



HAL
open science

Massive Somatic and Germline Chromosomal Integrations of Polydnviruses in Lepidopterans

Camille Heisserer, Héloïse Muller, Véronique Jouan, Karine Musset, Georges Periquet, Jean-Michel Drezen, Anne-Nathalie Volkoff, Clément Gilbert

► **To cite this version:**

Camille Heisserer, Héloïse Muller, Véronique Jouan, Karine Musset, Georges Periquet, et al.. Massive Somatic and Germline Chromosomal Integrations of Polydnviruses in Lepidopterans. *Molecular Biology and Evolution*, 2023, 40 (3), pp.msad050. 10.1093/molbev/msad050 . hal-04068886

HAL Id: hal-04068886

<https://hal.inrae.fr/hal-04068886>

Submitted on 14 Apr 2023

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Massive Somatic and Germline Chromosomal Integrations of Polydnviruses in Lepidopterans

Camille Heisserer,^{†,1,2} H lo se Muller,^{†,1} V ronique Jouan,³ Karine Musset,² Georges Periquet,² Jean-Michel Drezen ², Anne-Nathalie Volkoff ³, and Cl ment Gilbert ^{*,1}

¹Universit  Paris-Saclay, CNRS, IRD, UMR  volution, G nomes, Comportement et  cologie, Gif-sur-Yvette, France

²UMR 7261 CNRS, Institut de Recherche sur la Biologie de l'Insecte, Facult  des Sciences et Techniques, Universit  de Tours, Tours, France

³DGIMI, INRAE, University of Montpellier, Montpellier, France

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: clement.gilbert1@universite-paris-saclay.fr.

Associate editor: Mary O'Connell

Abstract

Increasing numbers of horizontal transfer (HT) of genes and transposable elements are reported in insects. Yet the mechanisms underlying these transfers remain unknown. Here we first quantify and characterize the patterns of chromosomal integration of the polydnvirus (PDV) encoded by the Campopleginae *Hyposoter didymator* parasitoid wasp (HdIV) in somatic cells of parasitized fall armyworm (*Spodoptera frugiperda*). PDVs are domesticated viruses injected by wasps together with their eggs into their hosts in order to facilitate the development of wasp larvae. We found that six HdIV DNA circles integrate into the genome of host somatic cells. Each host haploid genome suffers between 23 and 40 integration events (IEs) on average 72 h post-parasitism. Almost all IEs are mediated by DNA double-strand breaks occurring in the host integration motif (HIM) of HdIV circles. We show that despite their independent evolutionary origins, PDV from both Campopleginae and Braconidae wasps use remarkably similar mechanisms for chromosomal integration. Next, our similarity search performed on 775 genomes reveals that PDVs of both Campopleginae and Braconidae wasps have recurrently colonized the germline of dozens of lepidopteran species through the same mechanisms they use to integrate into somatic host chromosomes during parasitism. We found evidence of HIM-mediated HT of PDV DNA circles in no less than 124 species belonging to 15 lepidopteran families. Thus, this mechanism underlies a major route of HT of genetic material from wasps to lepidopterans with likely important consequences on lepidopterans.

Key words: polydnvirus, horizontal transfer, host–parasite relationships, insects, parasitoid wasps, evolutionary genomics.

Introduction

Parasitoid wasps are a paraphyletic group of Hymenoptera that are classified as ectoparasites or endoparasite insects, depending on whether they develop on or within an arthropod host, respectively (Beckage and Drezen 2012). To ensure the developmental success and survival of their eggs and larvae within hosts, many parasitoid wasps use viral particles that are akin to gene or protein-delivery agents (Herniou et al. 2013). These particles are injected together with wasp's eggs in parasitized hosts, of which they alter the physiology, thus enabling the development of wasp larvae within the host body (Beckage and Gelman 2004). The structural components of these viral particles are encoded by genes originating from molecular domestication of viral genomes that were integrated into the genome of wasp's ancestors (fig. 1).

Such domestication events occurred multiple times independently in different hymenopteran lineages, yielding

viral particles that differ in terms of structure, function, and content (Volkoff et al. 2010; B liveau et al. 2015; Pichon et al. 2015; Burke 2019; Di Giovanni et al. 2020; Burke et al. 2021; Mao et al. 2022). Given their viral origin, the particles packaging DNAs encoded by parasitoid wasps have been referred to as viruses and classified in their own viral family, the polydnviridae, or PDVs (Stoltz et al. 1984; Herniou et al. 2013). The two currently recognized genera of PDVs are the bracoviruses (BVs) and the ichnoviruses (IVs), respectively carried by wasps of the Braconidae and Ichneumonidae families that parasitize mainly lepidopteran hosts (B zier et al. 2009; B liveau et al. 2015; Burke et al. 2018; Volkoff and Cusson 2020). BVs result from the domestication of a nudivirus integrated in a braconid ancestor ~100 Ma (B zier et al. 2009), and are found today in all wasps belonging to the "Microgastroid complex" estimated to contain about 50,000 species (Murphy et al. 2008). IVs derive from an unknown viral

  The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

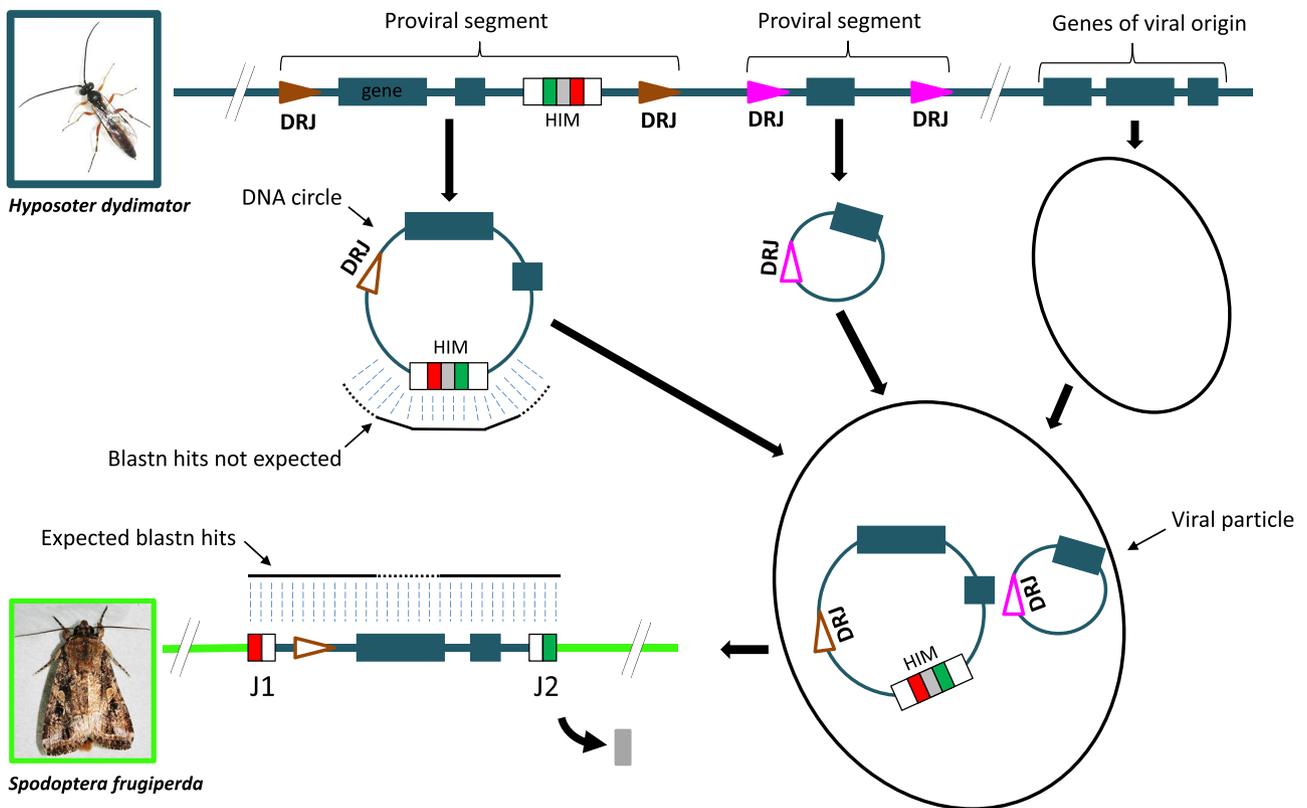


Fig. 1. Genome structure and chromosomal integration of PDVs. The genome and genes of the wasp (here *H. didymator*, Ichneumonidae) are shown in blue whereas the lepidopteran host (here *S. frugiperda*) is shown in green. The viral particle proteins of PDVs are encoded by genes of viral origin that have been coopted by parasitoid wasps. DNA circles packaged into viral particles originate from proviral segments, which are amplified and circularized in the calyx of female wasps, likely through recombination at the DRJ motifs (filled triangles). Circularization leads to the formation of a single recombined DRJ, or DRJ circle (Gauthier et al. 2021) made of a portion of the 5' DRJ and a portion of the 3' DRJ (empty triangles). Contrary to BVs in which DRJ motifs are similar between proviral segments, DRJ motifs differ between IVs segments, hence they are depicted with different colors (pink and brown). While ovipositing into larvae of their lepidopteran hosts, female wasps inject large amounts of PDV viral particles, which enter host somatic cells and deliver DNA circles to host cell nuclei. DNA circles possessing a HIM then undergo chromosomal integration catalyzed by wasp and host integrases (Wang et al. 2021). Integration involves a double-strand break within the HIM and linearized circles newly integrated into host DNA are bordered by junction 1 (J1 in red) and junction 2 (J2 in green) motifs. The short intervening region between J1 and J2 (in grey; about 50 bp in most circles) is lost during the process. We predicted that if HT of PDV sequences occur recurrently in lepidopteran hosts through HIM-mediated germline genome integration, few or no PDV sequence integrated in lepidopteran genomes should contain the intervening region. Thus, we should recover few or no blastn hit covering this region. By contrast, most or all blastn hits covering the HIM region should end at the J1 or J2 motif, which lie at the predicted junction with host DNA. The picture of *H. didymator* and *S. frugiperda* were taken by Marie Frayssinet, INRAE DGIMI lab.

family (Volkoff et al. 2010) and are present in two related Ichneumonidae subfamilies containing about 4,000 species. It was questioned whether their distribution in Ichneumonidae is due to one or two independent integration events (Béliveau et al. 2015) but recent phylogenetic analyses favor the second hypothesis (Santos et al. 2022).

A remarkable feature of BVs and IVs is that they both package circular DNA molecules (Volkoff and Huguet 2021; fig. 1). These DNA circles arise from amplification and excision of so-called “proviral segments” located in the wasp genome. The segments contain genes of different origins that encode virulence factors and are expressed in the parasitized host throughout the development of the wasp larvae. In braconid wasp genomes, most BV segments are arranged in tandem in large loci. For example, *Cotesia congregata* harbors 35 proviral segments, 17 of which are clustered in an ~2-Mb macrolocus located on chromosome

5 (Gauthier et al. 2021). Each of the segments is flanked by direct repeat junctions (DRJs) (filled triangles in fig. 1), sequence motifs that all contain an AGCT tetramer motif at which circularization through site-specific recombination occurs (Drezen et al. 1997; Desjardins et al. 2008; Burke et al. 2015). Ichneumonid IV segments are generally more numerous than BV segments and are highly dispersed in wasp genomes. For example, *Hyposoter didymator* harbors 57 proviral segments that are all separated by megabase-long portions of wasp genome (Legeai et al. 2020). Like BV segments, IV segments are flanked by DRJs but contrary to BVs, which segments share similar extremities, the sequence of IV DRJs are specific of each segment. Circularization is thought to occur through homologous recombination between left and right DRJs, with breakpoints located at varying positions along the DRJs (Legeai et al. 2020). For both BVs and IVs, circularization leads to the

formation of a single DRJ composed of a portion of the 5' DRJ and a portion of the 3' DRJ of proviral segments (empty triangles in [fig. 1](#)).

Although the different genomic organization of BV and IV proviral segments imply differences in the mechanisms regulating their amplification and circularization, DNA circles are produced in cells of the calyx, a special region of the female gonad, for both PDV types. They are then packaged into viral particles, released in the oviduct lumen and injected into the host during oviposition. Early cell line-based studies suggested that once in host cells, some DNA circles could persist as chromosomally integrated forms ([fig. 1](#)) ([McKelvey et al. 1996](#); [Volkoff et al. 2001](#); [Gundersen-Rindal and Lynn 2003](#)). This was confirmed using targeted Sanger and Illumina sequencing by showing that at least a subset of BV circles integrates into host hemocytes, and that integration likely involves an enzymatically regulated process through recognition of a specific motif of the DNA circles called host integration motif (HIM) ([Beck et al. 2011](#); [Chevignon et al. 2018](#)). In agreement with this, a recent study found by functional analysis that in some lepidopteran hosts, a fraction of DNA circle integration events are catalyzed by a series of viral integrases, whereas host integrases are also involved in the integration of some circles ([Wang et al. 2021a](#)). Using bulk Illumina sequencing of parasitized Mediterranean corn borer larvae (*Sesamia nonagrioides*, Lepidoptera, Noctuidae) we showed that chromosomal integration of BV circles from *Cotesia typhae* occurs systematically in multiple host tissues and that only BV DNA circles containing HIM motifs integrate at measurable levels. Our quantitative approach also allowed us to estimate that each host haploid genome suffers between 12 and 85 BV DNA circle integration events depending on the tissue. Despite this high level of integration activity, we found that non-integrated circles also persist as much as integrated circles in parasitized hosts during at least half the development of wasp larvae ([Muller et al. 2021](#)). Recent bulk Illumina sequencing of parasitized diamondback moths (*Plutella xylostella*, Lepidoptera, Plutellidae) larvae also revealed that IV DNA circles from *Diadegma semiclausum* (DsIV) undergo massive integration into host hemocyte genomes during parasitism ([Wang et al. 2021b](#)). Four DNA circles were found to integrate via a HIM motif, as observed for BV, but contrary to BV DNA circles, IV circles devoid of HIM were also found to integrate at substantial levels through DNA double-strand breaks occurring at varying locations along the circles ([Wang et al. 2021b](#)). The finding that both IV and BV DNA circles integrate via a shared HIM-mediated mechanism despite their independent origin is remarkable and several scenarios have been proposed to explain the possible evolutionary ties between these two PDV lineages ([Muller et al. 2021](#); [Wang et al. 2021b](#)).

In the present study, we first apply the approach developed in [Muller et al. \(2021\)](#) on the *C. typhae*/*S. nonagrioides* system to comprehensively characterize and quantify DNA circle integration of yet another PDV, the

Hyposoter didymator IV (HdIV), in two tissues of parasitized fall armyworms (*Spodoptera frugiperda*, Lepidoptera, Noctuidae). *Hyposoter didymator* is a Campopleginae wasp (Ichneumonidae) showing a marked host preference for *Helicoverpa armigera* in the wild, but it is able to parasitize a large variety of other noctuid moths, including *S. frugiperda* ([Frayssinet et al. 2019](#)). Our results further underline the striking similarities in integration patterns between IVs and BVs, despite that these domesticated endogenous viruses are derived from viruses belonging to different families. We then assess whether HIM-mediated IV and BV chromosomal integration occurred in the germline of their hosts in the past by screening the genome of 775 lepidopteran species. We found sequences from IV and BV DNA circles in no less than 124 species belonging to 15 different lepidopteran families, suggesting that HIM-mediated wasp-to-host horizontal transfer (HT) of PDV DNA circles occurred recurrently during the evolutionary history of lepidopterans.

Results

Quantifying *S. frugiperda* and *H. didymator* Genomic Material

In order to characterize genome-wide patterns of HdIV integration during parasitism of the fall armyworm (*S. frugiperda*), we Illumina-sequenced whole DNA extracted from fat bodies and hemolymph of parasitized larvae. We obtained from 321 to 403 million trimmed reads depending on the tissue and time post-parasitism (p.p.) (hemolymph 24 and 72 h [H24 and H72], fat body 24 and 72 h [FB24h and FB72h]) ([table 1](#)).

The majority of these reads (from 77% to 81%) aligned to the *S. frugiperda* genome (384 Mb), which was covered almost entirely (98%) in all samples. The average sequencing depth on the *S. frugiperda* genome varied from 101× for the H72h sample to 132× for the FB24h sample ([fig. 2](#)). By contrast, only 0.3–15% of the 226-Mb *H. didymator* genome was covered depending on the sample, at very low average sequencing depths (1–2×), most likely corresponding to the DNA of *H. didymator* eggs. By contrast, the average sequencing depth of the *H. didymator* wasp regions annotated as ichnovirus proviral segments by [Legai et al. \(2020\)](#) was much higher than the rest of the wasp's genome: 914× and 631× in the hemolymph (24 and 72 h p.p.) and 1,220× and 386× for the fat body (24 and 72 h p.p.) ([fig. 2](#)). This higher depth is consistent with the presence of many integrated and/or non-integrated forms of HdIV circles in the caterpillar after parasitization, since large amounts of virus particles are introduced with the parasite egg.

