Genome-Wide Profiling of *Diadegma semiclausum* Ichnovirus Integration in Parasitized *Plutella xylostella* Hemocytes Identifies Host Integration Motifs and Insertion Sites

Ze-hua Wang[1,2], Yue-nan Zhou[1,2], Jing Yang[1,2], Xi-qian Ye[1,3], Min Shi[1,2], Jian-hua Huang[1,2,3] and Xue-xin Chen[1,2,3,4]*

[1] Institute of Insect Sciences, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China, [2] Ministry of Agriculture Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Zhejiang University, Hangzhou, China, [3] Zhejiang Provincial Key Lab of Biology of Crop Pathogens and Insects, Zhejiang University, Hangzhou, China, [4] State Key Laboratory of Rice Biology, Zhejiang University, Hangzhou, China

Polydnaviruses (PDVs), classified into two genera, bracoviruses (BVs) and ichnoviruses (IVs), are large, double-stranded DNA viruses, which are beneficial symbionts of parasitoid wasps. PDVs do not replicate in their infected lepidopteran hosts. BV circles have been demonstrated to be integrated into host genomic DNA after natural parasitization. However, the integrations of IV circles *in vivo* remain largely unknown. Here, we analyzed the integration of *Diadegma semiclausum* ichnovirus (DsIV) in the genomic DNA of parasitized *Plutella xylostella* hemocytes. We found that DsIV circles are present in host hemocytes with non-integrated and integrated forms. Moreover, DsIV integrates its DNA circles into the host genome by two distinct strategies, conservatively, and randomly. We also found that four conserved-broken circles share similar motifs containing two reverse complementary repeats at their breaking sites, which were host integration motifs (HIMs). We also predicted HIMs of eight circles from other ichnoviruses, indicating that a HIM-mediated specific mechanism was conserved in IV integrations. Investigation of DsIV circle insertion sites of the host genome revealed the enrichment of microhomologies between the host genome and the DsIV circles at integration breakpoints. These findings will deepen our understanding of the infections of PDVs, especially IVs.

Keywords: host integration motif, *Plutella xylostella*, *Diadegma semiclausum* ichnovirus, integration, Polydnavirus

INTRODUCTION

Endoparasitoid wasps, one of the most species-rich animal groups on Earth, have developed various strategies to regulate their host's physiology and development to ensure successful parasitism (Pennacchio and Strand, 2006; Ye et al., 2018). Polydnaviruses (PDVs) are obligatory symbionts of parasitoid wasps and are essential for the successful parasitism of thousands of species of parasitoid wasps. PDV particles with multiple segments of double-stranded, superhelical DNAs were

first observed about 50 years ago (Rotheram, 1967; Vinson and Scott, 1975). PDVs were classified into two genera, bracoviruses (BVs) and ichnoviruses (IVs), respectively, associated with the two largest parasitoid groups, Braconidae and Ichneumonidae (Francki et al., 1991). PDVs are present in the wasp genome as integrated proviruses (Webb, 1998; Strand and Burke, 2014). The assembly and replication of PDV virions that are injected into lepidopteran hosts during parasitoid oviposition occur only in the nuclei of ovarian calyx cells of female wasps (Gruber et al., 1996). PDVs infect most of the host immune cells and many other tissue cells as well after parasitization (Strand, 1994; Beck et al., 2007; Bitra et al., 2011), but they do not propagate themselves in the host cells.

So far, 13 PDVs, including eight BVs (Wyder et al., 2002; Espagne et al., 2004; Choi et al., 2005; Webb et al., 2006; Desjardins et al., 2008; Chen et al., 2011; Jancek et al., 2013; Yu et al., 2016) and five IVs (Webb et al., 2006; Lapointe et al., 2007; Tanaka et al., 2007; Djoumad et al., 2013; Doremus et al., 2014), have been fully sequenced since the report of the first PDV, *C. congregata* bracovirus (CcBV) (Espagne et al., 2004). The encapsidated genomes of PDVs do not contain genes coding for replication and particle production, thereby impeding particle replication in caterpillar hosts during parasitism. However, PDV circles are reported to integrate into genomic DNA of host hemocytes after natural parasitization (Chevignon et al., 2018). Virulence genes are transcribed in the infected host cells, resulting in the expression of virulence proteins that can suppress immune responses and disrupt the development of the parasitized caterpillar hosts thereby ensuring wasp development (Theilmann and Summers, 1986; Strand and Burke, 2013, 2014; Ye et al., 2018). PDV-produced miRNA is also reported to arrest host growth by modulating expression of the host ecdysone receptor (Wang et al., 2018).

Polydnaviruses do not replicate within their lepidopteran hosts, but viral genes are expressed throughout parasitism, which raised the questions of whether and how they persist in the hosts. Some of these PDV circles were suggested to persist as episomes (Strand et al., 1992; Webb and Strand, 2005) in the caterpillar host, but recent studies also revealed the presence of chromosomally integrated forms in host-derived cultured cells (Gundersen-Rindal and Dougherty, 2000; Volkoff et al., 2001; Gundersen-Rindal and Lynn, 2003; Doucet et al., 2007; Beck et al., 2011). Studies performed on BVs, especially *Glyptapanteles indiensis* BV (GiBV), *Microplitis demolitor* BV (MdBV), and CcBV, identified a motif, named host integration motif (HIM), that mediates the insertion of viral circles into the genome of the lepidopteran hosts *Lymantria dispar*, *Pseudoplusia includens*, and *Manduca sexta*, respectively (Gundersen-Rindal and Lynn, 2003; Beck et al., 2011; Herniou et al., 2013; Chevignon et al., 2018). Two MdBV circles and eight CcBV circles have been formally demonstrated to be integrated into host genomic DNA after natural parasitization (Beck et al., 2011; Chevignon et al., 2018). As to the integrations of IVs, only *Tranosema rostrale* ichnovirus (TrIV) circle F was reported to integrate into the genomic DNA of the host *Choristoneura fumiferana* CF-124T cells *in vitro* (Doucet et al., 2007). However, the integrations of IV circles *in vivo* remain largely unknown.

In this study, we analyzed the integration of *Diadegma semiclausum* ichnovirus (DsIV), a PDV of the wasp *D. semiclausum* (Haliday) that is a larval parasitoid of the diamondback moth, *Plutella xylostella* (Linnaeus), one of the most important pests of cruciferous crops worldwide. We found that DsIV circles persist in parasitized *P. xylostella* hemocytes with two different forms, circular and integrated, through high-throughput sequencing analysis. The integrated DsIV circles integrate their DNA into the host genome by two distinct strategies, conservatively, and randomly. We identify four HIMs from DsIV and predict eight HIMs from other ichnoviruses and show that the HIMs of ichnoviruses have two pairs of boundary sequences forming reverse complementary repeats. We further found that the integrations of DsIV circles show a preference for the host genome regions that contained overlapping sequences of their HIMs.

MATERIALS AND METHODS

Insect Rearing and Parasitization

Plutella xylostella and its endoparasitoid *D. semiclausum* were reared as previously described (Huang et al., 2008). They were maintained at 25 ± 1°C, with 65% relative humidity, and a 14-h light/10-h dark cycle. Adult wasps were fed with 20% honey/water (V/V). Late 3rd instar *P. xylostella* host larvae were individually exposed to a single *D. semiclausum* female within a 10 mm × 80 mm tube to ensure 100% parasitization.

