins included in the alignment diverged considerably, we selected blocks of conserved sites using the gblocks algorithm parametrized with less stringent options (allowing smaller final blocks, gaps within final blocks, and less strict flanking positions; Castresana 2000). Phylogenetic reconstruction was then performed using PhyML (parameters: -d aa -m LG -b -4 -v e -c 4 -a e -f m). The branch supports were computed using aLRT. The accession numbers of the sequences used in the phylogenies are reported in supplementary table S5, Supplementary Material online.

PCR Amplification of ORF96

Based on the sequences of *L. boulardi*, *L. heterotoma*, and *L. clavipes*, we designed primers for the orthologs of LbFVORF96. The primer sequences are ATTGGTGAAATT CAATCGTC and TCATTCATTCGCAATAATTGTG. They amplified a 411-bp internal fragment of the coding sequence. PCR reaction was performed in a 25 μl volume containing 0.2 μM primers, 0.2 mM dNTPs, 1 mM $MgCl_2$, and 0.5 U of Taq DNA polymerase with the following cycling conditions: 95 °C 30″, 54 °C 30″, 72 °C 60″ (33 cycles).

dN/dS Calculation

The coding sequences of "universal arthropod" BUSCO genes identified in the three *Leptopilina* species were extracted and we first aligned the corresponding proteins using MUSCLE v3.8.1551 (Edgar 2004). Then, the sequences were reverse-aligned using the protein alignments as a guide (pal2nal v14; Suyama et al. 2006). We then computed the value of ω (dN/dS) in all branches of the species tree using CodeML (Yang 2007) as implemented in the Python package ete3 (Huerta-Cepas et al. 2016). We used the free-ratio model to estimate the ω ratio for tree branches. Each model was run three times with different starting values and we selected the model with the highest likelihood (starting ω values = 0.2, 0.7, 1.2). Then, a global ω ratio for all the tree was calculated as the tree length for d*N* divided by the tree length for d*S*. To compare the average d*N*/d*S* obtained for the 13 virally derived genes, we followed a bootstrap approach. We randomly sampled 13 d*N*/d*S* values among the 958 BUSCO d*N*/d*S* (with replacement), calculated the average, and repeated this procedure 500 times. The empirical *P*-value presented in the text corresponds to the proportion of simulations giving a similar or higher mean value compared with the observed mean for the 13 "viral" d*N*/d*S*.

Expression in the Venom Gland and Other Tissues

We studied the expression of genes during the pupal stage of *L. boulardi*, at days 11, 14, 16, 18, and 21. The wasp strain used is not infected by the behavior-manipulating virus LbFV. Eleven days corresponds to the beginning of the pupal stage, whereas 21 days corresponds to the emergence time. Wasps were gently extirpated from the *Drosophila* puparium, and venom gland (+ its reservoir), ovaries, rest of the body of

L. boulardi females were dissected in a droplet of PBS + 0.01% tween and deposited in the RLT+B-mercaptoethanol buffer of the Qiagen RNAeasy extraction kit. Males were also prepared as a control, in a similar way. The tissues extracted from 20 individuals were then pooled together and tissues were disrupted in a Qiagen homogenizer (3 min, 25 Hz). Two biological replicates were performed for each condition, except for day 11 where only one sample was obtained. Total RNA, previously DNAse treated (Turbo DNAse I, Ambion), was reverse transcribed using random primers in a 20 μl final volume (SuperscriptIII kit, ThermoFisher). After a 1/10 dilution, 4 μl of cDNA was used as template in a 10 μl final volume real-time PCR assay on a Biorad CFX-96 machine (SYBR green, SsoAdvanced Universal SYBR Green Supermix, Biorad). We quantified the number of copies of each target cDNA using a serial dilution standards (see supplementary table S7, Supplementary Material online, for details). Because we obtained only tiny quantities of RNA from this experiment (because of the very small size of the tissues dissected), we were not able to test numerous genes. We thus choose to use only one control gene (actin gene). As a counterpart, we were able to test all 13 virally derived genes and the *RhoGAP* gene. The primer sequences are given in supplementary table S6, Supplementary Material online.

