A horizontally transferred autonomous Helitron became a full polydnavirus segment in *Cotesia vestalis*

Pedro Heringer[1], Guilherme B. Dias[1], Gustavo C. S. Kuhn[1]*

[1] Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

* Corresponding author

E-mail: gcskuhn@ufmg.br

Abstract

Bracoviruses (BVs) and Ichnoviruses (IVs) belong to the family Polydnaviridae (PDV), which associates symbiotically with parasitoid wasps, working as vectors of virulence genes, and allowing the development of wasp larvae within hosts. One of the viral segments (c35) of *Cotesia vestalis* bracovirus (CvBV) contains an ORF that has been previously described as a helicase of unknown origin. Here we demonstrate that this gene is a Rep/Helicase from an autonomous Helitron transposable element (TE), which covers the viral segment almost entirely. This description of a PDV-Helitron fusion points to a new type of relationship between TEs and viruses. Our results also suggest that this TE underwent two consecutive horizontal transfer (HT) events: first from a Drosophila host ancestor to the genome of *C. vestalis* and its BV (CvBV), and then from *C. vestalis* to a lepidopteran host (*Bombyx mori*); the latter probably facilitated by viral segments injected during parasitization. We discuss this Helitron–BV fusion, the apparent TE exaptation by a PDV and its function, and implications for the study of wasp–host interaction and evolution.

Introduction

The family Polydnaviridae is composed of symbiotic viruses exclusively associated with more than 40,000 parasitoid wasp species from two families: Ichneumonidae and Braconidae (superfamily Ichneumonoidea). Polydnaviruses (PDVs) exist either as proviral copies in the wasp genome, or in their functional form as multiple dsDNA circles packaged into infective particles (reviewed in Gundersen-Rindal et al. 2013). During wasp oviposition, PDVs are injected into the host, where they express virulence genes that alter insect physiology and allow the development of parasitoid larvae (Strand and Burke 2013, Drezen et al. 2014). PDVs are classified in two genera, Ichnovirus (IV) and Bracovirus (BV), which are in turn associated with ichneumonid and braconid wasps, respectively. Despite similarities between these two groups of viruses, they originated from independent viral integration events on each of these two wasp lineages (Herniou et al. 2013).

There are more than 18,000 braconid species described to date, but some estimates point to a total number exceeding 40,000 (Quicke 2015). Braconid wasps are parasitoids of a wide range of insect orders, and are one of the most successful insects used in biological control programs (Wharton 1993). An important species from this family is *Cotesia vestalis*, who has been found parasitizing Lepidoptera

51 species in various regions of Asia, Europe, Africa and the Americas (Furlong et al. 2013). Also, *C. vestalis*
52 is one of the main parasitoids of the diamondback moth *Plutella xylostella*, which is a pest of *Brassica*
53 plants that costs annually US$4-5 billion worldwide in crop loss and management (Zalucki et al. 2012).
54 For those reasons, *C. vestalis* has been successfully used in numerous biological control introductions
55 (Furlong et al. 2013).

56 The complete genome sequencing of *Cotesia vestalis* bracovirus (CvBV) revealed 157 ORFs
57 distributed in 35 encapsidated segments, with most of these genes also having homologs in other BVs
58 (Chen et al. 2011). In the same work, the authors noted that a segment from CvBV (CvBV_c35) encodes
59 a protein displaying similarity with the human Pif1 helicase, which has no sequence homology with any
60 other gene described in BVs. The Pif1 family is a member of the superfamily 1B helicases and is involved
61 in many replication related functions in eukaryotes (Bochman, et al. 2010). Because all PDVs analyzed to
62 date only express replication genes in the calyx cells of wasps, and do not replicate in host cells (Bézier
63 et al. 2009, Gundersen-Rindal et al. 2013), the function of this helicase in a CvBV encapsidated segment
64 is worth of investigation, as noted by Chen et al. (2011).

65 In the present work, we found that segment CvBV_c35 is almost entirely made by a rolling-circle
66 Helitron transposon, which has fused to the CvBV genome. Although there has been reports of
67 transposable elements (TE) (including Helitrons) in BV genomes (e.g. Drezen et al. 2006, Thomas et al.
68 2010, Dupuy et al. 2011, Coates 2015, Guo et al. 2014), we describe for the first time a case where an
69 autonomous TE took over a viral segment. Based on our results we propose a scenario to explain the
70 presence of this Helitron in the virus, and suggest two consecutive horizontal transfer (HT) events
71 involving this TE across species from three insect orders: Diptera (oriental Drosophila), Hymenoptera (*C.*
72 *vestalis*), and Lepidoptera (*Bombyx mori*). Finally, we discuss the conservation of the Helitron as a PDV
73 segment, the possibility of a co-option, and its implications for the study of HT events.

74

75 ## Results and Discussion

76 ### Segment c35 from *Cotesia vestalis* bracovirus is a Helitron autonomous TE

77 The segment c35 (CvBV_c35) (gb: HQ009558.1) from *Cotesia vestalis* bracovirus (CvBV) has
78 5,667 bp and contains a single ORF, described as a DNA helicase that has never been reported on PDVs
79 (Chen et al. 2011). Indeed, its annotated amino acid sequence (AEE09607.1) has a C-terminal Pif1-like

helicase domain which belongs to the P-loop containing Nucleoside Triphosphate Hydrolases (P-loop_NTPase) superfamily. Because the Pif-like domain covers only a relatively small portion of the total amino acid sequence and it is absent in all known PDVs, we further investigated the identity of this protein.

We found that the N-terminal half of the protein is a Rep domain which is typical of the Rep-Helicases (Rep/Hel) from Helitron TEs (Kapitonov and Jurka 2001). Subsequent analysis revealed the presence of all three conserved motifs present in the Rep catalytic core (Kapitonov and Jurka 2007), confirming the identity of this protein as a Rep/Hel (Fig 1). This group of proteins presents a basic structure consisting of a N-terminal HUH endonuclease motif (Rep) and a C-terminal helicase motif (Hel) and is used by Helitrons during their rolling-circle (RC) transposition (reviewed in Thomas and Pritham 2015).