Genome Resequencing of Parasitized *P. xylostella* Hemocytes

Hemocytes of about 500 parasitized *P. xylostella* larvae at 24 h post parasitization (pp) were collected as one group. Genomic DNA from three independent replicates was isolated using the Puregene Core kit (Qiagen). DNA concentration was assessed by NanoDrop® spectrophotometers (Thermo Fisher, MA, United States). A total amount of 1 μg genomic DNA per group was used as input for the library preparation. The sequencing libraries were generated using the VAHTS Universal DNA Library Prep Kit for Illumina® (Vazyme, Nanjing, China) following the manufacturer's recommendations and were sequenced on an Illumina HiSeq X Ten platform with 150 bp paired-end module.

Read Mapping and Data Analysis

A total of 1.84 billion clean reads were obtained from the three Illumina runs (**Supplementary Table 1**). All clean reads were mapped against the DsIV genome (GenBank No. KF156214–KF156260) using BLASTN (E-value $<10^{-5}$). A total of 1,256,800 reads were mapped to 47 DsIV circles (**Supplementary Table 2**). In a stringent analysis, only reads mapped with at least 20 nucleotides were kept to avoid incorrect mapping due to short alignments. DsIV-related reads were then mapped against the *P. xylostella* genome (You et al., 2013) to identify specifically chimeric reads. Again, only reads that mapped to the *P. xylostella* genome with a size above 20 nucleotides were kept. A qualified

chimeric read must contain both DsIV and *P. xylostella* sequences, and the two sequences are on the opposite sides. The sum of the base numbers of their mapped sequences should be more than 140 bp. A total of 1,435 chimeric reads corresponding to DsIV circles were obtained based on the strict filtering criteria (**Table 1**). We also performed the mapping analysis by using Burrows–Wheeler alignment (BWA) tools, version 0.7.10, and only 797 chimeric reads were sorted out. While comparing the dataset with the results from BLASTN analysis, the 797 chimeric reads were all covered. Since the 1,435 chimeric reads were confirmed to have both the virus and the caterpillar genome sequences, we decided to use the dataset with the large amount to do the further analysis. Reads were then mapped again to the DsIV genome to identify junction sites of DsIV circles and to the *P. xylostella* genome to identify integration sites. For an example, if the 1–51 bases of a chimeric read were mapped to the circle of DsIV-15 (2048–1988 nt) and the 50–150 bases were mapped to the scaffold_206 of *P. xylostella* genome (80256–80156 nt), it means that DsIV-15 viral circle is linearized at the position of 1987–1988 nt and it integrates into host genome at the site of 80256–80257 nt within the scaffold_206. The sequences of DsIV and *P. xylostella* host within the same chimeric read are shown in **Supplementary Data 1**.

To test the robustness of the method and its ratio of false positives, we additionally analyzed the reads from a non-parasitized host with the same algorithm. We screened the chimeric reads among five datasets of the whole genome of non-parasitized *P. xylostella* with NCBI SRA numbers: ERR2508315,

ERR2508316, ERR2508317, ERR2508318, and ERR2512126. The results showed only three reads were mapped to DsIV genome, which confirms that our data is accurate and the BLASTN method is suitable in this study.

The Percent of Each Integrated DsIV Circle

We first estimated the depth for each DsIV circle according to the DsIV-related reads before calculating the percent of each integrated DsIV circle. The depth was calculated as follows: x = (the number of each DsIV-related reads × 150 bp) / the size of each DsIV circle. When one DsIV circle integrates into a host genome, two junction sites will be produced. So, the percent of each integrated DsIV circle is calculated as follows: % = (the number of chimeric reads / 2) / the depth of each circle × 100. The threshold used to differentiate circle forms was defined as 1%. For each integrated circle, the ratio of conservatively integrated form = the number of chimeric reads indicating the same junction site of each DsIV circle / the total chimeric reads of each DsIV circle.
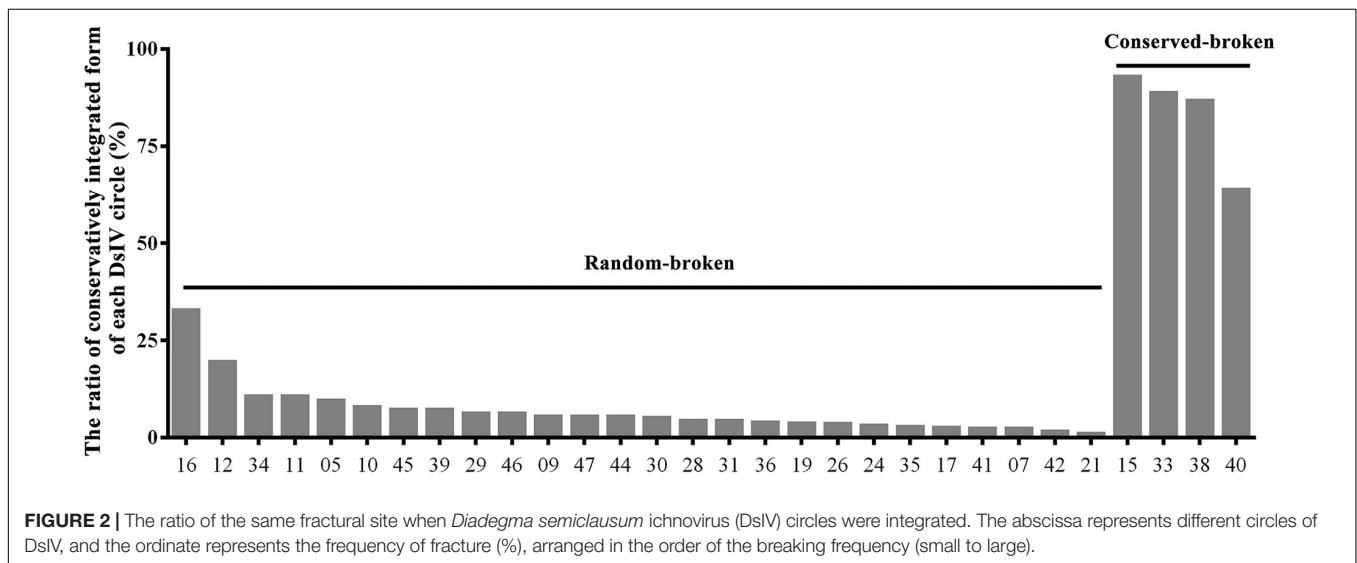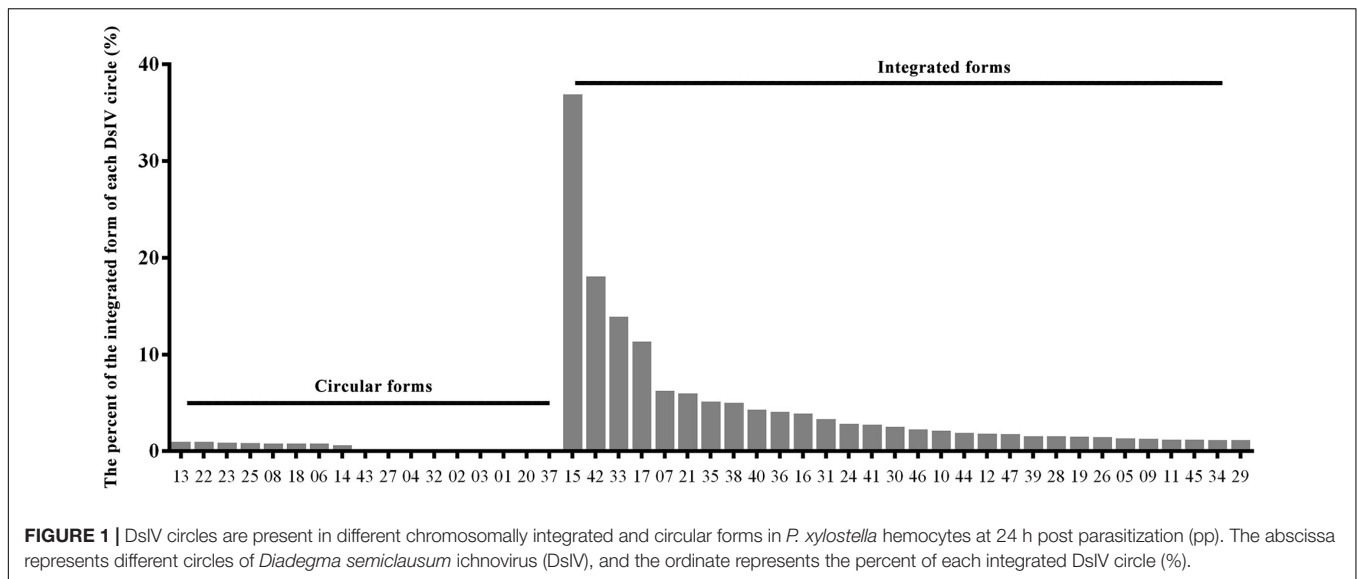
Verification of DsIV Host Integration Motifs

