# Cotesia congregata Bracovirus Circles Encoding PTP and Ankyrin Genes Integrate into the DNA of Parasitized Manduca sexta Hemocytes

Germain Chevignon, Georges Périquet, Gabor Gyapay, Nathalie Vega-Czarny, Karine Musset, Jean-Michel Drezen, Elisabeth Huguet

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

VIRUS-CELL INTERACTIONS

# Cotesia congregata Bracovirus Circles Encoding *PTP* and *Ankyrin* Genes Integrate into the DNA of Parasitized *Manduca sexta* Hemocytes

**Germain Chevignon,[a]* Georges Periquet,[a] Gabor Gyapay,[b] Nathalie Vega-Czarny,[b] Karine Musset,[a] Jean-Michel Drezen,[a] Elisabeth Huguet[a]**

[a]Institut de Recherche sur la Biologie de l'Insecte, CNRS UMR 7261, Université de Tours, Tours, France

[b]Commissariat à l'Energie Atomique et aux Energies Alternatives, Genoscope (Centre National de Séquençage), Evry, France

**ABSTRACT** Polydnaviruses (PDVs) are essential for the parasitism success of tens of thousands of species of parasitoid wasps. PDVs are present in wasp genomes as proviruses, which serve as the template for the production of double-stranded circular viral DNA carrying virulence genes that are injected into lepidopteran hosts. PDV circles do not contain genes coding for particle production, thereby impeding viral replication in caterpillar hosts during parasitism. Here, we investigated the fate of PDV circles of Cotesia congregata bracovirus during parasitism of the tobacco hornworm, *Manduca sexta*, by the wasp *Cotesia congregata*. Sequences sharing similarities with host integration motifs (HIMs) of Microplitis demolitor bracovirus (MdBV) circles involved in integration into DNA could be identified in 12 CcBV circles, which encode *PTP* and *VANK* gene families involved in host immune disruption. A PCR approach performed on a subset of these circles indicated that they persisted in parasitized *M. sexta* hemocytes as linear forms, possibly integrated in host DNA. Furthermore, by using a primer extension capture method based on these HIMs and high-throughput sequencing, we could show that 8 out of 9 circles tested were integrated in *M. sexta* hemocyte genomic DNA and that integration had occurred specifically using the HIM, indicating that an HIM-mediated specific mechanism was involved in their integration. Investigation of BV circle insertion sites at the genome scale revealed that certain genomic regions appeared to be enriched in BV insertions, but no specific *M. sexta* target site could be identified.

**IMPORTANCE** The identification of a specific and efficient integration mechanism shared by several bracovirus species opens the question of its role in braconid parasitoid wasp parasitism success. Indeed, results obtained here show massive integration of bracovirus DNA in somatic immune cells at each parasitism event of a caterpillar host. Given that bracoviruses do not replicate in infected cells, integration of viral sequences in host DNA might allow the production of PTP and VANK virulence proteins within newly dividing cells of caterpillar hosts that continue to develop during parasitism. Furthermore, this integration process could serve as a basis to understand how PDVs mediate the recently identified gene flux between parasitoid wasps and Lepidoptera and the frequency of these horizontal transfer events in nature.

**KEYWORDS** viral symbionts, polydnavirus, virus integration, parasitoid wasps, virulence

Endoparasitoid wasps that live during their larval stage within the body of a developing insect have been shown to finely tune host physiology to ensure parasitism success (1). In the case of parasitoid wasps harboring polydnaviruses (PDVs), this host manipulation may involve direct interactions with the genome of the para-

**Received** 14 March 2018 **Accepted** 4 May 2018

**Accepted manuscript posted online** 16 May 2018

**Citation** 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:e00438-18. https://doi.org/10.1128/JVI.00438-18.

**Editor** Jae U. Jung, University of Southern California



Address correspondence to Elisabeth Huguet, elisabeth.huguet@univ-tours.fr.

* Present address: Germain Chevignon, Department of Entomology, University of Georgia, Athens, Georgia, USA.

sitized host. Indeed, DNA virus sequences produced by the parasite may integrate into the genomic DNA of infected host cells (2).

These endogenous viruses, named polydnaviruses, are double-stranded DNA (dsDNA) viruses obligatorily associated with parasitoid wasps that develop within the body of lepidopteran larvae. These viruses have an original life cycle that is divided in two phases: (i) a replicative phase occurring in specialized cells in the wasp ovaries where the virus particles are produced and (ii) an infective phase that begins by the injection of the virus particles along with the wasp eggs into the host caterpillar. PDV infections of host tissues during parasitism induce regulation of several components of host biology, leading to parasitism success with the emergence of the wasp progeny from the host at the end of larval development (3–5).

PDVs are classified into two genera, ichnoviruses (IV) and bracoviruses (BV), belonging to two and six wasp subfamilies of *Ichneumonidae* and *Braconidae*, respectively. The two PDV genera originated from distinct viruses integrated in the respective wasp subfamily genomes. Following their integration, the endogenous viruses have evolved by convergent evolution and became tools used by parasitoid wasps for transferring virulence genes to the parasitized hosts (6). While the genes involved in ichnovirus production are specific to a new virus family described only for endogenous viruses (7, 8), the genes ensuring bracovirus particle production belong to nudiviruses, a sister group of baculoviruses infecting insects and crustaceans, indicating that the associations of braconid wasps with bracoviruses originated from the ancestral integration of a nudivirus into an ancestor wasp (9–11). These ancestor braconid wasps lived 100 million years ago (9), and their descendants now form a hyperdiversified monophyletic group, named the microgastroid complex, that is estimated to comprise up to 46,000 species based on recent molecular taxonomic studies (12).

BV genome organization has been characterized during the past few years. These genomes are composed of two components both residing in the wasp genomes. The first component corresponds to genes of nudiviral origin coding for proteins involved in particle production. The second component is composed of proviral segments used to produce circular dsDNA molecules (here named bracovirus circles) that are packaged in viral particles and contain genes coding for proteins involved in host physiology alterations (13, 14).

In particular, we have shown that the *Cotesia congregata* wasp genome contains 35 proviral segments organized in clusters that are amplified in the wasp ovaries within 12 replication units (RU) (15). Circles are produced from the amplified molecules by a complex process that has not been completely elucidated yet (5) and are packaged into viral particles that are injected in *Manduca sexta* larvae. BV proviral segments are delineated in the wasp genome by short direct repeats (16) that have been named either DRJs, for direct repeat junctions (17), or WIMs, for wasp integration motifs (2). The core motif of these sequences contains a tetramer (AGCT) that allows the circularization of the DNAs packaged into viral particles, which are injected into the lepidopteran host along with the wasp eggs. The genes carried by bracovirus circles are virulence genes that are expressed after their injection in the host, leading to suppression of the immune encapsulation response and effects on host development (18–20). Cotesia congregata bracovirus (CcBV) circles altogether code for 229 genes, among which 88 were shown to be expressed in *M. sexta* larva 24 h after parasitism in hemocytes and/or in fat body (18).

The bracovirus circles do not contain genes coding for replication and particle production, thereby impeding particle production in the caterpillar host during parasitism, which lasts more than 10 days. The fact that bracoviruses do not replicate in the caterpillar host therefore raised the questions of whether and how they persist in the host. Several lines of evidence indicate that at least part of these bracoviruses persist in insect cells, suggesting they play a role throughout the length of parasitism. Indeed, PDV DNA was shown to persist in cell cultures (2, 21–24) and was shown to be present as chromosomally integrated forms in host-derived cultured cells (2, 25–27). *In vivo*, persistent expression of certain genes of Glypapanteles indiensis bracovirus (GiBV),

CcBV, and Chelonus inanitus bracovirus (CiBV) within their lepidopteran hosts also suggested bracoviruses persist in their hosts after injection by the wasps, in some cases as chromosomally integrated forms (28, 29). Finally and most convincingly, two Microplitis demolitor bracovirus (MdBV) circles were formally demonstrated to persist as integrated forms in host caterpillar hemocyte genomic DNA after natural parasitization (2). Taken together, these data suggest virus circle integration in the DNA of parasitized host cells are a conserved feature of the bracovirus life cycle.

The study performed on MdBV identified a motif that mediates the insertion of viral circles into the genome of the lepidoteran host *Pseudoplusia includens* (2). This sequence was named HIM, for host integration motif (2), and one HIM (or, rarely, two) was identified in almost all of the MdBV circles (14). The HIM and DRJ are generally located in close proximity (109 to 168 bp); however, sometimes they are located further apart, for example, they are separated by 800 bp in MdBV circles C and G and by 7 kb in circle O. During integration the MdBV circles are opened within the HIM, a short 40-bp to 53-bp sequence of the circle is lost, and the remaining HIM sequences flanking this lost sequence become the extremities of the viral sequences integrated in the host DNA. For clarity, these viral extremities here are named J1 and J2 (for junctions 1 and 2; see Fig. 1).

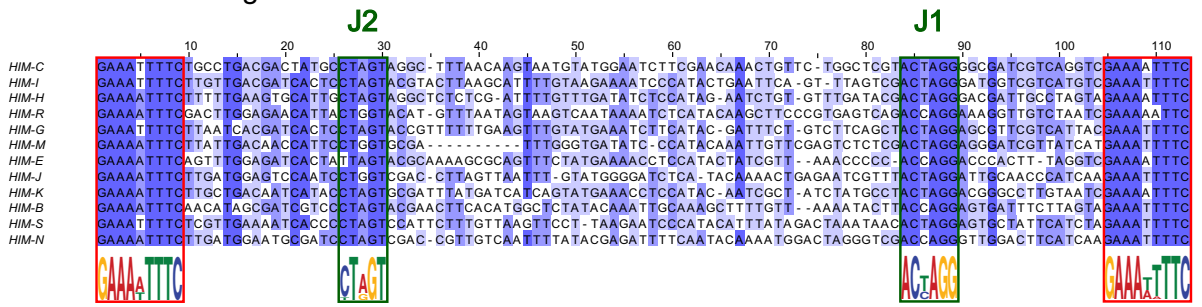
To date, the presence of HIMs has been reported only in BVs associated with wasps of the genus *Microplitis* (2). However, studies on Cotesia sesamiae BV genomic sequences (30) have strongly suggested that BV circle reintegration occasionally occurs in the wasp genome involving a motif resembling HIMs. Indeed, comparisons of 2 circles (CvBV27 and CvBVS2) and related viral sequences in the genomes of *Cotesia sesamiae* indicated that in addition to the proviral form, three additional viral forms, having a different organization and that are not fixed in the species, may be found (or not) in genomes of wasps depending on their geographic origin. These reintegrated forms contained an almost complete circle sequence but lacked a short stretch of nucleotides (40 to 53 bp), as observed following MdBV circle integration. Moreover, the extremities of the virus sequences in the wasp DNA (30) were similar to the ones reported for MdBV sequences integrated into host DNA, suggesting that their integration had involved an HIM-mediated mechanism leading to the production of an integrated form having J1 and J2 extremities.

In the present study, we have identified putative HIMs in 12 CcBV circles (out of 35). Notably, all of these circles (except one, circle 12) encode the well-described PTP and VANK gene families involved in lepidopteran host physiology disruption (31–34). Furthermore, we have shown, by a PCR approach assessing the presence of different viral forms on 4 of these circles, that they persist in naturally parasitized *M. sexta* host hemocytes as linear, probably integrated forms. We then analyzed the junction sites of 9 circles integrated within *M. sexta* hemocyte genomic DNA using integration site capture and high-throughput sequencing (35), thus experimentally demonstrating that HIMs were actually involved in the integration process of 8 circles. Moreover, the recent release of the *M. sexta* genome (36) enabled us to study, for the first time, the localization of bracoviral insertions in host DNA on a large genomic scale. The identification of HIMs in several bracovirus species opens the question of the role of integration in braconid parasitoid wasp parasitism success and supports the hypothesis that the same mechanism is used for BV circle integration, whether in caterpillar hosts (2, 25, 26, and this study) or in wasp genomes (4, 11, 30).