After we determined the Rep/Hel identity of this coding sequence, we verified if the rest of the TE was present in segment c35. By comparing CvBV_c35 and Helitron sequences retrieved from Repbase (Jurka et al. 2005) we could determine the structure of the whole element, with its 5'- and 3'-termini flanking the Rep/Hel ORF. This Helitron, hereafter named Hel_c35, has 5,294 bp, and covers most (~ 93.4%) of the viral segment c35. Because Hel_c35 contains an intact ORF and all the structural hallmarks of Helitrons (see Materials and Methods), we suggest that it should be classified as an autonomous TE. Only a small 373 bp region within segment CvBV_c35, but outside the Hel_c35 sequence, lacks significant homology with Helitrons. Moreover, we found the same 373 bp sequence (100% identity) as part of a different CvBV segment (CvBV_c19), indicating that this sequence is probably of viral nature. This 373 bp viral sequence does not display significant similarity with any known transposable elements or insect genome available to date.

The analysis performed with the CENSOR tool (Kohany et al. 2006) against the whole Repbase repeat library, using Hel_c35 as a query, indicated that this element has high sequence similarity (70-97%) with Helitrons from Diptera, especially Drosophila species (S1 Fig). Although Hel_c35 displays high sequence identity with these Helitrons, they share less than 80% similarity in their last 30 bp at the 3'-terminus. Therefore Hel_c35 does not belong to any described family, according to the proposed classification of Helitrons (Thomas and Pritham 2015).

Helitron insertions in PDV genomes have been reported before, including in CvBV (Thomas et al. 2010, Guo et al. 2014, Coates 2015). It is worth mentioning that some of these studies describe Helitron

insertions in the genomes of *Cotesia plutellae* and *Cotesia plutellae* bracovirus (CpBV), which are junior synonyms of *C. vestalis* and CvBV, respectively (Shaw 2003). However, these elements are non-autonomous, and display low sequence identity (< 80%) with Hel_c35, including on their terminal sequences (data not shown), thus belonging to a distinct Helitron family.

Searching the CvBV_c35 proviral locus

There are currently two lineages of *C. vestalis* with sequenced genomes: isolate ANU101 (BioSample: SAMN03273265) from South Korea and isolate 20120220 (BioSample: SAMN04378091) from China. In order to determine the site of Hel_c35 insertion within the proviral locus, we Blast-searched segment CvBV_c35 in these two *C. vestalis* sequenced genomes. Both lineages retrieved only one hit each, with > 40% query coverage and > 90% sequence identity, indicating that a single entire copy of Hel_c35 is present in *C. vestalis*. The best hit (gb: JZSA01007369.1) corresponds to contig 7377 from isolate ANU101 and contains a Helitron copy with ~ 99.9% nucleotide identity to the Hel_c35 element present in CvBV_c35 (Fig 2). Despite the high sequence similarity, this putative CvBV_c35 proviral sequence contains two deletions, one of 136 bp (between positions 4,780-4,915 of the Hel_c35 query sequence), and another corresponding to the whole 373 bp viral sequence within segment CvBV_c35. Thus, the similarity of this proviral copy with CvBV_c35 is limited to the Hel_c35 sequence. Curiously, Blast searches in the genomes of both wasp lineages, using this 373 bp viral sequence as a query, retrieved one hit (100% identity) in a locus that corresponds to CvBV_c19, only in the genome of isolate 20120220. Hence, this 373 bp sequence appears to be absent in the isolate ANU101, even though proviral segment CvBV_c19 (where this sequence was also found) is present in this genome.

Instead of the 373 bp viral sequence flanking its termini, this Hel_c35 copy in the *C. vestalis* genome presents a 396 bp sequence which is tandemly repeated 6 times after the Helitron 3'-end (Fig 2). These repeats are not abundant in the *C. vestalis* genome, being present as less than 100 copies distributed over ~ 20 short arrays containing on average four repeats. We did not find these repeats in other hymenopteran genomes, including *Microplitis demolitor*, which belongs to the same subfamily as *C. vestalis* (Microgastrinae), and currently is the closest wasp species with a sequenced genome. Moreover, we noticed that upstream the Helitron 5'-terminus there are two long direct repeated sequences with 376 bp each and ~ 99.5% identity, separated by 426 bp, and with similarity to BV sequences (Fig 2 and S1 Table). Those differences prompted us to analyze the flanking regions of this Hel_c35 copy up to several kbp upstream and downstream, to identify other possible PDV-related

sequences.

We found eight motifs that are conserved in many PDV-related sequences, including several BV circles and proviral regions flanking segments (Fig 2 and S1 Table). Five of the eight conserved motifs are also found within or nearby nudiviral-like genes from *Cotesia congregata* bracovirus (CcBV). Despite not being encapsidated, these genes are involved in BV replication and production of structural viral proteins (Strand and Burke 2013, Drezen et al. 2014). The fact that some of these motifs are found in both segment and nudiviral loci, and others are exclusive to segment loci, might be related to their common or specific functions in BVs, respectively. For example, common motifs could have a role in proviral replication, and specific motifs in circle excision, encapsidation and transcription. All the conserved sequences analyzed here have an average AT content of 71% (62.7-80%), and contain internal palindromic regions that could form hairpin folds with variable sizes. Additionally, two of these motifs around the Hel_c35, one upstream and the other downstream, contain long inverse repeats that could also form stem-loop structures (S2 Fig). Similar features have been described in flanking and intermediate sequences of other BV proviral loci, and are thought to be involved in viral replication (Louis et al. 2013, Burke et al. 2015).