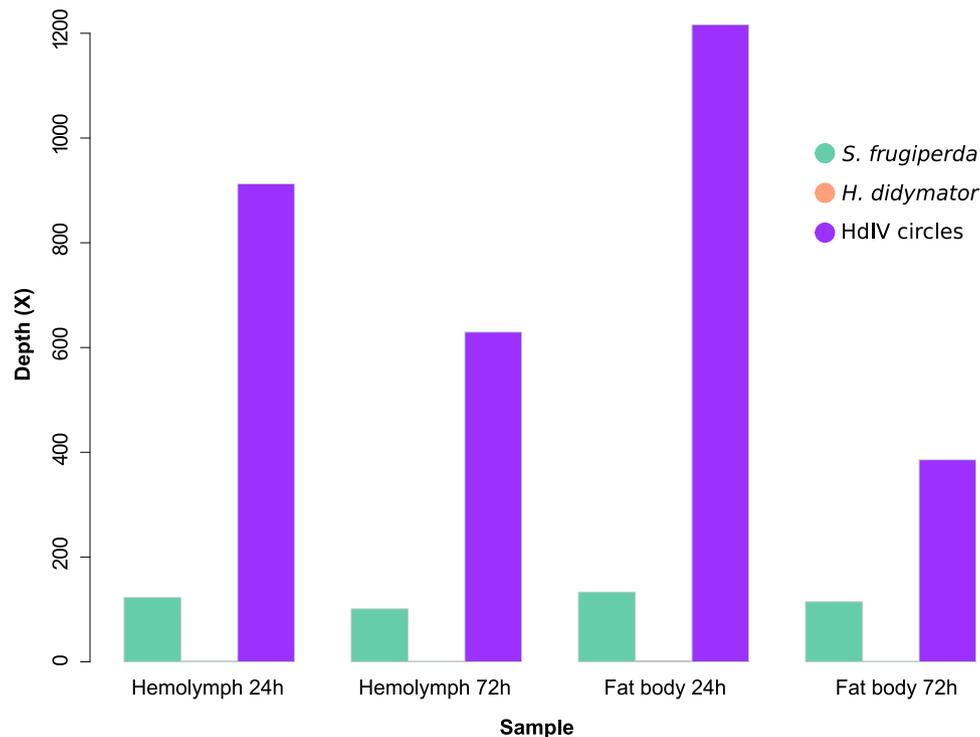
Quantification and Characterization of Integrated HdIV into *S. frugiperda* Somatic Genomes

To quantify chromosomal integration of HdIV circles into parasitized *S. frugiperda* cells, we searched for chimeric reads containing HdIV–*S. frugiperda* junctions (see

Table 1. Summary of the Number of Reads and Chimeric Reads Obtained from *S. frugiperda* Larvae Parasitized by *H. didymator*.

| | Accession Number (SRA) | Total Number of Trimmed Reads | Reads Mapped on Caterpillar Genome | Number of Chimeric Reads | RPM on Caterpillar Genome | Number of IEs |
|-------|-----------------------------|-------------------------------|------------------------------------|--------------------------|---------------------------|---------------|
| H24h | SRX15279414 | 371 695 498 | 320 038 242 (86%) | 2,861 | 9 | 2,449 |
| H72h | SRX15279415 | 321 231 671 | 263 127 889 (82%) | 4,903 | 19 | 4,398 |
| FB24h | SRX15279416 | 403 147 682 | 346 759 613 (86%) | 2,516 | 7 | 2,065 |
| FB72h | SRX15279417 | 347 218 922 | 299 552 940 (86%) | 3,325 | 11 | 2,908 |

NOTE. H, hemolymph; FB, fat body. Numbers between brackets are percentages of total sequencing reads mapped on the two reference genomes (that of *S. frugiperda* and that of *H. didymator*). IEs: Integrations events (host–wasp junctions with the same coordinate in the host genome but covered by more than one chimeric read are counted as one integration event). RPM: “number of chimeric reads per million reads mapping” on the caterpillar genome.

**Fig. 2.** Mean sequencing depth over parasitized host (*S. frugiperda*) and parasitoid wasp (*H. didymator*) genomes.

Materials and Methods). Depending on the sample, we detected from 2,516 to 4,903 chimeric reads (table 1). Even after discarding PCR duplicates, a same integration event (IE), as defined by a given set of coordinates in the HdIV and *S. frugiperda* genome, might be covered by more than one chimeric read. This may occur when a given IE is duplicated through cell division. We thus estimated the number of independent IEs by counting as one event all reads with identical or nearly identical coordinates for both genomes. We found that the vast majority of IEs were covered by only one read and in total, we counted between 2,065 and 4,398 IEs (table 1; fig. 3). The number of HdIV junctions covered by more than one reads went from 1 to 121 depending on the tissue (highest number of reads covering a junction = 3).

We found that seven out of the 57 HdIV circles had more than ten chimeric reads mapping to them in all four DNA samples, strongly suggesting these seven segments underwent chromosomal integration. The seven

segments form two groups depending on whether they have generated more or less than one IE per million reads mapped on the host (*S. frugiperda*) genome (respectively, group 1: Hd12, Hd16 and Hd27 and group 2: Hd13, Hd44.2, Hd47 and Hd53) (fig. 3). Comparing normalized IEs per segment revealed a higher number of IEs in the hemolymph than in the fat body and a higher number of IEs at 72 h than at 24 h p.p. These trends hold for all seven integrated segments and are in agreement with earlier findings on BVs (Beck et al. 2007; Muller et al. 2021).

Next, we followed the rationale exposed in Muller et al. (2021) and assessed how many IEs, on average, occurred per sequenced haploid genome. To do so, we divided the relative number of IEs (IEs per million reads mapped to the host genome) for each integrated circle by the read length and we multiplied it by the *S. frugiperda* genome size. We counted a total of 18, 40, 13, and 23 IEs per cell, on average, in the Hemocyte 24 and 72 h, and the Fat body 24 and 72 h samples, respectively.

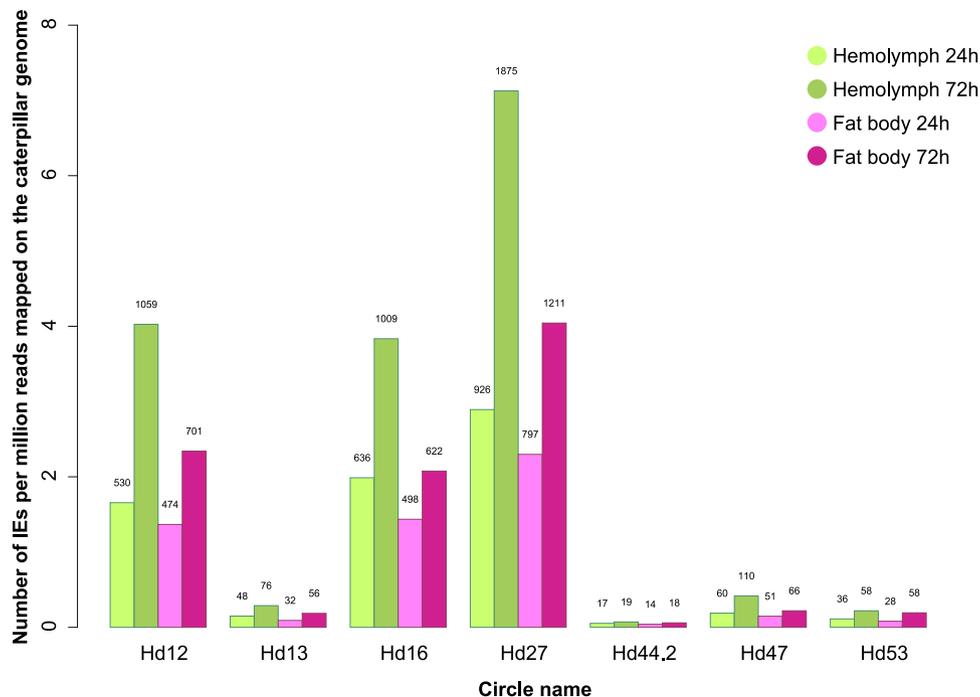


FIG. 3. Number of HdIV chromosomal IEs per HdIV circle in parasitized *S. frugiperda* larvae. Integration events (host–wasp junctions with the same coordinate in the host genome but covered by more than one chimeric read are counted as one IE).

Mechanisms Underlying Integration of HdIV Circles

To assess whether HdIV circle integration involves recombination within HIM as observed for BVs, and/or other mechanisms, we generated plots of chimeric read depths along all seven circles found integrated into *S. frugiperda* genomes (fig. 4; supplementary fig. S1, Supplementary Material online). The vast majority of chimeric reads (96.3–100% depending on the circle) map to a narrow region (in white in fig. 4) and most of the positions along each HdIV circle are not mapped by any chimeric reads. Depending on the circle, the length of the mapped region is between 92 and 120 bp. We arbitrarily delineated these specific regions in the HdIV circles following Muller et al. (2021), by identifying the two positions along the circles with the highest number of chimeric reads and by selecting 30 bp upstream of the left-most one and 30 bp downstream of the right-most one. We then verified that these regions align to the recently described HIM motifs in various IVs (Wang et al. 2021b) (fig. 4). Zooming on the HIM also confirmed the presence of two peaks of chimeric reads, as observed in BVs (Muller et al. 2021) and corresponding to the J1 and J2 motifs, at which double strand breaks most often occur during linearization and integration of the circles (Chevignon et al. 2018) (fig. 4; supplementary fig. S1, Supplementary Material online). Thus, our results show that as for BVs, most if not all chromosomal integrations of HdIV circles are mediated by ichnovirus HIMs.

To further characterize the mechanism involved in the integration within the HIM regions and in particular whether homology between viral and host sequences could be involved in the integration process, we studied

the distribution of microhomology lengths detected at the junction between the HdIV and the *S. frugiperda* genome. We only investigated chimeric reads mapping to the HIM because the remaining chimeric reads represent only a very small fraction of all chimeric reads (0.8%). As previously done in our study on *C. typhae* BV (Muller et al. 2021), we studied separately the major part of chimeric reads which map to the J1 and J2 motifs from those mapping outside of these motifs (but still within the HIM). We artificially generated chimeric reads following the method described in Peccoud et al. (2018) to compare the number of observed microhomology lengths with the number of expected microhomology lengths if junctions occurred randomly between HdIV HIMs and the *S. frugiperda* genome. For chimeric reads mapping to the J1 or J2 motifs, the analysis revealed a strong excess of 2- and 3-bp microhomologies and a slight excess of 4-bp microhomologies (fig. 5). The pattern was similar for chimeric reads mapping outside J1 and J2, though the excess of 3-bp microhomologies is less marked and that of 4-bp microhomologies no longer present. Interestingly, we observe a strong depletion of 0- and 1-bp microhomologies compared to the pattern expected by chance. This indicates that junction involving blunt-ended host–wasp sequences or 1-bp microhomology between them almost never occur.

Persistence of Integrated versus Non-integrated Circles

We evaluated the quantity of integrated versus non-integrated HdIV circles during parasitism, as approximated by sequencing depth. We found that except for circle Hd27

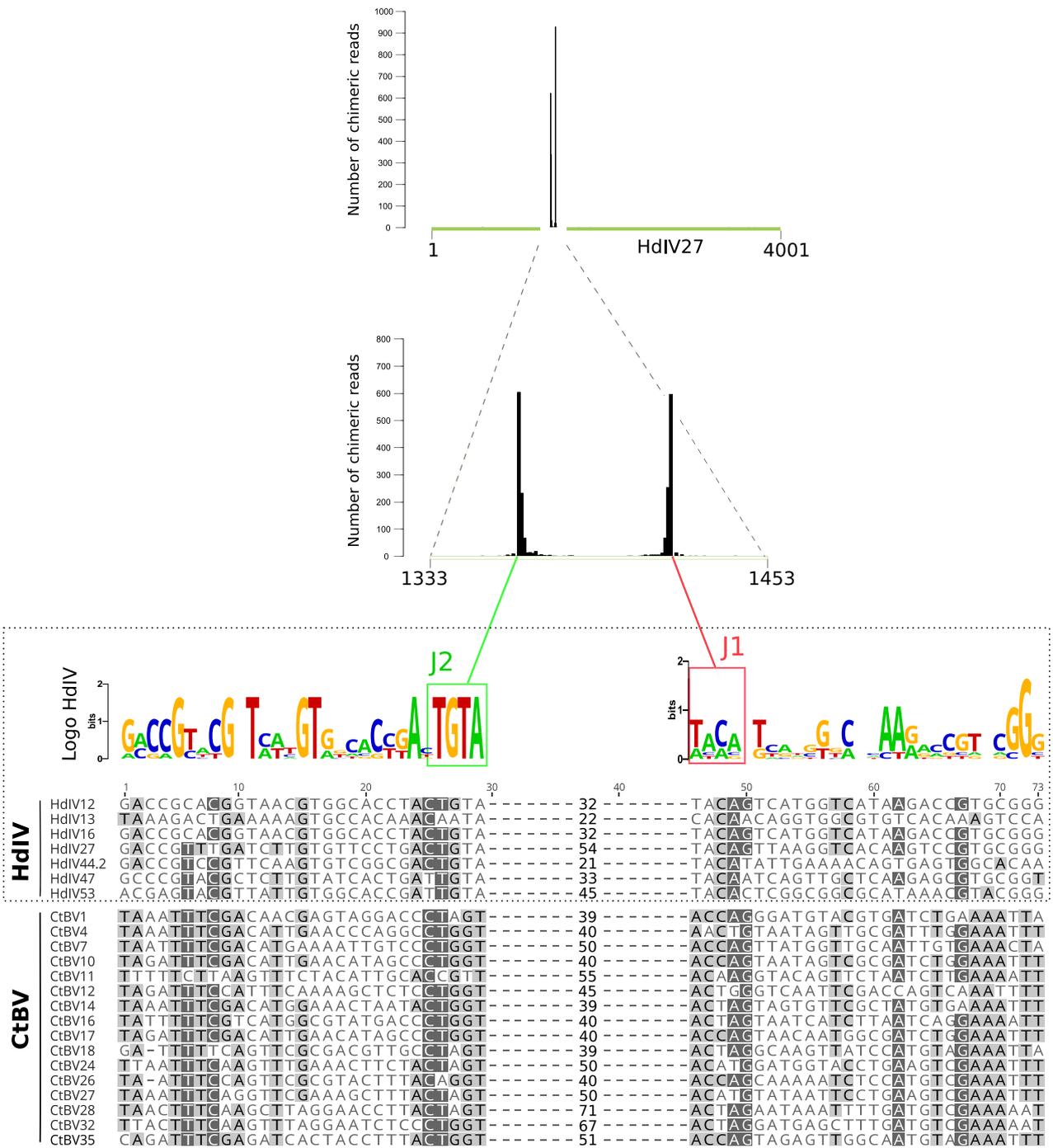


Fig. 4. Polynavirus DNA circles undergo HIM-mediated chromosomal integration. Top graph: map of chimeric reads along *Hyposoter didymator* ichnovirus circle 27 (Hd27) (green) in hemocytes showing that almost all chimeric reads map to the HIM (white). The circle is oriented from the 5' DRJ to the 3' DRJ. Below is a magnified view containing the 120-bp HIM, showing the two regions with many chimeric reads, called J2 (left) and J1 (right), corresponding both to the borders of Hd27 sequences integrated in parasitized host DNA. Below is shown the sequence logo of J2 and J1 from HdIV circles only (not including CtBV circles, as indicated by the dashed rectangle) generated with weblogo.berkeley.edu, and the alignment of the HIMs of the seven HdIV circles that integrated into the *S. frugiperda* genome. An alignment of the CtBV circles described in Muller et al. (2021) is also shown below to allow comparison between BV and IV HIM. Grey shading indicates the level of sequence conservation. Numbers in the middle of the alignment indicate the number of nucleotides that are present between the J1 and J2 motifs and that are lost upon integration of PDV circles in host genomes. This region is not conserved between circles and was thus removed to facilitate the reading of the figure.

which showed a very high sequencing depth (normalized depth > 15X in three samples), the ranges of sequencing depths of integrated HdIV circles were similar to that of

non-integrated HdIV circles (fig. 6). For example, in hemolymph 24 h p.p., integrated circles (in blue in fig. 6) were sequenced at depths varying between 243X and 3,400X