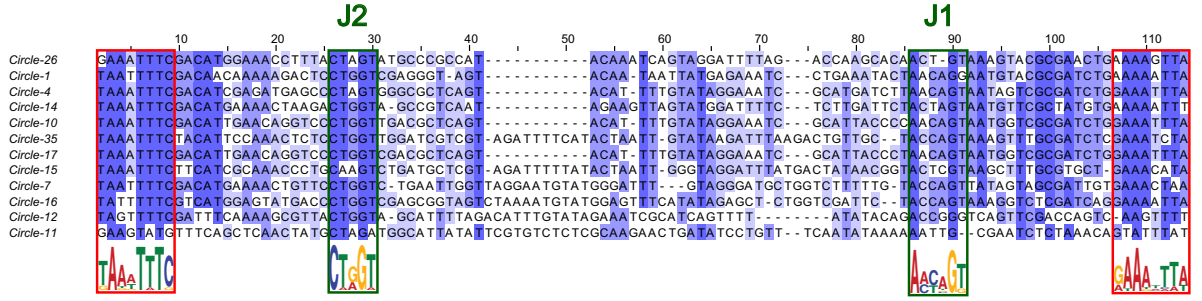
## RESULTS

**CcBV circles display sequences similar to HIMs previously identified for MdBV.** We investigated whether HIMs could be identified in CcBV circles. BLAST analyses against the NCBI database using MdBV HIM and *C. sesamiae* J1 and J2 sequences led to the identification of a conserved region present on 12 CcBV circles. These sequences correspond to a 100- to 110-bp domain consisting of two 30-bp-long imperfect inverted repeats flanking a 50-bp sequence (Fig. 1A). They are present as a single copy in each circle and corresponding proviral segment.

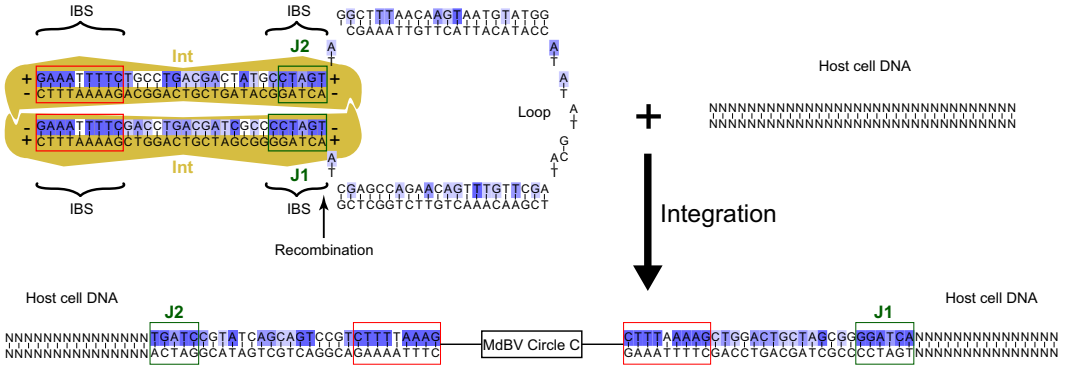
## A. **MdBV** Host Integration Motif




## B. **CcBV** Host Integration Motif




## C. Hypothetical model of interactions between **MdBV** circle-C HIM and integrase involved in circle integration into host DNA




## D. Hypothetical model of interactions between **CcBV** circle-C4 HIM and integrase involved in circle integration into host DNA

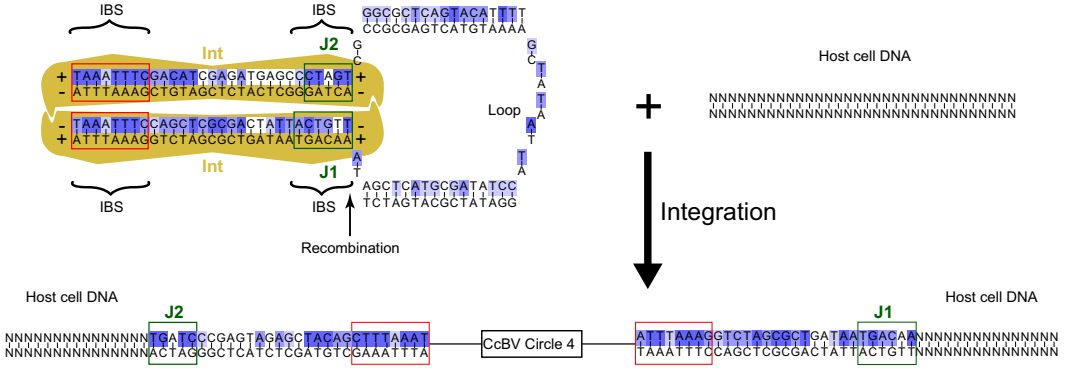



**FIG 1** Alignment of host integration motifs (HIMs) of 12 MdBV circles and of CcBV homologous sequences. Alignment of HIM sequences with similarity for each site colored in shades of blue for MdBV (A) and CcBV (B). All these sequences show boundary sequences forming indirect repeats of 9 nucleotides in MdBV and 8 nucleotides in CcBV. In MdBV HIMs, the sites ACtAGG, forming the J1 border of the insertion with the

(Continued on next page)

CcBV HIM sequences appeared to be fairly similar in their structure to MdBV HIMs, although CcBV HIMs are more divergent from one another (Fig. 1A and B). In CcBV, as in MdBV, a high conservation is found between the first and last 30 nucleotides (nt) of the motifs. In particular, both CcBV and MdBV HIMs share similar boundary sequences consisting of palindromic sequences of 9 bp (GAAAaTTTC and GAAAtTTTC) in MdBV and 8 bp (TAaaTTTC and gAAA.TTa) in CcBV (uppercase letters indicate bases conserved among 80% of different HIMs; see the legend to Fig. 1). These nucleotides and the whole 30-bp repeats were predicted to base pair to form a stem-loop structure for some MdBV HIMs (2). Using Mfold 3.9 software (37), stem-loop structures could be predicted for HIMs of five CcBV circles (C4, C10, C14, C16, and C35) with variable loop regions. However, the biological relevance of these structures is uncertain, since the viral circular DNA is double stranded. Moreover, based on known integration mechanisms (38, 39), these palindromic sequences more probably correspond to similar protein-binding sites on each of the 30-bp repeats, resulting in the assembly of a nucleoprotein complex (see a proposed model of interaction between HIM and integrase monomers in Fig. 1C and D). Indeed, so-called synaptic complexes allowing integration of mobile elements or viruses into targeted DNA often involve homodimers or heterodimers of integrase/transposase proteins interacting with binding sites present in repeats that interact with dimers of integrases/transposases present on the target sites to form tetramers (40). In contrast to the inverted repeats, the internal sequence of about 50 bp is much less well conserved. This is probably related to the fact that this sequence is deleted after insertion into host DNA, as found for MdBV (2), and might be used only to confer flexibility to the DNA, allowing the two repeats of the HIM to come into contact.

In MdBV, the borders of this deleted sequence, J1 (ACtAGG) and J2 (CTaGT), correspond to the extremities of the viral insertions with the lepidopteran host genome. In CcBV circles, conserved candidate J1 (AccaGT) and J2 (CTgGT) sites were identified that are similar to those at the extremities of reintegrated circles in *C. sesamiae* (30) (Fig. 1A and B). In the same way as that for MdBV HIMs, CcBV HIM putative sequences were generally located in the vicinity (144 to 822 bp) of the DRJ in the majority of circles. However, these motifs were 1,934, 3,227, and 4,224 bp distant from the DRJ in three circles, C11, C15, and C35, respectively.

J1 and J2 sequences had previously been identified in bracoviruses of other *Cotesia* species and the closely related *Glytapanteles* wasp genus (3, 11, 30). Here, we show that HIM-like sequences sharing at least 75% identity with those of *C. congregata* could be identified in 6, 12, 11, and 9 bracovirus circles of *C. sesamiae*, *C. vestalis*, *Glyptapanteles indiaensis*, and *G. flavicoxis*, respectively (data not shown). Notably, HIM-like sequences were confirmed in *C. sesamiae* circles highly similar to CvBV27 and CvBVS2 and GiBV segment F (also known as circle GiBV25), corresponding to reintegrated circles in the genomes of two *C. sesamiae* wasp populations (30) and to an integrated circle in *Lymantria dispar* lepidopteran cell cultures (25, 26). The reintegrated forms of CvBV27-like and CvBVS2-like circles present HIM sequences that are 85% identical to those of CcBV circles C10/C17 and C26, respectively. However, no homologous HIMs could be detected in BV circles associated with *Chelonus inanitus*, a wasp belonging to another subfamily of the microgastroid complex, the Cheloninae, which diverged 85 Mya (41).

**FIG 1** Legend (Continued)
host genome, and CTaGT, forming the J2 border, are shown. For CcBV HIMs, similar conserved sites AccaGT and CTgGT were identified. (C and D) Model of integrase binding on HIM sites. HIM sites are composed of 29 (CcBV) or 30 (MdBV) nucleotide palindromic extremities separated by a stretch of 40 to 50 nucleotides (depending on the circle) that is lost during integration. This organization is compatible with a model in which a monomer of an integrase protein (here named Int) binds to each extremity. The two monomers then dimerize to form a synaptic complex, allowing J1 and J2 to come into contact. During circle integration into host DNA, one of the strand exchanges results in the loss of the loop. The bases conserved in MdBV or CcBV HIMs are shown in dark or light blue (bases conserved among 80% and 60% of the different HIMs, respectively). Nucleotides interacting directly with the integrase (IBS, integration binding site) are likely to be the ones that are the most conserved between circles. (C) Note that in MdBV HIMs most of the bases conserved in the J1 plus strand correspond to those conserved in the J2 minus strand, which is compatible with the binding of monomers from the same protein on both extremities. (D) Several positions are conserved between CcBV and MdBV HIMs, which suggests a common origin of the integration mechanism; however, CcBV extremities are more divergent from one another, which questions whether they still bind a single protein or two different monomers.

This suggests that HIM sequences have either diverged to the extent that they are difficult to identify by similarity or that the ability of circles to integrate has been lost in this lineage or was not inherited from the ancestral integrated nudivirus but acquired later specifically by the Microgastrinae (the subfamily to which *Cotesia*, *Glyptapanteles*, and *Microplitis* wasp species belong).

Altogether, the motifs found in the 12 CcBV circles appeared to be good HIM candidates. Before testing whether they were actually implicated in insertion of CcBV circles in *M. sexta* DNA, we first determined whether the sequences of circles containing putative HIMs could persist in *M. sexta* hemocytes during parasitism either as circular DNAs or as the linear form predicted after their insertion into host DNA.

**CcBV circles are present in different circular and linear forms in *M. sexta* hemocytes during parasitism.** To determine whether CcBV sequences persisted in *M. sexta* hemocytes during parasitism, primer pairs were designed for four of the circles in which a candidate HIM could be detected (circles C4, C10, C14, and C26) (see Table S1 in the supplemental material). For each circle, primers allowed amplification of a region in specific genes present on the circles (primer pair *a*), a region that includes only the DRJ motif (primer pair *b*), and a region that spans the HIM, including the DRJ (primer pair *c*) (Fig. 2A). Significant results are shown in Fig. 2C for circles C10 and C26, and the results obtained for circles C4 and C14 were similar (data not shown).

In female wasps that produce particles containing bracovirus circles, primer pairs *b* and *c* allowed amplification of products corresponding to the viral circular forms, while primer pair *a* allowed amplification of a gene-containing region common to both the circular and the integrated form (Fig. 2B and C), as expected.

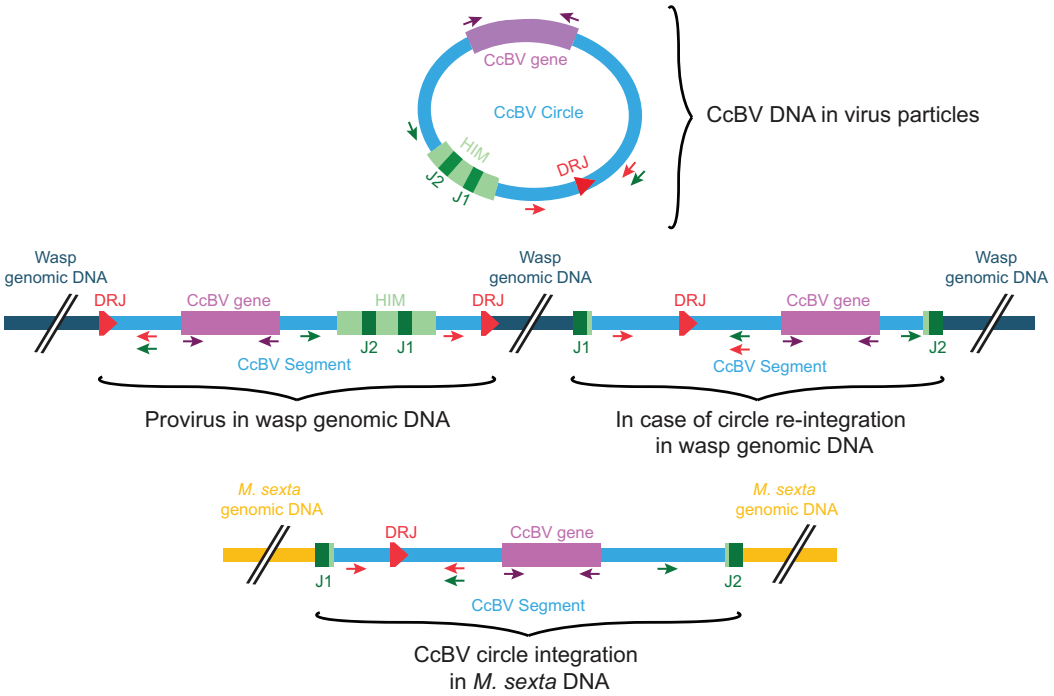
In male wasps that do not produce bracovirus particles, only primer pair *a* allowed amplification of a PCR product, corresponding to the amplification of the integrated form. Lack of amplification was expected with the third set of primers because there is no detectable production of viral circles in males. A PCR product could be obtained in theory with primer pair *b*, spanning the DRJ in case of a circle reintegration event in the DNA of the wasp (Fig. 2B and C), as has been described in *C. sesamiae* (30), but no amplification was obtained with *C. congregata* males, indicating that no reintegrated forms of the circles tested were present in the wasp genome.