Two DsIV circles (DsIV-15 and DsIV-40) were chosen to confirm the locations of HIMs identified by chimeric reads according to PCR-based detection (Beck et al., 2011); DsIV-13 and DsIV-21 lacking HIMs were used as controls. Briefly, DsIV-13, 15, 21, and 40 were divided into 4, 4, 5, and 5 amplicons, respectively, by designing overlapping primer pairs (**Supplementary Table 4**) that specifically amplified regions of different DsIV circles (**Figure 3A**). Genomic DNA from hemocytes of parasitized *P. xylostella* (24 h pp) was used as a template, while genomic DNA isolated from female *D. semiclausum* ovaries was used as a control. PCRs were then run in 20-µl reaction mixtures containing 0.2 µM of each specific primer, 10 ng of template DNA, and 1 unit of LA Taq polymerase (TaKaRa, Tokyo, Japan). Cycling conditions were as follows: initial denaturation step at 94°C for 2 min, followed by 30 cycles of denaturation at 94°C for 20 s, annealing at 55°C for 20 s, and extension at 72°C for 2 min, with a final extension step at 72°C for 10 min. The amplified products were analyzed on 1.0% agarose gels.

Identification of Candidate HIMs in Other Ichnovirus Circles

Host integration motif sequences of four DsIV circles (DsIV-15, 33, 38, and 40) were used to search for similar motifs among *Tranosema rostrale* IV (TrIV), *Hyposoter fugitivus* IV (HfIV), *Glypta fumiferanae* IV (GfIV), *Campoletis sonorensis* IV (CsIV), and *Apophua simplicipes* IV (AsIV) circles by BLASTN analyses (*E*-value < 10, identity >80%, and length >15). The database used for BLASTN analyses contained 41 segments of TrIV, 56 segments of HfIV, 106 segments of GfIV, 24 segments of CsIV, and 226 segments of AsIV, which were downloaded from NCBI and the accession numbers are listed in **Supplementary Table 5**. Alignment analyses of HIM regions were performed with MEGA

TABLE 1 | The numbers of chimeric reads for different DsIV circles.

DsIV ID	Related reads	DsIV ID	Related reads
DsIV-01	0	DsIV-25	14
DsIV-02	8	DsIV-26	25
DsIV-03	3	DsIV-27	1
DsIV-04	12	DsIV-28	21
DsIV-05	10	DsIV-29	15
DsIV-06	17	DsIV-30	18
DsIV-07	36	DsIV-31	21
DsIV-08	5	DsIV-32	1
DsIV-09	17	DsIV-33	129
DsIV-10	12	DsIV-34	9
DsIV-11	9	DsIV-35	31
DsIV-12	5	DsIV-36	23
DsIV-13	1	DsIV-37	0
DsIV-14	7	DsIV-38	47
DsIV-15	541	DsIV-39	13
DsIV-16	3	DsIV-40	42
DsIV-17	33	DsIV-41	35
DsIV-18	14	DsIV-42	48
DsIV-19	24	DsIV-43	3
DsIV-20	0	DsIV-44	17
DsIV-21	69	DsIV-45	13
DsIV-22	6	DsIV-46	15
DsIV-23	17	DsIV-47	17
DsIV-24	28		

FIGURE 1 | DsIV circles are present in different chromosomally integrated and circular forms in *P. xylostella* hemocytes at 24 h post parasitization (pp). The abscissa represents different circles of *Diadegma semiclausum* ichnovirus (DsIV), and the ordinate represents the percent of each integrated DsIV circle (%).

FIGURE 2 | The ratio of the same fractural site when *Diadegma semiclausum* ichnovirus (DsIV) circles were integrated. The abscissa represents different circles of DsIV, and the ordinate represents the frequency of fracture (%), arranged in the order of the breaking frequency (small to large).

7.0 software, and the visualizations of the alignments were made using Jalview (Waterhouse et al., 2009).
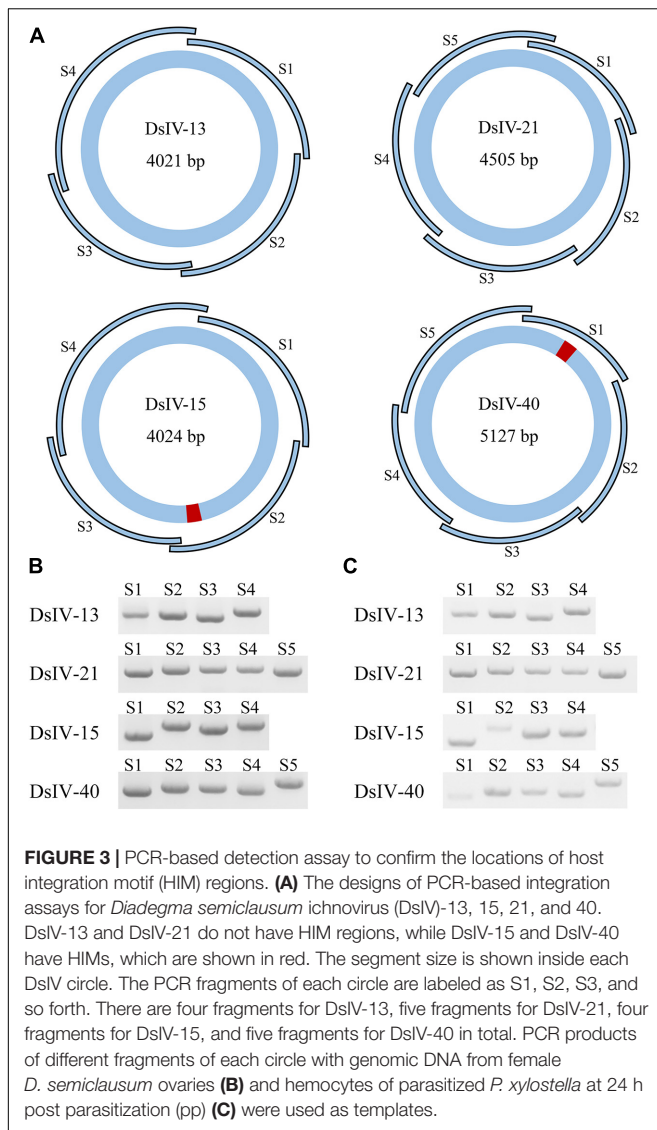
Data Availability

The raw data for genome resequencing of parasitized *P. xylostella* hemocytes have been deposited at the SRA database of NCBI with accession numbers SRR11880655, SRR11880656, and SRR11880657. The sequences of each chimeric read are shown in **Supplementary Data 1**.