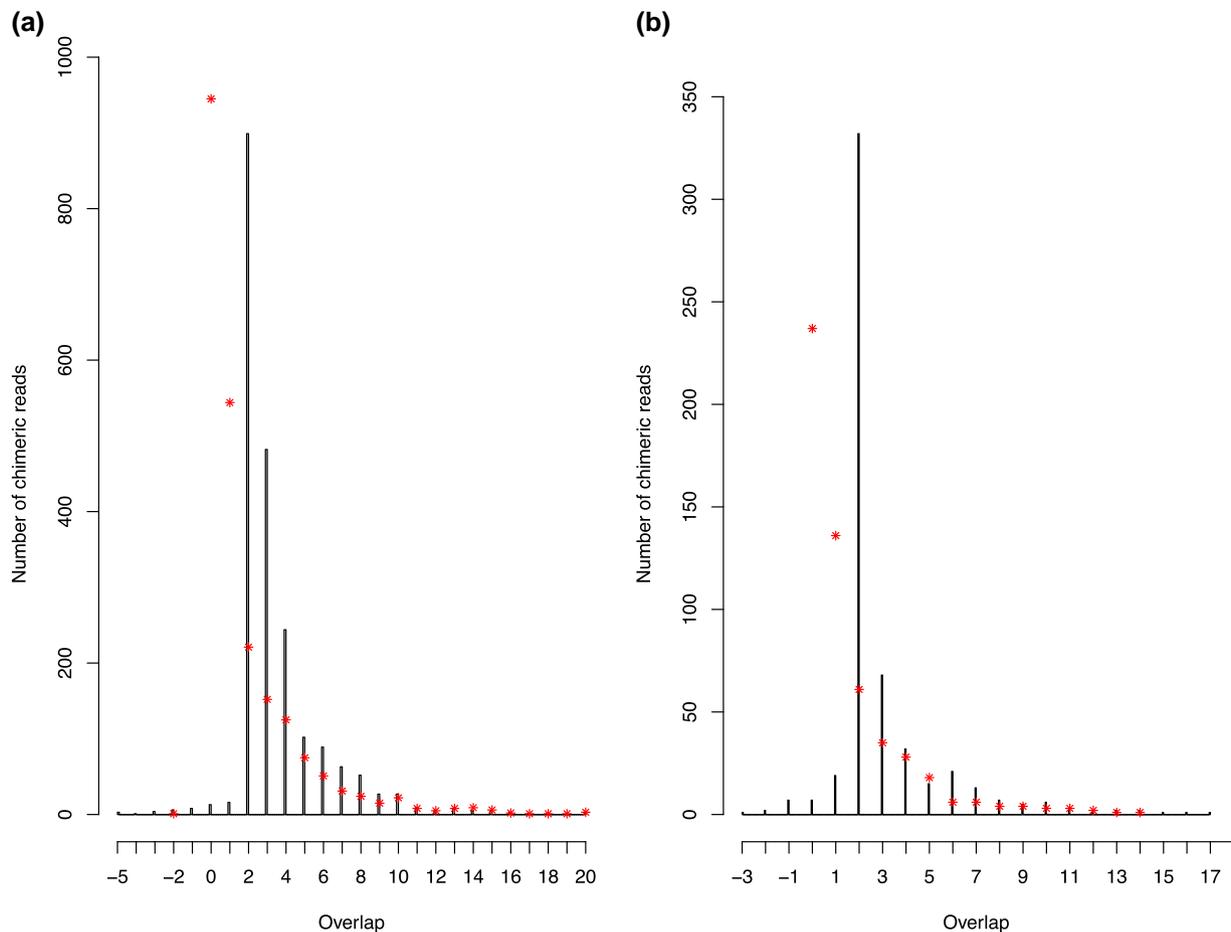


FIG. 5. Distribution of microhomology lengths at wasp–host junctions in chimeric reads. Black bars correspond to the numbers of observed chimeric reads for each microhomology length. Red asterisks correspond to the expected numbers of chimeric reads for each microhomology length. (a) Distribution of microhomology lengths for HdIV–host junctions mapped in J1 or J2. (b) Distribution of microhomology lengths for HdIV–host junctions mapped within HIM but outside J1 or J2.

whereas non-integrated circles (in red) were sequenced at depths varying between 93 \times and 4,164 \times . At the later time-point during parasitism (72 h p.p.), the overall quantity of circles (integrated and non-integrated) decreases, with, for example, about three times less circles in the fat body at 72 h p.p. than at 24 h p.p. (fig. 2). This decrease holds true even accounting for the fact that overall read count was lower at 72 h p.p. than at 24 h p.p. (table 1). Interestingly, this decrease is overall stronger for non-integrated circles than for integrated ones.

This is well-illustrated by the fact that sequencing depth over Hd16 is lower than that of some non-integrated circles at 24 h but it becomes higher than non-integrated circles at 72 h p.p., a trend holding in both hemolymph and fat body (fig. 6). The capacity to integrate for a circle may thus increase its ability to persist in large amount during parasitism.

Distribution of Wasp Circles Throughout the Genome of *S. frugiperda*

We investigated whether the integrations of *H. didymator* viral circles occur randomly along the caterpillar genome

or whether there is an integration bias. To achieve this, we have split the *S. frugiperda* genome into 100,000-bp windows and assessed whether some windows were subject to more integrations than expected by chance. We found that the numbers of HdIV integrations per window did not follow a Poisson distribution with P -values <0.001 for all samples indicating that these integrations are not randomly distributed along the *S. frugiperda* genome. Under the Poisson distribution, we do not expect any windows with more than six or seven IEs depending on the samples. However, for example, we detected in the hemolymph samples a window with 14 IEs (H24h) and another with 18 (H72h) IEs. These results suggest that segments do not integrate completely randomly in the caterpillar genome, where IEs seem to be slightly concentrated in specific genomic regions.

HIM-mediated Integrations of Polydnvirus DNA Circles in Multiple Lepidopteran Genomes

This analysis together with that of Wang et al. (2021b) on DsIV, that of Wang et al. (2021a) on CvBV, and our earlier study on CtBV (Muller et al. 2021) shows that PDV circles

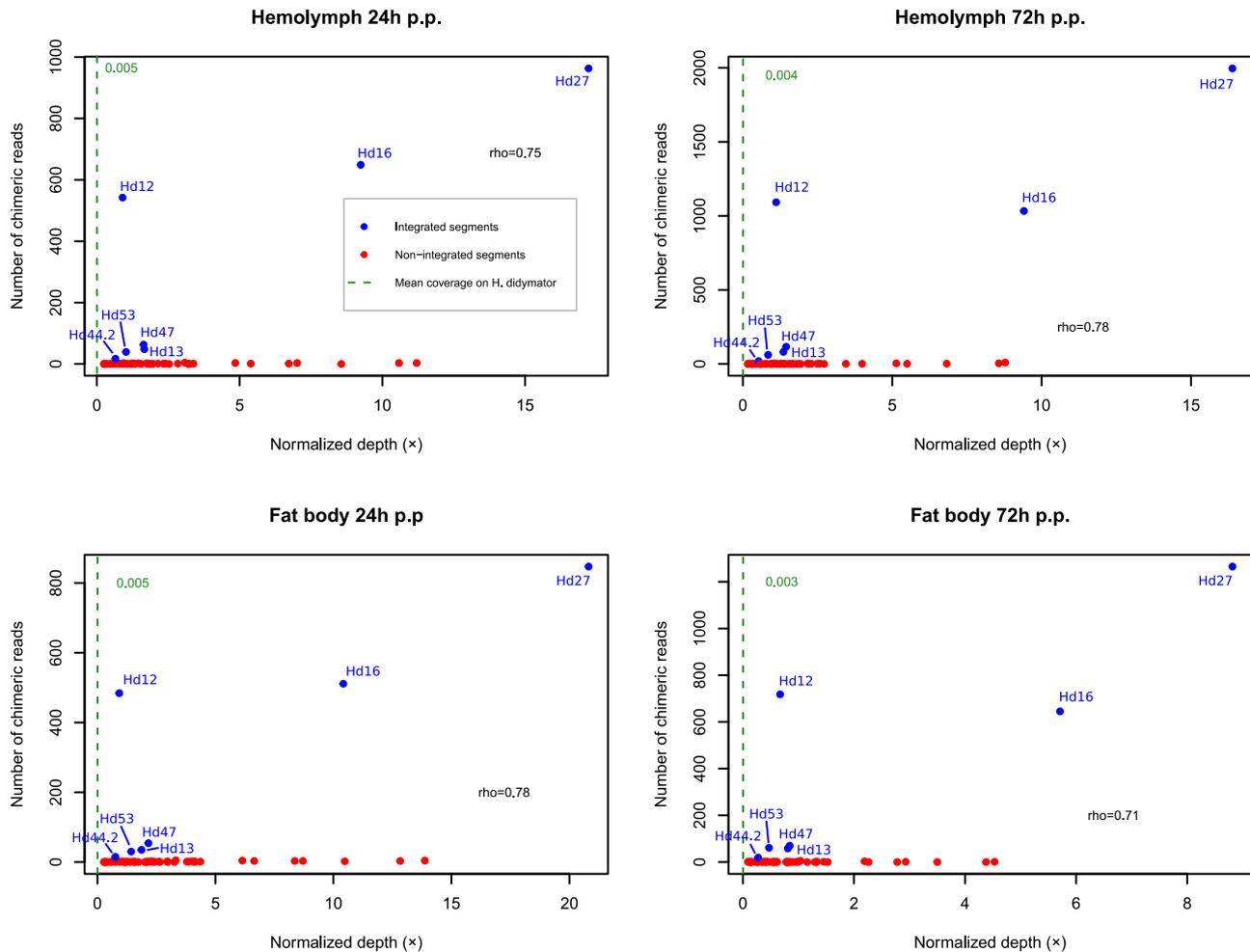


Fig. 6. Plot of the sequencing depth versus the number of chimeric reads for each of the HdIV circle in hemolymph 24 and 72 h post parasitism (p.p.), and in fat body 24 and 72 h p.p. Blue dots represent circles that do integrate into the *S. frugiperda* genome, and red dots represent circles that do not integrate. The identification numbers of the segments are shown near each blue dot. The vertical green dashed line shows the average sequencing depth over the *H. didymator* genome (the precise value is indicated at the top of the line). The Spearman rho value indicates the correlation between sequencing depth and the number of IEs for circles that do integrate into the *S. frugiperda* genome. The x-axis shows normalized depth, that is sequencing depth per million reads.

undergo massive HIM-mediated chromosomal integration in host tissues during parasitism. We thus sought to assess whether HT of PDV circles occurred from wasp to lepidopteran through HIM-mediated chromosomal integration in the germline of lepidopteran hosts, followed by vertical transmission in host populations. For this, we used HIM-containing HdIV and CtBV circles to perform blastn similarity searches on 775 lepidopteran genomes available in Genbank as of December 2021. Our search yielded 4,648 and 49,079 lepidopteran sequences longer than 300 bp that showed similarity to HIM-containing HdIV and CtBV circles, respectively, with an e-value lower than 0.0001. We reasoned that if some of these sequences were integrated through double-strand breaks within HIM, then they were bordered by the J1 and J2 motifs (fig. 1) and may have since conserved these extremities, or at least one of them (J1 or J2). Moreover, the region found between J1 and J2 in the circular PDV molecule before integration (in grey in fig. 1), which is lost during

integration, should not be retrieved during the analyses (see above and Cheignon et al. (2018) and Muller et al. (2021)). In agreement with these expectations, no less than 2,213 of the 4,648 IV sequences were found in lepidopteran genomes starting or ending within the HIM, and none of them contained the entire region lying between the J1 and J2 motifs. Regarding the 49,079 blastn hits involving CtBV, we found 174 sequences starting or ending within the HIM and only three containing the entire region lying between the J1 and J2 motifs. Our blastn searches did not allow us to directly retrieve full length PDV circles bordered by J1 on one side and J2 on the other side. Importantly however, we manually curated several lepidopteran PDV sequences and were able to reconstruct several examples of full-length integrated circles bordered by J1 and J2 and containing a single recombined DRJ sequence (supplementary Dataset 1, Supplementary Material online). We found that all lepidopteran PDV sequences were highly rearranged compared with our query

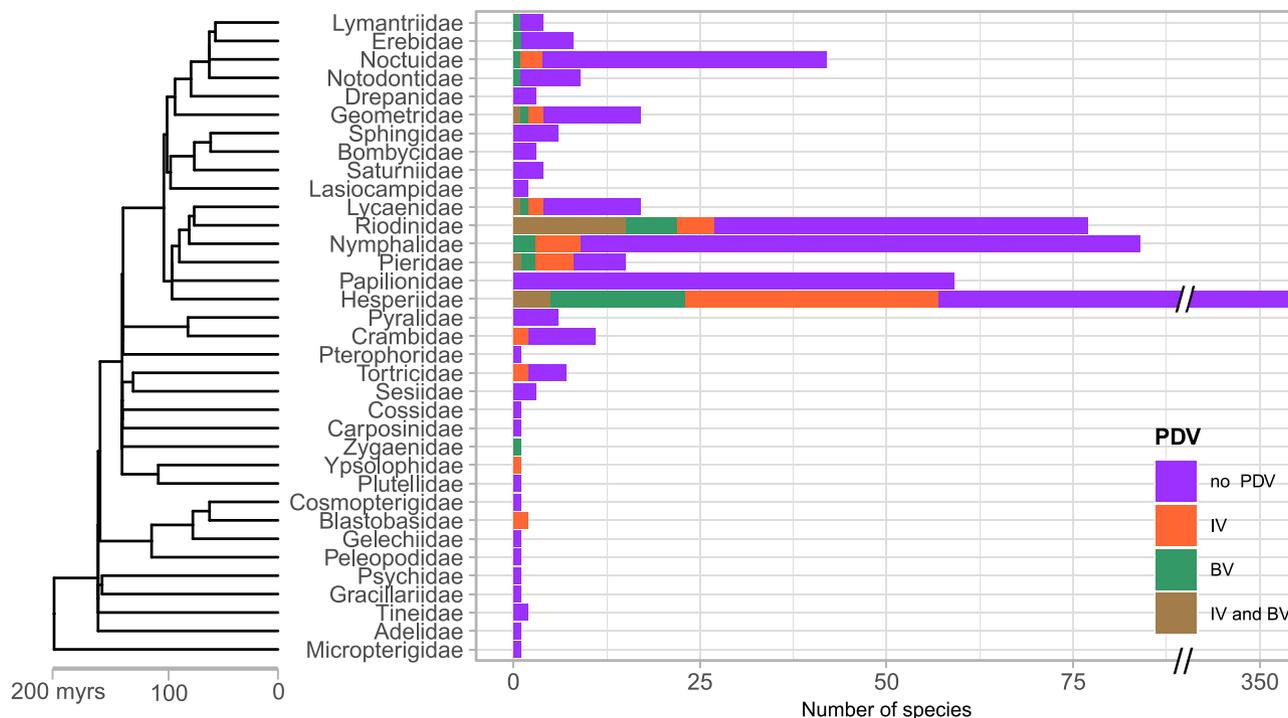


Fig. 7. Numbers of genomes per lepidopteran families in which J1- or J2-bordered BV and IV fragments were found. Details on the numbers and type of PDV circles in each lepidopteran species can be found in [figure 8](#) and [supplementary tables S3 and S4, Supplementary Material](#) online. The tree and divergence times were recovered from [timetree.org](#) (Kumar et al. 2017).

wasp PDV sequences. This may be because integrations are ancient and were followed by rearrangements and/or degradation of the circles. Another reason behind the fragmentary nature of the blastn hits may be that the donor wasp species was distantly related to *H. didymator* or *C. typhae*. Although complete, circles from such distant species may be too divergent to align over their entire length with no interruption to those of *H. didymator* or *C. typhae*. In the following, we refer to fragments of IV or BV sequences as retrieved by our blastn search, that is fragments longer than 300 bp and bordered either by the J1 or J2 motif, to describe patterns of HIM-mediated integration of PDV circles into lepidopteran genomes.