No amplification products were obtained when reactions were performed on hemocyte genomic DNA extracted from nonparasitized *M. sexta*, whereas all three primer pairs amplified products from genomic DNA extracted from *M. sexta* organisms parasitized for 24 h. However, primer pair *c*, spanning the HIM, consistently led to a barely detectable level of amplification compared to the other pairs, a result consistently obtained using 5 caterpillars, while the PCR products obtained using female wasp DNA was abundant with the three primer pairs. Strikingly, this amplification was no longer detected in caterpillars parasitized for 12 days (8 caterpillars tested). This result is expected if a large proportion of injected circles are integrated in *M. sexta* hemocyte genomic DNA using the HIM. Indeed, in the linear form (Fig. 2A) produced after circle integration, pair *c* primers are no longer in an orientation enabling PCR amplification. However, it should be noted that PCR amplification with primer pair *c* could be obtained using a more processive DNA polymerase (GoldStar; Eurogentec), indicating that some sequences remain as a circular form (data not shown).

Taken together, these results indicate that sequences corresponding to the four circles tested persisted in *M. sexta* hemocytes during the whole course of parasitism and that the CcBV circles tested are likely to integrate into the DNA of host hemocytes early during parasitism in such a way that the quantity of circles might drop below the threshold allowing PCR amplification. This led us to analyze more precisely the characteristics of circle integration sites by using a high-throughput approach (Fig. 3).

**CcBV circles are inserted into *M. sexta* genome during parasitism.** An adaptation of the primer extension capture (PEC) method (35) was used to capture and sequence CcBV DNA integrated into the genome of *M. sexta* hemocytes using oligonucleotides located in

## A. Circular and Integrated CcBV form




## B. Expected PCR amplification

| Insect DNA | CcBV form | PCR amplification *a* CcBV gene | *b* DRJ region spanning DRJ motif | *c* J1 / J2 region spanning DRJ and HIM motif |
|---|---|---|---|---|
| *C. congregata* female wasp | Circular + Proviral + Re-integrated | Yes | Yes | Yes |
| *C. congregata* male wasp | Proviral | Yes | No | No |
| | Re-integrated | Yes | Yes | No |
| *M. sexta* NP | No virus | No | No | No |
| *M. sexta* P 24h | Integrated only | Yes | Yes | No |
| | Circular | Yes | Yes | Yes |
| *M. sexta* P 12 days | Circular | Yes | Yes | Yes |
| | Integrated only | Yes | Yes | No |
| | Absent | No | No | No |

## C. Representative results of PCR amplifications for two CcBV segments

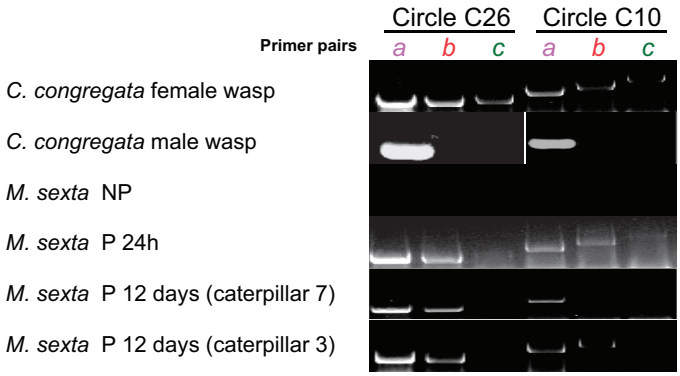



**FIG 2** PCR-based detection method of CcBV circular or linear forms. Different forms of CcBV DNA are expected. CcBV circles containing one direct repeat junction (DRJ) and an HIM are produced in female wasps and encapsidated in viral particles that are injected into the host. (A) These circles originate from a proviral segment integrated in the wasp genome that is flanked

(Continued on next page)

the vicinity of J1 and J2 sequences deduced from CcBV HIM candidates. The approach aimed to identify and study junction regions between lepidopteran and viral DNA.

From the roughly 63.7 million reads obtained by specific DNA-probe capture and Illumina sequencing and by using stringent mapping criteria (see Materials and Methods), approximately 35.3 million reads (55%) could be aligned to the studied regions in 9 CcBV circles (Table S2). The remaining reads corresponded either to low-quality reads, lepidopteran sequences, or nontargeted regions in CcBV circles. However, overall the capture was specific; indeed, using a given probe, only a few thousand reads corresponded to nontargeted circles, whereas several million reads corresponded to targeted circles (Table S2).

The 35.3 million reads were then mapped to the *M. sexta* genome, allowing us to identify 1.7 million chimeric reads containing at least 30 nucleotides of CcBV sequence and 30 nucleotides of *M. sexta* sequence. Out of these chimeric reads, 0.9 million could be positioned at 12,552 different insertion sites in *M. sexta* single-copy genomic regions (Table 1 and Table S3). We interpreted these chimeric reads as corresponding to the junctions of CcBV circles with *M. sexta* DNA produced following circle integration into the DNA of parasitized host hemocytes.

We were able to validate our primer extension capture by the identification of 45 locations in the *M. sexta* genome corresponding to 2 insertion events (IE) from the same circle with the two different CcBV junctions (J1 or J2) and separated by less than 7 bp. More precisely, 38 locations corresponded to IE from C1, 3 from C4, and 2 from C10 and C17. As described in Materials and Methods, for C1 and C4 captures we used two different probes targeting the two junctions (J1 and J2), so we interpreted these 41 locations from C1 and C4 as potentially corresponding to the boundaries of the same integrated molecule. Regarding C10 and C17, the higher overall IE numbers associated with these circles (see below) along with nonspecific capture and deep sequencing could explain the identification of IE corresponding to both J1 and J2, despite the fact that only J1 probes had been used for the capture experiments.

**The numbers of insertions detected in hemocyte DNA differ depending on the circles.** The number of reads mapping to the different CcBV circles (Table 1) varied, which can be due to their relative abundance in the particles injected during oviposition and/or to different capture efficiencies and sequencing run depths (Table S2). For example, earlier results have shown, by quantitative PCR, that in this wasp population circles C1 and C26 are produced at higher abundance than C4 and C14 in wasp ovaries (18), which might explain the high number of reads obtained for these circles in the capture experiment. Here, in general, the number of chimeric reads between targeted circles and *M. sexta* DNA (Table 1) correlated with the number of reads mapping to CcBV circles. For example, high numbers of reads (over 2.9 million to more than 7.6 million) aligned to C1, C10, C17, and C26, leading to high numbers of chimeric reads (over 165,000 to more than 1 million), whereas fewer reads (900,000 to 2 million) aligned to C4, C14, and C35, resulting in fewer chimeric reads (below 16,000) (Table 1). There were two notable exceptions, namely, high numbers of reads aligned to C15 and C16 (over 5 million), resulting in only 1,881 and 20,304 chimeric reads, respectively

**FIG 2** Legend (Continued)
by DRJ sequences. (B) Reintegrated forms of CcBV circles may also be present in wasp genomic DNA and in lepidopteran host DNA that are flanked by J1 and J2 sequences contained within HIMs. To detect the different CcBV forms, primers were designed to allow amplification of a region in a specific gene of the circle (set *a*), a region that included the DRJ motif (set *b*), and a region that spanned both the DRJ and the J1/J2 junctions (set *c*). Lack of amplification of a PCR product was expected with the third set of primers in the case of integration of the circles in *M. sexta* genomic DNA at the J1/J2 borders and also for male wasps, because there is no detectable production of viral circles in male wasps. Lack of amplification of a PCR product was expected with the second set of primers in male wasps because there is no detectable production of viral circles in male wasps unless a reintegrated form of the circle is present in the genome, as has been described for *C. sesamiae* strains. (C) Representative results obtained for the different primer couples allowing amplification of circles C10 and C26 are shown. Lack of amplification of primer couples (set *c*) in *M. sexta* organisms parasitized for 12 days suggest CcBV circles C10 and C26 are integrated in the *M. sexta* genome and are no longer detectable as circular forms. Note that amplification of primer couples (set *c*) could be obtained using a more processive polymerase, indicating that not all circles are integrated at this time point (data not shown).

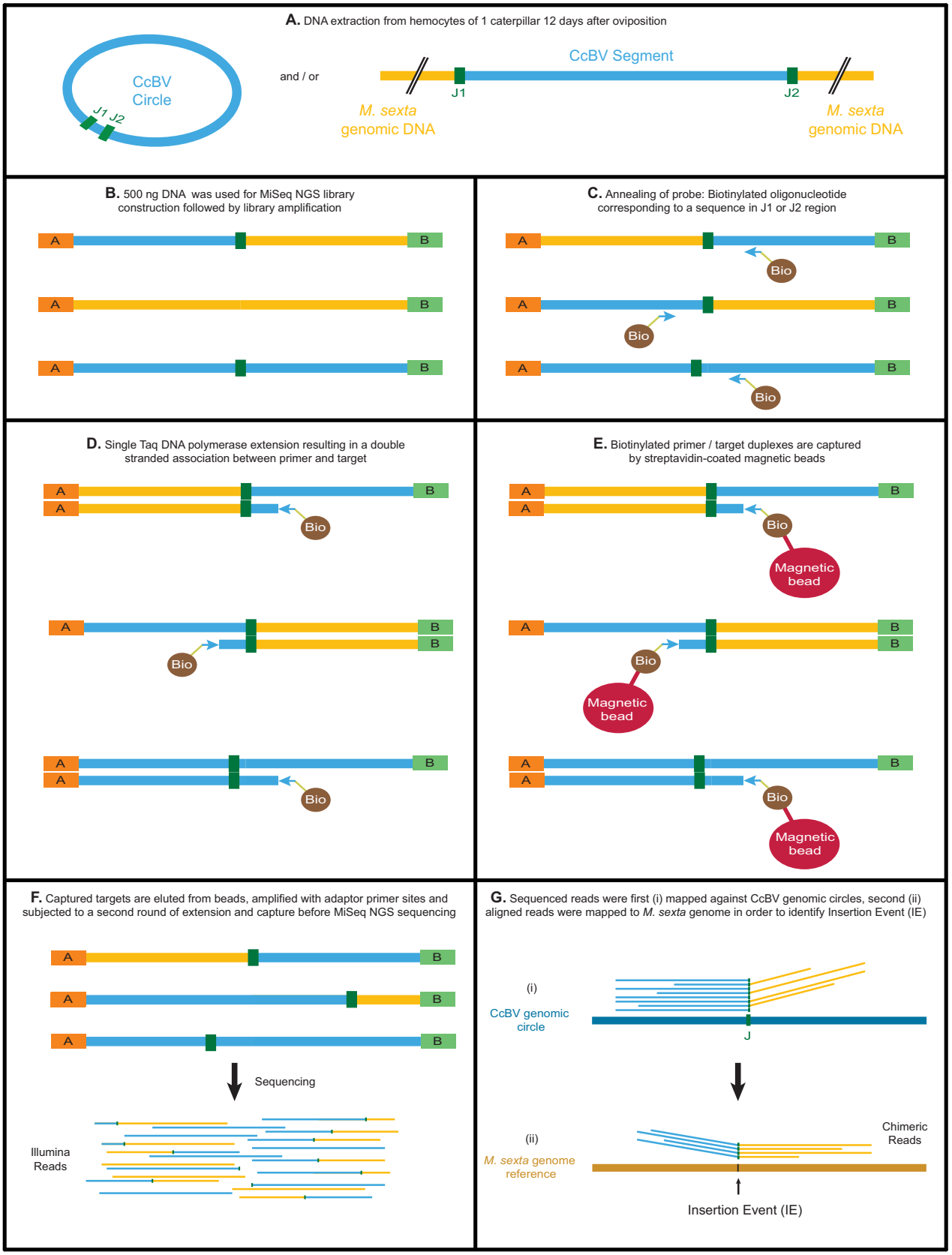



**FIG 3** Principle of the primer extension capture (PEC) method used to detect insertions of CcBV circles in the *M. sexta* genome. (A) DNA was extracted from hemocytes of one caterpillar 12 days after oviposition. This DNA potentially consists of circular CcBV DNA and *M. sexta* DNA with integrated CcBV DNA. (B) The DNA was used to construct an MiSeq library, which was amplified. Biotinylated primers corresponding to sequences near viral J1 and/or J2 junctions

(Continued on next page)

(Table 1). We therefore calculated the expected number of insertion events for each circle, based on the number of reads that aligned on each circle, as a measure of their relative abundance and with the null hypothesis that all circles integrated with the same efficiency in order to compare it with the number of observed insertions (12,552 IE for a total of 35.3 million aligned reads) (Table 1). Comparisons between the numbers of obtained and expected reads suggest that circles C1, C10, C17, and C26 integrated with higher efficiencies, whereas C4, C14, C15, C16, and C35 integrated with lower efficiencies than expected from the null hypothesis. Note that due to high levels of similarities between circles C10 and C17 and the stringent parameters we used for read alignments, our pipeline probably underestimated the number of IE detected for C10 and C17. Another factor, which could also influence the number of chimeric reads obtained, is the distance of the probe to the J1 or J2 site (Table S4). In the case of C1, for example, many more IE and chimeric reads were observed for J2 (probe distance of 3 bp) than for J1 (probe distance of 96 bp) (Table 1 and Fig. 4). The probes used to capture C4, C14, and C16 were over 100 bp from the respective J sites, which might explain the apparent lower integration efficiencies of these circles. In the case of C4, on top of probe distances (probes are 139 bp distant from J1 and 101 bp distant from J2; Table S4), the fact that sequencing read sizes were of only 150 bp (Table S2) may also have contributed to the smaller amount of observed IE for this circle.