Genomic Amplification

Using a similar assay, we extracted the DNA of *L. boulardi*, at days 11, 14, 16, 18, and 21, using an uninfected strain (no LbFV present). The genomic DNA of 15 pooled individuals was extracted using the Nucleospin tissue Macherey-Nagel kit following provider's instructions. Three biological replicates per condition were done. Real-time PCR assays were then performed with SYBR green using standard procedures on a Biorad CFX-96 machine. We quantified the number of copies of each target genes using a serial dilution standards. The primer sequences are given in supplementary table S6, Supplementary Material online. For an unknown reason, the amplification with DNA extracted from ovaries was particularly difficult, in particular when the ovaries were mature (at day 21). We thus had to remove this tissue from the statistical analysis because Cqs were too high to be reliable. For the same reason, most data for ovaries at day 21 were removed from figure 5. Shake and actin genes were chosen as single-copy genes. This was checked by looking at the BLAST results using each primer set (a single 100% match was observed for both pairs of primers). Accordingly, a single band of the expected size was observed on a gel and the expected sequence was obtained after Sanger-sequencing for both loci.

Statistical Analysis

For both the transcriptomic and genomic analyses, we calculated the absolute copy number of each gene of interest and divided it by the absolute copy number of the actin control gene. This ratio was then analyzed in an ANOVA framework with time, tissue, and time:tissue interaction as factors. The effects were tested by likelihood ratio tests of full model versus reduced one. Contrasts between tissues were also calculated at each time point (corresponding to the star in figs. 4 and 5). Residuals of the models were judged as unstructured and had an overall normal distribution.

Morphogenesis and Electron Microscopy of the Venom Gland

To follow the morphogenesis of the venom gland, we dissected *L. boulardi* pupae at days 11, 14, 16, 18, and 21, in a similar design used for transcriptomics. Wasps were gently extirpated from the *Drosophila* puparium, and the venom gland of females was dissected in a droplet of PBS + 0.01% tween. Venom glands were either directly mounted on a glass slide for further examination under a light microscope or transferred into a solution of 2% glutaraldehyde in PBS for further examination under the TEM. For TEM, the tissues were then postfixed 1 h in 2% osmium tetroxide in the same buffer, thoroughly rinsed in distilled water, stained "en bloc" with a 5% aqueous uranyl acetate solution, dehydrated in a series of graded ethanol, and embedded in Epon's medium. Ultrathin sections were cut on an LKB ultratome and double stained in UranyLess and lead citrate. Samples were examined with a Jeol 1200 Ex transmission microscope at 80 kV. Images were taken with a Quemesa 11-megapixel Olympus camera and analyzed with ImageJ software (https://imagej.nih.gov/ij/).

Proteomics

VLP Purification

VLPs were purified from the venom gland and reservoir extracted from 50 (sample 1) or 150 (sample 2) *L. boulardi* females as in Pichon et al. (2015). The venom glands and reservoir were dissected in 1× PBS and gently disrupted in order to release the VLPs. The sample was then centrifuged (200 × g, 10 min, 4 °C) to remove the disrupted venom glands and reservoirs. The supernatant, containing VLPs, was transferred to a new tube and recentrifuged (20,000 × g, 30 min, 4 °C). The pellet containing purified VLPs was eluted in a Laemmli solution and frozen at −80 °C.

Protein Sequencing

Proteins extracted from purified VLPs were stacked in the top of an SDS-PAGE gel (4–12% NuPAGE, Life Technologies), stained with Coomassie blue R-250 and in-gel digested using modified trypsin (Promega, sequencing grade) as previously described (Salvetti et al. 2016). Resulting peptides were analyzed by online nanoliquid chromatography coupled to tandem mass spectrometry (UltiMate 3000 RSLCnano and Q-Exactive HF, Thermo Scientific). Peptides were sampled on a 300 μm × 5 mm PepMap C18 precolumn and separated on a 75 μm × 250 mm C18 column (Reprosil-Pur 120 C18-AQ, 1.9 μm, Dr Maisch) using a 120-min gradient. MS and MS/MS data were acquired using Xcalibur (Thermo Scientific). Peptides and proteins were identified using Mascot (version 2.6) through concomitant searches against the homemade *L. boulardi* database (see above for details), classical contaminant database, and the corresponding reversed databases. The Proline software (http://proline.profiproteomics.fr) was used to filter the results: conservation of rank 1 peptides,

peptide identification false discovery rate < 1% as calculated on peptide scores by employing the reverse database strategy and minimum of one specific peptide per identified protein group. Proline was then used to perform a compilation, grouping, and spectral counting-based comparison of the protein groups identified in the different samples. Proteins from the contaminant database were discarded from the final list of identified proteins.