Numbers and Distribution of Polydnvirus DNA Circles in the Lepidopteran Tree

IV fragments bordered by the J1 or J2 motif were found for 4 out of the 7 HIM-containing HdIV circles in a total of 87 lepidopteran species (out of 775) that belong to 11 out of the 35 lepidopteran families included in our search ([supplementary table S1, Supplementary Material](#) online; [fig. 7](#)). BV fragments bordered by the J1 or J2 motif were found for 14 out of 16 HIM-containing CtBV circles in a total of 60 lepidopteran species belonging to 11 families, including 7 (Geometridae, Hesperidae, Lycaenidae, Noctuidae, Nymphalidae, Pieridae, and Riodinidae), which also had IV fragments integrated in their genome ([supplementary table S2, Supplementary Material](#) online; [fig. 7](#)). Altogether, PDV fragments (IV + BV) were retrieved

in a total of 124 species with 23 having both IV and BV in their genome. These 124 species are from 15 lepidopteran families, which belong to eight superfamilies (Noctuoidea, Geometroidea, Papilionoidea, Pyraloidea, Tortricoidea, Zygaenoidea, Yponomeutoidea, Gelechioidea) that diverged more than 100 Ma, and cover a large diversity of lepidopterans ([fig. 7](#)) (Kawahara et al. 2019).

Lepidopteran IV fragments retained for this study are between 300 and 4,110 bp (median = 621 bp) and show between 65% and 96% nucleotide identity to HdIV circles (median = 78.5%). BV fragments are between 304 and 2860 bp (median = 515 bp) and show between 65% and 88% nucleotide identity to CtBV circles (median = 77%). All PDV fragments bordered by the J1 or J2 motifs as well as their alignment coordinates and percent identity to CtBV or HdIV are provided in [supplementary tables S1 and S2, Supplementary Material](#) online.

Numbers of PDV fragments are heterogenous both in terms of circles and lepidopteran species. In terms of lepidopteran species, numbers of IV fragments are generally relatively low, with 52 species having only one or two fragments and 76 out of the 87 species having less than 10 ([supplementary table S3, Supplementary Material](#) online). However, three species have high (81 in *Blastobasis adustella* and 80 in *B. laticollela*, Blastobasidae) to very high (1,756 in the wainscot hooktip *Ypsolopha scabrella*, Ypsolophidae) numbers of IV fragments ([supplementary table S3, Supplementary Material](#) online, [fig. 8](#)). Numbers of BV fragments are also generally relatively low, with 48 out of the 60 species having only one or two J1- or

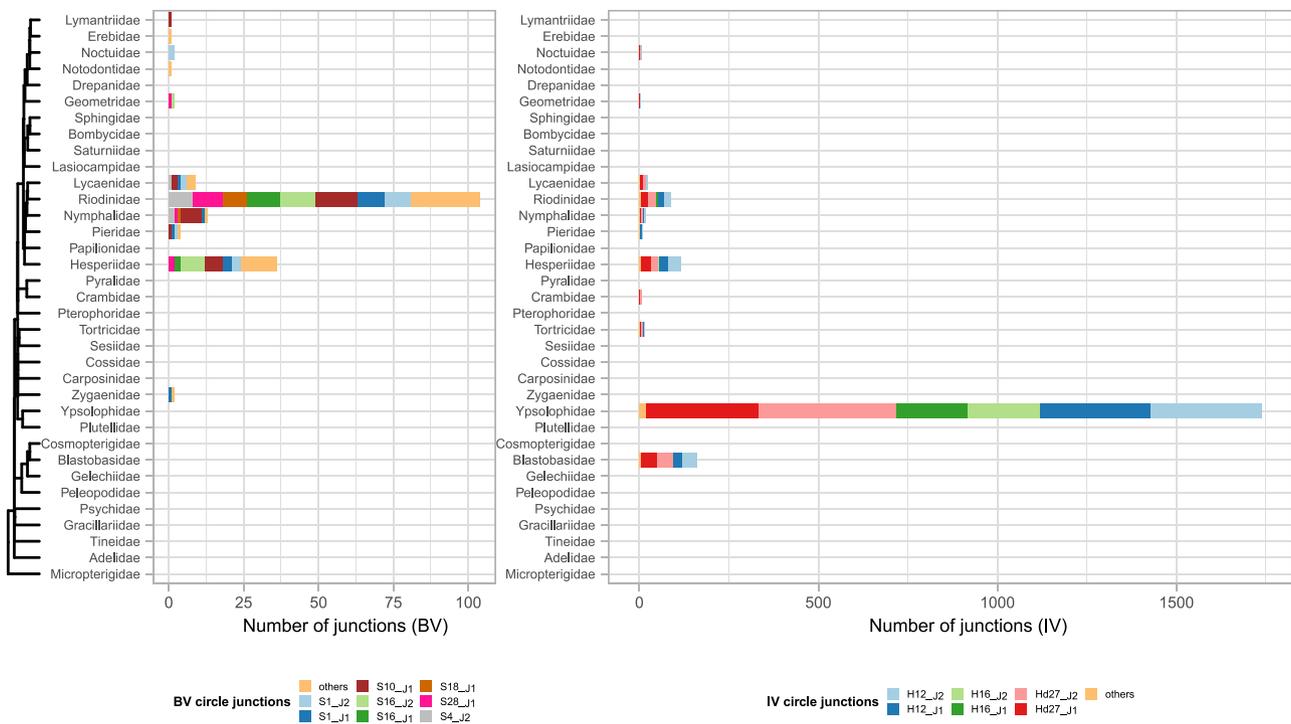


Fig. 8. Numbers of J1- or J2-bordered BV and IV fragments per lepidopteran families. Details on the numbers and type of PDV circles in each lepidopteran species can be found in [supplementary tables S3 and S4, Supplementary Material](#) online. Junctions' types for which less than 8 (BV) or 10 (IV) fragments were found in all families were grouped together in the "others" category.

J2-bordered BV fragments ([supplementary table S4, Supplementary Material](#) online). Four species have ten or more (up to 17) BV fragments (*Calephelis nemesis*, *C. perditalis*, *Apodemia duryi*, and *Boloria selene*) ([supplementary table S4, Supplementary Material](#) online, [fig. 8](#)). The number of both IV and BV fragments per lepidopteran family was positively correlated with the number of genomes surveyed per family (both Pearson's $r = 0.98$; P -values < 0.000001).

In terms of circles, we found a much higher number of IV fragments corresponding to Hd12/16 (634 J1 and 595 J2) and Hd27 (498 J1 and 422 J2) than to Hd47 (28 J1 and 35 J2) and Hd44.2 (1 J2) ([fig. 8; supplementary table S3, Supplementary Material](#) online). This mirrors the pattern observed for somatic integrations, with Hd12, Hd16, and Hd27 being highly integrated and Hd47 being the mostly highly integrated among lowly integrated circles ([fig. 3](#)). Numbers of IV J1 and J2 junctions are similar, which is consistent with the fact that during HIM-mediated integration, each integrated IV circle is linearized through double-strand break within the HIM and thus ends flanked by the J1 motif on one side and J2 motif on the other side ([fig. 1](#)). In terms of BV circles, the number of fragments found in lepidopteran genomes did not match with the relative abundance of integrated circles in parasitized hosts. For example, although we found the highest number of BV fragments for CtBV circles 16 (21 J2 and 13 J1) and 10 (31 J2) ([supplementary table S4, Supplementary Material](#) online, [fig. 8](#)), these two circles are among the least integrated circles in somatic

genomes of parasitized *S. nonagrioides* individuals ([fig. 6](#) in [Muller et al. 2021](#)). CtBV circle 1, which is by far the most highly integrated circle in parasitized host tissues is among the most integrated circles in lepidopteran genomes but the number of C1 fragments (16 J1 and 17 J2) are close to those of several other circles ([supplementary table S4, Supplementary Material](#) online, [fig. 8](#)). Contrasting with IV, for which similar amounts of J1-bordered and J2-bordered fragments were recovered for each circle in lepidopteran genomes, we often found different numbers of J1 and J2 BV fragments. In fact, we found only J1-bordered fragments (no J2-bordered fragments) for nine BV circles ([supplementary table S4, Supplementary Material](#) online, [fig. 8](#)). We believe that this disequilibrium is unlikely to have biological underpinnings because when we lowered the length threshold used to filter blastn hits to 200 bp, that is when we retain all BV-like sequences longer than 200 bp instead of 300 bp, the number of J1- and J2-bordered BV fragments are more similar for some circles (not shown). Yet we chose to keep the 300 bp threshold to retain blastn hits in order to maximize specificity and to ensure recovering large enough PDV fragments to conduct phylogenetic analyses.

Sequencing Depth Supports Germline Integration of Polydnavirus DNA Circles

Several features of the lepidopteran PDV fragments strongly suggest that they do not result from contamination or from

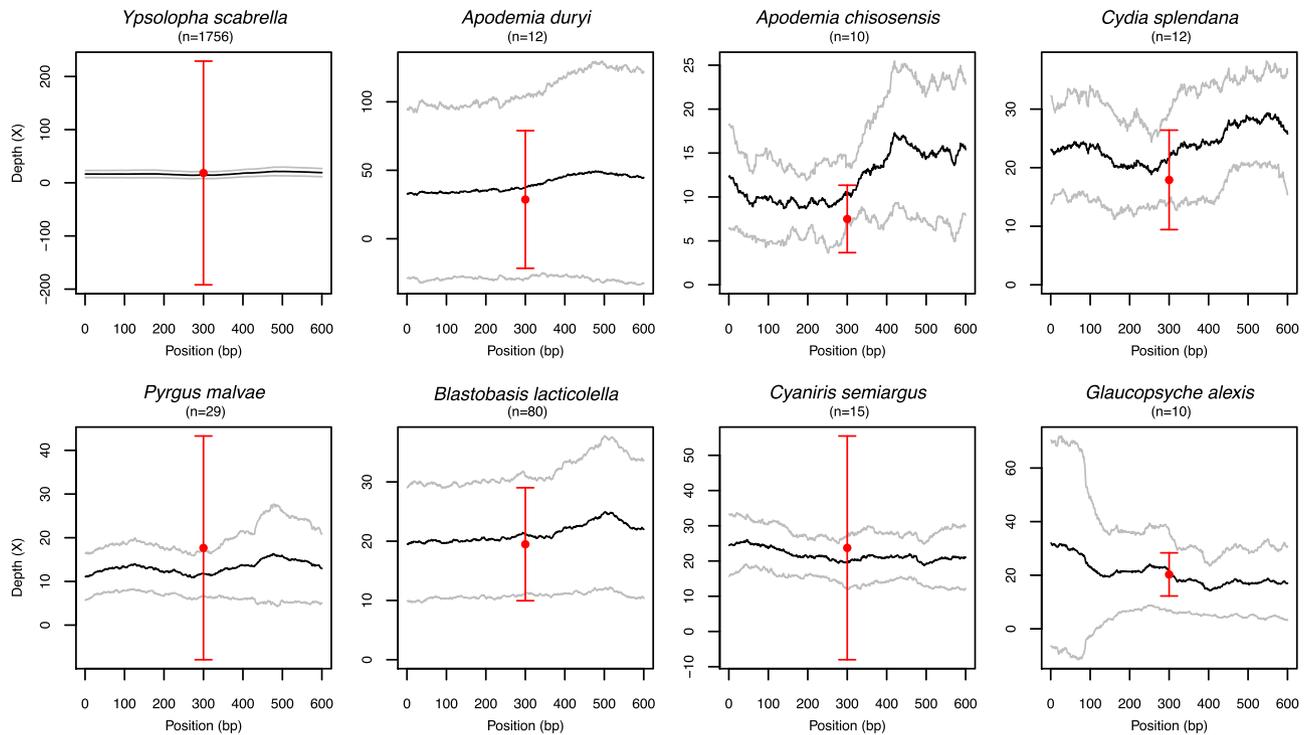


FIG. 9. Sequencing depth and numbers of chimeric reads support HIM-mediated chromosomal integration of IV fragments in the germline of lepidopteran species. Eight lepidopteran species in which 10 or more J1- or J2-bordered IV fragments were found were selected to compute average sequencing depth over all junctions. Number of junctions for each species are indicated between brackets. The x-axis indicates the genome position along the junction, the IV fragments being located between position 1 and 300 bp and the flanking lepidopteran genome regions being between 301 and 600 bp. The black line indicates average sequencing depth over 300 bp upstream and downstream of the junctions, with grey lines indicating standard deviation. Red dots indicate average numbers of chimeric reads supporting junctions in each species. Standard deviation is also shown.

chromosomal integrations in somatic genomes of parasitized individuals. Polydnavirus DNA circles are known to occur in three forms (fig. 1): 1) proviral segments flanked by DRJs in the genome of wasps, in which J1 and J2 motifs are next to each other, separated by a short sequence (about 50 bp) in the HIM, 2) circular sequences in the wasp ovaries and in host larvae, containing a single, recombinant DRJ, and in which J1 and J2 motifs are also next to each other in the HIM, and 3) sequences bordered by J1 and J2 motifs integrated in parasitized hosts' somatic genomes. The fact that PDV fragments we report here in lepidopteran genomes are bordered at one of their extremities by the J1 or J2 motif, unlike in the wasp genome in which these motifs lie next to each other, argues against the possibility that these fragments result from contamination by wasp DNA that could have occurred during DNA extraction and/or sequencing. However, these fragments could correspond to somatic integrations that could have occurred in the individuals that were used for whole genome sequencing (many lepidopteran genomes are obtained from a single individual). This would imply that these individuals were parasitized by a wasp prior to sequencing. An important difference between somatic and germline integrations of PDV circles is that although the former is present only in a subset of somatic cells, the later should be present in all cells of an individual that would have received them from its parents. In agreement with the presence of

PDV IEs in a subset of somatic cells during parasitism, 98% of the CtBV IEs we identified in parasitized *S. nonagrioides* larvae were covered by only one chimeric read (Muller et al. 2021). The pattern is similar here for HdIV, with on average 96.6% of HdIV IEs covered by only one read across samples. In contrast with somatic IEs, germline integrations should be covered by multiple reads because they are expected to be present in all cells of the sequenced lepidopteran individual. Furthermore, sequencing depth over the integration should be similar between the PDV and lepidopteran flanking genomic region. To verify this, we recovered raw Illumina reads for the eight lepidopteran species having ten or more IV fragments, and for the six species having five or more BV IEs, and we computed average sequencing depth and number of chimeric reads for all IV or BV junctions in each species. In agreement with our predictions, we found that the average number of chimeric reads per species was always higher than one (figs. 9 and 10). In fact, it was higher than five for five out of six species harboring BV fragments (fig. 10) and higher than ten for seven out of eight species harboring IV fragments (fig. 9). Importantly, the average number of chimeric reads per IE was always close to the average sequencing depth at the PDV integration point. Thus, variations in numbers of chimeric reads between species have no biological underpinnings but are due to variation in sequencing depth. Furthermore, we found that for most species, average

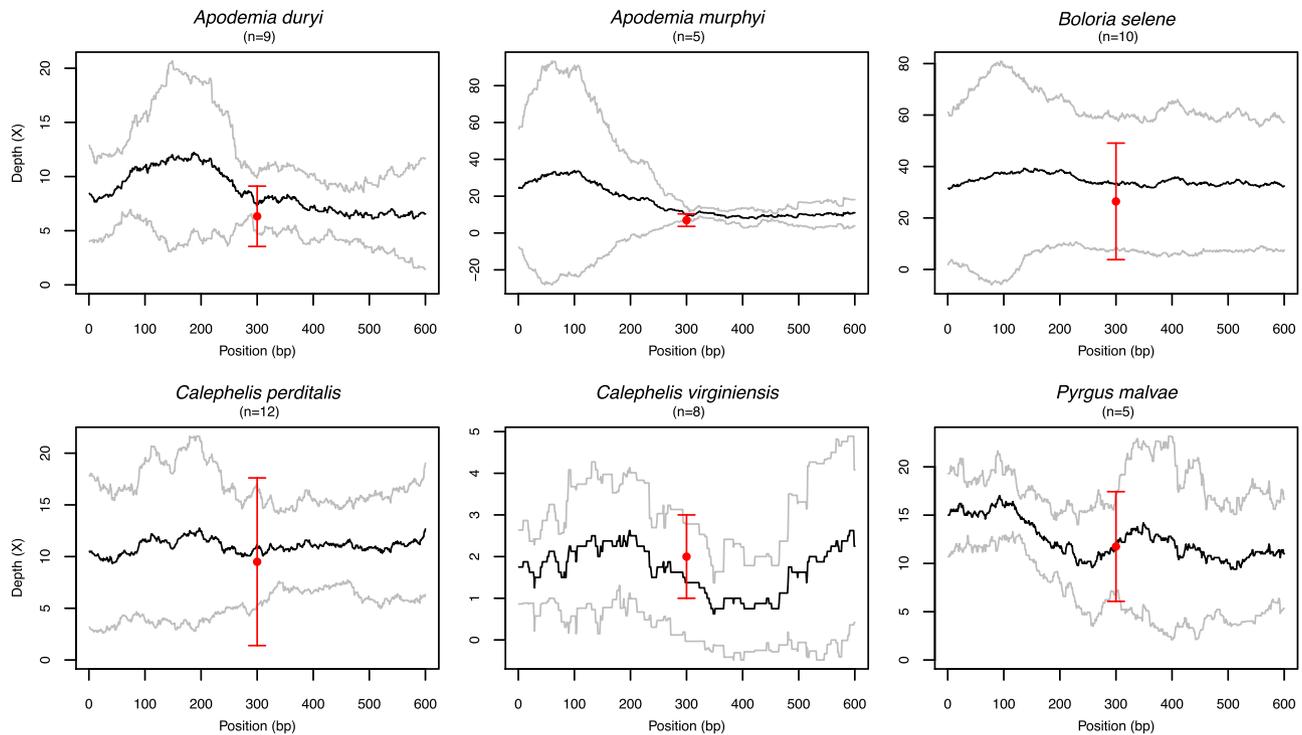


FIG. 10. Sequencing depth and numbers of chimeric reads support HIM-mediated chromosomal integration of BV fragments in the germline of lepidopteran species. Six lepidopteran species in which five or more J1- or J2-bordered BV fragments were found were selected to compute average sequencing depth over all junctions. Number of junctions for each species are indicated between brackets. The x-axis indicates the genome position along the junction, the BV fragments being located between position 1 and 300 bp and the flanking lepidopteran genome regions being between 301 and 600 bp. The black line indicates average sequencing depth over 300 bp upstream and downstream of the junctions, with grey lines indicating standard deviation. Red dots indicate average numbers of chimeric reads supporting junctions in each species. Standard deviation is also shown.

sequencing depth per species was homogeneous across PDV/lepidopteran junctions (figs. 9 and 10).