RESULTS

DsIV Circles Are Present in Two Different Forms in *P. xylostella* Hemocytes

Genomic DNAs were isolated from hemocytes of parasitized *P. xylostella* larvae and were deep-sequenced to analyze the

integration pattern of each DsIV circle. A total of 1.84 billion clean reads were obtained from three independent experiments (**Supplementary Table 1**). Among them, 1,256,800 reads were aligned to 47 DsIV circles (**Supplementary Table 2**). The number of DsIV-related reads mapping to each DsIV circle was variable (**Supplementary Table 2**). For example, 214,609 reads were aligned to DsIV-02, while only 877 reads were aligned to DsIV-16 (**Supplementary Table 2**). DsIV-related reads were then mapped against the *P. xylostella* genome to identify chimeric reads. A total of 1,435 chimeric reads containing both nucleotides of DsIV sequence and *P. xylostella* sequence were sorted out (**Table 1** and Supplementary Data 1). We found that the numbers of chimeric reads mapping to different circles varied widely (**Table 1**). For example, the number of chimeric reads mapping to DsIV-15 was 541, while no chimeric reads mapped to DsIV-01, 20, or 37 (**Table 1**). The circles for which a higher number of chimeric reads was found did not correspond to circles that are

FIGURE 3 | PCR-based detection assay to confirm the locations of host integration motif (HIM) regions. **(A)** The designs of PCR-based integration assays for *Diadegma semiclausum* ichnovirus (DsIV)-13, 15, 21, and 40. DsIV-13 and DsIV-21 do not have HIM regions, while DsIV-15 and DsIV-40 have HIMs, which are shown in red. The segment size is shown inside each DsIV circle. The PCR fragments of each circle are labeled as S1, S2, S3, and so forth. There are four fragments for DsIV-13, five fragments for DsIV-21, four fragments for DsIV-15, and five fragments for DsIV-40 in total. PCR products of different fragments of each circle with genomic DNA from female *D. semiclausum* ovaries **(B)** and hemocytes of parasitized *P. xylostella* at 24 h post parasitization (pp) **(C)** were used as templates.

more abundant within the particles (**Supplementary Table 2**). As the number of chimeric reads is related to the integration efficiency of the circles, we determined the integrations of DsIV circles by analyzing the percentage of each integrated DsIV circle. We found that DsIV circles are present in two different forms in *P. xylostella* hemocytes at 24 h pp. There were 17 "non-integrated" circles (**Figure 1**). On the contrary, the other 30 DsIV circles were present in both circular and integrated forms (**Figure 1**).

Two Distinct Integration Strategies of DsIV Circles

Based on the alignment of chimeric reads on the DsIV genome, we could easily figure out the junction sites for integrated DsIV circles (**Supplementary Figure 1**). The integrated circles can be divided into two categories according to the ratio of conservatively integrated forms of each DsIV circle. One type we referred as "conserved-broken circle" (ratio >50%), which means

that the DsIV circles are linearized at a particular site of their sequence, and the other type we called "random-broken circle" (ratio <50%), which means that the DsIV circles are linearized randomly (**Figure 2**). Specifically, four DsIV circles (15, 33, 38, and 40) had the particular site of their circles, which were named as conserved-broken circles (CBCs) (**Figure 2**). The remaining 26 integrated circles were linearized completely randomly during integration, and therefore, they were named "random-broken circles" (RBCs) (**Figure 2**). CBCs can also integrate randomly into the host genome (**Supplementary Table 3**). The data combined with the results from alignments of chimeric reads and the DsIV genome show that CBC integration is associated with the deletion of a stretch of nts (32 to 311 bp) in each DsIV circle (**Table 2**).

Validation of HIM Locations Using a PCR-Based Detection Assay

Previous studies performed on BVs identified a motif, HIM, that mediated the insertion of viral circles into their host genomes. We selected two CBCs (DsIV-15 and DsIV-40) and used a PCR-based assay to validate the locations of their HIMs identified by chimeric reads. One "non-integrated" circle (DsIV-13) and one RBC (DsIV-21) were selected as negative controls (devoid of HIM). Primers were designed to amplify specific regions from each circle (**Figure 3A**). When using the genomic DNA of wasp ovaries as a template, PCR products were obtained for all pairs of primers (**Figure 3B**). In contrast, very few amplicons were obtained for DsIV-15-S2 and DsIV-40-S1 regions using templates isolated from host hemocytes 24 h pp (**Figure 3C**). The results suggest that DsIV-15 is disrupted into linear DNAs at its S2 region, and DsIV-40 is disrupted at its S1 region during integration, which was expected because DsIV-15 S2 and DsIV-40 S1 contain the HIM sequences.