**The HIM is involved in circle integration.** Thanks to the alignment of chimeric reads to the 9 different CcBV circles, we could confirm experimentally that CcBV sequence integration involves specific sites on the circles (Fig. 4). To experimentally determine J1 and/or J2 sites, we reasoned that, in a chimeric read, the first nucleotide after the J1 or J2 junction can be by chance (with a probability of 0.25) the same in the *M. sexta* DNA insertion site. Therefore, we determined J1 and/or J2 sites by setting a threshold to exclude these reads (Fig. 4). With this approach we were able to experimentally define CcBV J1 and J2 sequences and compare them to candidate sequences predicted on the basis of sequence similarity (Fig. 1). The experimental J2 sequence was found to coincide exactly with the candidate sequence (Fig. 4), whereas the experimental J1 sequence was identified as tAccaGT, that is, one nucleotide longer than the predicted J1 sequence (Fig. 4). The positions of J1 and J2 could be confirmed for C1 and C4. For the circles for which integration was tested at a single extremity (J1 or J2), J1 could be confirmed for C10, C14, C16, C17, and C26, and J2 could be confirmed for C35. In contrast, the relatively few chimeric reads containing C15 captured sequences have various extremities, which suggests low levels of HIM-mediated integration of this circle (Table 1 and Fig. 4).

**CcBV IE are widespread in *M. sexta* hemocyte DNA but not evenly distributed in the genome.** To determine whether CcBV circle integration occurred randomly or in preferential regions of the *M. sexta* genome, we analyzed *M. sexta* regions in which insertions were detected.

To obtain a global view of the distribution of insertions throughout the *M. sexta* genome, we analyzed *M. sexta* genomic scaffolds using 100-kb-length windows and counted IE within these windows (Fig. 5). A global analysis using these 100-kb windows, covering the entire *M. sexta* genome, revealed the presence of IE throughout the *M. sexta* genome (Fig. 5). It is worth noting that the data did not reveal a specific shared *M. sexta* motif near the different insertion sites. In contrast, we could identify windows enriched in insertions in *M. sexta* hemocyte DNA. For example, over a total of 4,195 windows of *M. sexta* genome we found that there were 567 windows with at least 7 IE, whereas under 150 windows were expected according to a Poisson distribution (Table S5). More generally we observed a much higher number of 100-kb windows with one

**FIG 3** Legend (Continued)
were hybridized to the DNA (C) and extended by amplification (D). (E) Biotinylated primer/target duplexes were captured with streptavidin-coated magnetic beads. (F) Captured targets were eluted, amplified with adaptor primer sites, and subjected to a second round of extension capture before MiSeq NGS sequencing. (G) Sequenced reads were mapped against the CcBV genomic circle, and aligned reads were then mapped to the *M. sexta* genome for IE identification.

**TABLE 1** Global insertion event statistics

| Circle name | Circle size (bp) | CcBV reference region size (bp) | Reads that map to | | IE no. | Expected IE no.[a] | Read no. constituting IE | IE localization | | | IE ratio in *M. sexta* genome[b] (bp) | IE from: | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | CcBV genome | CcBV plus *M. sexta* genome | | | | Intergenic | Intron | Exon | | J1 | J2 |
| C1 | 27,312 | 946 | 3,973,196 | 165,312 | 3,774 | 1,410 | 114,854 | 2,045 | 1,435 | 294 | 111,139 | 829 | 2,945 |
| C4 | 15,452 | 851 | 2,012,151 | 5,336 | 310 | 714 | 5,027 | 165 | 115 | 30 | 1,353,028 | 127 | 183 |
| C10 | 14,286 | 857 | 3,683,180 | 269,843 | 1,455 | 1,307 | 89,564 | 804 | 513 | 138 | 288,274 | 1,388 | 67 |
| C14 | 31,972 | 902 | 953,904 | 4,217 | 81 | 339 | 2,061 | 48 | 31 | 2 | 5,178,255 | 82 | 0 |
| C15 | 8,706 | 1,058 | 5,552,605 | 1,881 | 7 | 1,970 | 168 | 2 | 2 | 3 | 59,919,811 | 7 | 0 |
| C16 | 12,375 | 1,057 | 6,845,992 | 20,304 | 168 | 2,429 | 14,733 | 91 | 62 | 15 | 2,496,659 | 167 | 1 |
| C17 | 15,158 | 857 | 3,149,277 | 288,451 | 1,573 | 1,118 | 86,852 | 887 | 542 | 144 | 266,649 | 1,499 | 74 |
| C26 | 12,673 | 872 | 7,684,464 | 1,004,703 | 3,983 | 2,727 | 544,584 | 2,208 | 1,465 | 310 | 105,307 | 0 | 3,983 |
| C35 | 11,226 | 1,057 | 1,516,929 | 15,248 | 206 | 538 | 8,666 | 98 | 89 | 19 | 2,036,110 | 1 | 205 |
| C10/C17 | | | | | 995 | | 110,564 | 531 | 348 | 116 | 421,546 | 993 | 2 |
| Total | 149,160 | 8,457 | 35,371,698 | 1,775,295 | 12,552 | | 977,073 | 6,879 | 4,602 | 1,071 | 33,416 | 5,093 | 7,460 |
| Avg | 16,573 | 940 | 3,930,189 | 197,255 | 1,255 | | 97,707 | 688 | 460 | 107 | | 509 | 746 |

[a]Expected IE number based on number of reads mapped on CcBV genome for each circle.
[b]*M. sexta* genome size, 419.4 Mb.

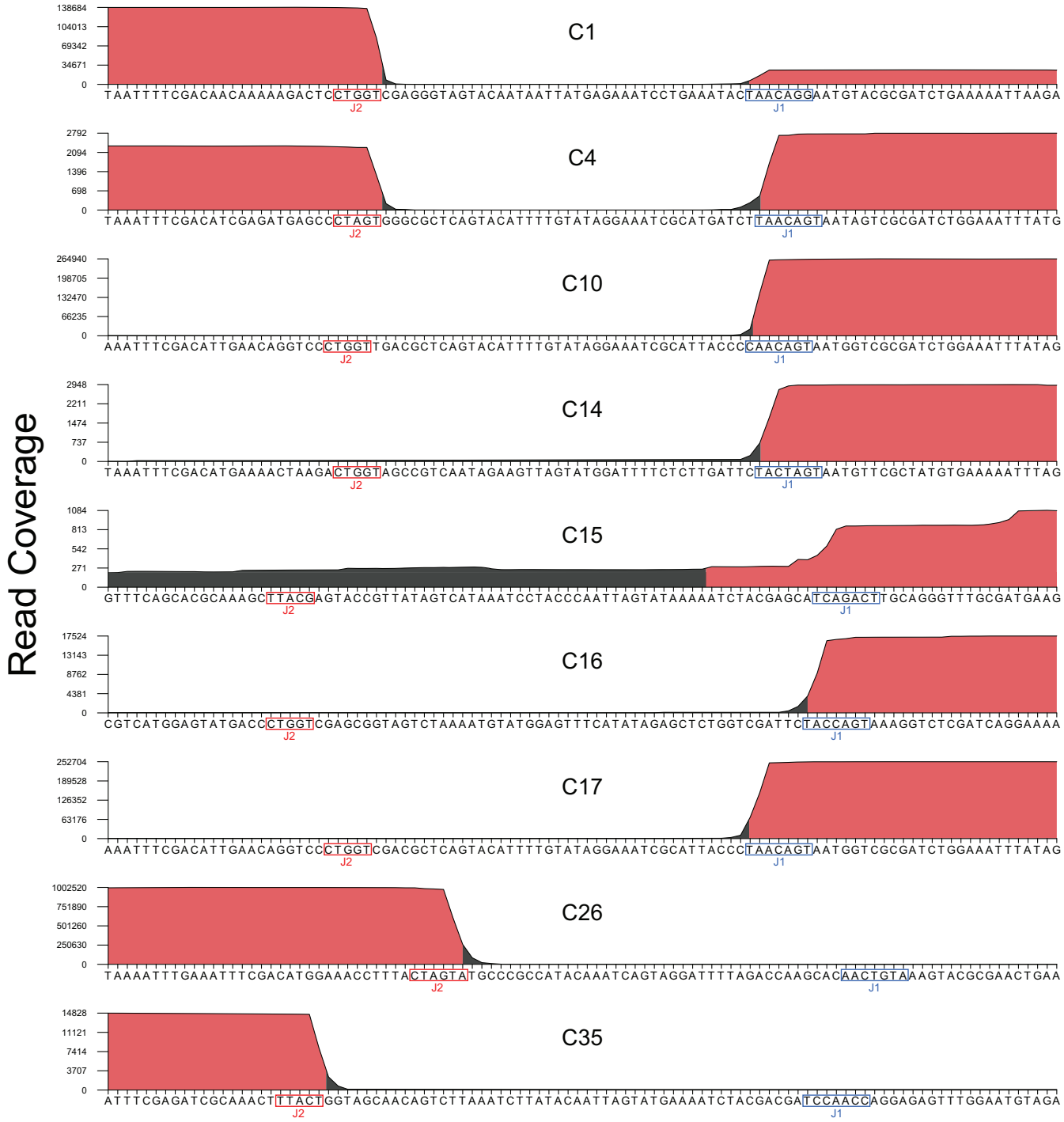



**FIG 4** Chimeric read coverage of CcBV HIM regions for each of the nine CcBV circles. J1 and J2 junction sites are indicated by red and blue boxes. The *x* axis corresponds to a 100-bp sequence spanning the HIM region with J1 and J2 sites. The *y* axis corresponds to read coverage for each base. To experimentally define J1 and J2 sequences, colored areas differentiate nucleotides above (pink) or below (gray) a chosen threshold that corresponds to one-fourth of the maximum read coverage for each circle at each junction J1 or J2. Indeed, in a chimeric read the first nucleotide after the J1 or J2 junction can be by chance (with a probability of 0.25) the same in the *M. sexta* DNA insertion site. These reads were therefore excluded to define J1 and J2 sites.