Annotation of Viral Genes

We used a PSI-BLAST approach (BLOSUM45) using the 13 LbFV proteins as queries, on refseq (accessed July 15, 2019). In case no hit was found, a second round was performed using nr as database. In addition, we searched for the presence of conserved domains in the 13 LbFV proteins horizontally transferred to *Leptopilina* species using the hmmer webserver (https://www.ebi.ac.uk/Tools/hmmer/, accessed May 5, 2018).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by a grant from the Agence Nationale de la Recherche (ANR) to J.V. (11-JSV7-0011 Viromics). We thank the LABEX ECOFECT (ANR-11-LABX-0048) for financial support for the internship of D.D.G. and B. Bitner-Mathé and the Brazilian programs FAPERJ and CAPES/PVE-2012 for visits to UFRJ. Proteomic experiments were partly supported by the Agence Nationale de la Recherche (ProFI Grant ANR-10-INBSQ6). The bioinformatic work was performed using the computing facilities of the CC LBBE/PRABI. We thank Shubha Govind for kindly providing the *Ganaspis brasiliensis* strain, Elzemiek Geuverink for providing *Leptopilina clavipes* samples, Ken Kraaijeveld for providing access to raw Illumina reads of *L. clavipes*, D. Colinet and N. Mortimer for sharing mass-spec data and analysis, D. Kahn for advices on data analysis, and N. Burlet for help in the lab. We also thank PCI Evol Biol reviewers for helpful comments on an earlier draft which has been reviewed and recommended by Peer Community in Evolutionary Biology (https://dx.doi.org/10.24072/pci.evolbiol.100062). Scripts used for this publication are available at https://doi.org/10.5281/zenodo.1889392.

References

Abd-Alla AMM, Cousserans F, Parker AG, Jehle JA, Parker NJ, Vlak JM, Robinson AS, Bergoin M. 2008. Genome analysis of a *Glossina pallidipes* salivary gland hypertrophy virus reveals a novel, large, double-stranded circular DNA virus. *J Virol*. 82(9):4595–4611.

Abd-Alla AMM, Vlak JM, Bergoin M, Maruniak JE, Parker A, Burand JP, Jehle JA, Boucias DG; Hytrosavirus Study Group of the ICTV. 2009. Hytrosaviridae: a proposal for classification and nomenclature of a new insect virus family. *Arch Virol*. 154(6):909–918.

Béliveau C, Cohen A, Stewart D, Periquet G, Djoumad A, Kuhn L, Stoltz D, Boyle B, Volkoff A-N, Herniou EA, 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.

Bezier A, Annaheim M, Herbiniere J, Wetterwald C, Gyapay G, Bernard-Samain S, Wincker P, Roditi I, Heller M, Belghazi M, et al. 2009. Polydnaviruses of braconid wasps derive from an ancestral nudivirus. *Science* 323(5916):926–930.

Bitra K, Zhang S, Strand MR. 2011. Transcriptomic profiling of *Microplitis demolitor* bracovirus reveals host, tissue and stage-specific patterns of activity. *J Gen Virol*. 92(9):2060–2071.

Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. 2015. msa: an R package for multiple sequence alignment. *Bioinformatics* 31(24):3997–3999.

Boto L. 2014. Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc R Soc B*. 281(1777):20132450.

Burke GR, Simmonds TJ, Sharanowski BJ, Geib SM. 2018. Rapid viral symbiogenesis via changes in parasitoid wasp genome architecture. *Mol Biol Evol*. 15(2):137.

Carton Y, Boulétreau M, van Alphen JJM, Van Lenteren JC. 1986. The *Drosophila* parasitic wasps. In: Ashburner M, Carson HL, and Thompson JN, editors. The genetics and biology of Drosophila. Vol. 3. London: Academic Press Inc. p. 348–394.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17(4):540–552.

Cheng R-L, Xi Y, Lou Y-H, Wang Z, Xu J-Y, Xu H-J, Zhang C-X. 2014. Brown planthopper nudivirus DNA integrated in its host genome. *J Virol*. 88(10):5310–5318.

Chevignon G, Theze J, Cambier S, Poulain J, Da Silva C, Bézier A, Musset K, Moreau SJM, Drezen J-M, Huguet E. 2014. Functional annotation of *Cotesia congregata* bracovirus: identification of viral genes expressed in parasitized host immune tissues. *J Virol*. 88(16):8795–8812.