The Structure of HIMs From Ichnoviruses

We investigated the structure of DsIV HIMs from four CBCs. Alignment analysis showed that they had two pairs of boundary sequences forming reverse complementary repeats of 11 nts (CCGTACGCTCT and AGAGCGTACGG) and 6 nts (ACTGTA and TACAGT) constituting the borders of the insertions (**Figure 4A**). Further, the DsIV HIM sequences were used to identify the candidate HIMs in other IV circles from TrIV, HfIV, GfIV, CsIV, and AsIV. Finally, we identified candidate HIMs from five HfIV and three TrIV circles (**Table 3**). The distance between two putative junction sites of each circle ranges from 32 to 1,781 bp (**Table 3**), which is similar to what is observed in DsIV. Among circles containing predicted HIMs, TrIV-F1 was reported to be integrated into *Choristoneura fumiferana* CF-124T cells (Doucet et al., 2007), and the breaking site of TrIV-F1 is consistent with our results (**Figure 4B**). Alignment analysis of eight predicted HIMs showed that they also had two pairs of boundary sequences forming similar reverse complementary repeats (**Figure 4B**).
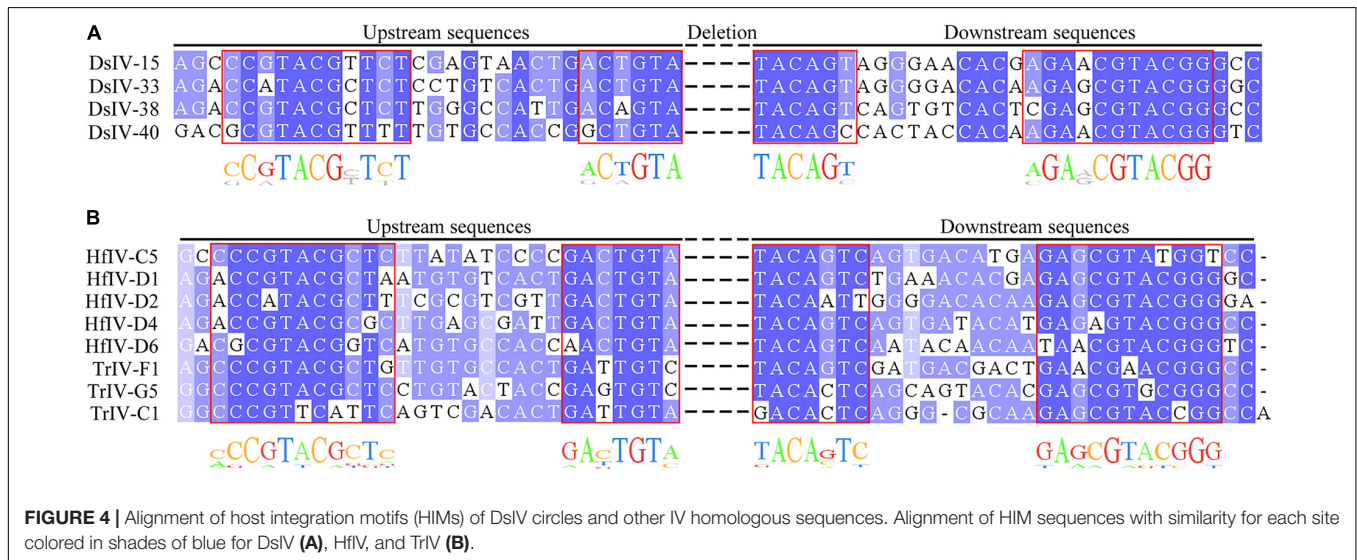
The Integration Sites of CBCs

To determine whether DsIV circle integration occurred randomly or in preferential regions of the *P. xylostella* genome,

TABLE 2 | Sites where DsIV circles are linearized.

DsIV ID	DsIV size (bp)	Accession numbers	Upstream DsIV nt position	Downstream DsIV nt position	Length (nt)
DsIV-15	4,024	KF156228.1	1943–1944	1997–1998	54
DsIV-33	5,277	KF156246.1	1058–1059	1090–1091	32
DsIV-38	4,756	KF156251.1	2391–2392	2702–2703	311
DsIV-40	5,127	KF156253.1	373–374	449–450	76

TABLE 3 | Information of the predicted host integration motif of other IVs.

Ichnoviruses	Wasp hosts	Circle ID	Accession numbers	Putative junction sites		Putative deletion
				Upstream nt position	Downstream nt position	(nt)
HfIV	*H. fugitivus*	C5	AB291183.1	920–921	952–953	33
		D1	AB291196.1	854–855	886–887	33
		D2	AB291197.1	1768–1769	1800–1801	33
		D4	AB291199.1	4556–4557	4600–4601	45
		D6	AB291200.1	364–365	441–442	78
TrIV	*T. rostrale*	F1	AF421353.1	2194–2195	2226–2227	33
		G5	AB291163.1	5094–5095	6874–6875	1,781
		C1	AY940454.1	1456–1457	1487–1488	32

FIGURE 4 | Alignment of host integration motifs (HIMs) of DsIV circles and other IV homologous sequences. Alignment of HIM sequences with similarity for each site colored in shades of blue for DsIV **(A)**, HfIV, and TrIV **(B)**.

we analyzed the integration sites of four CBCs, which can be obtained according to the chimeric reads. However, we did not observe any regions of the *P. xylostella* genome in which DsIV circle integration preferentially occurred. We anticipated whether there is a specific shared motif in the *P. xylostella* genome near the different insertion sites for any DsIV circles. We analyzed 499 chimeric reads of DsIV-15, 115 chimeric reads of DsIV-33, 38 chimeric reads of DsIV-38, and 27 chimeric reads of DsIV-40. As shown in **Figure 5A**, we found that there was generally an overlapping sequence in the middle of each chimeric read. We counted and classified the overlapping sequences of chimeric reads. The results showed that these four circles preferred to integrate into the host genome regions that contained overlapping sequences of their HIM regions (**Figure 5**).

DISCUSSION

Polydnaviruses are large, double-stranded DNA viruses that constitute a unique virus family, Polydnaviridae, and are classified into two genera: bracoviruses (BVs) and ichnoviruses (IVs) (Francki et al., 1991). PDVs influence two major areas of host biology: immunity and development (Kim et al., 2013; Ignesti et al., 2018; Wang et al., 2018; Ye et al., 2018), which are essential for the successful parasitism of parasitoid wasps. Though the functions of BVs and IVs injected into their hosts during parasitism are similar, there are many differences between them, including morphology and origins (Gundersen-Rindal et al., 2013; Strand and Burke, 2014; Ye et al., 2014).

The DNA circles in BV or IV virions injected into hosts do not contain genes coding for particle production, thereby
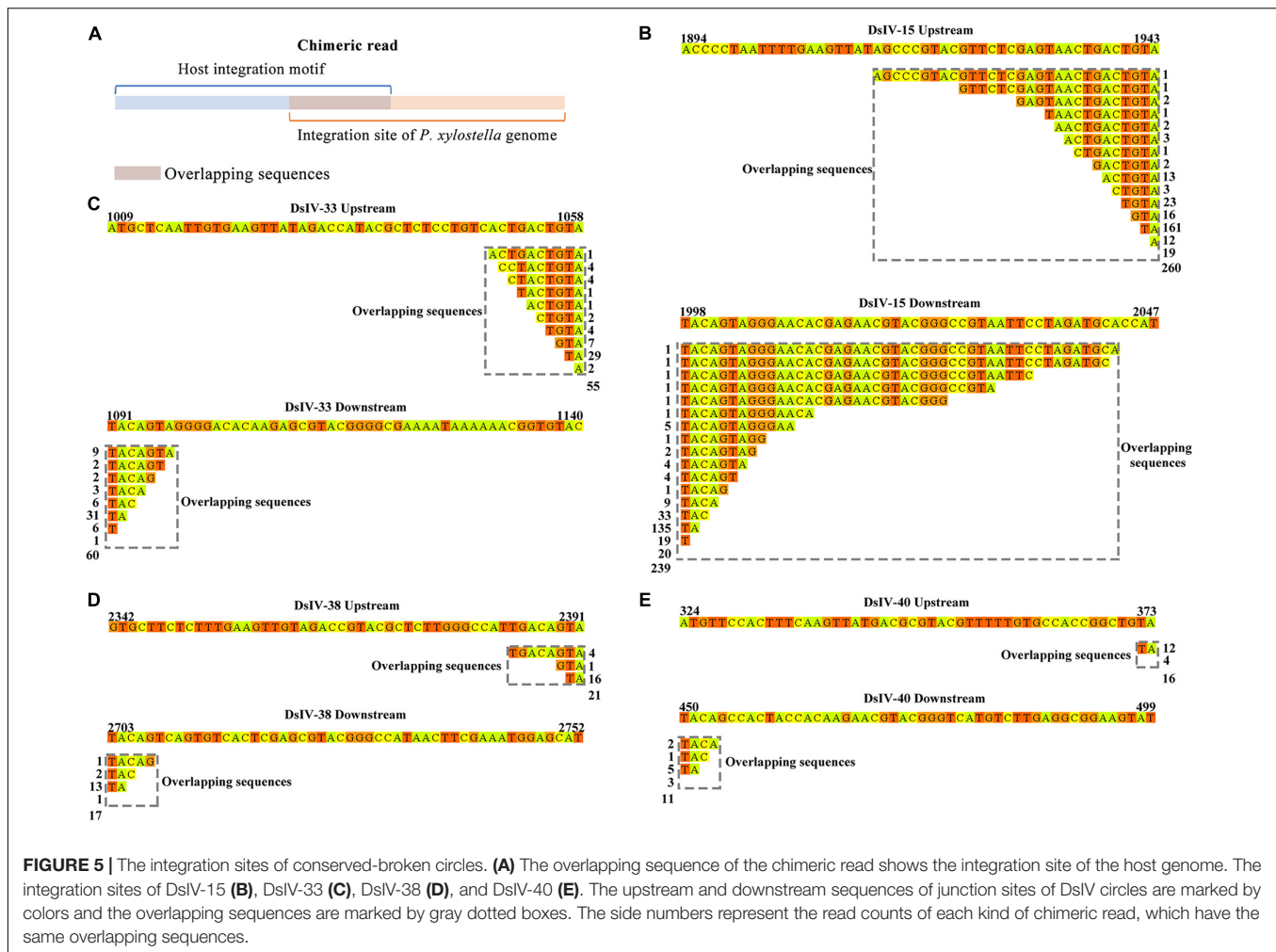
FIGURE 5 | The integration sites of conserved-broken circles. **(A)** The overlapping sequence of the chimeric read shows the integration site of the host genome. The integration sites of DsIV-15 **(B)**, DsIV-33 **(C)**, DsIV-38 **(D)**, and DsIV-40 **(E)**. The upstream and downstream sequences of junction sites of DsIV circles are marked by colors and the overlapping sequences are marked by gray dotted boxes. The side numbers represent the read counts of each kind of chimeric read, which have the same overlapping sequences.