IE or with more than 7 IE than expected by chance, resulting in a lower number of 100-kb windows with 2 to 6 IE (Table S5). We also investigated whether different IE could impact the same locations (defined as a region of under 100 bp). We were able to find 438 locations in the genome where 2 or 3 IE were separated by a maximum of 100 bp. Interestingly, in 24 locations, 2 different IE (corresponding to different circles or

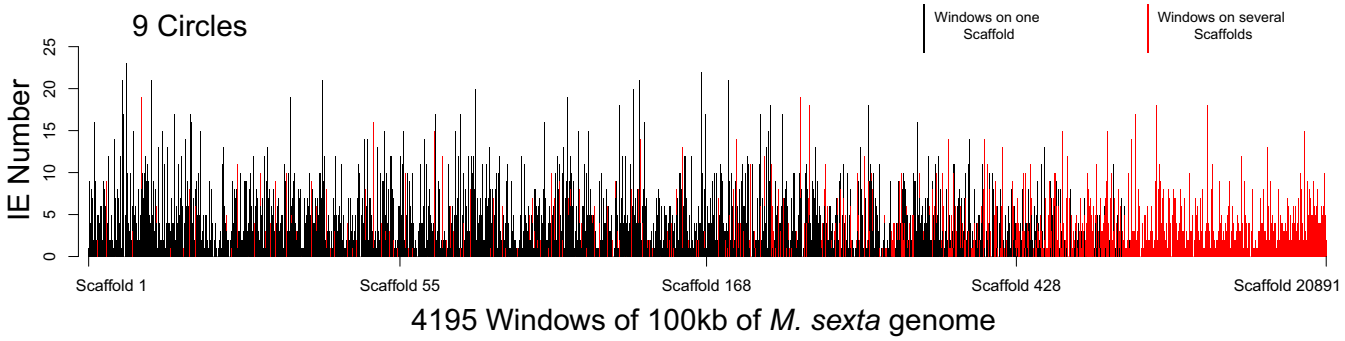



**FIG 5** IE distribution over the *M. sexta* genome. The *x* axis corresponds to 4,195 windows of 100 kb of *M. sexta* genome. The *y* axis corresponds to the cumulative IE number counted for each of the nine circles inside each 100-kb window. Because of the fragmentation of the *M. sexta* genome assembly (20,891 scaffolds from 3.2 Mb to 500 bp), the 100-kb windows spanning several *M. sexta* genome scaffolds were highlighted in red, whereas windows fully present in one scaffold are in black. The number of 5 representative scaffolds is indicated below the *x* axis. Note that no specific integration sites could be detected in the *M. sexta* DNA sequence.

different junctions [J1 or J2]) were found to be at the exact same position, which is a higher value than expected if the IE occur randomly (*P* value of $<0.0001$ according to Fisher's exact test).

**CcBV IE show a slight preference for expressed coding sequences.** In *M. sexta*, we calculated that intergenic regions correspond to 57.5% of the genome, whereas genes (untranslated regions, introns, and exons) represent 42.5%, with 5% corresponding to exons (36). In our experiment, we observed a slight overrepresentation of IE in gene-containing regions (45.2%), particularly in exons (8.5%), suggesting there is a relative preference for IE to occur in exons (Table 1). Moreover, after identification of *M. sexta* genes expressed in hemocytes collected from nonparasitized caterpillars, according to Zhang et al. (42), an enrichment analysis showed a slight overrepresentation of expressed genes being impacted by IE (*P* value of $<0.001$ according to Fisher's exact test) (Table S6).

Overall, we found that 3,165 genes were impacted by at least one IE; among them, 159 genes were impacted by at least 5 IE and 10 genes by at least 10 IE (Table S6). For example, 5 genes, Msex206398, Msex213781, Msex210051, Msex211008, and Msex200449, which code for proteins involved in cell cytoskeleton dynamics (dynein, kinesin, and regulator of cytokinesis), contained 20, 9, 9, 6, and 6 IE, respectively. To obtain a global view of functions that might be impacted by IE, we performed cluster of orthologous group (COG) enrichment analyses to determine whether certain COG were more impacted by IE than all other functions. COG categories corresponding to signal transduction and nutrient transport and metabolism functions (such as amino acid/nucleotide/carbohydrate/coenzyme and inorganic ion transport and metabolism) were found to be significantly enriched in genes affected by IE (*P* value of $<0.05$ according to Fisher's exact test) (Table S7). In contrast, COG categories corresponding to transcription or translation processes appeared to be less impacted by IE.

## DISCUSSION

Polydnaviruses are virus symbionts associated with parasitoid wasps that ensure wasp parasitism success. Lack of viral replication in hosts bearing developing parasitic wasps for several days (typically 10 days for *M. sexta* parasitized by *C. congregata*) raised the question of whether viral DNAs persist during the course of parasitism and how they may avoid being diluted in cells undergoing successive divisions. Indeed, the wasps are koinobiont parasites, meaning that parasitized caterpillars develop throughout parasitism. In the MdBV-*P. includens* interaction, it was formally demonstrated for two MdBV circles that they do integrate into hemocyte genomic DNA of parasitized hosts (2), and a specific sequence mediating integration was identified in almost all MdBV circles, suggesting integration is a common property of all these circles. Indeed,

sequence analysis of host-virus junctions showed that the integration of MdBV circular molecules occurred in association with a sequence named the host integration motif (HIM). HIMs were initially identified in bracoviruses associated with the genus *Microplitis* (2), but BV circle reintegration events were later detected in the genomes of two strains of *C. sesamiae*. Strikingly, the integrated sequences had extremities (J1 and J2) similar to the ones identified for MdBV integrated circles (30). This led to the hypothesis that BVs associated with wasps of the *Cotesia* genus had the same integrative properties as MdBV, particularly allowing them to persist as integrated forms within parasitized lepidopteran host cell DNA.

Here, we carried out a study to determine whether CcBV DNA circles do integrate within host DNA during parasitism of *M. sexta* by *C. congregata*. We found that 12 out of 35 CcBV circles harbor an HIM-like motif. Furthermore, we have shown by a PCR approach assessing the presence of different viral forms that all 4 circles tested persist in naturally parasitized *M. sexta* host hemocytes as a linear form (as expected after integration), which could be detected early during parasitism (24 h postoviposition) as well as after wasp larva emergence from the host body. Furthermore, by using a primer extension capture (PEC) method based on the HIMs of 9 circles, we could show that 8 circles were integrated in *M. sexta* hemocyte genomic DNA 12 days after oviposition and that integration had occurred specifically using their HIM. Moreover, the high-throughput approach enabled us to perform, for the first time, an investigation of bracovirus circle insertion sites on the scale of an entire host genome.

To identify CcBV circle integration sites within the *M. sexta* genome, we adapted a PEC method (35) that combines high specificity of capture with high-throughput sequencing. This approach has been used previously to identify retroviral insertions in order to monitor the history of a viral infection decimating koala populations (43). The identification of HIM-like sequences in CcBV circles allowed us to deduce the potential extremities of viral insertions named J1 and J2 (30). Primers were designed in the vicinity of J1 and/or J2 in order to capture chimeric sequences corresponding to junctions between viral and *M. sexta* hemocyte DNA. These chimeric reads then were mapped to the recently released *M. sexta* genome (36) to determine the positions of the insertions within parasitized host hemocyte DNA.

This PEC method allowed us to show experimentally that 8 (C1, C4, C10, C14, C16, C17, C26, and C35) out of 9 of the circles tested do integrate using HIM sites and to determine experimentally the J1 and/or J2 extremities of integrated CcBV circles. J1 differed slightly from the predicted sequence, being one nucleotide longer. In the same way as MdBV HIMs, CcBV HIMs corresponded to a 100- to 110-nucleotide domain consisting of two imperfect inverted repeats. We suggest that these repeats contain binding sites for monomers of an integrase that would then dimerize and form a nucleoprotein complex involved in circle integration into host cell DNA (Fig. 1). The J1-J2 extremities of integrated sequences are separated in the circle by an internal sequence of ~50 nucleotides, which is lost during integration. This loss is, to our knowledge, a unique feature among integration mechanisms described for viruses and mobile elements. Transposons and viruses having latent phases need to maintain their integrity after integration into host DNA for further transposition or infectious cycles, which is not the case for bracovirus circles owing to their exclusive vertical transmission, thus the loss of a noncoding DNA stretch is not expected to be counterselected. In contrast, the few chimeric reads of circle C15 did not have common junction sites, only a small subset of the reads corresponded to the J1 junction and none to the J2 junction (both tested for this circle), suggesting a low level of HIM-mediated integration for circle C15 and that these reads correspond to rare nonspecific integration events detected because of sequencing depth. Note that circle C15 J2 (CAAGT) is more divergent from other J2 sequences (CTA/GGT), which might impair HIM-mediated integration.

It was observed that numbers of reads mapping on the different CcBV circles tested differed (Table 1). Some of these differences might be due to relative capture efficiencies of the probes and/or to the relative abundance of the circles in the particles (44).

Indeed, bracovirus circles are known to be individually packaged into virions and virions containing certain virus circles are more abundant than others, leading to nonequimolar ratios of circles being produced by the wasp and injected into the host (18, 45).

In our analysis, HIM sequences were identified in 12 CcBV circles out of 35, whereas in MdBV 23 out of the 25 circles have been reported to contain HIMs (14). Interestingly, CcBV circles harboring HIMs have characteristics in common. Apart from segments S15, S16, and S35, which are localized in the second half of PL2, a part of the "macrolocus" containing 70% of the proviral segments, most segments producing these circles (C1, C4, C7, C10, C11, C12, C14, C17, and C26) are widespread in the *C. congregata* genome, in the so-called isolated loci that contain 1 to 3 segments. Strikingly, a common feature of all these circles (except circle C12), including those of the macrolocus, is their gene content. Indeed, all members of protein tyrosine phosphatase (*PTP*; 27 genes) and viral ankyrin (*VANK*; 9 genes) gene families are localized within these circles. Along the same lines, the J1/J2-containing integrating GiBV segment F (circle GiBV25) also encodes 8 PTPs and one ankyrin (46). Some *PTP*s and *VANK*s have been shown to act as virulence factors within the host (33, 34, 47), probably disrupting important signaling pathways involved in immune responses and development (48). Unlike most proteins from the CcBV macrolocus, PTPs and VANKs lack secretion signals and probably act directly within infected cells. Their genes (except *VANK9*, which is highly expressed) have been shown to be expressed at low levels in *M. sexta* (18), in sharp contrast with virulence genes encoding secreted products, such as *Early Protein 1* (*EP1*), which represents up to 5% of the hemolymph protein content 24 h after oviposition (49), or *cystatin*, the mRNA levels of which are 50-fold higher than that of *actin* 12 h postoviposition (50). In this context, while most MdBV circles may still have the ability to integrate, this ability might have been lost by CcBV circles, except for circles harboring *PTP* and *VANK* genes, because their products have to operate on signaling pathways within each cell while virulence proteins secreted in the hemolymph can be produced from any cell. Along these lines, it will be of great interest to study the dynamics of expression of CcBV virulence genes throughout parasitism to determine whether circle integration allows persistence of *PTP* and *VANK* genes expression. Indeed, their production in the later stages of parasitism could also be important for the successful development of the wasp. *In vivo* persistent expression of PTPs and VANK encoded by the integrated segment F (GiBV25) circle has indeed already been demonstrated, sustaining the hypothesis that in the absence of viral replication, integration may allow these proteins to be produced throughout parasitism (26, 28).

On top of possibly allowing viral gene expression persistence, integration may also impact the lepidopteran host genome by modifying or disrupting expression of host genes that would negatively impact wasp development. In this study, thanks to the recent sequencing of the *M. sexta* genome (36) and to the PEC approach taken, we were able to identify the position of 12,552 CcBV insertions in *M. sexta* DNA. By analyzing integration in 100-kb windows spanning the *M. sexta* genome, insertions appeared to be widespread. However, this distribution was not totally random, as an excess of 100-kb windows harboring 7 or more insertions (up to 23) could be detected. These regions enriched in IE could represent zones of DNA that are more readily accessible, such as DNA sequences undergoing transcription, for example. Genes impacted by IE could be identified within these enriched regions. Also, by identifying *M. sexta* genes expressed in nonparasitized hemocytes according to Zhang et al. (42), we showed that genes expressed before parasitization are slightly more impacted by IE than genes that are not expressed. Moreover, 159 coding DNA sequences were impacted by at least 5 IE from at least two different CcBV segments (Table S6). Among the impacted genes, our enrichment analysis for COG function showed that encoded proteins could be involved in general nutrient transport and metabolism as well as signal transduction. Although parasitism is known to involve metabolic changes (51) and to impact different signaling pathways involved in development and immunity (52, 53), further work is required to investigate whether disruption of expression of these genes by IE is

advantageous to wasp development. Indeed, because millions of hemocytes are likely to have been impacted by the integration of CcBV circles, hemocytes could each be impacted in different genomic locations in such a way that a gene impacted in one hemocyte may not be impacted in the other and therefore allow functional compensation. In MdBV, two integrations were shown to be present in the same region of the host genome, suggesting a specificity of integration site (2), but this does not appear to be the case for CcBV.

It should be noted that because the aim of this study was to detect insertion events without ambiguities, we used very stringent criteria and only analyzed chimeric reads containing at least 30 bp of viral and lepidopteran genomic DNA. The insertions analyzed here therefore represent only a subset of the total integration events. Moreover, most of the reads obtained in the capture experiment corresponded to nonintegrated circles (Table 1), which does not fit with results from the initial PCR approach (Fig. 2), suggesting that the number of integration events is greater than the number of the remaining circular forms. One explanation for this discrepancy between the two approaches might be a bias in favor of viral circular DNA during initial library preparation and/or during the capture experiment.