For two species (*Apodemia duryi* and *A. murphyi*), we noticed a particularly high increase in average sequencing depth over BV circles compared to the flanking lepidopteran genomic sequence (fig. 10), which could be interpreted as resulting from the presence of unintegrated BV circles in these species. However, reads produced by unintegrated circles are expected to cover such circles relatively homogeneously, which should here translate into a sharp increase in depth on the circle side starting from the very junction. Yet, here the increase in depth is progressive. We note that the quality of the assembly of the two species is low as both have N50 lower than 500 bp. It is thus possible that the increase in sequencing depth over the circle side in *Apodemia* species may be caused by the fact that several integrated circles have not been included in the assembly. Furthermore, the average number of chimeric reads in the two species (6.3 and 7.0) is consistent with the presence of most BV IEs being present in all cells. Although we are unable to provide a definitive explanation for the higher depth on BV circles compared to flanks for these two species, our observations are not consistent with the presence of unintegrated circles. Overall, we contend that these results indicate that most J1/J2-bordered PDV fragments we found in lepidopteran genomes result from germline integration that were then transmitted vertically in host populations.

An independent confirmation of this reasoning was obtained by assessing experimentally the presence of J1 and J2 extremities of insertions related to the HdIV12 circle and to the *C. congregata* BV circles 1 and 17 in different individuals of two very common butterflies: the small white and the green vein white (*Pieris rapae* and *Pieris napi*). Twelve individuals of each species were collected from three different locations in France. We obtained specific bands by PCR amplification of the junctions between flanking and viral sequences for all individuals (fig. 11).

Sequencing of the PCR products confirmed that J1 and J2 constituted actually the extremities of the four insertions identified from *P. rapae* and *P. napi* genomes, which thus appear to be fixed in the population and not a particular feature of the individuals used for genome sequencing. All sequences are provided in [supplementary Dataset 2, Supplementary Material](#) online. The *P. napi* HdIV 12 J2 sequence corresponds to the blastn hit number 167 in [supplementary table S1, Supplementary Material](#) online. The other five verified junctions were not recovered in the bioinformatic search because this experimental study was performed independently by three of us (G.P., J.M.D., and K.M.) in 2018. The study differed from the bioinformatic survey performed in 2022 in that it used CcBV as a query bracovirus instead of CtBV and it included two genomes of *P. napi* instead of one (see Materials and Methods and [supplementary Dataset 2, Supplementary Material](#) online).

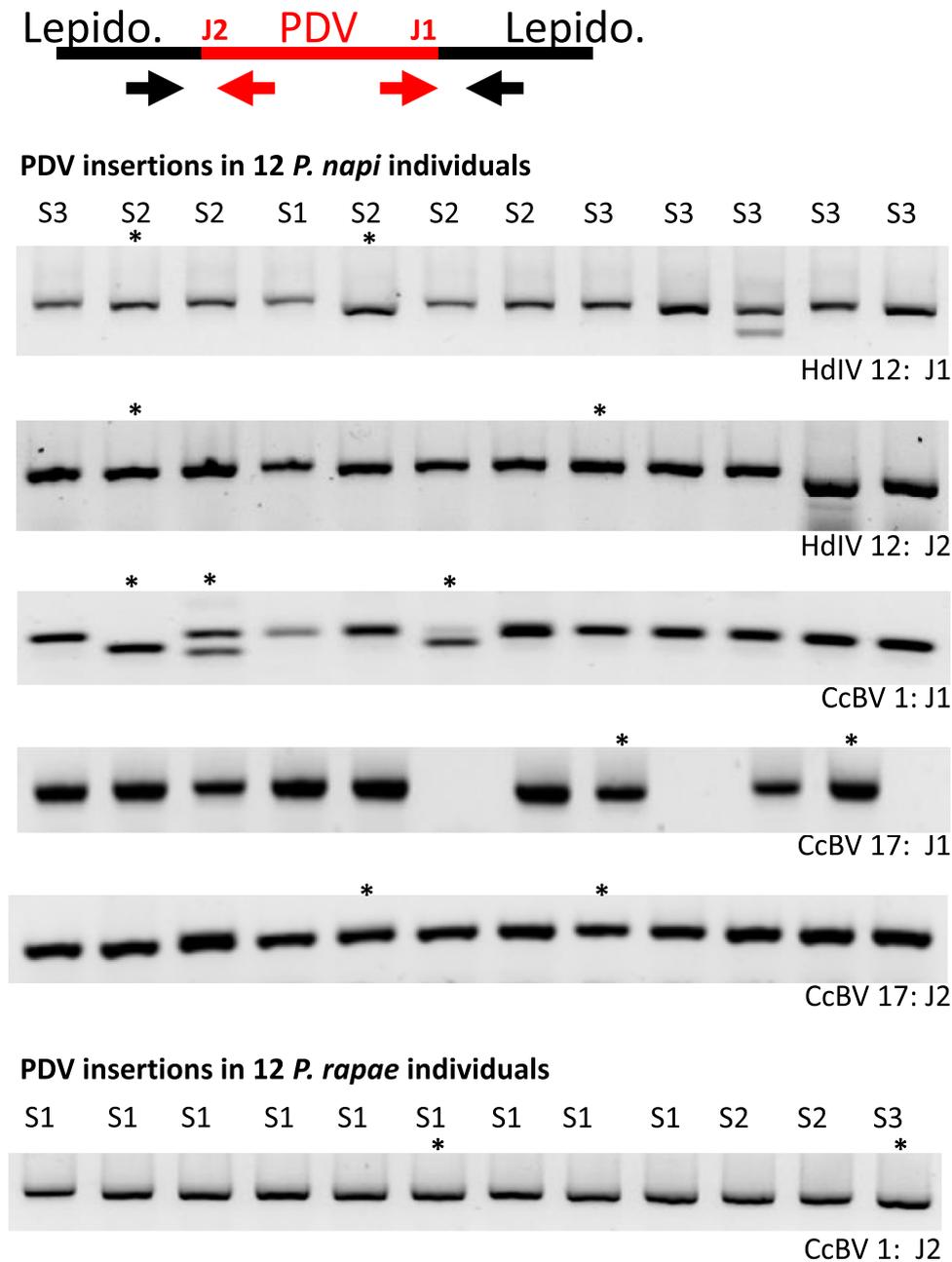


Fig. 11. Experimental verification of PDV-lepidopteran junctions in populations of two butterfly species (*Pieris napi* and *P. rapae*). The top illustration shows a schematic HIM-mediated insertion of a PDV circle bordered by the J1 and J2 motifs, as well as the position of the PCR primers (arrows at the bottom) designed on each side of the junction. Below are pictures of electrophoresis gels showing presence or absence of PCR bands obtained with primers targeting five PDV-*P. napi* junctions and one PDV-*P. rapae* junction in 12 individuals of each species sampled in three different geographical locations (S1, S2, S3, see Materials and Methods). Asterisks indicate PCR products that were Sanger-sequenced. The sequences are provided in [supplementary Dataset 2, Supplementary Material](#) online. For each verified junction, the type of PDV circle is indicated at the bottom right of the gel. HdIV, *Hyposoter didymator* ichnovirus; CcBV, *C. congregata* BV. The genomic coordinate of each junction is provided in [supplementary Dataset 2, Supplementary Material](#) online. Note that polymorphism was observed in *P. napi* for CcBV1 J1 containing PCR products (two individuals having two bands) and in CcBV17 J1 (amplification was reproducibly not obtained from the DNA of three individuals) probably reflecting a still ongoing erosion of these regions. CcBV1 insertions in *P. napi* and *P. rapae* have different flanking regions (see [supplementary Dataset 2, Supplementary Material](#) online).

Integration Dynamics of Polydnvirus Fragments During Lepidopteran Evolution

Some PDV circle integrations are likely to be ancient and thus shared with other species as found for a 5-million years old bracovirus integrations shared by the monarch

and related species of the Danaina subtribe (Gasmi et al. 2015). To gain insights into the timing of different IEs of PDV fragments in lepidopteran germlines, we first searched for PDV fragments that are orthologous between species, that is fragments located at the same position in

Table 2. Characteristics of J1- or J2-bordered PDV Fragments Shared at Orthologous Loci Between Two or More Lepidopteran Species.

| PDV Type | Number ^a | Junction | Species 1 | Species 2 | %ID | Ali. Length ^b | PDV Size ^c | Flank Size |
|----------|---------------------|----------|----------------------------|-------------------------|--------|--------------------------|-----------------------|------------|
| IV | 1 ^(A) | Hd27_J2 | <i>A. duryi</i> | <i>A. virgulti</i> | 97.652 | 1,533 | 546 | 987 |
| IV | 5 ^(A) | Hd27_J2 | <i>A. mejicanus</i> | <i>A. virgulti</i> | 99.547 | 1,546 | 546 | 1000 |
| IV | 7 ^(A) | Hd27_J2 | <i>A. mormo</i> | <i>A. virgulti</i> | 95.111 | 1,493 | 553 | 940 |
| IV | 2 ^(B) | Hd27_J1 | <i>A. duryi</i> | <i>A. virgulti</i> | 99.367 | 1,421 | 421 | 1,000 |
| IV | 4 ^(B) | Hd27_J1 | <i>A. mejicanus</i> | <i>A. duryi</i> | 99.766 | 856 | 421 | 435 |
| IV | 6 ^(B) | Hd27_J1 | <i>A. mormo</i> | <i>A. duryi</i> | 95.148 | 742 | 421 | 321 |
| IV | 3 | H12_J1 | <i>A. duryi</i> | <i>A. virgulti</i> | 91.442 | 1,297 | 564 | 733 |
| IV | 19 | Hd27_J2 | <i>C. nemesis</i> | <i>C. perditalis</i> | 90.94 | 1,181 | 465 | 716 |
| IV | 20 | H16_J2 | <i>C. nemesis</i> | <i>C. virginiensis</i> | 82.149 | 1,238 | 490 | 748 |
| IV | 23 | Hd27_J1 | <i>Doberes anticus</i> | <i>Telegonus cellus</i> | 95.878 | 1,407 | 407 | 1,000 |
| IV | 24 | Hd27_J1 | <i>Ostrinia furnacalis</i> | <i>O. nubilalis</i> | 99.142 | 1,049 | 433 | 616 |
| IV | 25 | H12_J2 | <i>P. macdunnoughi</i> | <i>P. napi</i> | 98.541 | 754 | 1745 | 298 |
| BV | 2 | S10_J1 | <i>C. nemesis</i> | <i>C. perditalis</i> | 93.57 | 1,377 | 479 | 898 |
| BV | 3 | S1_J1 | <i>C. nemesis</i> | <i>C. perditalis</i> | 92.62 | 1,628 | 629 | 999 |
| BV | 4 | S1_J2 | <i>C. nemesis</i> | <i>C. perditalis</i> | 93.14 | 1,890 | 901 | 989 |
| BV | 6 | S16_J2 | <i>E. lupina</i> | <i>E. tegula</i> | 94.18 | 969 | 429 | 547 |
| BV | 7 | S7_J1 | <i>O. hopfferi</i> | <i>O. stangelandi</i> | 99.07 | 1,288 | 487 | 1,000 |

^aOrthologous J1- or J2-bordered PDV fragments are listed by pair of species, number refer to their number label in sequence alignments provided in [supplementary Datasets 3 and 4, Supplementary Material](#) online. Letters between brackets indicate pairs of species sharing the same orthologous fragment (e.g., PDV fragments 1, 5, and 7 are shared at the same orthologous locus in *Apodemia duryi*, *A. virgulti*, *A. mejicanus* and *A. mormo*).

^bAli. size indicate the length of the PDV + flank alignment provided in [supplementary Datasets 3 and 4, Supplementary Material](#) online.

^cPDV size indicate the size of the largest PDV fragment among a group of orthologous fragments.

the genome of two or more species. Our approach allowed us to identify five BV fragments and eight IV fragments shared by two or more species at the same genomic position ([table 2](#)).

One IV orthologous fragment, found in *Calephelis metalmarks* may be as old as the monarch integration as it shows only 82.1% nucleotide identity between *C. nemesis* and *C. virginiensis* and the split between the two species has been dated at 4.9 Ma ([table 2, supplementary Dataset 3, Supplementary Material](#) online) ([Cong et al. 2017](#)). By contrast, all other cases involve species for which divergence time is unknown but is likely recent because the species are congeneric and/or nucleotide identity between orthologous fragments is high (from 90.9% to 99.3%; mean = 96.6%). Regarding IV orthologous fragments, two are shared between four very closely related species of metalmark butterflies (genus *Apodemia*, Riodinidae) ([Zhang et al. 2019](#)), one of them is shared between the green-veined white (*Pieris napi*, Pieridae) and its close relative *P. macdunnoughi* ([Chew and Watt 2006](#)) and another one is shared between the European corn borer (*Ostrinia nubilalis*, Crambidae) and the Asian corn borer (*O. furnacalis*), which diverged recently ([table 2; supplementary Dataset 3, Supplementary Material](#) online) ([Bourguet et al. 2014](#)). In terms of BV orthologous fragments, three are shared between *C. nemesis* and *C. perditalis* metalmark butterflies, one is shared between two species of firetips (*Oxynetra hopfferi* and *O. stangelandi*) and one is shared between *Emesis lupina* and *E. tegula* ([table 2; supplementary Dataset 4, Supplementary Material](#) online) ([Cong et al. 2017](#)).

In addition to orthologous fragments shared between species, we found several paralogous IV fragments sharing the same immediate flanking genome region ([supplementary table S5 and supplementary Dataset 5,](#)

[Supplementary Material](#) online). Numbers of such paralogous fragments vary from two in the Kamehameha butterfly *Vanessa tameamea* (Nymphalidae) to 83 in the wainscot hooktip *Y. scabrella* (Ypsolophidae), with a maximum of three sequences in a given paralogous group ([supplementary table S5, Supplementary Material](#) online). The finding of paralogous sequences sharing flanks indicates that some IV fragments were duplicated after integration. Most duplications are likely recent as identity levels within these paralogous groups are generally high (from 82.7% to 100%; mean = 97.2%; median = 99.2%).