The presence of several motifs flanking this Hel_c35 in the *C. vestalis* genome suggests that this copy was probably inserted on a proviral loci. If true, why this Hel_c35 copy is different from the one present in the viral genome? This inconsistency might be explained if this Hel_c35 insertion does not correspond to the main proviral CvBV_c35, but to a paralogous segment that could be either active (it only produces a small part of c35 circles) or inactive (such as a 'pseudo-segment'). A large portion of PDV genes found in viral segments belong to multigenic families, and segment reintegration, locus duplication/deletion, and gene gain/loss, are common features during PDV evolution (Burke and Strand 2012a, Herniou et al. 2013).

Another possibility, which perhaps is more plausible, is that CvBV_c35 is polymorphic in the different lineages used for the genome sequencing of *C. vestalis* (BioSample: SAMN03273265) and CvBV (HQ009558.1). In addition to different CvBV_c35 segments in distinct lineages, this polymorphism could also include lineages with no CvBV_c35 encapsidated circles. Before the work conducted by Chen et al. (2011), the CvBV (referred as CpBV) genome was partially sequenced from a wasp strain of South Korea (Choi et al. 2009), and no circles containing a Rep/Hel gene were detected. Although this is a partial genome sequence, the missing CvBV_c35 could be a consequence of its real absence in the CvBV particles from this wasp strain. As we described above, the best candidate for a proviral CvBV_c35

sequence also belongs to a wasp lineage from South Korea, in contrast to the CvBV_c35 sequence (HQ009558.1), which is derived from a wasp strain of China. Hence, the differences between the wasp and BV genomes might reflect a polymorphic CvBV_c35 which, depending on the lineage, exists as a proviral segment that generates encapsidated viral circles, or as an 'ancestral' form with the Hel_c35 in a proviral region, but not able to produce viral particles. Although the confirmation of this hypothesis of polymorphism would require the genome sequencing of different *C. vestalis* strains and their respective BVs, it is worth mentioning that there is evidence of significant variation between BVs from different wasp populations of the same species (Rincon et al. 2006, Branca et al. 2011).

Rep/Hel transcripts in a lepidopteran host parasitized by *Cotesia chilonis*

The data described here so far indicate a Helitron-BV fusion. In this context, it is important to investigate when, during the evolution of BV, this event happened. Apart from CvBV, we did not find the Hel_c35 Rep/Hel ORF in any other PDV sequenced genome available to date, including those from two other bracoviruses present in two *Cotesia* species: *Cotesia congregata* bracovirus (CcBV) and *Cotesia sesamiae* bracovirus (CsBV). Although there are no other *Cotesia* sequenced genomes available, Wu et al. (2013) described the influence of *Cotesia chilonis* on its lepidopteran host *Chilo suppressalis* during parasitization, by analyzing the transcriptome of fat body and hemocytes in the larvae of this moth. In addition to host genes, they identified 19 unique sequences associated with PDVs, which were classified as *Cotesia chilonis* BV (CchBV) transcripts; so we further searched for sequences similar to Hel_c35 on their Short Read Archive (SRA) (accession: SRR651040) and Transcriptome Shotgun Assembly (TSA) (gb: GAJS00000000) databases in order to pinpoint the approximate phylogenetic position of the Hel_c35 integration into the ancestral BV genome.

We detected ~ 70 reads with > 80% similarity, and ~ 50 reads with > 90% similarity with the Hel_c35 ORF in the SRA, covering different regions of its sequence (Fig 3). These reads were mainly located around two Hel_c35 ORF regions extending 223 and 608 bp, which were assembled as Unigene42046 (gb: GAJS01040222.1), and Unigene57509 (gb: GAJS01055664.1), respectively (Fig 3). Interestingly, searches on the genomes of *C. suppressalis* and *Amyelois transitella* (both from the Pyraloidea superfamily) using Unigene42046 as a query retrieved sequences with only 74-85% identity. On the other hand, the same Unigene has 99% sequence identity with the Rep/Hel ORF from CvBV_c35 (S3 Fig). The high similarity of this transcript with the CvBV segment, and its likely absence from the genome of *C. suppressalis*, suggests that this transcript is derived from CchBV.

Based on the distribution of the Rep/Hel in the phylogeny of *Cotesia* species (Fig 4) we might assume that a homologous bracovirus segment with a Rep/Hel gene was independently lost twice during *Cotesia* evolution, one in the *C. congregata* and the other in the *C. sesamiae* clade (Michel-Salzat and Whitfield 2004); or that two independent Helitron integration events occurred, one in CvBV and the other in CchBV. The hypothesis of two integration events involving two different Helitron families would explain the disparity of the Hel_c35 ORF sequence similarity with Unigene42046 (99%) and Unigene57509 (71%). In this case, both Unigenes would be part of one single ORF from a Helitron family that is distinct from Hel_c35. This hypothesis is also supported by our date estimate of when Hel_c35 was first inserted in the *C. vestalis* genome (< 1 mya) (see topic "Tracing an evolutionary pathway of Hel_c35 in insect genomes"), which appears to be too recent to have happened before the divergence of the two species. Nevertheless, the genome sequencing of *C. chilonis* and CchBV will be necessary to test (i) the identity between Hel_c35 and Unigene42046, (ii) the presence of this TE within CchBV, and (iii) the homology of this insertion with the one found in segment CvBV_c35.

Horizontal transfer of Hel_c35 between three insect orders

To better characterize the evolution of the Helitron Hel_c35, we searched for sequences displaying high similarity with the ORF's Rep domain from CvBV_c35 (S1 Data) in all arthropod sequenced genomes available to date. A nucleotide alignment containing all obtained sequences is shown in S4 Fig. We then used these sequences to construct a phylogenetic tree (Fig 5, see Materials and Methods).

The resultant tree revealed some strikingly incongruences with the species' phylogeny (Fig 5). Most notably, Hel_c35 from *Bombyx mori*, *Drosophila rhopaloa* and *Drosophila ficusphila* were allocated in the same clade as the Hel_c35 from *Cotesia vestalis* bracovirus (CvBV). It is important to mention that we did not find Hel_c35 in any other hymenopteran species, even though more than forty genomes from this insect order were used in our search. Because the marked phylogenetic incongruences and the extreme discontinuous or "patchy" distribution of Hel_c35 are indicative of horizontal transfer (HT) events (Silva et al. 2004, Wallau et al. 2012), we decided to further investigate this possibility.