Data obtained on retroviral integration are likely to be an interesting source of inspiration to study the mechanisms of BV integration into lepidopteran host genomic DNA. Indeed, parallels can be drawn between the two systems. First, integration of retroviruses is catalyzed by a viral integrase (Int) protein. In the case of BVs, two tyrosine recombinases belonging to the lambda integrase family could be involved in integration of BV circles into lepidopteran host genomes (10, 54–56). The knockdown of both *vlf-1* and *int-1* genes encoding these proteins in *Microplitis demolitor* led to near-complete inhibition of excision of BV segments in wasp ovaries, suggesting these proteins are indeed involved in recombination reactions leading to excision (56). The fact that these proteins are detected in BV virions (56, 57) suggests these proteins also have important functions in the parasitized host, possibly by mediating BV circle integration. Second, retroviral integration apparently is not a random process, with different retroviral genera favoring different chromatin environments for integration. The integrase protein interacts with tethering proteins that guide the preintegration viral complex into specific chromatin contexts. The final integration site, usually in bendable DNA sites on nucleosomal DNA, shows weak nucleotide sequence preference (58, 59). Our finding that CcBV circle integration is not totally random but does not appear to involve a precise integration site thus is similar to results observed for retroviruses.

Bracoviruses evolved from integrated nudiviruses by several dramatic genomic transformations after integration of the ancestral nudivirus genome into the wasp genome (10, 60). The possibility to maintain virulence genes within host cells by the integration of viral circles could also constitute another crucial event conferring to wasps the ability to develop in their lepidopteran hosts. The maintenance of this ability might have been selected, and we could expect the conservation of HIMs in all lineages. In the present study, HIMs were identified in 12 CcBV circles. HIM-like domains were also identified in several other BVs: Cotesia sesamiae BV (CsBV), Cotesia vestalis BV (CvBV), Glyptapanteles indiensis BV (GiBV), and Glyptapanteles flavocoxis BV (GfBV). Moreover, the extremities of GiBV-integrated sequences in cultured cells previously described (for 1 segment, segment F or GiBV25) (26) corresponded to predicted J1 and J2 of *Glypapanteles indiensis* segment F, suggesting that HIMs are also functional in GiBV. In particular, the presence of HIM-like sequences was also confirmed in *C. sesamiae* circles highly similar to CvBV27 and CvBVS2 that correspond to reintegrated circles with J1 and J2 extremities in the genomes of two *C. sesamiae* wasp populations (30). These findings strongly suggest that the same HIM-mediated mechanism is involved both in circle integration in lepidopteran DNA and circle reintegration in wasp DNA.

The inverted repeats and the junction sites within the HIMs of these BVs are clearly similar but the loop domains are more divergent, which might be explained if this region is solely required to form a loop enabling the inverted repeats of the HIM to

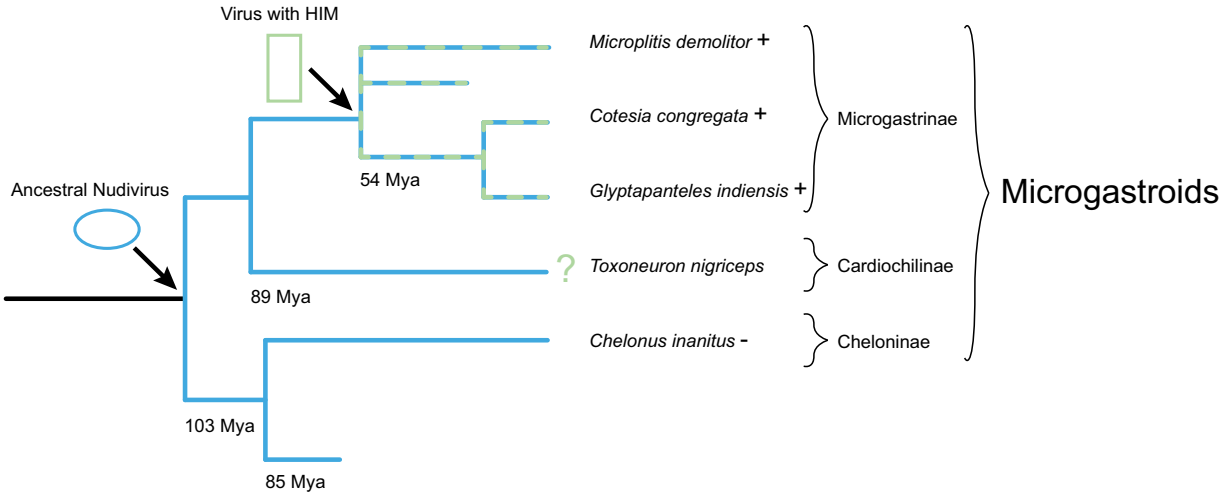



**FIG 6** Proposed evolutionary scenario for the acquisition of HIMs in bracoviruses. Key dates (in millions of years) for the origin of braconid microgastroid subfamilies are from Murphy et al. (9). So far HIMs have only been identified in bracoviruses belonging to the Microgastrinae. No HIMs have been identified so far in *Cheloninus inanitus* (Cheloninae), and whether they are present in *Toxoneuron nigriceps* (Cardiochilinae) awaits investigation. A proposed evolutionary scenario is that acquisition of HIMs by BV occurred before the divergence of the Microgastrinae 54 Mya.

come into contact. The wasp species harboring these BVs all belong, like *C. congregata* and *Microplitis* spp., to the Microgastrinae subfamily of wasps that diversified 54 Mya (9), whereas no homologous HIMs could be detected in the BV associated with *Chelonus inanitus*, a wasp belonging to another subfamily of the microgastroid complex, the Cheloninae, that diverged 85 Mya (Fig. 6). This is in contrast with the DRJ sequences involved in circle production that are found among BVs from wasps in all subfamilies of the microgastroid complex, from *Chelonus inanitus* to *Cotesia* species (13), and questions whether integration ability was inherited from the captured nudivirus ancestor or acquired later specifically in the Microgastrinae lineage. A possible evolutionary scenario is that acquisition of HIMs by BV occurred before the divergence of the Microgastrinae 54 Mya (9), leading to braconid wasps harboring BVs with HIMs (Fig. 6). These viruses containing integration motifs might have enhanced wasp parasitic success, leading these wasp species to perform better than wasp species harboring viruses devoid of these motifs. In braconids, such an event might have favored species diversification, explaining the impressive radiation of the Microgastrinae (37,500 to 46,400 species) compared to all other taxa, including Cardiochilinae and Cheloninae (7,000 species) (12 and J. Whitfield, personal communication).

Unraveling the mechanisms and extent of PDV integration events (i.e., investigating integration of other circles and integration in other tissues, including germinal cells) will not only allow us to better understand virulence strategies developed by parasitoid wasps to successfully parasitize their lepidopteran hosts but might also provide the mechanism of recently uncovered horizontal gene transfer between parasitic wasps, PDVs, and the genomes of moths and butterflies. Indeed, in some cases CcBV DNA containing genes of wasp origin have been found to be present in the genomes of many lepidopteran species, including the iconic Monarch species (61–63). However, unlike BV sequences identified in *C. sesamiae* wasps, we could not find a complete circle inserted in the genome of Lepidoptera, which could be due to the fact that the insertions are ancient and have undertaken many rearrangements after circle integration. The integrated genes are not solely nonfunctional remnants, as results suggest that certain genes integrated in notorious caterpillar pest species can play a protective role against other pathogenic viruses (baculoviruses) present in nature and also used in biocontrol programs. Investigating the extent of these transfers will be important to evaluate the risks associated with the introduction of exotic parasitoid wasps associated with PDV in classical biological control programs. Also, with the advent of gene modification technologies the risk of transfer of modified or new genes in nature will

have to be evaluated if modified parasitoid wasps are envisaged for biological control (64).

## MATERIALS AND METHODS

**Insect rearing and parasitization.** Larvae of the tobacco hornworm, *M. sexta* (Lepidoptera, Sphingidae), were reared on an artificial diet at 27°C under a 17-h light and 7-h dark photoperiod as previously described (65). At the 4th instar, *M. sexta* larvae were parasitized by exposing them to *C. congregata* wasps (Hymenoptera, Braconidae) until two oviposition events were observed. Dozens of wasp eggs are injected during wasp oviposition, larvae develop within the caterpillar host body and egress from the host on day 10 to 12 to pupate within a silken cocoon, and adults emerge 5 to 6 days later.

**Identification of candidate HIMs in CcBV circles.** MdBV HIM and CsBV J1 and J2 motifs were used to search for similar motifs among CcBV (13, 66) circles by BLAST analyses (67). Alignments between MdBV and candidate CcBV HIMs were refined using Multalin (68), MAFFT 7 (69), and DIALIGN-TX (70). The visualizations and figures of the alignments were made using Jalview (71).

**DNA isolation and PCR-based detection of CcBV circular and linear forms.** Genomic DNA from whole adult female and male *C. congregata* wasps and from hemocytes from parasitized *M. sexta* larvae (24 h and 12 days postoviposition [$n = 8$]) was isolated by using the QIAamp DNA minikit (Qiagen, France). It should be remarked that at 12 days postoviposition, wasp larvae had emerged from *M. sexta* host and spun their cocoons. Genomic DNA isolated from unparasitized *M. sexta* served as controls.

To detect the different forms (circular and/or integrated) of the CcBV, primer pairs were designed for four circles in which a candidate HIM could be detected (circles C4, C10, C14, and C26) (see Table S1 in the supplemental material). For each circle, primers allowed amplification of a region in a specific gene of the circle (primer pair *a*), a region that included the DRJ motif (primer pair *b*), and a region that spanned the HIM containing J1, J2, and the DRJ (primer pair *c*) (Fig. 2A). A PCR product was expected for all samples with the first set of primers (positive control). Lack of amplification of a PCR product was expected with the second set of primers in male wasps because there is no detectable production of viral circles in male wasps (Fig. 2B). Lack of amplification of a PCR product was expected with the third set of primers for male wasps and in the case of integration in *M. sexta* genomic DNA (Fig. 2B) of circles originally present in the particles infecting hemocytes. Indeed, during integration the HIM is separated in two pieces, with the J1 and J2 sequences becoming the extremities of the integrated form, and consequently the primers are no longer in the correct orientation for PCR amplification.

PCRs were performed on genomic DNA extracted from individuals using GoTaq (Promega, France) in a final volume of 25 $\mu$l containing 50 ng of genomic DNA (wasp or *M. sexta* hemocyte DNA), 1.25 U of GoTaq, 3 mM $MgCl_2$, and 20 pmol of each specific primer with the following cycling conditions: 4 min of initial denaturation at 94°C, followed by 35 cycles of denaturation at 94°C for 40 s, primer hybridization at 58°C for 40 s, extension at 72°C for 60 s, and final extension at 72°C for 10 min.

**Capture and high-throughput sequencing of CcBV sequences integrated in *M. sexta* genome.** An adaptation of the PEC method (35) was developed to capture and determine CcBV sequences that have been integrated in the lepidopteran genome (Fig. 3).

Oligonucleotide sequences used for the capture were complementary to the targeted CcBV circles and close to J1 or J2 sequences (Table S4) in order to capture junctions between viral and parasitized host DNA. Oligonucleotide distances from J1 or J2 sequences ranged from 3 to 159 bp (Table S4). Oligonucleotides were designed in nine circles out of the 12 in which a candidate HIM had been detected. Oligonucleotides in the vicinity of J1 and J2 viral sequences were designed for circles C1, C4, and C15. Only one of the two extremities was tested for the other circles (J1 for circles C10, C14, C16, C17, and C26 and J2 for C35) (Table S4). A spacer sequence, CAAGGACATCCG, to which a biotin molecule was attached, was included in the probes to allow capture using streptavidin beads (Fig. 3C and Table S4). These biotinylated probes were synthesized by Eurogentec (France).

The method described in detail in the legend to Fig. 3 consisted first in constructing a next-generation sequencing (NGS) library starting from 500 ng of DNA isolated from the whole population of hemocytes collected from one parasitized caterpillar 12 days postoviposition (Fig. 3A). We then amplified this library using the attached adaptors as the priming site (Fig. 3B).

We then carried out a first sequence capture using the biotinylated probes (Fig. 3C to F and Table S4). For each probe we determined the optimal hybridization temperature. Indeed, all of the probes were tested using PCR amplification products of the targeted regions in viral circles (data not shown). After hybridization the probes were extended in order to stabilize the interaction, and duplexes were captured using streptavidin beads (Fig. 3D and F). Captured targets, corresponding to the templates on which extension occurred (Fig. 3F), were amplified again, and we carried out a second capture using the same probes, slightly increasing the temperature in order to obtain a higher hybridization specificity (Fig. 3). The material obtained after the second capture was paired-end sequenced using Illumina MiSeq. The size of the reads was 150 bp or 250 bp depending on the Illumina sequencing runs (Table S2).