Cobbs C, Heath J, Stireman JO, Abbot P. 2013. Carotenoids in unexpected places: gall midges, lateral gene transfer, and carotenoid biosynthesis in animals. *Mol Phylogenet Evol*. 68(2):221–228.

Colinet D, Cazes D, Belghazi M, Gatti J-L, Poirié M. 2011. Extracellular superoxide dismutase in insects. *J Biol Chem*. 286(46):40110–40121.

Colinet D, Dubuffet A, Cazes D, Moreau S, Drezen J, Poirié M. 2008. A serpin from the parasitoid wasp *Leptopilina boulardi* targets the *Drosophila* phenoloxidase cascade. *Dev Comp Immunol*. 33(5):681–689.

Colinet D, Schmitz A, Cazes D, Gatti J-L, Poirié M. 2010. The origin of intraspecific variation of virulence in an eukaryotic immune suppressive parasite. *PLoS Pathog*. 6(11):e1001206.

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.

Cornelis G, Funk M, Vernochet C, Leal F, Tarazona OA, Meurice G, Heidmann O, Dupressoir A, Miralles A, Ramirez-Pinilla MP, et al. 2017. An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental *Mabuya* lizard. *Proc Natl Acad Sci U S A*. 114(51):E10991–E11000.

Danchin EGJ, Rosso M-N, Vieira P, de Almeida-Engler J, Coutinho PM, Henrissat B, Abad P. 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc Natl Acad Sci U S A*. 107(41):17651–17656.

David JR. 1962. A new medium for rearing *Drosophila* in axenic conditions. *Drosophila Info Service*. 36:128.

Di Lelio I, Illiano A, Astarita F, Gianfranceschi L, Horner D, Varricchio P, Amoresano A, Pucci P, Pennacchio F, Caccia S. 2019. Evolution of an insect immune barrier through horizontal gene transfer mediated by a parasitic wasp. *PLoS Genet*. 15(3):e1007998.

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.

Dupas S, Brehelin M, Frey F, Carton Y. 1996. Immune suppressive virus-like particles in a *Drosophila* parasitoid: significance of their intraspecific morphological variations. *Parasitology* 113(3):207–212.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5(1):113.