impeding viral replication, which raised the question of how viral DNA persists in lepidopteran hosts. In addition, the question of persistence was also raised by the expression of the genes throughout parasitism. Some PDV circles were suggested to persist as episomes (Strand et al., 1992; Webb and Strand, 2005), and other pieces of evidence indicate that part of these PDV circles integrate into the genomes of host insect cells *in vitro* or *in vivo* (Gundersen-Rindal and Dougherty, 2000; Gundersen-Rindal and Lynn, 2003; Doucet et al., 2007; Beck et al., 2011; Chevignon et al., 2018). Eight out of nine tested circles of CcBV were integrated in *M. sexta* hemocyte genomic DNA (Beck et al., 2011; Chevignon et al., 2018), which indicates that some other circles of CcBV may be present in lepidopteran hosts as non-integrated forms. We analyzed the integration of 47 DsIV circles in parasitized *P. xylostella* hemocytes using high-throughput sequencing and found that DsIV circles persist in *P. xylostella* at 24 h pp in two forms, i.e., non-integrated and integrated forms.

It should be noted that there were huge variations in the numbers of aligned reads mapping to each DsIV circle. This situation indirectly reflects the various abundance of DsIV circles in wasp ovaries as reported for MdBV and CcBV (Beck et al., 2007; Chevignon et al., 2014). Though no relationship between

circle abundance and existing forms in lepidopteran hosts was observed, we found that the top nine most abundant DsIV circles persisted in the non-integrated form at 24 h pp. We only detected the integrations of DsIV circles in the host hemocytes at 24 h pp, which neglects the integrations in other tissues or at other time points.

It was reported that integrations of BV circles involve a HIM (Beck et al., 2007; Chevignon et al., 2014). Interestingly, we found that 26 DsIV circles integrated into the *P. xylostella* genome by two distinct strategies, conservatively and randomly. However, the HIMs only mediated the integrations of four CBCs while the remaining 26 DsIV circles had no HIMs and were broken randomly during integration, which we named RBCs. In this study, the number of chimeric reads from 4 CBCs is 759 and the number of chimeric reads from 26 RBCs is 567, which suggest that the efficiency of HIM-mediated integrations of CBCs is higher than that of RBCs. In addition, we also get extra 109 chimeric reads for the remaining 17 DsIV circles. Due to the limitations of the few number of the reads, we cannot predict the integration models for those 17 circles. Alternatively, we did try Manta (Version1.6.0) for identifying the integration sites in host genome, but it failed probably because of the very low frequency

of integration events of DsIV circles. Further studies may be needed to figure out why DsIV has such a low insertion frequency and what are the integration patterns for all DsIV circles. Till now, HIMs were identified in several BV species, including GiBV, MdBV, and CcBV (Gundersen-Rindal and Lynn, 2003; Beck et al., 2011; Herniou et al., 2013; Chevignon et al., 2018). However, as to the integrations of IVs, only one TrIV circle was reported as integrated into its host genome (Doucet et al., 2007), which makes it impossible to reveal the conserved structure of IV HIMs. We identified four HIMs of DsIV circles and thus found that the HIMs had two pairs of boundary sequences forming reverse complementary repeats constituting the borders of the insertions according to the alignment analysis. The reverse complementary repeats also exist in the HIMs from BVs (Beck et al., 2011; Chevignon et al., 2018).

However, the reverse complementary repeats of BV HIMs are different from those of IV HIMs (**Supplementary Figure 2**). In particular, both CcBV and MdBV HIMs share similar reverse complementary repeats consisting of palindromic sequences of 9 bp (GAAAATTTC and GAAATTTTC) and 5 bp (CTAGT and ACTAG) in MdBV and 8 bp (TAAATTTC and GAAATTTA) and 5 bp (CTGGT and ACCAG) in CcBV (Chevignon et al., 2018). However, the reverse complementary repeats in the HIMs of DsIV circles consist of 11 bp (CCGTACGCTCT and AGAGCGTACGG) and 6 bp (ACTGTA and TACAGT). During this study, we also identified eight candidate HIMs from HfIV and TrIV. Moreover, the HIM of TrIV-F1 mediates the integration that was reported (Doucet et al., 2007). It is unsurprising that similar HIMs are not found for GfIV and AsIV, whose associated wasps belong to Banchinae. However, it is inexplicable that HIMs are not found for CsIV, whose associated wasp belongs to Campopleginae. The genome resequencing of MdBV identified 10 additional circles (Burke et al., 2014). Thus, the resequencing of the CsIV genome may find circles containing similar HIMs. However, the biological relevance of the structures of HIMs is uncertain. It is hypothesized that the HIMs more probably correspond to similar protein-binding sites, resulting in the assembly of a nucleoprotein complex (Chevignon et al., 2018). Further work is required to uncover the mechanism of PDV integrations mediated by HIMs.

Data obtained on the integration of CcBV shows that the insertion events are widespread in *M. sexta* hemocyte DNA (Chevignon et al., 2018), which is consistent with results found in DsIV based on the analysis of chimeric reads. However, we reveal a specific shared motif in the *P. xylostella* genome near the different insertion sites for DsIV circles, which was not found in the integrations of CcBV (Chevignon et al., 2018). In our dataset, we observed a significant enrichment of microhomologies between the host genome and the DsIV circles at integration breakpoints, which is similar to what was observed in human papillomavirus integration (Hu et al., 2015).