We next generated multiple alignments of PDV fragments inserted in lepidopteran genomes and PDV circles from various wasps, that we submitted to phylogenetic analyses. It is important to note that these analyses were not performed to infer the direction of HT of PDV fragments ([supplementary Text 1, Supplementary Material](#) online). The direction can indeed be unambiguously inferred to be from wasp-to-lepidopteran based on the wasp origin of PDVs, the presence of J1 and/or J2 motif at the extremities of PDV fragments and the presence of recombined DRJs as explained in the preceding sections ([supplementary Text 1, Supplementary Material](#) online). As expected in these trees, PDV insertions found to be orthologous between two or more lepidopteran species group together. This is the case, for example, of the four Hd27 J2 fragments flanked by the same genomic region in four *Apodemia* species ([table 2; fig. 12](#)). By contrast, when multiple PDV fragments were found in genomes of a given lepidopteran family, genus, or species, they generally do not cluster together in a monophyletic group ([fig. 12; supplementary fig. S2, Supplementary Material](#) online). For example, Hd27 J2 fragments found in various Hesperidae species (dark green in [fig. 12](#)) fall in at least

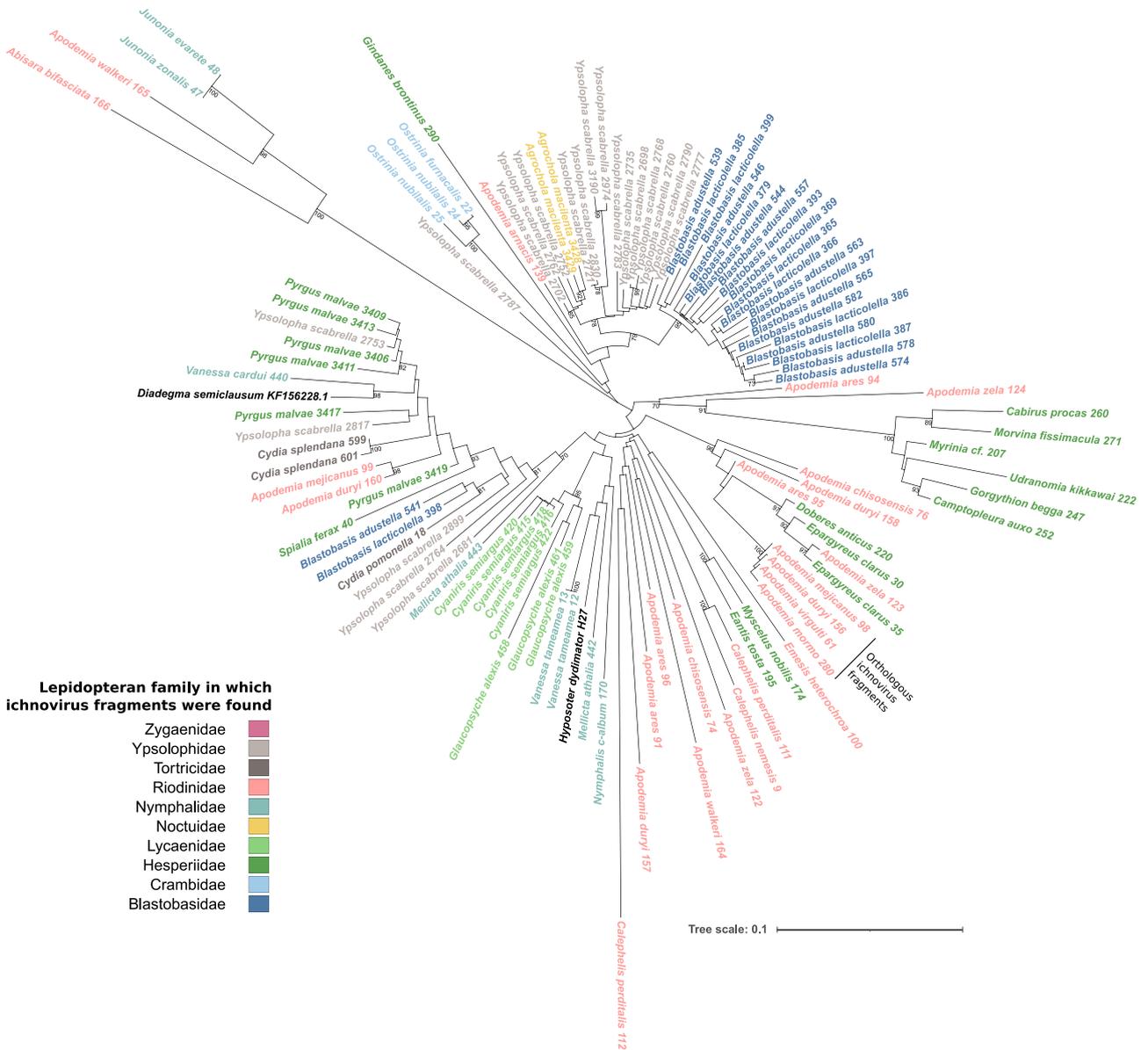


Fig. 12. Unrooted phylogeny of IV fragments homologous to the J2 junction of the HdIV27 circle found in lepidopteran genomes. All fragments similar to the J2 extremity of the HdIV27 circle found in lepidopteran genomes were aligned together with HdIV27 and a homologous circle from *D. semiclausum* (both are in black). Prior to alignment, the numerous fragments found in *Y. scabrella*, *B. adustella* and *Blastobasis lacticolella* were clustered at 95% nucleotide identity threshold. The name of the species in which PDV fragments were found are colored according to lepidopteran family and the number associated to each species name is a unique identifier that allows retrieving the sequence in [supplementary table S1, Supplementary Material](#) online (column “Number”). Numbers on branches are bootstrap values higher than 70%. The HdIV27-like orthologous fragments found in *Apodemia* (table 2) are indicated. The tree was built using the Maximum Composite Likelihood model of substitution available in MEGA 11 (Tamura et al. 2021) with 100 bootstrap replicates.

six different clusters intermingled with fragments found in other lepidopteran families. Similarly, Hd27 J2 fragments from various species of the *Apodemia* genus, as well as those from the species *Y. scabrella* are polyphyletic, forming multiple clusters in the tree (fig. 12).

Furthermore, the few PDV sequences available only from different contemporary wasps are also generally scattered in the trees. This is the case, for example, of the Hd27 J2 sequence from the wasp *D. semiclausum*, which falls more closely related to lepidopteran Hd27 J2 than to the Hd27 J2 sequence of *H. didymator*

(fig. 12). Similarly, H12/H16 J1 sequences from *H. fugitivus* and *D. semiclausum* are scattered in the phylogeny, falling closer to lepidopteran H12/H16 J1 fragments than to H12 and H16 J1 sequences of *H. didymator* (supplementary fig. S2, Supplementary Material online). Altogether, these results suggest that lepidopteran species have suffered IV integrations from multiple, relatively distantly related donor wasps (and at different times of the evolutionary history of wasp species). We predict that repeating such an analysis when PDVs from more Campopleginae and Braconidae wasps (which contain

thousands of species) are available will allow inferring the wasp source of all or most PDV insertions found in lepidopteran genomes.

Discussion

Preferential Integration of HdIV in Hemolymph Cells of *S. frugiperda*

In this study, we have shown that seven out of the 57 IV circles of the Ichnumonidae wasp *H. didymator* undergo chromosomal integration into two tissues (hemolymph and fat body) of parasitized fall armyworms (*S. frugiperda*). Three of the circles integrate massively into these somatic cells and we estimate that on average, each host haploid genome suffers between 13 and 40 IEs 72 h p.p. These numbers are similar to those reported earlier for *C. typhae* bracovirus (CtBV) which typically undergoes 12–85 IEs per host haploid genome on average, depending on the tissue (Muller et al. 2021). As observed for CtBV, HdIV integrations are more numerous in the hemolymph than in the fat body. This is also in line with earlier observations made for *Microplitis demolitor* BV (Beck et al. 2007). Given that hemocytes are individual cells that circulate through the entire body of caterpillars, they may be more accessible to PDVs than cells of other tissues, such as the fat body, that are akin to cell aggregates in which only external layers may be immediately reachable by PDVs. Alternatively, or in addition, PDVs may have evolved to preferentially target hemocytes, the main mediators of insect immune response (Lavine and Strand 2002; Jiang et al. 2010).

What Role for PDV Circle Integration in Parasitism Success?

A question emerging from earlier studies of PDVs relates to the role of circle integration during parasitism (Beck et al. 2011; Benoist et al. 2017; Chevignon et al. 2018; Muller et al. 2021). In our study of CtBV, we found that similar quantities of integrated circles (i.e., circles possessing a HIM motif) and non-integrated circles (no HIM motif) can persist over at least 7 days after oviposition, that is about half the time needed for *C. typhae* larvae to emerge from their host (Muller et al. 2021). Thus, it seemed that integration did not play so much of a role in ensuring persistence of BV circles in the context of this earlier study. However, host tissues parasitized by *C. typhae* were sequenced at only one time point in Muller et al. (2021), such that we could not assess how the overall quantity of integrated versus non-integrated circles evolves during parasitism. Here, we show that the number of IEs increases with time after oviposition. Although we did not perform replicates to test whether this increase is statistically supported, we observe that it holds both for the hemolymph and fat body, which may be considered pseudo-replicates. We deduce that this increase in IEs through time is due to continuous integration of circles from 24 to 72 h p.p., rather than to caterpillar cell replication, since we did not detect more

IEs covered by multiple reads (as a result of cell division) at 72 h p.p. than at 24 h p.p. Mirroring this increase in integrated circles, we found that the sequencing depth over HdIV, and thus the overall quantity of IV circles, decreases through time (fig. 2). The decrease is particularly striking in the fat body, whereby HdIV sequencing depth is divided by three between 24 and 72 h p.p. (fig. 2). The combined observation of an increase in integrated IV circles and a decrease in overall circle quantity implies that the quantity of non-integrated circles decreases quite sharply through time after oviposition, and faster than that of integrated circles. In this context, integration may be seen as a way for the PDVs to persist in larger amounts in host cells throughout the entire duration of wasp progeny development.

Remarkable Similarity in Ichnovirus and Bracovirus Chromosomal Integration Patterns

This study is the second one investigating integration of IV DNA circles using bulk Illumina sequencing of parasitized host tissues. Like in Wang et al. (2021b), we show that the majority of IEs occur through double-strand breaks located in the HIM. We recently investigated chromosomal integration of a bracovirus (CtBV) during parasitism using the same approach as here and found that CtBV circles also almost exclusively integrated through double-strand breaks occurring within the HIM (Muller et al. 2021). This confirms the central role of these motifs in mediating circle integration in host somatic cells and ensuring parasitism success. Applying the same method to characterize circle integration of an IV and a BV (Muller et al. 2021) allows us to directly compare the two systems. Patterns of chromosomal integrations turn out to be highly similar between HdIV and CtBV, in terms of nature—all IEs are mediated by a HIM motif having a similar structure with conserved inverted repeats separated by a short stretch of sequence not conserved—, tissue tropism (hemocytes as preferential target), and numbers of IEs per host cell. This similarity suggests the mechanisms involved in integration could be highly related, which is unexpected given that IV and BV have an independent origin, deriving from domestication of viruses belonging to different viral families (Gimenez et al. 2020; Gauthier et al. 2021; Gilbert and Belliardo 2022).

The mechanisms underlying PDV circle integration are not fully understood. In the case of BV, three candidate wasp genes of nudiviral origin coding for Tyrosine recombinases (*vlf-1* and *int-2* homologues) have been proposed to operate together ensuring the integration of HIM-containing circles (Chevignon et al. 2018) and were shown by functional assays to be involved in the integration of some circles of *Cotesia vestalis* bracovirus (Wang et al. 2021a). Moreover, there is evidence showing that some BV circles also rely on different lepidopteran retroviral integrases for chromosomal integration (Wang et al. 2021a). In the case of IV, circle integration has not been studied at the functional level yet, but the viral machinery

does not include a gene encoding a protein with a known integrase domain (Legeai et al. 2020), which suggests that host integrases could be involved. Indeed, site-specific recombinases/integrases such as retroviral integrases and transposases are abundant in eukaryote genomes. Another possibility is that among the genes of unknown function deriving from the IV ancestor, there is a recombinase from a still uncharacterized family that has not yet been recognized as such. It is noteworthy that the proteins encoded by the conserved gene set of ichnoviruses do not contain any recognizable conserved domain (Volkoff et al. 2010). The striking similarities between BV and IV circles integration mechanisms could be explained by the fact that they were shaped by similar structural constraints, despite their independent origin. In fact, the J1 and J2 motifs are much alike inverted terminal repeats of DNA transposons (ITRs). However, unlike ITRs which lie at the extremities of DNA transposon copies, these inverted repeats are located next to each other, only separated by the short stretch of nucleotides deleted during integration. Once a transposase binds to the ITRs, the formation of the transpososome relies on properties of the DNA separating the ITRs, including a minimum length (Hickman et al. 2018) (supplementary fig. S3, Supplementary Material online). Similarly, to form an integration complex starting from a PDV circle, a minimum sequence length between the ITRs is most probably required, which would explain the presence of the intervening DNA between the J1 and J2 motifs (supplementary fig. S3, Supplementary Material online). The fact that the IV and BV HIM only show very low similarity in primary sequence may be due to the involvement of recombinases belonging to distantly related families. Thus, the action of similar constraints could be sufficient to explain the similarity of the HIM structure of BV and IVs. Confirmation of this hypothesis requires to better characterize the molecular mechanisms involved in BV and IV circles integration. Another possibility would be that recombinases of the same family, which remain to be identified, may act as major players in circle integration of both BVs and IVs. In this case, HT events of HIM sites might be invoked to explain that the two PDVs genera share the same mechanism despite their different origin. The similarity in J1 and J2 motifs as well as in other positions of the HIM between BVs and IVs may be viewed as supporting this scenario (fig. 4). HT events have been proposed as an explanation of the common phylogenetic signature shared by the virulence factors V-ANKs of BV and IVs (Falabella et al. 2007; Ceirqueira de Araujo et al. forthcoming).

Finally, though as for HdIV, Wang et al. (2021b) found a majority of DsIV IEs occurring at HIM in *P. xylostella* hemocytes, they also identified large numbers of IEs involving random positions along both HIM-bearing circles and circles devoid of HIM. This led them to conclude that two distinct mechanisms were involved in DsIV chromosomal integration. Here the number of IE falling outside HIM or in circle devoid of HIM is so low that the biological significance of these integrations in terms of parasitism success

in the *H. didymator/S. frugiperda* system is questionable. Interestingly, whatever the system (CtBV, DsIV, HdIV), there is an enrichment for microhomologies between the host genome and IV circles at the virus–host junction site, suggesting that host DNA repair mechanisms may be involved in at least a subset of IEs (Muller et al. 2021; Wang et al. 2021b). In fact, the fraction of junctions involving microhomology is higher for IV than for BV as contrary to BV, for IVs we observe almost no junction involving blunt-ended wasp and host sequences or 1-bp microhomology between them.

Widespread Germline Infiltrations of Polydnaviruses in Multiple Lepidopteran Families

Earlier studies uncovered a number of BV circle fragments in the genome of various species of lepidopterans and proposed that these sequences were horizontally transferred from wasp through HIM-mediated integration in the germline genome of lepidopterans (Schneider and Thomas 2014; Gasmı et al. 2015; Di Lelio et al. 2019). However, none of the horizontally transferred BV sequences reported in these studies contained the J1 or J2 motif of the HIM. This could be explained either because the transfers were ancient and the integrated PDV sequences (including the HIM) were degraded, or because these transfers occurred through a mechanism not involving a double-strand break within the HIM. The finding that CcBV (Chevignon et al. 2018), DsIV (Wang et al. 2021b), CvBV (Wang et al. 2021a), CtBV (Muller et al. 2021), and HdIV (this study) undergo massive HIM-mediated chromosomal integration in host somatic cells during parasitism, together with the recent availability of hundreds of lepidopteran whole genome sequences led us to assess the extent to which this mechanism fostered wasp-to-lepidopteran HT during evolution. We uncovered dozens of J1- or J2-bordered IV and BV fragments in a total of 124 lepidopteran species belonging to 15 families. In addition to show that sequencing depth was homogenous over lepidopteran/PDV circle junctions for most of these fragments, we also found that they were supported by multiple chimeric reads, contrasting with somatic PDV circle insertions that are almost all covered by only one chimeric read. We also found that many of the PDV fragments integrated into lepidopteran genomes contained a single, recombined DRJ motif (supplementary tables S1 and S2; supplementary Dataset 1, Supplementary Material online; fig. 1), consistent with the transfer of wasp PDV circles that were then linearized and integrated into host genomes. Furthermore, we PCR-validated typical extremities of HIM-mediated junctions of two IV and two BV integrations in multiple individuals of two species of *Pieris* butterflies and uncovered multiple integrations shared at orthologous loci between two to four lepidopteran species. Altogether these results indicate that HIM-mediated germline infiltration of PDV circles is widespread in lepidopterans, the main hosts of braconid and ichneumonid parasitoid wasps.