Espagne E, Dupuy C, Huguet E, Cattolico L, Provost B, Martins N, Poirié M, Periquet G, Drezen JM. 2004. Genome sequence of a polydnavirus: insights into symbiotic virus evolution. *Science* 306(5694):286–289.
Ferrarese R, Morales J, Fimiarz D, Webb BA, Govind S. 2009. A supracellular system of actin-lined canals controls biogenesis and release of virulence factors in parasitoid venom glands. *J Exp Biol*. 212(14):2261–2268.
Forbes AA, Bagley RK, Beer MA, Hippee AC, Widmayer HA. 2018. Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. *BMC Ecol*. 18(1):21.
Gandon S, Rivero A, Varaldi J. 2006. Superparasitism evolution: adaptation or manipulation? *Am Nat*. 167(1):E1–E22.
Gauthier J, Drezen J-M, Herniou EA. 2018. The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps. *Parasitology* 145(6):713–723.
Goecks J, Mortimer NT, Mobley JA, Bowersock GJ, Taylor J, Schlenke TA. 2013. Integrative approach reveals composition of endoparasitoid wasp venoms. *PLoS One* 8(5):e64125.
Gokhman VE, Johnston JS, Small C, Rajwani R, Hanrahan SJ, Govind S. 2011. Genomic and karyotypic variation in *Drosophila* parasitoids (Hymenoptera, Cynipoidea, Figitidae). *Comp Cytogenet*. 5(3):211–221.
Goubert C, Modolo L, Vieira C, ValienteMoro C, Mavingui P, Boulesteix M. 2015. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome Biol Evol*. 7(4):1192–1205.
Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, Osborne EJ, Dermauw W, Ngoc PCT, Ortego F, et al. 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479(7374):487–492.
Gueguen G, Rajwani R, Paddibhatla I, Morales J, Govind S. 2011. VLPs of *Leptopilina boulardi* share biogenesis and overall stellate morphology with VLPs of the heterotoma clade. *Virus Res*. 160(1–2):159–165.
Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59(3):307–321.
Guy L, Roat Kultima J, Andersson SGE. 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26(18):2334–2335.
Heavner ME, Ramroop J, Gueguen G, Ramrattan G, Dolios G, Scarpati M, Kwiat J, Bhattacharya S, Wang R, Singh S, et al. 2017. Novel organelles with elements of bacterial and eukaryotic secretion systems weaponize parasites of *Drosophila*. *Curr Biol*. 27(18):2869–2877.e6.
Herniou EA, Huguet E, Theze J, Bézier A, Periquet G, Drezen J-M. 2013. When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. *Philos Trans R Soc B*. 368(1626):20130051.
Hoff KJ, Stanke M. 2013. WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Res*. 41(W1):W123–W128.
Huerta-Cepas J, Serra F, Bork P. 2016. Ete 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol*. 33(6):1635–1638.
Husnik F, McCutcheon JP. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol*. 16(2):67–79.
Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer—a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15(1):182.
Kariithi HM, Ince IA, Boeren S, Vervoort J, Bergoin M, van Oers MM, Abd-Alla AMM, Vlak JM. 2010. Proteomic analysis of *Glossina pallidipes* salivary gland hypertrophy virus virions for immune intervention in tsetse fly colonies. *J Gen Virol*. 91(Pt 12):3065–3074.
Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet*. 9(8):605–618.
Kraaijeveld K, Anvar Y, Frank J, Schmitz A, Bast J, Wilbrandt J, Petersen M, Ziesmann T, Niehuis O, Knijff PD, et al. 2016. Decay of sexual trait genes in an asexual parasitoid wasp. *Genome Biol Evol*. 8(12):3685–3695.
Labrosse C, Carton Y, Dubuffet A, Drezen JM, Poirié M. 2003. Active suppression of *D. melanogaster* immune response by long gland products of the parasitic wasp *Leptopilina boulardi*. *J Insect Physiol*. 49(5):513–522.
Labrosse C, Eslin P, Doury G, Drezen JM, Poirié M. 2005. Haemocyte changes in *D. melanogaster* in response to long gland components of the parasitoid wasp *Leptopilina boulardi*: a Rho-GAP protein as an important factor. *J Insect Physiol*. 51(2):161–170.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–359.
Lavialle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, Heidmann T. 2013. Paleovirology of "syncytins", retroviral env genes exapted for a role in placentation. *Philos Trans R Soc B*. 368(1626):20120507.
Leclercq S, Theze J, Chebbi MA, Giraud I, Moumen B, Ernenwein L, Grève P, Gilbert C, Cordaux R. 2016. Birth of a W sex chromosome by horizontal transfer of *Wolbachia* bacterial symbiont genome. *Proc Natl Acad Sci U S A*. 113(52):15036–15041.
Leobold M, Bézier A, Pichon A, Herniou EA, Volkoff A-N, Drezen J-M. 2018. The domestication of a large DNA virus by the wasp *Venturia canescens* involves targeted genome reduction through pseudogenization. *Genome Biol Evol*. 10(7):1745–1764.
Lepetit D, Gillet B, Hughes S, Kraaijeveld K, Varaldi J. 2016. Genome sequencing of the behavior manipulating virus LbFV reveals a possible new virus family. *Genome Biol Evol*. 8(12):3718–3739.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
Malmstrøm M, Matschiner M, Tørresen OK, Jakobsen KS, Jentoft S. 2017. Whole genome sequencing data and de novo draft assemblies for 66 teleost species. *Sci Data*. 4(1):160132.
Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6):764–770.
Martinez J, Fleury F, Varaldi J. 2012. Heritable variation in an extended phenotype: the case of a parasitoid manipulated by a virus. *J Evol Biol*. 25(1):54–65.
Martinez J, Fleury F, Varaldi J. 2015. Competitive outcome of multiple infections in a behavior-manipulating virus/wasp interaction. *Ecol Evol*. 5(24):5934–5945.
Martinson EO, Martinson VG, Edwards R, Mrinalini, Werren JH. 2016. Laterally transferred gene recruited as a venom in parasitoid wasps. *Mol Biol Evol*. 33(4):1042–1052.
Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun*. 7(1):9.
Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346(6210):763–767.
Morales J, Chiu H, Oo T, Plaza R, Hoskins S, Govind S. 2005. Biogenesis, structure, and immune-suppressive effects of virus-like particles of a *Drosophila* parasitoid, *Leptopilina victoriae*. *J Insect Physiol*. 51(2):181–195.
Moran N, Jarvik T. 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* 328(5978):624–627.
Mortimer NT. 2013. Parasitoid wasp virulence. *Fly* 7(4):242–248.
Mortimer NT, Goecks J, Kacsoh BZ, Mobley JA, Bowersock GJ, Taylor J, Schlenke TA. 2013. Parasitoid wasp venom SERCA regulates *Drosophila* calcium levels and inhibits cellular immunity. *Proc Natl Acad Sci U S A*. 110(23):9427–9432.
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.
Patot S, Allemand R, Fleury F, Varaldi J. 2012. An inherited virus influences the coexistence of parasitoid species through behaviour manipulation. *Ecol Lett*. 15(6):603–610.
Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420–1428.