In summary, our results demonstrated that the DsIV circles integrate into the host genome by two distinct strategies. HIMs were identified from IVs, which could mediate integration of these particular circles. Furthermore, the enrichment of microhomologies between the host genome and the DsIV circles at integration breakpoints was observed. These

findings will deepen the understanding of how PDV circles persist in the hosts.

SUMMARY

Parasitoid wasps are a species-rich group of animals that live in or on other arthropods. Polydnaviruses (PDVs), divided into two genera, bracoviruses, and ichnoviruses, are double-stranded DNA viruses associated with parasitic wasps (primary hosts), which do not replicate in their infected caterpillar hosts (secondary hosts). During parasitoid oviposition, PDVs enter infected secondary hosts, triggering expression of virulence genes, which manipulate multiple biological processes of hosts to fulfill all the requirements of parasitoid offspring. Bracovirus circles are integrated into host genomic DNA after natural parasitization. For ichnoviruses, the ability of the viral molecules to integrate in lepidopteran cell lines has also been shown, but there was no knowledge on what succeeds *in vivo*. We found that ichnovirus circles were present in hosts with non-integrated and integrated forms, and ichnoviruses integrated DNA circles into the host genome by two distinct strategies, conservatively and randomly. We identified the breaking sites of ichnovirus circles and the integration sites in the host genome. Our work shows, for the first time, the integration of ichnovirus molecules in the genome of the lepidopteran host following parasitism, which will deepen our understanding of how PDV circles persist in hosts and integrate into host genomic DNA.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, SRR11880655, https://www.ncbi.nlm.nih.gov/, SRR11880656, https://www.ncbi.nlm.nih.gov/, SRR11880657.

AUTHOR CONTRIBUTIONS

X-xC, J-hH, and MS designed the study. Z-hW, Y-nZ, and X-qY performed the experiments and analyzed the data. Z-hW and X-xC wrote the manuscript. All authors reviewed the manuscript, contributed to the article, and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

REFERENCES

Beck, M. H., Inman, R. B., and Strand, M. R. (2007). *Microplitis demolitor* bracovirus genome segments vary in abundance and are individually packaged in virions. *Virology* 359, 179–189. doi: 10.1016/j.virol.2006.09.002

Beck, M. H., Zhang, S., Bitra, K., Burke, G. R., and Strand, M. R. (2011). The encapsidated genome of *Microplitis demolitor* bracovirus integrates into the host *Pseudoplusia includens*. *J. Virol.* 85, 11685–11696. doi: 10.1128/jvi.05726-11

Bitra, K., Zhang, S., and Strand, M. R. (2011). Transcriptomic profiling of *Microplitis demolitor* bracovirus reveals host, tissue and stage-specific patterns of activity. *J. Gen. Virol.* 92, 2060–2071. doi: 10.1099/vir.0.032680-0

Burke, G. R., Walden, K. K. O., Whitfield, J. B., Robertson, H. M., and Strand, M. R. (2014). Widespread genome reorganization of an obligate virus mutualist. *PLoS Genet.* 10:e1004660. doi: 10.1371/journal.pgen.1004660

Chen, Y. F., Gao, F., Ye, X. Q., Wei, S. J., Shi, M., Zheng, H. J., et al. (2011). Deep sequencing of *Cotesia vestalis* bracovirus reveals the complexity of a polydnavirus genome. *Virology* 414, 42–50. doi: 10.1016/j.virol.2011.03.009

Chevignon, G., Periquet, G., Gyapay, G., Vega-Czarny, N., Musset, K., Drezen, J. M., et al. (2018). *Cotesia congregata* Bracovirus circles encoding PTP and Ankyrin genes integrate into the DNA of parasitized *Manduca sexta* hemocytes. *J. Virol.* 92:15. doi: 10.1128/JVI.00438-18

Chevignon, G., Theze, J., Cambier, S., Poulain, J., Da Silva, C., Bezier, A., et al. (2014). Functional annotation of *Cotesia congregata* Bracovirus: Identification of viral genes expressed in parasitized host immune tissues. *J. Virol.* 88, 8795–8812. doi: 10.1128/jvi.00209-14

Choi, J. Y., Roh, J. Y., Kang, J. N., Shim, H. J., Woo, S. D., Jin, B. R., et al. (2005). Genomic segments cloning and analysis of *Cotesia plutellae* polydnavirus using plasmid capture system. *Biochem. Biophys. Res. Commun.* 332, 487–493. doi: 10.1016/j.bbrc.2005.04.146

Desjardins, C. A., Gundersen-Rindal, D. E., Hostetler, J. B., Tallon, L. J., Fadrosh, D. W., Fuester, R. W., et al. (2008). Comparative genomics of mutualistic viruses of *Glyptapanteles* parasitic wasps. *Genome Biol.* 9, 1–17.

Djoumad, A., Stoltz, D., Béliveau, C., Boyle, B., Kuhn, L., and Cusson, M. (2013). Ultrastructural and genomic characterization of a second banchine polydnavirus confirms the existence of shared features within this ichnovirus lineage. *J. Gen. Virol.* 94, 1888–1895. doi: 10.1099/vir.0.052506-0

Doremus, T., Cousserans, F., Gyapay, G., Jouan, V., Milano, P., Wajnberg, E., et al. (2014). Extensive transcription analysis of the *Hyposoter didymator* ichnovirus genome in permissive and non-permissive lepidopteran host species. *PLoS One* 9:8. doi: 10.1371/journal.pone.0104072

Doucet, D., Levasseur, A., Beliveau, C., Lapointe, R., Stoltz, D., and Cusson, M. (2007). In vitro integration of an ichnovirus genome segment into the genomic DNA of lepidopteran cells. *J. Gen. Virol.* 88, 105–113. doi: 10.1099/vir.0.82314-0

Espagne, E., Dupuy, C., Huguet, E., Cattolico, L., Provost, B., Martins, N., et al. (2004). Genome sequence of a polydnavirus: insights into symbiotic virus evolution. *Science* 306, 286–289. doi: 10.1126/science.1103066

Francki, R. I. B., Fauquet, C. M., Knudson, D. L., and Brown, F. (1991). *Classification and nomenclature of viruses:fifth report of the international committee on taxonomy of viruses for virology division of the international union of microbiological societies*. Berlin: Springer-Verlag Berlin.