**Read mapping and data analyses.** A total of 63.7 million reads were obtained from the five Illumina runs (Table S2). Reads from different runs were pooled and were aligned against targeted regions in CcBV circles (their positions and ENA accession numbers are indicated in Table S8) that correspond to roughly 1 kb of circle sequences spanning J1 and J2 predicted motifs and the position of the probe used for the DNA capture. Mapping was performed using the Burrows-Wheeler alignment (BWA) tool, version 0.7.10, with the "mem" algorithm and with parameters "-c 1" in order to remove reads with multiple alignments (72). A total of 35.3 million reads were aligned to the nine CcBV regions (Table S2). In a stringent analysis, to avoid incorrect mapping due to short alignments, only reads which aligned with at least 30 nucleotides of one of the CcBV regions analyzed were kept. CcBV-aligned reads were then mapped

against the *M. sexta* genome (36). We chose to align reads as single-end reads against the *M. sexta* genome in order to identify specifically chimeric reads (i.e., reads with CcBV and *M. sexta* sequences instead of one read of the pair fully aligned to CcBV and the other one fully aligned to *M. sexta*). Again, only reads that aligned to the *M. sexta* genome with a size above 30 nucleotides were kept. Reads were then mapped again to the CcBV regions in order to experimentally identify J1 and J2 boundaries in chimeric reads. We used a custom-written Bash script along with specific tools from the BEDtools suite (73) to localize, count, and identify the orientation of the chimeric reads corresponding to insertion events (IE). Finally, in order to focus on high-quality and unambiguous IE, we removed from the analysis IE detected in low-complexity regions (microsatellite-like DNA and repetitive elements) and IE detected in multiple positions (i.e., due to *M. sexta* genome duplications or inaccurate genome assemblies).

**Analysis of IE in coding sequences.** Lastly, when IE were detected in coding sequences, the putative functions of encoded proteins were assessed by Hmmer search against the reference proteome and Pfam database (74). The involvement of encoded proteins in putative metabolic pathways was performed by searches against the KEGG database using the blastKOALA tool (75).

In order to determine if certain host functions were particularly targeted by IE, cluster of orthologous group (COG) categories were assigned for each *M. sexta* protein (18) by eggNOG-mapper (76), and enrichment analyses were performed to determine whether certain COG were more impacted by IE than all other functions. COG enrichment analyses for IE were performed by obtaining contingency tables for each COG category followed by a Fisher exact test (Table S7).

Finally, in order to assess a potential link between IE detected in genes and gene expression in hemocytes, we performed a reanalysis of the 454 hemocyte transcriptome obtained by Zhang et al. (42). The reanalysis consisted of retrieving the control hemocyte (CH) library from NCBI SRA (accession number SRS167319) followed by mapping of reads to the *M. sexta* genome (36) with the BWA tool, version 0.7.10, with the "mem" algorithm and with parameters "-B 1 -O 1,1 -E 1,1 -L 1,1." Read counts for each *M. sexta* gene were then generated by HTSeq-count (77). Enrichment analyses were performed to determine whether expressed genes affected by IE were overrepresented compared to expressed genes not affected by IE. Enrichment analyses were performed by obtaining a contingency table for genes with or without IE known to be expressed or not in hemocytes, followed by a Fisher exact test.

**Accession number(s).** Reads were deposited in the NCBI Sequence Read Archive under accession numbers ERR2391708, ERR2391709, ERR2391710, ERR2391711, and ERR2391712.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JVI.00438-18.

**SUPPLEMENTAL FILE 1,** XLSX file, 1.9 MB.

## ACKNOWLEDGMENTS

We thank the students from the Master course "Ecologie comportementale, evolution et biodiversité" (years 2013 to 2015) at the University of Tours for contributing to the development of the PCR-based detection method and the Biologie Animale et Génétique teaching department for financial support of the PCR-based experiment. Funding was provided by recurrent funds from the Centre National de la Recherche Scientifique (CNRS) and the University of Tours and by grant COTEBIO ANR 17-CE-32-015-02.

We thank Cindy Menoret for insect rearing. In particular, many thanks to Hélène Boulain for designing primers for CcBV circles 10 and 26. We gratefully thank members of the Réseau Intégration de l'ADN Viral et Dynamique Chromatinienne for stimulating discussions and advice. Many thanks to Jim Whitfield for fruitful discussions on braconid wasp phylogeny and dating of phylogenetic trees.

## REFERENCES

1. Pennacchio F, Strand MR. 2006. Evolution of developmental strategies in parasitic hymenoptera. Annu Rev Entomol 51:233–258. https://doi.org/10.1146/annurev.ento.51.110104.151029.
2. Beck MH, Zhang S, Bitra K, Burke GR, Strand MR. 2011. The encapsidated genome of Microplitis demolitor bracovirus integrates into the host *Pseudoplusia includens*. J Virol 85:11685–11696. https://doi.org/10.1128/JVI.05726-11.
3. Strand MR, Burke GR. 2013. Polydnavirus-wasp associations: evolution, genome organization, and function. Curr Opin Virol 3:587–594. https://doi.org/10.1016/j.coviro.2013.06.004.
4. Strand MR, Burke GR. 2014. Polydnaviruses: nature's genetic engineers. Annu Rev Virol 1:333–354. https://doi.org/10.1146/annurev-virology-031413-085451.
5. Drezen J-M, Chevignon G, Louis F, Huguet E. 2014. Origin and evolution of symbiotic viruses associated with parasitoid wasps. Curr Opin Insect Sci 6:35–43. https://doi.org/10.1016/j.cois.2014.09.008.
6. Drezen J-M, Leobold M, Bézier A, Huguet E, Volkoff A-N, Herniou EA. 2017. Endogenous viruses of parasitic wasps: variations on a common theme. Curr Opin Virol 25:41–48. https://doi.org/10.1016/j.coviro.2017.07.002.
7. Volkoff A-N, Jouan V, Urbach S, Samain S, Bergoin M, Wincker P, Demettre E, Cousserans F, Provost B, Coulibaly F, Legeai F, Béliveau C, Cusson M, Gyapay G, Drezen J-M. 2010. Analysis of virion structural components reveals vestiges of the ancestral ichnovirus genome. PLoS Pathog 6:e1000923. https://doi.org/10.1371/journal.ppat.1000923.
8. Béliveau C, Cohen A, Stewart D, Periquet G, Djoumad A, Kuhn L, Stoltz

D, Boyle B, Volkoff A-N, Herniou EA, Drezen J-M, Cusson M. 2015. Genomic and proteomic analyses indicate that banchine and campoplegine polydnaviruses have similar, if not identical, viral ancestors. J Virol 89:8909–8921. https://doi.org/10.1128/JVI.01001-15.
9. 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. https://doi.org/10.1016/j.ympev.2008.01.022.
10. Bézier A, Annaheim M, Herbinière J, Wetterwald C, Gyapay G, Bernard-Samain S, Wincker P, Roditi I, Heller M, Belghazi M, Pfister-Wilhem R, Periquet G, Dupuy C, Huguet E, Volkoff A-N, Lanzrein B, Drezen J-M. 2009. Polydnaviruses of braconid wasps derive from an ancestral nudivirus. Science 323:926–930. https://doi.org/10.1126/science.1166788.
11. Herniou EA, Huguet E, Theze J, Bézier A, Periquet G, Drezen J-M. 2013. When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. Philos Trans R Soc Lond B Biol Sci 368:20130051–20130051. https://doi.org/10.1098/rstb.2013.0051.
12. Rodriguez JJ, Fernández-Triana JL, Smith MA, Janzen DH, Hallwachs W, Erwin TL, Whitfield JB. 2012. Extrapolations from field studies and known faunas converge on dramatically increased estimates of global microgastrine parasitoid wasp species richness (Hymenoptera: Braconidae). Insect Conserv Divers 6:530–536. https://doi.org/10.1111/icad.12003.
13. Bézier A, Louis F, Jancek S, Periquet G, Thézé J, Gyapay G, Musset K, Lesobre J, Lenoble P, Dupuy C, Gundersen-Rindal D, Herniou EA, Drezen J-M. 2013. Functional endogenous viral elements in the genome of the parasitoid wasp *Cotesia congregata*: insights into the evolutionary dynamics of bracoviruses. Philos Trans R Soc Lond B Biol Sci 368:20130047.
14. Burke GR, Walden KKO, Whitfield JB, Robertson HM, Strand MR. 2014. Widespread genome reorganization of an obligate virus mutualist. PLoS Genet 10:e1004660. https://doi.org/10.1371/journal.pgen.1004660.
15. Louis F, Bézier A, Periquet G, Ferras C, Drezen J-M, Dupuy C. 2013. The bracovirus genome of the parasitoid wasp *Cotesia congregata* is amplified within 13 replication units, including sequences not packaged in the particles. J Virol 87:9649–9660. https://doi.org/10.1128/JVI.00886-13.
16. Gruber A, Stettler P, Heiniger P, Schumperli D, Lanzrein B. 1996. Polydnavirus DNA of the braconid wasp Chelonus inanitus is integrated in the wasp's genome and excised only in later pupal and adult stages of the female. J Gen Virol 77:2873–2879. https://doi.org/10.1099/0022-1317-77-11-2873.
17. Savary S, Beckage N, Tan F, Periquet G, Drezen J-M. 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(Part 12):3125–3134. https://doi.org/10.1099/0022-1317-78-12-3125.
18. Chevignon G, Thézé J, Cambier S, Poulain J, Da Silva C, Bézier A, Musset K, Moreau SJM, Drezen J-M, Huguet E. 2014. Functional annotation of Cotesia congregata bracovirus: identification of the viral genes expressed in parasitized host immune tissues. J Virol 88:8795–8812. https://doi.org/10.1128/JVI.00209-14.
19. Chevignon G, Cambier S, Da Silva C, Poulain J, Drezen J-M, Huguet E, Moreau SJM. 2015. Transcriptomic response of *Manduca sexta* immune tissues to parasitization by the bracovirus associated wasp *Cotesia congregata*. Insect Biochem Mol Biol 62:86–99. https://doi.org/10.1016/j.ibmb.2014.12.008.
20. Bitra K, Zhang S, Strand MR. 2011. Transcriptomic profiling of Microplitis demolitor bracovirus reveals host, tissue and stage-specific patterns of activity. J Gen Virol 92:2060–2071. https://doi.org/10.1099/vir.0.032680-0.
21. Kim MK, Sisson G, Stoltz D. 1996. Ichnovirus infection of an established gypsy moth cell line. J Gen Virol 77(Part 9):2321–2328. https://doi.org/10.1099/0022-1317-77-9-2321.
22. Volkoff A-N, Rocher J, Cerutti P, Ohresser M, d'Aubenton-Carafa Y, Devauchelle G, Duonor-Cerutti 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. https://doi.org/10.1099/0022-1317-82-4-963.
23. Gundersen-Rindal D, Lynn DE, Dougherty EM. 1999. Transformation of lepidopteran and coleopteran insect cell lines by Glyptapanteles indiensis polydnavirus DNA. In Vitro Cell Dev Biol Anim 35:111–114. https://doi.org/10.1007/s11626-999-0010-z.
24. 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. https://doi.org/10.1006/bbrc.1996.1248.
25. Gundersen-Rindal D, Dougherty EM. 2000. Evidence for integration of Glyptapanteles indiensis polydnavirus DNA into the chromosome of Lymantria dispar in vitro. Virus Res 66:27–37. https://doi.org/10.1016/S0168-1702(99)00125-2.
26. Gundersen-Rindal DE, Lynn DE. 2003. Polydnavirus integration in lepidopteran host cells in vitro. J Insect Physiol 49:453–462. https://doi.org/10.1016/S0022-1910(03)00062-3.
27. Doucet D, Levasseur A, Béliveau C, Lapointe R, Stoltz D, Cusson M. 2007. In vitro integration of an ichnovirus genome segment into the genomic DNA of lepidopteran cells. J Gen Virol 88:105–113. https://doi.org/10.1099/vir.0.82314-0.
28. Gundersen-Rindal DE, Pedroni MJ. 2006. Characterization and transcriptional analysis of protein tyrosine phosphatase genes and an ankyrin repeat gene of the parasitoid Glyptapanteles indiensis polydnavirus in the parasitized host. J Gen Virol 87:311–322. https://doi.org/10.1099/vir.0.81326-0.
29. Weber B, Annaheim M, Lanzrein B. 2007. Transcriptional analysis of polydnaviral genes in the course of parasitization reveals segment-specific patterns. Arch Insect Biochem Physiol 66:9–22. https://doi.org/10.1002/arch.20190.
30. Serbielle C, Dupas S, Perdereau E, Héricourt F, Dupuy C, Huguet E, Drezen J-M. 2012. Evolutionary mechanisms driving the evolution of a large polydnavirus gene family coding for protein tyrosine phosphatases. BMC Evol Biol 12:253. https://doi.org/10.1186/1471-2148-12-253.
31. Provost B, Varricchio P, Arana E, Espagne E, Falabella P, Huguet E, La Scaleia R, Cattolico L, Poirié M, Malva C, Olszewski JA, Pennacchio F, Drezen J-M. 2004. Bracoviruses contain a large multigene family coding for protein tyrosine phosphatases. J Virol 78:13090–13103. https://doi.org/10.1128/JVI.78.23.13090-13103.2004.
32. Suderman RJ, Pruijssers AJ, Strand MR. 2008. Protein tyrosine phosphatase-H2 from a polydnavirus induces apoptosis of insect cells. J Gen Virol 89:1411–1420. https://doi.org/10.1099/vir.0.2008/000307-0.
33. Falabella P, Varricchio P, Provost B, Espagne E, Ferrarese R, Grimaldi A, de Eguileor M, Fimiani G, Ursini MV, Malva C, Drezen J-M, Pennacchio F. 2007. Characterization of the IkappaB-like gene family in polydnaviruses associated with wasps belonging to different Braconid subfamilies. J Gen Virol 88:92–104. https://doi.org/10.1099/vir.0.82306-0.
34. Thoetkiattikul H, Beck M, Strand M. 2005. Inhibitor kappa B-like proteins from a polydnavirus inhibit NF-kappa B activation and suppress the insect immune response. Proc Natl Acad Sci U S A 102:11426–11431. https://doi.org/10.1073/pnas.0505240102.
35. Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Schmitz R, Doronichev VB, Golovanova LV, la Rasilla de M, Fortea J, Rosas A, Paabo S. 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science 325:318–321. https://doi.org/10.1126/science.1174462.
36. Kanost MR, Arrese EL, Cao X, Chen Y-R, Chellapilla S, Goldsmith MR, Grosse-Wilde E, Heckel DG, Herndon N, Jiang H, Papanicolaou A, Qu J, Soulages JL, Vogel H, Walters J, Waterhouse RM, Ahn S-J, Almeida FC, An C, Aqrawi P, Bretschneider A, Bryant WB, Bucks S, Chao H, Chevignon G, Christen JM, Clarke DF, Dittmer NT, Ferguson LCF, Garavelou S, Gordon KHJ, Gunaratna RT, Han Y, Hauser F, He Y, Heidel-Fischer H, Hirsh A, Hu Y, Jiang H, Kalra D, Klinner C, König C, Kovar C, Kroll AR, Kuwar SS, Lee SL, Lehman R, Li K, Li Z, Liang H, Lovelace S, Lu Z, Mansfield JH, McCulloch KJ, Mathew T, Morton B, Muzny DM, Neunemann D, Ongeri F, Pauchet Y, Pu L-L, Pyrousis I, Rao X-J, Redding A, Roesel C, Sanchez-Gracia A, Schaack S, Shukla A, Tetreau G, Wang Y, Xiong G-H, Traut W, Walsh TK, Worley KC, Wu D, Wu W, Wu Y-Q, Zhang X, Zou Z, Zucker H, Briscoe AD, Burmester T, Clem RJ, Feyereisen R, Grimmelikhuijzen CJP, Hamodrakas SJ, Hansson BS, Huguet E, Jermiin LS, Lan Q, Lehman HK, Lorenzen M, Merzendorfer H, Michalopoulos I, Morton DB, Muthukrishnan S, Oakeshott JG, Palmer W, Park Y, Passarelli AL, Rozas J, Schwartz LM, Smith W, Southgate A, Vilcinskas A, Vogt R, Wang P, Werren J, Yu X-Q, Zhou J-J, Brown SJ, Scherer SE, Richards S, Blissard GW. 2016. Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. Insect Biochem Mol Biol 76:118–147. https://doi.org/10.1016/j.ibmb.2016.07.005.
37. Zuker M. 2003. Mfold web server for Nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415. https://doi.org/10.1093/nar/gkg595.
38. Augé-Gouillou C, Brillet B, Hamelin M-H, Bigot Y. 2005. Assembly of the