Another line of evidence used to evaluate the occurrence of HT is the high sequence similarity between DNA sequences from distantly related taxa, when compared with more closely related ones (Silva et al. 2004, Wallau et al. 2012). We found that the nucleotide identity between Hel_c35 from CvBV, and Hel_c35 from *B. mori*, *D. rhopaloa* and *D. ficusphila* sequences are ~ 99.7%, ~ 96% and ~ 94%,

231 respectively.

232 For a HT event to be inferred, the species distributions also should be considered, as the
233 candidate taxa must overlap geographically in order to a HT event to occur (Loreto et al. 2008, Carareto
234 2011). Thus, we classified each taxon in the phylogeny to one of eight major geographical regions, based
235 on their known distributions (see Materials and Methods). The addition of this information revealed
236 interesting aspects that helped to explain some of the main inconsistencies found in the phylogeny (Fig
237 5). First, among the ten closest taxa with Hel_c35, seven are native from eastern/southestern Asia,
238 including the three species that group immediately with the Hel_c35 from CvBV (*B. mori*, *D. rhopaloa*
239 and *D. ficusphila*). Second, the positioning of *Bactrocera dorsalis* (Tephritidae) within a well-supported
240 clade of Drosophila species, instead of a position closer to other tephritid fruit flies, also coincides with
241 their common distribution in southeastern Asia. Third, the incongruent grouping of *Calycopis cecrops*
242 (Lepidoptera) with *Drosophila willistoni* (Diptera), involve two species with geographic overlapping in
243 southeastern United States. Fourth, despite the fact that *Drosophila mojavensis* and the lepidopteran
244 *Amyelois transitella* belong to different insect orders, they are grouped together in a well-supported
245 clade and overlap geographically in southwestern United States. Thus, the geographical distribution of
246 the analyzed taxa supports the HT hypothesis as an explanation for at least some of the phylogenetic
247 incongruences, especially the one directly related to CvBV. Other incongruences outside the immediate
248 CvBV clade might also represent bona fide HTs; however, each of them requires a careful analysis that is
249 beyond the scope of the present work.

250 For HT events, the spatial overlapping of candidate species must be associated with at least
251 some degree of ecological overlapping (Loreto et al. 2008, Carareto 2011). Accordingly, braconid wasps
252 interact very closely with several insect host orders, including lepidopteran and Drosophila species
253 (Wharton 1993, Quicke 2015) and as a rule, these connections reach the cellular and even to the
254 chromosomal level during wasp oviposition and bracovirus infection of host cells (Gundersen-Rindal et
255 al. 2013). It is noteworthy that these extremely direct interactions are not occasional, but a fundamental
256 part of the braconid wasps' life cycle.

257 In summary, our results suggest at least two HT events involving Hel_c35: one between a
258 Drosophila species and *C. vestalis* and another between *C. vestalis* and *B. mori*, which probably occurred
259 in this respective order, as indicated by the phylogeny. This hypothesis is supported by four lines of
260 evidence: (i) the marked incongruence between host and TE phylogeny, (ii) the patchy distribution of
261 this TE on the main taxa involved, (iii) the high sequence identity between Hel_c35 copies from different

insect orders, and (iv) the spatial/ecological overlap among the candidate species.

Tracing an evolutionary pathway of Hel_c35 in insect genomes

Because our constructed phylogeny with Hel_c35 sequences suggests HT events between *D. rhopaloa*, *C. vestalis* and *B. mori*, we estimated when these events occurred using the equation for divergence time ($T = K/2r$) on non-coding sequences (~ 750 bp) from Hel_c35 copies of these species. We also set minimum and maximum date thresholds, supposing different evolutionary constrains acting on these sequences (see Materials and Methods). This analysis can also help to infer the order of the two HT events which, according to the phylogeny, appear to have occurred from *D. rhopaloa* to *C. vestalis* (represented by CvBV), and then from *C. vestalis* to *B. mori*. The results give an approximate date of 0.862 mya (0.574 – 1.15 mya) for the *D. rhopaloa* – *C. vestalis* HT, and 0.211 mya (0.141 – 0.282 mya) for the *C. vestalis* – *B. mori* HT.

Based on our results, we suggest the following scenario to explain the observed distribution of Hel_c35 copies among insect genomes. The transposable element Hel_c35 belongs to a Helitron family originally present in southeastern Asian Drosophila species and, more specifically, within the oriental subgroup of the *Drosophila melanogaster* species group. This taxon comprises the largest number of species with Hel_c35 copies clustering with CvBV, and displays a topology roughly coherent with the group's phylogeny (Seetharam and Stuart 2013). Initially, a Hel_c35 element from a Drosophila species in the *rhopaloa* subgroup was horizontally transferred into the ancestral genome of *C. vestalis* ~ 0.862 mya, probably during parasitization. After genome integration, Hel_c35 transposed into the CvBV proviral genome in a locus responsible for segment production, also becoming part of the encapsidated genome of this bracovirus. Then, a second HT event occurred ~ 0.211 mya, this time involving the transfer of Hel_c35 from *C. vestalis* to *B. mori*, through CvBV_c35 circles (Fig 6). If after entering the proviral locus, Hel_c35 'fused' gradually with CvBV, the observed sequence inconsistencies between CvBV_c35 from different *C. vestalis* lineages could represent incomplete stages of this fusion on different wasp strains. In that case, the formation of segment CvBV_c35 would have been completed after geographical divergence of the lineages used for wasp (S. Korea) and BV (China) genome sequencing.