Limitations of the Current Study and Perspectives

The number of J1- or J2-bordered PDV circle fragments we report here is necessarily underestimated because we only used HdIV and CtBV circles, for which HIM were annotated using the same approach (Muller et al. 2021; this study), as queries to perform our similarity searches. In addition, we used relatively conservative criteria to retain blast hits for downstream analysis (at least 300-bp long and e -value < 0.0001) and many genomes included in our search are of poor quality (only 210 out of 775 genomes included have a N50 > 10 kb). However, our phylogenetic analyses reveal widespread polyphyly of PDV fragments at the lepidopteran family, genus, and even species level, indicating that a given lepidopteran species or multiple species within a family or a genus have received PDV fragments from various wasp donor species. We found that the number of PDV fragments uncovered per lepidopteran family was positively correlated with the number of genomes available in each family, with no apparent biological factor further explaining the distribution of these fragments in lepidoptera. It is, however, noteworthy that some species have remarkably high numbers of PDV fragments (e.g., 81 and 80 PDV fragments in 2 *Blastobasis* species and 1,756 PDV fragments in *Y. scabrella*) that were acquired from multiple donor wasp species (fig. 12; supplementary fig. S2, Supplementary Material online). We are unaware of how frequently these species are targeted by parasitoid wasps in the wild. By contrast, all lepidopteran species that are known hosts of CtBV (host: *S. nonagrioides*), CcBV (host: *Manduca sexta*), CvBV (host: *P. xylostella*), HdIV (host: *H. armigera*) and DsIV (host: *P. xylostella*) are devoid of HIM-mediated PDV insertions. It will be interesting to assess what factors may foster HIM-mediated germline integration of PDV circles in different hosts or non-hosts lepidopteran species using a dedicated and comprehensive sampling of parasitoid wasp and lepidopteran species.

This study is only a first step in the large-scale characterization of wasp-to-host HT of PDV circles. We chose to only focus on J1- or J2-bordered PDV fragments because these fragments contain the molecular signature typical of well-characterized chromosomal integrations occurring during parasitism. We could thus formulate explicit predictions regarding expected patterns of sequencing depth and number of chimeric reads. However, our similarity search uncovered many more PDV-like sequences not bordered by the J1 or J2 motifs of the HIM in lepidopteran genomes. For example, our initial blastn search yielded 2,214 hits longer than 300 bp with an e -value < 0.0001 to HdIV circles devoid of HIM. This number was even much higher ($n = 19,558$) when using CtBV devoid of HIM as queries. Though these sequences likely underwent some form of HT involving lepidopterans and parasitoid wasps, another full dedicated study will be necessary to determine and quantify which scenarios best explain these transfers. For example, it is possible that some of these sequences originate from donor wasp species in which these

PDV circles bear a HIM motif but that this motif is not present in the homologous PDV circles of the wasps we used as queries to perform our searches (*H. didymator* or *C. typhae*). Alternatively, the fact that nucleocapsids of PDVs are able to enter the nuclei may also favor integration of circles by a more general mechanism such as DNA repair. Although we did not observe a significant level of circle integration not involving HIM sites during parasitism, such rare events could occur in the germline at the time scale of evolution. It is also possible that many of these sequences correspond to transposable elements present in HdIV and CtBV (Dupuy et al. 2011) and that these TEs were horizontally transferred either together with PDV circles or independently, as proposed by Heringer and Kuhn (2022). Another limitation of this study is that it is focused only on lepidopterans, the major hosts of PDV-encoding braconid and campoplegine parasitoid wasps (Gauld 1988). Yet, some wasps within the two families are known to parasitize non-lepidopteran hosts (Robin et al. 2019) and it will be interesting to extend the search for HIM-mediated HT of PDV circles to a larger diversity of hosts. Extending the search to species not known to be hosts of parasitoid wasps may also reveal unexpected wasp–host interactions.

A Major Route of Horizontal Transfer of Genetic Material Among Insects

HT of genetic material is widespread and a major force shaping prokaryote evolution (Soucy et al. 2015). In eukaryotes, the importance of HT is increasingly recognized (Husnik and McCutcheon 2018; Sibbald et al. 2020; Van Etten and Bhattacharya 2020), but the extent to which it influenced genome evolution remains debated, especially in animals and other multicellular taxa (Martin 2017; Salzberg 2017). Several studies reported spectacular individual cases of HT of genes in various animals, often with functional evidence supporting an important role played by horizontally transferred genes in the recipient species (Moran and Jarvik 2010; Acuna et al. 2012; Wybouw et al. 2014, 2016; Danchin et al. 2016; Leclercq et al. 2016; Gasmi et al. 2021; Xia et al. 2021; Cummings et al. 2022). Large-scale analyses of HT of genes and TEs in insects tend to show that these transfers occurred recurrently in most insect lineages and likely had important consequences on insect genome evolution (Peccoud et al. 2017; Li et al. 2022). So far however, the mechanisms through which these transfers occurred remain unclear and no mechanism dedicated to HT is known in animals. Together with previous studies on CcBV (Chevignon et al. 2018), DsIV (Wang et al. 2021b), CvBV (Wang et al. 2021a), and CtBV (Muller et al. 2021), our work on HdIV shows that parasitoid wasps use massive HIM-mediated HT of PDV circles to hijack host somatic cells, ensuring parasitism success. We also show that though not dedicated to germline integration, this mechanism fostered many accidental HT of PDV sequences in the germline of a large number of lepidopteran hosts that survived to

parasitism (or to the injection of PDV only) and transmitted these sequences to their offspring. HIM-mediated HT of PDV sequences may thus be viewed as a major route of HT in insects. The extent to which it may have facilitated the HT of non-PDV sequences integrated into PDV DNA circles or copackaged with PDV circles into PDV viral particles remains to be assessed. Two studies have provided functional evidence that a PDV gene (*Sl gasmin*), acquired by a noctuid moth through HT, plays an essential role in the antibacterial immune response of the moth (Gasmi et al. 2015; Di Lelio et al. 2019). Our work suggests that many PDV genes have been acquired by lepidopterans through HIM-mediated HT. It opens new avenues to further quantify this phenomenon and the impact it had on insect evolution.

Materials and Methods

Insects Rearing, Parasitization and DNA Sequencing

The *S. frugiperda* laboratory colony is maintained on a semi-synthetic maize diet under stable conditions (24 ± 2 °C; 75–65% relative humidity; 16 h light:8 h dark photoperiod). The wasp *H. didymator* laboratory colony is reared on *S. frugiperda* at 26 ± 2 °C with a 16 h light:8 h dark photoperiod. Fourth instar *S. frugiperda* larvae were each parasitized by exposing them to one *H. didymator* female wasp until one oviposition event was observed. Two tissues were sampled: the hemolymph, which is the most targeted by BVs (Muller et al. 2021), and the fat body, because it is easy to isolate large quantity of this tissue, in turn facilitating DNA extractions. Hemolymph and fat body of 20 parasitized caterpillars were collected after 24 h or 72 h after oviposition (i.e., post-parasitism). Hemolymph was collected from the caterpillar proleg and stocked into 1.5 ml of ATL buffer from the Qiagen DNeasy Blood & Tissue Kit. The fat body was collected, washed with PBS buffer and also stored into 1.5 ml of ATL buffer after caterpillars were dissected and the digestive tract removed. The genomic DNA from the four caterpillar pools was extracted using the DNeasy Blood & Tissue Kit. DNA of the four samples from the parasitized caterpillars *S. frugiperda* was quantified on a Qubit apparatus and was sent to Novogene for Illumina sequencing in a 2×150 bp paired-end mode.

Reference Genomes Used to Characterize Somatic Insertions of *H. didymator* Ichnovirus

The genome of *S. frugiperda* (corn strain) was sequenced by Gimenez et al. (2020) using the long-read PacBio RSII (Pacific Biosciences) technology and assembled with Platanus. It is available on the BIPAA platform (<https://bipaa.genouest.org/is/>; Rennes, France) and under accession number PRJNA662887 in NCBI. This genome is 384.46 Mb, it has an N50 (i.e., minimum contig length that covers 50% of the genome) of 13.15 Mb and it is composed of 125 contigs.

The genome of *H. didymator* was sequenced by Legeai et al. (2020) using the short-read Illumina HiSeq

technology and assembled with Platanus. It is available in the NCBI database under accession number PRJNA 589497. The genome is 226.9 Mb, it has an N50 of 3.3 Mb and it is composed of 131,161 sequences. HdIV regions were annotated by Legeai et al. (2020) using the genome annotation editor Apollo browser. Fifty-seven viral segments localized in 66 viral loci were annotated in 32 scaffolds.

Measuring Sequencing Depth on *H. didymator* and *S. frugiperda* Genomes

The four datasets of Illumina reads we obtained (two parasitized caterpillar tissues at two timepoints post-parasitism; SRA accession numbers provided in table 1) were quality-trimmed with Trimmomatic v.0.38 (LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:36 options) (Bolger et al. 2014) and the quality was evaluated with FastQC v.0.11.8 (Wingett and Andrews 2018). To get statistics on sequencing depth, we aligned the four datasets of trimmed paired-end reads using Bowtie2 v.2.3.4.1 (Langmead and Salzberg 2012) in end-to-end mode on both *S. frugiperda* and *H. didymator* reference genomes. We converted the output SAM files into sorted BAM files with samtools v.1.9. The sequencing depth for each of our samples was estimated with bedtools genomecov v2.26.0 (Quinlan and Hall 2010) on both reference genomes (-ibam and -d options). To get the sequencing depth of *H. didymator* proviral segments, we used bedtools coverage v2.26.0 (-d option) and provided as input a bed file containing the proviral segment coordinates and the bam file of the Illumina reads aligned on the *H. didymator* genome.

Characterizing Integrations of Wasp Ichnovirus DNA Circles in Somatic Host Genomes

To identify integrations of HdIV DNA circles in the *S. frugiperda* somatic genomes, we searched for chimeric reads, that is reads for which a portion aligns exclusively on HdIV and another portion aligns exclusively on the *S. frugiperda* genome. Reverse and forward raw fastq files were converted into fasta files. The resulting fasta files were aligned on the wasp genome with blastn version 2.6.0 (-task megablast -max_target_seqs 1 -outfmt 6). Reads aligning on the wasp genome were extracted from the fasta file using seqtk subseq (<https://github.com/lh3/seqtk>) and aligned with blastn (same options as above) on the *S. frugiperda* genome.

To further find chimeric reads most likely to result from the integration of HdIV into *S. frugiperda* genome, we used the approach described in Muller et al. (2021). Briefly, we used the tabulated outputs of the two successive blastn searches as entries of an R pipeline initially written to identify artificial chimeras generated during deep sequencing library preparation (Peccoud et al. 2018). This script identifies reads as chimeric reads if 1) at least 90% of the read aligns on one or the other genome, 2) at least 16 bp align only on one of the two genomes, 3) with a maximal of 20 bp overlap between the two read portions aligning on

a different genome, and 4) a maximal of random 5 bp insert at the chimeric junction. Then, we estimated the number of independent IEs by counting as one event all reads with identical or nearly identical coordinates in the two genomes. To account for possible sequencing errors and/or alignment differences between reads covering the same HdIV—*S. frugiperda* junction (resulting from the fact that the length of the chimeric read region corresponding to each genome varies between reads), we allowed the coordinate of chimeric reads to vary by 5-bp in the two genomes, that is all reads aligning at the same position ± 5 bp in the two genomes were considered as reflecting the same integration events.

To be able to compare the numbers of integration events between samples, we normalized values by the number of sequenced reads mapping to the *S. frugiperda* genome, as in Muller et al. (2021). We calculated a scaling factor by dividing the number of reads mapped on the *S. frugiperda* genome by 1,000,000. Then, to obtain the number of IEs per million reads mapping on *S. frugiperda* for each HdIV circle, we divided the absolute number of IEs by this scaling factor. We used samtools view (version 1.9, option -c -F 4) to determine the number of reads mapped on the *S. frugiperda* genome.

Searching for Polydnavirus DNA Circles in Lepidopteran Genomes

To search for HIM-mediated HT of PDV circles in lepidopteran genomes available in Genbank, we used the seven HIM-containing HdIV circles and the 16 HIM-containing *C. typhae* BV (CtBV) circles previously described in Muller et al. (2021) as queries to perform similarity searches using blastn (-task blastn) on lepidopteran genomes. All HdIV and CtBV DNA circle sequences are provided in [supplementary Dataset 6, Supplementary Material](#) online and the HIM coordinates within these circles are provided in [supplementary Dataset 7, Supplementary Material](#) online. The sequences of each HIM motifs of HdIV and CtBV circles are also provided as a multiple alignment in [supplementary Dataset 8, Supplementary Material](#) online as well as in [figure 4](#). A total of 844 lepidopteran genomes were available in Genbank as of December 2021. When more than one genome per species was available, we only retained the largest one, resulting in 775 genomes that were submitted to our search. Accession numbers, size, and N50 of these genomes are provided in [supplementary table S6, Supplementary Material](#) online. Blastn outputs were filtered in R (R Core Team) based on hit size, *e*-value, and coordinates. Some PDV circles are similar to each other. For example, Hd12 and Hd16 are similar to each other over most of their length. To avoid counting multiple times the same lepidopteran genome region as resulting from HIM-mediated HT, alignment coordinates of blastn hits were thus merged using bedtools v2.26.0 merge (-s -c 4 -o distinct) (Quinlan and Hall 2010). Merging was performed independently for IV-like and BV-like sequences. The script written for this part of the study is available on GitHub: <https://github.com/>

HeloiseMuller/Chimera/tree/master/scriptsArticles/Heisserer 2022. DRJ, recombined DRJ (also called “DRJ circle”) and HIM motifs were searched manually or by blastn in lepidopteran PDV insertions using DRJ and HIM motifs previously annotated in CcBV and HdIV as queries (Legeai et al. 2020; Gauthier et al. 2021).

Sequencing Depth Surrounding Junctions Between PDV DNA Circles and Lepidopteran Genomes

To compute sequencing depths on junctions between PDV DNA circles and lepidopteran genomes, raw Illumina reads were downloaded from Genbank for nine species in which more than four junctions were found. Accession numbers of these reads are provided in [supplementary table S7, Supplementary Material](#) online. Reads were mapped on the genome of the nine species using bowtie2 (default options) (Langmead and Salzberg 2012). Read depth over 300 bp upstream and downstream of each junction was obtained using bedtools v2.26.0 coverage (Quinlan and Hall 2010).

Search for Orthologous PDV Fragments in Lepidopteran Genomes

We extracted all PDV fragments ending in the HIM together with 1,000 bp flanking the J1 or J2 motif of the HIM. We filtered out fragments showing similarity to PDV sequences over >100 bp in their flank, which may result from integration of PDV circles next to (or within) each other. We then used blastn to align all PDV fragments plus flanks on themselves. Self blastn hits as well as hits involving sequences corresponding to different circles were filtered out. We retained blastn hits involving the PDV fragments plus at least 200 bp of flanking regions in both sequences as candidate orthologous sequences. These candidate orthologous sequences were then all submitted to manual inspection to only retain PDV sequences integrated at the very same position in two species.

PCR Verifications

To experimentally verified that PDV are present in natural populations of butterflies, 12 individuals of *Pieris napi* and 12 individuals of *P. rapae* were collected in 2018 in three sites in France: S1, Pénerf (Morbihan) 47.511463, -2,622753; S2, Vernou-sur Brenne (Indre-et-Loire) 47.415418 0.858978; S3, Tours (Indre-et-Loire) 47.355069 0,702746. PDV junctions were identified in the genome of *P. napi* and *P. rapae* using similarity searches (blastn). This study was performed in 2018, before the large-scale bioinformatic study reported in this article. It was performed on two different assemblies of the *P. napi* genome (CAJQFU010000000 and DWAF000000000) whereas the large-scale bioinformatic study only included one (DWAF000000000). Although ichnovirus circles used as queries for the blastn searches were those of *H. didymator* (the same as those used in the large-scale bioinformatic study), bracovirus circles were those of *C. congregata* BV (instead of CtBV in the large-scale bioinformatic study).

This explains why only one junction (out of six) verified by PCR was also found in the large-scale bioinformatic search. PCR primers were designed on each side of each PDV–butterfly junction. DNA was extracted from all 24 butterfly and a standard PCR protocol was used to screen for the presence/absence of six junctions in these individuals. Junctions found in at least two individuals were Sanger-sequenced and are provided in [supplementary Dataset 2, Supplementary Material](#) online, together with the genomic coordinates of each junction in the genomes of *P. napi* and *P. rapae*.

Phylogenetic Analyses

Sequences homologous to PDV DNA circles and starting or ending within HIM were extracted from lepidopteran genomes using seqtk subseq (<https://github.com/lh3/seqtk>) and are provided in [supplementary tables S1 and S2, Supplementary Material](#) online. These sequences were then aligned, together with homologous DNA circles from various parasitoid wasps (*D. semiclausum*, *H. didymator*, and *H. fugitivus* for IV and *Cotesia sesamiae*, *C. vestalis* and *C. congregata* for BV) using MUSCLE (Edgar 2004). Given the large number of IV junctions found in some species, we clustered these sequences for all species in which more than five sequences were found for a given junction using Trimal (-maxidentity 0.95) (Capella-Gutierrez et al. 2009). Alignments were then trimmed with Trimal (-automated1) and submitted to Neighbor-joining analysis in MEGA 11 (Maximum Composite Likelihood model of substitution, uniform rates among sites and lineages, pairwise deletion) (Tamura et al. 2021). Bootstrap values were obtained through 100 replicates of the analysis as implemented in the MEGA software.

Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

Acknowledgment

This work was supported by Agence Nationale de la Recherche, project ANR-18-CE02-0021-01 TranspHorizon.

Data Availability

All data produced during this study are available in public repositories. Genbank accession numbers of raw reads produced from *S. frugiperda* larvae parasitized by *H. didymator* are provided in [table 1](#). Scripts are provided in GitHub: <https://github.com/HeloiseMuller/Chimera/tree/master/scriptsArticles/Heisserer2022>. All wasp PDV circles and all lepidopteran PDV fragments bordered by HIM are provided in [supplementary Datasets 6, Supplementary Material](#) online and [supplementary tables S1 and S2, Supplementary Material](#) online. Accession numbers of lepidopteran genomes surveyed in this study are provided in [supplementary table S7, Supplementary Material](#) online.

References

- Acuna R, Padilla BE, Florez-Ramos CP, Rubio JD, Herrera JC, Benavides P, Lee SJ, Yeats TH, Egan AN, Doyle JJ, et al. 2012. Adaptive horizontal transfer of a bacterial gene to an invasive pest of coffee. *Proc Natl Acad Sci U S A*. **109**:4197–4202.
- Beck MH, Inman RB, Strand MR. 2007. Microplitis demolitor bracovirus genome segments vary in abundance and are individually packaged in virions. *Virology* **359**:179–189.
- Beck MH, Zhang S, Bitra K, Burke GR, Strand MR. 2011. The encapsidated genome of Microplitis demolitor bracovirus integrates into the host pseudoplusia inclusions. *J Virol*. **85**:11685–11696.
- Beckage NE, Drezen J-M. 2012. *Parasitoid viruses: symbionts and pathogens*. 1st ed. London; New York: Elsevier/Academic Press.
- Beckage NE, Gelman DB. 2004. W ASP PARASITOID DISRUPTION OF HOST DEVELOPMENT: implications for new biologically based strategies for insect control. *Annu Rev Entomol*. **49**: 299–330.
- 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**:8909–8921.
- Benoist R, Chantre C, Capdevielle-Dulac C, Bodet M, Mougél F, Calatayud PA, Dupas S, Huguet E, Jeannette R, Obonyo J, et al. 2017. Relationship between oviposition, virulence gene expression and parasitism success in *Cotesia typhae* nov. sp. parasitoid strains. *Genetica* **145**:469–479.
- Bézier A, Herbinère J, Lanzrein B, Drezen J-M. 2009. Polydnavirus hidden face: the genes producing virus particles of parasitic wasps. *J Invertebr Pathol*. **101**:194–203.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120.
- Bourguet D, Ponsard S, Streiff R, Meusnier S, Audiot P, Li J, Wang Z-Y. 2014. ‘Becoming a species by becoming a pest’ or how two maize pests of the genus *Ostrinia* possibly evolved through parallel ecological speciation events. *Mol Ecol*. **23**:325–342.
- Burke GR. 2019. Common themes in three independently derived endogenous nudivirus elements in parasitoid wasps. *Curr Opin Insect Sci*. **32**:28–35.
- Burke GR, Hines HM, Sharanowski BJ. 2021. The presence of ancient core genes reveals endogenization from diverse viral ancestors in parasitoid wasps. *Genome Biol Evol*. **13**(7):evab105.
- Burke GR, Simmonds TJ, Sharanowski BJ, Geib SM. 2018. Rapid viral symbiogenesis via changes in parasitoid wasp genome architecture. *Mol Biol Evol*. **35**:2463–2474.
- 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**:9511–9523.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973.
- Chevignon G, Periquet G, Gyapay G, Vega-Czarny N, Musset K, Drezen J-M, Huguet E. 2018. *Cotesia congregata* bracovirus circles encoding *PTP* and *Ankyrin* genes integrate into the DNA of parasitized *Manduca sexta* hemocytes. *J Virol*. **92**(15): e00438–18. Available from: <http://jvi.asm.org/lookup/doi/10.1128/JVI.00438-18>
- Chew FS, Watt WB. 2006. The green-veined white (*Pieris napi* L.), its Pierine relatives, and the systematics dilemmas of divergent character sets (Lepidoptera, Pieridae): DIVERGENT CHARACTERS AND DNA PHYLOGENY. *Biol J Linn Soc*. **88**: 413–435.
- Cong Q, Shen J, Li W, Borek D, Otwinowski Z, Grishin NV. 2017. The first complete genomes of Metalmarks and the classification of butterfly families. *Genomics* **109**:485–493.
- Cummings TFM, Gori K, Sanchez-Pulido L, Gavriilidis G, Moi D, Wilson AR, Murchison E, Dessimoz C, Ponting CP,

- Christophorou MA. 2022. Citrullination was introduced into animals by horizontal gene transfer from Cyanobacteria. *Mol Biol Evol.* **39**:msab317.
- Danchin EG, Guzeeva EA, Mantelin S, Berepiki A, Jones JT. 2016. Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *globodera pallida* to feed on host-derived sucrose. *Mol Biol Evol.* **33**:1571–1579.
- Desjardins CA, Gundersen-Rindal DE, Hostetler JB, Tallon LJ, Fadrosch DW, Fuester RW, Pedroni MJ, Haas BJ, Schatz MC, Jones KM, et al. 2008. Comparative genomics of mutualistic viruses of *Glyptapanteles* parasitic wasps. *Genome Biol.* **9**:R183.
- Di Giovanni D, Lepetit D, Guinet B, Bennetot B, Boulesteix M, Couté Y, Bouchez O, Ravallec M, Varaldi J. 2020. A behavior-manipulating virus relative as a source of adaptive genes for *Drosophila* parasitoids. *Mol Biol Evol.* **37**:2791–2807.
- 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**:e1007998.
- Drezen JM, Periquet G, Savary S, Tan F, Beckage N. 1997. Excision of the polydnavirus chromosomal integrated EP1 sequence of the parasitoid wasp *Cotesia congregata* (Braconidae, Microgastinae) at potential recombinase binding sites. *J Gen Virol.* **78**:3125–3134.
- Dupuy C, Periquet G, Serbielle C, Bézier A, Louis F, Drezen J-M. 2011. Transfer of a chromosomal maverick to endogenous bracovirus in a parasitoid wasp. *Genetica* **139**:489–496.
- Edgar RC. 2004. [No title found]. *BMC Bioinformatics* **5**:113.
- Falabella P, Varricchio P, Provost B, Espagne E, Ferrarese R, Grimaldi A, de Eguileor M, Fimiani G, Ursini MV, Malva C, et al. 2007. Characterization of the IκB-like gene family in polydnaviruses associated with wasps belonging to different Braconid subfamilies. *J Gen Virol.* **88**:92–104.
- Frayssinet M, Audiot P, Cusumano A, Pichon A, Malm LE, Jouan V, Vabre M, Malavieille S, Delalande M, Vargas-Osuna E, et al. 2019. Western European populations of the ichneumonid wasp *hyposoter didymator* belong to a single taxon. *Front Ecol Evol.* **7**:20.
- Gasmi L, Boulain H, Gauthier J, Hua-Van A, Musset K, Jakubowska AK, Aury JM, Volkoff AN, Hugué E, Herrero S, et al. 2015. Recurrent domestication by lepidoptera of genes from their parasites mediated by bracoviruses. *PLoS Genet.* **11**:e1005470.
- Gasmi L, Sieminska E, Okuno S, Ohta R, Coutu C, Vatanparast M, Harris S, Baldwin D, Hegedus DD, Theilmann DA, et al. 2021. Horizontally transmitted parasitoid killing factor shapes insect defense to parasitoids. *Science* **373**:535–541.
- Gauld ID. 1988. Evolutionary patterns of host utilization by ichneumonoid parasitoids (Hymenoptera: Ichneumonidae and Braconidae). *Biol J Linn Soc.* **35**:351–377.
- Gauthier J, Boulain H, van Vugt JJFA, Baudry L, Persyn E, Aury J-M, Noel B, Bretaudeau A, Legeai F, Warris S, et al. 2021. Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. *Commun Biol.* **4**:104.
- Gilbert C, Belliardo C. 2022. The diversity of endogenous viral elements in insects. *Curr Opin Insect Sci.* **49**:48–55.
- Gimenez S, Abdelgaffar H, Goff GL, Hilliou F, Blanco CA, Hänniger S, Bretaudeau A, Legeai F, Nègre N, Jurat-Fuentes JL, et al. 2020. Adaptation by copy number variation increases insecticide resistance in the fall armyworm. *Commun Biol.* **3**:664.
- Gundersen-Rindal DE, Lynn DE. 2003. Polydnavirus integration in lepidopteran host cells in vitro. *J Insect Physiol.* **49**:453–462.
- Heringer P, Kuhn GCS. 2022. Multiple horizontal transfers of a Helitron transposon associated with a parasitoid wasp. *Mob DNA.* **13**:20.
- Herniou EA, Hugué E, Theze J, Bézier A, Periquet G, Drezen JM. 2013. When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. *Philos Trans R Soc Lond B Biol Sci.* **368**:20130051.
- Hickman AB, Voth AR, Ewis H, Li X, Craig NL, Dyda F. 2018. Structural insights into the mechanism of double strand break formation by Hermes, a hAT family eukaryotic DNA transposase. *Nucleic Acids Res.* **46**(19):10286–10301. Available from: <https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky838/5103951>
- Husnik F, McCutcheon JP. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol.* **16**:67–79.
- Jiang H, Vilcinskis A, Kanost MR. 2010. Immunity in lepidopteran insects. In: Söderhäll K, editor. *Invertebrate immunity*. Vol. 708. Advances in Experimental Medicine and Biology. Boston, MA: Springer. p. 181–204. Available from: http://link.springer.com/10.1007/978-1-4419-8059-5_10
- Kawahara AY, Plotkin D, Espeland M, Meusemann K, Toussaint EFA, Donath A, Gimnich F, Frandsen PB, Zwick A, dos Reis M, et al. 2019. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc Natl Acad Sci U S A.* **116**:22657–22663.
- Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* **34**:1812–1819.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* **9**:357–359.
- Lavine MD, Strand MR. 2002. Insect hemocytes and their role in immunity. *Insect Biochem Mol Biol.* **32**:1295–1309.
- Leclercq S, Thézé 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**:15036–15041.
- Legeai F, Santos BF, Robin S, Bretaudeau A, Dikow RB, Lemaitre C, Jouan V, Ravallec M, Drezen J-M, Tagu D, et al. 2020. Genomic architecture of endogenous ichnoviruses reveals distinct evolutionary pathways leading to virus domestication in parasitic wasps. *BMC Biol.* **18**:89.
- Li Y, Liu Z, Liu C, Shi Z, Pang L, Chen C, Chen Y, Pan R, Zhou W, Chen X, et al. 2022. HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell* **185**(16):2975–2987.
- Mao M, Strand MR, Burke GR. 2022. The complete genome of *Chelonus insularis* reveals dynamic arrangement of genome components in parasitoid wasps that produce bracoviruses. *J Virol.* **96**:e01573-21.
- Martin WF. 2017. Too much eukaryote LGT. *BioEssays* **39**:1700115.
- McKelvey TA, Lynn DE, Gundersen-Rindal D, Guzo D, Stoltz DA, Guthrie KP, Taylor PB, Dougherty EM. 1996. Transformation of gypsy moth (*Lymantria dispar*) cell lines by infection with glyptapanteles indiensis polydnavirus. *Biochem Biophys Res Commun.* **225**:764–770.
- Moran NA, Jarvik T. 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**:624–627.
- Muller H, Chebbi MA, Bouzar C, Périquet G, Fortuna T, Calatayud P-A, Le Ru B, Obonyo J, Kaiser L, Drezen J-M, et al. 2021. Genome-wide patterns of bracovirus chromosomal integration into multiple host tissues during parasitism. *J Virol.* **95**:e00684-21.
- 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**:378–395.
- Peccoud J, Lequime S, Moltini-Conclois I, Giraud I, Lambrechts L, Gilbert C. 2018. A survey of virus recombination uncovers canonical features of artificial chimeras generated during deep sequencing library preparation. *G3* **8**:1129–1138.
- Peccoud J, Loiseau V, Cordaux R, Gilbert C. 2017. Massive horizontal transfer of transposable elements in insects. *Proc Natl Acad Sci U S A.* **114**:4721–4726.
- Pichon A, Bézier A, Urbach S, Aury J-M, Jouan V, Ravallec M, Guy J, Cousserans F, Thézé J, Gauthier J, et al. 2015. Recurrent DNA virus

- domestication leading to different parasite virulence strategies. *Sci Adv.* **1**:e1501150.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842.
- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Robin S, Ravallec M, Frayssinet M, Whitfield J, Jouan V, Legeai F, Volkoff A-N. 2019. Evidence for an ichnovirus machinery in parasitoids of coleopteran larvae. *Virus Res.* **263**:189–206.
- Salzberg SL. 2017. Horizontal gene transfer is not a hallmark of the human genome. *Genome Biol.* **18**:85.
- Santos BF, Klopstein S, Whitfield JB, Sharanowski BJ. 2022. Many evolutionary roads led to virus domestication in ichneumonoid parasitoid wasps. *Curr Opin Insect Sci.* **50**:100861.
- Schneider SE, Thomas JH. 2014. Accidental genetic engineers: horizontal sequence transfer from parasitoid wasps to their lepidopteran hosts. *PLoS ONE* **9**:e109446.
- Sibbald SJ, Eme L, Archibald JM, Roger AJ. 2020. Lateral gene transfer mechanisms and pan-genomes in eukaryotes. *Trends Parasitol.* **36**:927–941.
- Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nat Rev: Genet.* **16**:472–482.
- Stoltz DB, Krell P, Summers MD, Vinson B. 1984. Polydnaviridae—a proposed family of insect viruses with segmented, double-stranded, circular DNA genomes. *Intervirology* **21**:1–4.
- Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. battistuzzi FU, editor. *Mol Biol Evol.* **38**:3022–3027.
- Van Etten J, Bhattacharya D. 2020. Horizontal gene transfer in eukaryotes: not if, but how much? *Trends Genet.* **36**:915–925.
- Volkoff A-N, Cusson M. 2020. The unconventional viruses of ichneumonid parasitoid wasps. *Viruses* **12**:1170.
- Volkoff A-N, Huguet E. 2021. Polydnaviruses (polydnaviridae), editors. *Encyclopedia of virology*: San Diego: Elsevier. p. 849–857.
- Available from: <https://linkinghub.elsevier.com/retrieve/pii/S09780128096338215562>
- Volkoff A-N, Jouan V, Urbach S, Samain S, Bergoin M, Wincker P, Demetree 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**:e1000923.
- Volkoff A-N, Rocher J, Cérutti P, Ohresser MCP, d'Aubenton-Carafa Y, Devauchelle G, Duonor-Cérutti M. 2001. Persistent expression of a newly characterized Hyposoter didymator polydnavirus gene in long-term infected lepidopteran cell lines. *J Gen Virol.* **82**:963–969.
- Wang Z, Ye X, Zhou Y, Wu X, Hu R, Zhu J, Chen T, Huguet E, Shi M, Drezen J-M, et al. 2021a. Bracoviruses recruit host integrases for their integration into caterpillar's genome. *PLoS Genet.* **17**:e1009751.
- Wang Z-H, Zhou Y, Yang J, Ye X, Shi M, Huang J, Chen X. 2021b. Genome-wide profiling of diadegma semiclausum ichnovirus integration in parasitized plutella xylostella hemocytes identifies host integration motifs and insertion sites. *Front Microbiol.* **11**:608346.
- Wingett SW, Andrews S. 2018. FastQ Screen: a tool for multi-genome mapping and quality control. *F1000Res.* **7**:1338.
- Wybouw N, Dermauw W, Tirry L, Stevens C, Grbic M, Feyereisen R, Van Leeuwen T. 2014. A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. *Elife* **3**:e02365.
- Wybouw N, Pauchet Y, Heckel DG, Van Leeuwen T. 2016. Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome Biol Evol.* **8**:1785–1801.
- Xia J, Guo Z, Yang Z, Han H, Wang S, Xu H, Yang X, Yang F, Wu Q, Xie W, et al. 2021. Whitefly hijacks a plant detoxification gene that neutralizes plant toxins. *Cell.* **184**:1693–1705.e17.
- Zhang J, Shen J, Cong Q, Grishin NV. 2019. Genomic analysis of the tribe emesidini (lepidoptera: riodinidae). *Zootaxa.* **4668**:475–488.