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, et al. 2017. Evolutionary history of the Hymenoptera. *Curr Biol*. 27(7):1013–1018.

Pichon A, Bézier A, Urbach S, Aury JM, Jouan V, Ravallec M, Guy J, Cousserans F, Theze J, Gauthier J, et al. 2015. Recurrent DNA virus domestication leading to different parasite virulence strategies. *Sci Adv*. 1(10):e1501150.

Poirié M, Colinet D, Gatti J-L. 2014. Insights into function and evolution of parasitoid wasp venoms. *Curr Opin Insect Sci*. 6:52–60.

Reineke A, Asgari S, Schmidt O. 2006. Evolutionary origin of *Venturia canescens* virus-like particles. *Arch Insect Biochem Physiol*. 61(3):123–133.

Rizki RM, Rizki TM. 1990. Parasitoid virus-like particles destroy *Drosophila* cellular immunity. *Proc Natl Acad Sci U S A*. 87(21):8388–8392.

Saeedi R, Li M, Frohlich J. 2015. A review on lecithin:cholesterol acyltransferase deficiency. *Clin Biochem*. 48(7–8):472–475.

Salvetti A, Couté Y, Epstein A, Arata L, Kraut A, Navratil V, Bouvet P, Greco A. 2016. Nuclear functions of nucleolin through global proteomics and interactomic approaches. *J Proteome Res*. 15(5):1659–1669.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19): 3210–3212.

Sun H, Ding J, Piednoël M, Schneeberger K. 2018. findGSE: estimating genome size variation within human and *Arabidopsis* using k-mer frequencies. *Bioinformatics* 34(4):550–557.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 34(Web Server):W609–W612.

Theze J, Bézier A, Periquet G, Drezen J-M, Herniou EA. 2011. Paleozoic origin of insect large dsDNA viruses. *Proc Natl Acad Sci U S A*. 108(38):15931–15935.

van Oers MM, Vlak JM. 2007. Baculovirus genomics. *Curr Drug Targets*. 8(10):1051–1068.

Varaldi J, Fouillet P, Ravallec M, Lopez-Ferber M, Boulétreau M, Fleury F. 2003. Infectious behavior in a parasitoid. *Science* 302(5652):1930.

Varaldi J, Lepetit D. 2018. Deciphering the behaviour manipulation imposed by a virus on its parasitoid host: insights from a dual transcriptomic approach. *Parasitology* 145(14):1979–1989.

Varaldi J, Ravallec M, Labrosse C, Lopez-Ferber M, Boulétreau M, Fleury F. 2006. Artifical transfer and morphological description of virus particles associated with superparasitism behaviour in a parasitoid wasp. *J Insect Physiol*. 52(11–12):1202–1212.

Volkoff A-N, Jouan V, Urbach S, Samain S, Bergoin M, Wincker P, Demettre E, Cousserans F, Provost B, Coulibaly F, et al. 2010. Analysis of virion structural components reveals vestiges of the ancestral ichnovirus genome. *PLoS Pathog*. 6(5):e1000923.

Wybouw N, Pauchet Y, Heckel DG, Van Leeuwen T. 2016. Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome Biol Evol*. 8(6):1785–1801.

Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S. 2012. FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS One* 7(12):e52249.

Xue W, Li J-T, Zhu Y-P, Hou G-Y, Kong X-F, Kuang Y-Y, Sun X-W. 2013. L_RNA_scaffolder: scaffolding genomes with transcripts. *BMC Genomics* 14(1):604–601.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591.

Zhang G, Schmidt O, Asgari S. 2006. A calreticulin-like protein from endoparasitoid venom fluid is involved in host hemocyte inactivation. *Dev Comp Immunol*. 30(9):756–764.