To hold true, the above hypothesis might require at least two assumptions that should be addressed. First, that *C. vestalis* is or at least was until recently capable to parasitize Drosophila species, in addition to their typical lepidopteran hosts. This hypothesis seems improbable at a first sight, because

all hosts of Microgastrinae wasps described to date are Lepidoptera larvae (Quicke 2015). Indeed, the major radiation dates for both taxa seem to coincide, reinforcing the idea that Microgastrinae have evolved as a group of specialized Lepidoptera parasitoids (Banks and Whitfield 2006). On the other hand, it is also important to note the remarkable host diversity and variability within the Braconidae family. For instance, the subfamilies Alysiinae and Opiinae are endoparasitoids of many cyclorrhaphous Diptera, including Drosophila and tephritid fruit flies (Carton et al. 1986, Wharton 1993), and several wasp species in the Braconinae subfamily can attack both dipteran and lepidopteran hosts, depending on the location and time of the year (Žikić et al. 2012, Gadallah and Ghahari 2015). The Exothecinae subfamily contains parasitoids of several insect orders: for instance, *Colastes braconius* can parasitize dipteran, lepidopteran, coleopteran and even hymenopteran species (Shaw and Huddleston 1991). Other ecological aspects are also relevant if we consider the host range and specificity within the *Cotesia* genus. The first is that, even though *Plutella xylostella* is commonly described as the main host of *C. vestalis* in the literature, this wasp is capable to parasitize a wide range of Lepidoptera families and superfamilies, which are extremely diverse in their ecology and morphology (Cameron and Walker 1997, Malysh et al. 2016).

It is also worth mentioning that, although to our knowledge there is no report of *C. vestalis* parasitization on *B. mori*, the close related species *Cotesia glomerata* can use this moth as a host (Sathe and Jadhav 2001), and *Cotesia dictyoplocae*, which is in the same species group of these two wasps, parasitizes moths of the Bombycoidea superfamily (Gupta et al. 2016). Additionally, the present work is not the first to report a HT between *C. vestalis* and *B. mori* (e.g. Coates 2015, Zhang et al. 2016a, 2016b), indicating that *C. vestalis* could indeed attack this lepidopteran species, and might also parasitize a wider variety of hosts.

Secondly, the known hosts of a parasitoid wasp do not necessarily correspond to their actual host range, and may simply represent the commonly attacked species of which successful parasitization is more likely to ensue. It has been suggested that unusual conditions could induce attacks outside the suitable host range, and even result in successful parasitization of unsuitable hosts, notwithstanding their rarity (Heimpel et al. 2003, Quicke 2015). It is also noteworthy that some Drosophila species phylogenetically close to *D. rhopaloa* can be resistant to attacks from the wasp *Asobara japonica*, probably because of a long-continued interaction with braconid parasitoids (Ideo et al. 2008, Furihata et al. 2016). Additionally, on southeast Asia, two Drosophila species closely related to *D. rhopaloa* breed on plants that are also commonly used as food by Lepidoptera larvae (Suwito et al. 2002), which could

facilitate an encounter between Cotesia wasps and unusual dipteran hosts. Hence, it is not difficult to conceive that *Cotesia* wasps attack, and sometimes can successfully parasitize Drosophila larvae, even if these interactions may be rare.

PDV-Helitron exaptation?

There are several reports of TE integrations in PDV (e.g. Drezen et al. 2006, Dupuy et al. 2011). Moreover, Helitrons have already been found in the genomes of Cotesia species and in their respective PDVs, and were also involved in HT events (Thomas et al. 2010, Guo et al. 2014, Coates 2015). However, all these insertions represent non-autonomous short or fragmented elements. To our knowledge, we describe the first instance of a putative autonomous Helitron that not only integrated the PDV genome, but effectively became a viral segment.

Exaptation refers to features originally evolved for some function but that were later co-opted for a different role (Gould and Vrba 1982). Repetitive DNAs like TEs can also take part on this process (Brosius and Gould 1992), as revealed by many studies in eukaryotes (recently reviewed by Chuong et al. 2017). Although there is evidence for the role of TEs on PDV evolution, including the exaptation of a gene apparently derived from from a retroelement (reviewed in Burke and Strand 2012b), we report the first intact TE that appears to have been co-opted by a bracovirus. The main reasons for this suggestion are: (i) this TE appears to be autonomous; (ii) it has conserved an intact structure in the genome of *C. vestalis* for the last ~ 1 my, despite the low number of copies; (iii) the element occupies almost 94% of segment CvBV_c35, which is one of the most replicated circles of this virus (Chen et al. 2011); (iv) it contains the only ORF in CvBV_c35, and all known PDV segments have coding sequences, with a few exceptions (Burke et al. 2014); (v) the CvBV_c35 portion outside Hel_c35 has only 373 bp, and does not contain any ORF or conserved sequences found in other PDVs (apart from segment CvBV_c19). Although it is possible that this short sequence has a functional role (e.g. as an encapsidation signal), the whole PDV segment CvBV_c35 is essentially a Helitron. Furthermore, we found a single full copy of Hel_c35 in the *C. vestalis* genome, even though *D. rhopaloa* and *B. mori* contain several highly similar partial copies (data not shown). This unique arrangement for a TE in a eukaryote genome reinforces the suggestion that segment CvBV_c35 has been kept as an intact Helitron by selective constraints.

Supposing that Hel_c35 is not just a selfish element within the CvBV genome, but also play an active role in this bracovirus, an import question emerges: what could be the advantage of having a Helitron in a PDV genome? We have three non-mutually excluding hypotheses to explain this question.

353 The first hypothesis is that this Helitron Rep/Hel was exaptated by CvBV for its helicase function,
354 to aid the amplification of viral replication units or circles in wasp calyx cells. Most viral genes thought to
355 be essential for BV replication have not being identified, with a few exceptions, like a nudiviral helicase
356 and a fen-like flap endonuclease (Herniou et al. 2013). The Pif1 family of helicases (which include the Hel
357 domain in Rep/Hel proteins) and the flap endonucleases are important to process DNA secondary
358 structures like hairpin or fold-back substrates during replication (Pike et al. 2010, Balakrishnan and
359 Bambara 2013). Interestingly, TA-rich inverted repeats and other palindromic motifs capable of forming
360 hairpins are found throughout the proviral genomes of BVs, and probably serve as replication origins
361 (Louis et al. 2013, Burke et al. 2015). Although possible, this hypothesis does not explain why the
362 Rep/Hel ORF is part of the CvBV encapsidated genome, as that would not be necessary for replication in
363 calyx cells.