mariner Mos1 synaptic complex. Mol Cell Biol 25:2861–2870. https://doi.org/10.1128/MCB.25.7.2861-2870.2005.
39. Hickman AB, Dyda F. 2015. Mechanisms of DNA transposition. Microbiol Spectr 3:MDNA3–0034–2014.
40. Normand C, Duval-Valentin G, Haren L, Chandler M. 2001. The terminal inverted repeats of IS911: requirements for synaptic complex assembly and activity. J Mol Biol 308:853–871. https://doi.org/10.1006/jmbi.2001.4641.
41. Annaheim M, Lanzrein B. 2007. Genome organization of the Chelonus inanitus polydnavirus: excision sites, spacers and abundance of proviral and excised segments. J Gen Virol 88:450–457. https://doi.org/10.1099/vir.0.82396-0.
42. Zhang S, Gunaratna RT, Zhang X, Najar F, Wang Y, Roe B, Jiang H. 2011. Pyrosequencing-based expression profiling and identification of differentially regulated genes from Manduca sexta, a lepidopteran model insect. Insect Biochem Mol Biol 41:733–746. https://doi.org/10.1016/j.ibmb.2011.05.005.
43. Cui P, Löber U, Alquezar-Planas DE, Ishida Y, Courtiol A, Timms P, Johnson RN, Lenz D, Helgen KM, Roca AL, Hartman S, Greenwood AD. 2016. Comprehensive profiling of retroviral integration sites using target enrichment methods from historical koala samples without an assembled reference genome. PeerJ 4:e1847. https://doi.org/10.7717/peerj.1847.
44. Beckage NE, Tan F, Schleifer KW, Lane RD, Cherubin LL. 1994. Characterization and biological effects of *Cotesia congregata* polydnavirus on host larvae of the tobacco hornworm, *Manduca sexta*. Arch Insect Biochem Physiol 26:165–195. https://doi.org/10.1002/arch.940260209.
45. 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. https://doi.org/10.1016/j.virol.2006.09.002.
46. Gundersen-Rindal D. 2012. Chapter 8. Integration of polydnavirus DNA into host cellular genomic DNA. *In* Beckage NE, Drezen J-M (ed), Parasitoid viruses. Elsevier, Academic Press, San Diego, CA.
47. Pruijssers AJ, Strand MR. 2007. PTP-H2 and PTP-H3 from Microplitis demolitor bracovirus localize to focal adhesions and are antiphagocytic in insect immune cells. J Virol 81:1209–1219. https://doi.org/10.1128/JVI.02189-06.
48. Moreau SJM, Huguet E, Drezen J-M. 2009. Chapter 9. Polydnaviruses as tools to deliver wasp virulence factors to impair lepidopteran host immunity. *In* Rolff J, Reynolds SE (ed), Insect infection and immunity. Oxford University Press, New York, NY.
49. Harwood SH, Grosovsky AJ, Cowles EA, Davis JW, Beckage NE. 1994. An abundantly expressed hemolymph glycoprotein isolated from newly parasitized *Manduca sexta* larvae is a polydnavirus gene product. Virology 205:381–392. https://doi.org/10.1006/viro.1994.1659.
50. Espagne E, Douris V, Lalmanach G, Provost B, Cattolico L, Lesobre J, Kurata S, Iatrou K, Drezen J-M, Huguet E. 2005. A virus essential for insect host-parasite interactions encodes cystatins. J Virol 79:9765–9776. https://doi.org/10.1128/JVI.79.15.9765-9776.2005.
51. Pruijssers AJ, Falabella P, Eum JH, Pennacchio F, Brown MR, Strand MR. 2009. Infection by a symbiotic polydnavirus induces wasting and inhibits metamorphosis of the moth Pseudoplusia includens. J Exp Biol 212:2998–3006. https://doi.org/10.1242/jeb.030635.
52. Falabella P, Caccialupi P, Varricchio P, Malva C, Pennacchio F. 2006. Protein tyrosine phosphatases of Toxoneuron nigriceps bracovirus as potential disrupters of host prothoracic gland function. Arch Insect Biochem Physiol 61:157–169. https://doi.org/10.1002/arch.20120.
53. Bitra K, Suderman RJ, Strand MR. 2012. Polydnavirus Ank proteins Bind NF-κB homodimers and inhibit processing of relish. PLoS Pathog 8:e1002722. https://doi.org/10.1371/journal.ppat.1002722.
54. Bézier A, Herbinière J, Lanzrein B, Drezen J-M. 2009. Polydnavirus hidden face: the genes producing virus particles of parasitic wasps. J Invertebr Pathol 101:194–203. https://doi.org/10.1016/j.jip.2009.04.006.
55. Beckage NE, Drezen J-M. 2012. Parasitoid viruses. Elsevier, Academic Press, San Diego, CA.
56. Burke GR, Thomas SA, Eum JH, Strand MR. 2013. Mutualistic polydnaviruses share essential replication gene functions with pathogenic ancestors. PLoS Pathog 9:e1003348. https://doi.org/10.1371/journal.ppat.1003348.
57. Wetterwald C, Roth T, Kaeslin M, Annaheim M, Wespi G, Heller M, Mäser P, Roditi I, Pfister-Wilhelm R, Bézier A, Gyapay G, Drezen J-M, Lanzrein B. 2010. Identification of bracovirus particle proteins and analysis of their transcript levels at the stage of virion formation. J Gen Virol 91:2610–2619. https://doi.org/10.1099/vir.0.022699-0.
58. Demeulemeester J, De Rijck J, Gijsbers R, Debyser Z. 2015. Retroviral integration: site matters: mechanisms and consequences of retroviral integration site selection. Bioessays 37:1202–1214. https://doi.org/10.1002/bies.201500051.
59. Lesbats P, Engelman AN, Cherepanov P. 2016. Retroviral DNA integration. Chem Rev 116:12730–12757. https://doi.org/10.1021/acs.chemrev.6b00125.
60. Gauthier J, Drezen J-M, Herniou EA. 2017. The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps. Parasitology 3457:1–13. https://doi.org/10.1017/S0031182017000725.
61. Schneider SE, Thomas JH. 2014. Accidental genetic engineers: horizontal sequence transfer from parasitoid wasps to their lepidopteran hosts. PLoS One 9:e109446. https://doi.org/10.1371/journal.pone.0109446.
62. Gasmi L, Boulain H, Gauthier J, Hua-Van A, Musset K, Jakubowska AK, Aury J-M, Volkoff A-N, Huguet E, Herrero S, Drezen J-M. 2015. Recurrent domestication by Lepidoptera of genes from their parasites mediated by bracoviruses. PLoS Genet 11:e1005470. https://doi.org/10.1371/journal.pgen.1005470.
63. Drezen J-M, Josse T, Bézier A, Gauthier J, Huguet E, Herniou EA. 2017. Impact of lateral transfers on the genomes of Lepidoptera. Genes (Basel) 8:E315. https://doi.org/10.3390/genes8110315.
64. Li M, Au LYC, Douglah D, Chong A, White BJ, Ferree PM, Akbari OS. 2017. Generation of heritable germline mutations in the jewel wasp Nasonia vitripennis using CRISPR/Cas9. Sci Rep 7:901. https://doi.org/10.1038/s41598-017-00990-3.
65. Harwood SH, McElfresh JS, Nguyen A, Conlan CA, Beckage NE. 1998. Production of early expressed parasitism-specific proteins in alternate sphingid hosts of the braconid wasp *Cotesia congregata*. J Invertebr Pathol 71:271–279. https://doi.org/10.1006/jipa.1997.4745.
66. Espagne E. 2004. Genome sequence of a polydnavirus: insights into symbiotic virus evolution. Science 306:286–289. https://doi.org/10.1126/science.1103066.
67. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
68. Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16:10881–10890. https://doi.org/10.1093/nar/16.22.10881.
69. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
70. Subramanian AR, Kaufmann M, Morgenstern B. 2008. DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. Algorithms Mol Biol 3:6. https://doi.org/10.1186/1748-7188-3-6.
71. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191. https://doi.org/10.1093/bioinformatics/btp033.
72. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.
73. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033.
74. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015. HMMER web server: 2015 update. Nucleic Acids Res 43:W30–W38. https://doi.org/10.1093/nar/gkv397.
75. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
76. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol 34:2115–2122. https://doi.org/10.1093/molbev/msx148.
77. Anders S, Pyl PT, Huber W. 2015. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169. https://doi.org/10.1093/bioinformatics/btu638.