Gruber, A., Stettler, P., Heiniger, P., Schümperli, D., and 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. doi: 10.1099/0022-1317-77-11-2873

Gundersen-Rindal, D. E., and Lynn, D. E. (2003). Polydnavirus integration in lepidopteran host cells in vitro. *J. Insect Physiol.* 49, 453–462. doi: 10.1016/s0022-1910(03)00062-3

Gundersen-Rindal, D., and Dougherty, E. M. (2000). Evidence for integration of *Glyptapanteles indiensis* polydnavirus DNA into the chromosome of *Lymantria dispar* in vitro. *Virus Res.* 66, 27–37. doi: 10.1016/s0168-1702(99)00125-2

Gundersen-Rindal, D., Dupuy, C., Huguet, E., and Drezen, J. (2013). Parasitoid polydnaviruses: evolution, pathology and applications. *Biocontrol Sci. Technol.* 23, 1–61. doi: 10.1080/09583157.2012.731497

Herniou, E. A., Huguet, E., Theze, J., Bezier, A., Periquet, G., and Drezen, J. M. (2013). When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.* 368:20130051. doi: 10.1098/rstb.2013.0051

Hu, Z., Zhu, D., Wang, W., Li, W., Jia, W., Zeng, X., et al. (2015). Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat. Genet.* 47, 158–163. doi: 10.1038/ng.3178

Huang, F., Cao, T. T., Shi, M., Chen, Y. F., and Chen, X. X. (2008). Parasitism-induced effects on host growth and metabolic efficiency in *Plutella xylostella* larvae parasitized by *Cotesia vestalis* or *Diadegma semiclausum*. *Insect Sci.* 15, 237–243. doi: 10.1111/j.1744-7917.2008.00206.x

Ignesti, M., Ferrara, R., Romani, P., Valzania, L., Serafini, G., Pennacchio, F., et al. (2018). A polydnavirus-encoded ANK protein has a negative impact on steroidogenesis and development. *Insect. Biochem. Mol. Biol.* 95, 26–32. doi: 10.1016/j.ibmb.2018.03.003

Jancek, S., Bezier, A., Gayral, P., Paillusson, C., Kaiser, L., Dupas, S., et al. (2013). Adaptive selection on bracovirus genomes drives the specialization of *Cotesia* parasitoid wasps. *PLoS One* 8:5. doi: 10.1371/journal.pone.0064432

Kim, J., Hepat, R., Lee, D., and Kim, Y. (2013). Protein tyrosine phosphatase encoded in *Cotesia plutellae* bracovirus suppresses a larva-to-pupa metamorphosis of the diamondback moth, *Plutella xylostella*. *Comp. Biochem. Physiol. Mol. Integr. Physiol.* 166, 60–69. doi: 10.1016/j.cbpa.2013.04.025

Lapointe, R., Tanaka, K., Barney, W. E., Whitfield, J. B., Banks, J. C., Béliveau, C., et al. (2007). Genomic and morphological features of a banchine polydnavirus: Comparison with bracoviruses and ichnoviruses. *J. Virol.* 81, 6491–6501. doi: 10.1128/jvi.02702-06

Pennacchio, F., and Strand, M. R. (2006). Evolution of developmental strategies in parasitic hymenoptera. *Annu. Rev. Entomol.* 51, 233–258. doi: 10.1146/annurev.ento.51.110104.151029

Rotheram, S. (1967). Immune surface of eggs of a parasitic insect. *Nature* 214, 700–700. doi: 10.1038/214700a0

Strand, M. R. (1994). *Microplitis demolitor* polydnavirus infects and expresses in specific morphotypes of *Pseudoplusia includens* hemocytes. *J. Gen. Virol.* 75, 3007–3020. doi: 10.1099/0022-1317-75-11-3007

Strand, M. R., and Burke, G. R. (2013). Polydnavirus-wasp associations: evolution, genome organization, and function. *Curr. Opin. Virol.* 3, 587–594. doi: 10.1016/j.coviro.2013.06.004

Strand, M. R., and Burke, G. R. (2014). Polydnaviruses: nature's genetic engineers. *Annu. Rev. Virol.* 1, 333–354. doi: 10.1146/annurev-virology-031413-085451

Strand, M. R., McKenzie, D. I., Grassl, V., Dover, B. A., and Aiken, J. M. (1992). Persistence and expression of *Microplitis demolitor* polydnavirus in *Pseudoplusia includens*. *J. Gen. Virol.* 73, 1627–1635. doi: 10.1099/0022-1317-73-7-1627

Tanaka, K., Lapointe, R., Barney, W. E., Makkay, A. M., Stoltz, D., Cusson, M., et al. (2007). Shared and species-specific features among ichnovirus genomes. *Virology* 363, 26–35. doi: 10.1016/j.virol.2006.11.034

Theilmann, D. A., and Summers, M. D. (1986). Molecular analysis of *Campoletis sonorensis* Virus DNA in the lepidopteran host *Heliothis virescens*. *J. Gen. Virol.* 67, 1961–1969. doi: 10.1099/0022-1317-67-9-1961

Vinson, S. B., and Scott, J. R. (1975). Particles containing DNA associated with the oocyte of an insect parasitoid. *J. Invertebr. Pathol.* 25, 375–378. doi: 10.1016/0022-2011(75)90097-x

Volkoff, A. N., Rocher, J., Cerutti, P., Ohresser, M. C. P., d'Aubenton-Carafa, Y., Devauchelle, G., et al. (2001). Persistent expression of a newly characterized *Hyposoter didymator* polydnavirus gene in long-term infected lepidopteran cell lines. *J. Gen. Virol.* 82, 963–969. doi: 10.1099/0022-1317-82-4-963

Wang, Z. Z., Ye, X. Q., Shi, M., Li, F., Wang, Z. H., Zhou, Y. N., et al. (2018). Parasitic insect-derived miRNAs modulate host development. *Nat. Commun.* 9:2205. doi: 10.1038/s41467-018-04504-1

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033

Webb, B. A. (1998). "Polydnavirus biology, genome structure, and evolution," in *The Insect Viruses*, eds L. K. Miller and L. A. Ball (Boston, MA: Springer), 105–139. doi: 10.1007/978-1-4615-5341-0_5

Webb, B. A., and Strand, M. R. (2005). "The biology and genomics of polydnaviruses," in *Comprehensive Molecular Insect Science*, ed. L. I. Gilbert (Amsterdam: Elsevier), 323–360. doi: 10.1016/b0-44-451924-6/00086-7

Webb, B. A., Strand, M. R., Dickey, S. E., Beck, M. H., Hilgarth, R. S., Barney, W. E., et al. (2006). Polydnavirus genomes reflect their dual roles as mutualists and pathogens. *Virology* 347, 160–174. doi: 10.1016/j.virol.2005.11.010

Wyder, S., Tschannen, A., Hochuli, A., Gruber, A., Saladin, V., Zumbach, S., et al. (2002). Characterization of *Chelonus inanitus* polydnavirus segments: sequences and analysis, excision site and demonstration of clustering. *J. Gen. Virol.* 83, 247–256. doi: 10.1099/0022-1317-83-1-247

Ye, X. Q., Shi, M., Huang, J. H., and Chen, X. X. (2018). Parasitoid polydnaviruses and immune interaction with secondary hosts. *Dev. Comp. Immunol.* 83, 124–129. doi: 10.1016/j.dci.2018.01.007

Ye, X., Shi, M., and Chen, X. (2014). Origin and characteristics of polydnaviruses carried by parasitoid wasps. *Sci. Sin. Vit.* 44:342. doi: 10.1360/052014-48

You, M., Yue, Z., He, W., Yang, X., Yang, G., Xie, M., et al. (2013). A heterozygous moth genome provides insights into herbivory and detoxification. *Nat. Genet.* 45, 220–225. doi: 10.1038/ng.2524

Yu, D. S., Chen, Y. B., Li, M., Yang, M. J., Yang, Y., Hu, J. S., et al. (2016). A polydnaviral genome of *Microplitis bicoloratus* bracovirus and molecular interactions between the host and virus involved in NF-kappa B signaling. *Arch. Virol.* 161, 3095–3124. doi: 10.1007/s00705-016-2988-3