364 The second hypothesis, is the use of the whole Helitron (with its non-coding sequences) as a
365 mechanism for specific segment CvBV_c35 amplification in calyx cells of wasps, which could explain why
366 CvBV_c35 constitutes one of the most abundant circles in CvBV, despite being a new segment. This
367 'auto-replication' would be useful if PDV capsids also play a role on virulence, independently of the
368 genes they carry. In that case, abundant segments could be selected for their ability to assemble more
369 capsids that are later injected into the host. This hypothesis is based on the observation that *Microplitis*
370 *demolitor* produces some BV encapsidated segments with no apparent coding sequences (Burke et al.
371 2014), and several parasitoid wasp species, including braconids, use virus-like particles (VLP) with no
372 detected PDV sequences as virulence factors against their hosts (Rizki and Rizki1990, Herniou et al.
373 2013, Furihata et al. 2016). In contrast to the first hypothesis, this scenario also predicts the
374 conservation of Hel_c35 non-coding sequences, as they would be at least partially necessary for
375 segment CvBV_c35 to take advantage of the Helitron rolling-circle replication (RCR) mechanism. That is
376 because Rep/Hel proteins recognize Helitron terminal sequence motifs to start and finish RCR correctly
377 (Grabundzija et al. 2016). Although the sequences used in our analysis display slightly higher nucleotide
378 conservation on coding regions in comparison with non-coding regions (data not shown), the exact level
379 of sequence stability necessary for Rep/Hel recognition and RCR viability of Helitrons is not well
380 understood.

381 The third hypothesis is that CvBV_c35 ORF could be used to replicate its own circle and other
382 CvBV circles, once within host cells. Like the second hypothesis, this explanation considers the fact that
383 Hel_c35 Rep/Hel is part of the BV encapsidated genome, but in this case, the Rep/Hel protein would

384 further amplify other CvBV genes after their injection in the host, in addition to CvBV_c35 circles. This
385 mechanism of circle replication in host cells could further increase the number of CvBV genes, and
386 therefore enhance virulence. This scenario would be particularly interesting to investigate, as all PDVs
387 analyzed to date only express replication genes in the calyx cells of wasps, and apparently do not
388 replicate after encapsidation and injection in host cells (Bézier et al. 2009, Gundersen-Rindal et al.
389 2013). Thus, the hypothesis of Hel_c35 exaptation by CvBV for some replication-related function is
390 plausible, even though the replication of this group of viruses is not fully understood.

391 In order to test the three hypotheses stressed above, more in-depth experimental assays with
392 parasitoid and host nucleotide sequencing will be necessary. Also, the extent and importance of the
393 putative Hel_c35 exaptation by CvBV could be further validated through a combination of genome and
394 transcriptome studies using *C. vestalis* and the associated CvBV from the same wasp strain. That is
395 because it is currently not possible to discard that CvBV_c35 is polymorphic within *C. vestalis* wasps
396 from different lineages.

397

398 Conclusion

399 There are several reports of TE integrations in PDVs. However, all these cases involve classical TE
400 insertions within PDV genomes, mostly of non-autonomous and fragmented elements. In contrast, our
401 study revealed an autonomous Helitron within a bracovirus segment, which is not only an integrated
402 copy, but a TE covering almost entirely a viral circle. The segment CvBV_c35 is effectively an intact
403 Helitron, and our results suggest that this TE might have been recently co-opted by CvBV, probably for
404 its helicase and/or rolling-circle replication function. This PDV-Helitron fusion points to a new type of
405 relationship between TEs and viruses. In addition, our data reinforces the idea of PDVs as effective
406 agents of horizontal transfer (HT), and of Helitrons as one of the TEs most commonly involved in those
407 events. Specifically, we reported two consecutive transfers of Hel_c35 across three insect orders.

408 The number of reported HT events in eukaryotes has been growing recently, although the
409 probability of their occurrence must be very low, as they require the presence of several biological
410 features and conditions. Notably, the tripartite parasitoid system composed by braconid wasps, PDVs
411 and hosts fulfill most of those requirements (Venner et al. 2017), being exceptional candidates for the
412 study of HT in eukaryotes. Furthermore, as pointed out by Quicke (2015), because the number of

413 estimated Microgastrinae species ranges between 16,000-40,000 and apparently, all of them have
414 associated PDVs, this could be the biggest known group of viruses. The future genome sequencing of
415 more PDVs and wasp species will help to understand the real importance of HT for the evolution of
416 parasitoids and their hosts. Furthermore, the investigation of events like this can contribute to
417 ecological analyses, by revealing potential hosts and cryptic interactions not detected in field studies.

418

419 ## Materials and Methods

420 ### Rep/Hel amino acid analysis

421 To confirm that the CvBV_c35 codes a Rep/Hel we first analyzed its amino acid sequence (AEE09607.1)
422 using the NCBI's conserved domain database (CDD) (Marchler-Bauer et al. 2015). Rep/Hel ORF
423 sequences were retrieved from Repbase (Jurka et al. 2005), either directly or using the CENSOR tool
424 (Kohany et al. 2006) with CvBV_c35 ORF as a query. Sequences were then aligned using MEGA7 (Kumar
425 et al. 2016). We determined the residues that characterize Rep domains by visually inspecting the
426 CvBV_c35 amino acid sequence, based on the consensus from Kapitonov and Jurka (2007) (Fig 1C).

427 ### Hel_c35 structure hallmarks

428 To determine the whole Helitron structure we first used the CvBV_c35 sequence as a query in a search
429 against the Repbase reference collection (Kohany et al. 2006) (S1 Fig). After visually establishing the
430 element's putative boundaries, we determined its precise termini by aligning CvBV_c35 with the best
431 results, and verified if the sequence contained the hallmarks of Helitrons, like the insertion between AT
432 nucleotides, the hairpin structure close to the 3'-end, and the conserved terminal nucleotides. To
433 further validate the TE limits, we used the putative Helitron sequence (named Hel_c35) as a query for a
434 Blastn search (Altschul et al. 1990) against the genome of *Cotesia vestalis* available on GenBank (Benson
435 et al. 2013), and then used the best results, together with their flanking sequences as a query to a
436 second Blastn search. Most results with multiple hits only display similarity with Hel_c35 sequences
437 within the terminal nucleotides, suggesting that our defined limits for the element encompass the whole
438 TE.

439 ### Search for CvBV_c35 proviral locus and analysis of PDV conserved sequences

440 To Blast-search (Altschul et al. 1990) the CvBV_c35 proviral locus we used its sequence (HQ009558.1) as

a query against the *C. vestalis* genome on GenBank (Benson et al. 2013), selecting the best result based on score, query cover and identity. We then scanned the selected contig (gb: JZSA01007369.1) searching for sequences displaying similarity with PDV genomes and other PDV-related sequences available on Genbank (Benson et al. 2013). To detect structural features on the PDV-like sequences found in the selected contig, we first used dotplots from the software Dotlet (Junier and Pagni 2000), followed by manual curation of sequences. Finally, the analysis of palindromic or inverted repeat motifs, and their predicted secondary structures, were made using the softwares mfold (Zuker2003) and Palindrome analyser (Brázda et al. 2016).

Transcriptome analysis of *Chilo suppressalis*

We used the Rep/Hel coding DNA sequence (CDS) of CvBV_c35 (HQ009558.1) as a query for Blastn searches against the Short Read Archive (SRA) (SRR651040) and Transcriptome Shotgun Assembly (TSA) (GAJS00000000) of *Chilo suppressalis* generated by Wu et al. (2013). Reads with identity > 90%, and Unigenes with score ≥ 200 were selected. The sequences of selected Unigenes were then used as queries for Blastn searches against genomes from all insect orders for comparison. The sequences from Unigene42046 and the corresponding region from the CvBV_c35 CDS were aligned using M-coffee (Moretti et. al. 2007; available in http://tcoffee.crg.cat/apps/tcoffee/do:mcoffee) (S3 Fig).

Search on Arthropoda genomes, alignment and phylogeny

We conducted Blastn searches (Altschul et al. 1990) on all Arthropoda genomes available on GenBank (Benson et al. 2013), using a query of 1,675 bp selected from the Hel_c35 Rep/Hel ORF located in the 'Rep' region that is exclusive from Helitrons (S1 Data). In contrast, the 'Hel' region is similar to the Pif1 family of helicases, which are involved in many cellular processes and pervasive in eukaryotes (Bochman, et al. 2010). For that reason, and because most of the initial sampling searches using the whole ORF (4,538bp) retrieved only fragmented hits, we decided to use a shorter query covering only the Rep motif. This selected nucleotide stretch correspond to the "Helitron_like_N" domain annotated on the NCBI's CDD (Marchler-Bauer et al. 2015), the rest of the catalytic core that we manually selected using the Helitron consensus (see 'Rep/Hel amino acid analysis' above), and an intermediate region between Rep and Hel (S1 Data). In order to include in our study only TEs from the same family (see Wicker et al. 2007 for a hierarchical proposal for transposable elements classification), results with > 70% cover and > 80% identity were selected based on their Max score, and then aligned using MUSCLE on MEGA7 (Kumar et al. 2016) (S4 Fig). Although we recognize that Helitrons families should be

471 classified in a manner distinct from most DNA TEs, due to their peculiar structure and evolutionary
472 dynamics (Thomas and Pritham 2015), our interest was to retrieve coding sequences, including the ones
473 from elements lacking their terminal 30 bp used in the mentioned classification. For the phylogeny of
474 aligned Rep sequences, we used MrBayes 3.2 (Ronquist et al. 2012) to run a mixed analysis, which
475 samples across the different nucleotide substitution models. The results were compared with the
476 phylogeny estimated using MEGA7 (Kumar et al. 2016), with the Maximum Likelihood method, and
477 using the Tamura 3-parameter model (Tamura 1992), which resulted a similar branch topology (data not
478 shown). The main geographical distributions of the taxa included in the phylogeny were retrieved from
479 various Web sources.

480 Date estimation of HT events

481 For the date estimation of the HT events between *Drosophila rhopaloa*, *Cotesia vestalis* and *Bombyx*
482 *mori*, we used the equation for divergence time given by

483 $T = K/2r$

484 where T is the number of generations, K is the number of substitutions per site, and r is the rate of
485 nucleotide substitution, which is equal to the mutation rate (μ) for neutral mutations (Graur and Li
486 2000). The sequences used comprised the 5' and 3' non-coding flanking regions of the Hel_c35 ORF,
487 which together sum ~ 750 bp. The copy from CvBV was used to represent *C. vestalis*, and as a query for
488 Blast searches in the genomes of *D. rhopaloa* and *B. mori*. Although we used non-coding sequences
489 flanking the Hel_c35 ORF for the analysis, assuming they evolve neutrally, the importance of these
490 regions for Helitron transposition are not fully understood, apart from the terminal 40 bp on each end
491 (Grabundzija et al. 2016). As the selective pressure on these sequences cannot be discarded, the
492 assumption of neutrality should be taken cautiously. For that reason, we used one equation ($T = K/2\mu$)
493 for a minimum date estimation, with both sequences evolving neutrally, and another ($T = K/\mu$) for a
494 maximum date estimation, with only one sequence evolving neutrally. This second equation (maximum
495 threshold) gives the divergence time for a single branch in the phylogeny (Cutter 2008), and represents
496 the hypothetical scenario where the whole Helitron sequence has been conserved (or 'static') since the
497 HT events. We considered an μ value of 3.0×10^{-9} mutations per generation per site per haploid genome
498 for *D. rhopaloa*, *C. vestalis* and *B. mori*, based on the direct measures conducted on four insect species
499 from three different orders (Keightley et al. 2014a, 2014b, Yang et al. 2015, Liu et al. 2016), and on the
500 possibility that insects might have the same mutation rate (Liu et al. 2016). For alignment and

estimation of evolutionary divergence between sequences (*K*) we used MEGA7 (Kumar et al. 2016). As the *T* values are given in number of generations, and we were interested in estimate the HT dates in millions of years (*mya*), two averages for the number of generations per year (gen/y) were used: one for the *D. rhopaloa* – *C. vestalis* (14 gen/y) and another for the *C. vestalis* – *B. mori* (11 gen/y) HT event. They were based on the known estimates for each taxon: ~ 10 gen/y for the *D. melanogaster* species group (McDonald and Kreitman 1991, Cutter 2008, Tochen et al. 2014, Asplen et al. 2015), ~ 18 gen/y for Braconidae (Nikam and Sathe 1983, Nikam and Pawar 1993), and ~ 4 gen/y for *B. mori* (Maekawa et al. 1988, Reddy et al. 1999).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol, 215(3), 403-410.

Asplen MK, Anfora G, Biondi A, Choi DS, Chu D, Daane KM, et al. 2015. Invasion biology of spotted wing Drosophila (*Drosophila suzukii*): a global perspective and future priorities. J Pest Sci, 88(3), 469-494.

Balakrishnan L, Bambara RA. 2013. Flap endonuclease 1. Annu. Rev. Biochem, 82, 119-138.

Banks JC, Whitfield JB. 2006. Dissecting the ancient rapid radiation of microgastrine wasp genera using additional nuclear genes. Mol. Phylogenet. Evol, 41(3), 690-703.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2013. GenBank. Nucleic Acids Res, 41(D1), D36-D42.

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

Bochman ML, Sabouri N, Zakian VA. 2010. Unwinding the functions of the Pif1 family helicases. DNA repair, 9(3), 237-249.

Branca A, Le Ru BP, Vavre F, Silvain JF, Dupas S. 2011. Intraspecific specialization of the generalist parasitoid Cotesia sesamiae revealed by polydnavirus polymorphism and associated with different Wolbachia infection. Mol. Ecol, 20(5), 959-971.

Brázda V, Kolomazník J, Lýsek J, Hároníková L, Coufal J, Šťastný J. 2016. Palindrome analyser – A new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. Biochem. Biophys. Res. Commun. http://dx.doi.org/10.1016/j.bbrc.2016.09.015

Brosius J, Gould SJ. 1992. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". Proc Natl Acad Sci U.S.A, 89(22), 10706-10710.

Burke GR, Strand MR. 2012a. Deep sequencing identifies viral and wasp genes with potential roles in replication of *Microplitis demolitor* bracovirus. J. Virol, JVI-06434.

Burke GR, Strand MR. 2012b. Polydnaviruses of parasitic wasps: domestication of viruses to act as gene delivery vectors. Insects, 3(1), 91-119.

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

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

Cameron PJ, Walker GP. 1997. Host specificity of *Cotesia rubecula* and *Cotesia plutellae*, parasitoids of white butterfly and diamondback moth. Proc 50th NZ Plant Protection Conf 1997. New-Zealand. 236–241.

Carareto CM. 2011. Tropical Africa as a cradle for horizontal transfers of transposable elements between species of the genera *Drosophila* and *Zaprionus*. Mob Genet Elements, 1(3), 179-186.

Carton Y, Boulétreau M, van Alphen JJM, van Lenteren JC. 1986. The Drosophila parasitic wasps. In: Ashburner M, Carson HL, Thompson JN Jr., editors. The genetics and biology of Drosophila. London: Academic Press; 1986. pp. 347–394

Chen YF, Gao F, Ye XQ, Wei SJ, Shi M, Zheng HJ, Chen XX. 2011. Deep sequencing of *Cotesia vestalis* bracovirus reveals the complexity of a polydnavirus genome. Virology, 414(1), 42-50.

Choi JY, Kwon SJ, Roh JY, Yang TJ, Yoon SH, Kim H, et al. 2009. Sequence and gene organization of 24 circles from the *Cotesia plutellae* bracovirus genome. Arch. Virol, 154(8), 1313-1327.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet 18(2):71-86.

Coates BS. 2015. Horizontal transfer of a non-autonomous Helitron among insect and viral genomes. BMC genomics, 16(1), 137.

Cutter AD. 2008. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. Mol. Biol. Evol, 25(4), 778-786.

Drezen JM, Bézier A, Lesobre J, Huguet E, Cattolico L, Periquet G, & Dupuy C. 2006. The few virus-like genes of Cotesia congregata bracovirus. Arch Insect Biochem Physiol, 61(3), 110-122.

Drezen JM, Chevignon G, Louis F, Huguet E. 2014. Origin and evolution of symbiotic viruses associated with parasitoid wasps. Curr Opin Insect Sci, 6, 35-43.

Dupuy C, Periquet G, Serbielle C, Bézier A, Louis F, Drezen JM. 2011. Transfer of a chromosomal Maverick to endogenous bracovirus in a parasitoid wasp. Genetica, 139(4), 489-496.

567 Furihata S, Matsumura T, Hirata M, Mizutani T, Nagata N, Kataoka M, et al. 2016. Characterization of
568 venom and oviduct components of parasitoid wasp *Asobara japonica*. PLoS One, 11(7), e0160210.

569 Furlong MJ, Wright DJ, Dosdall LM. 2013. Diamondback moth ecology and management: problems,
570 progress, and prospects. Annu. Rev. Entomol, 58, 517-541.

571 Gadallah NS, Ghahari H. 2015. An annotated catalogue of the Iranian Braconinae (Hymenoptera:
572 Braconidae). Entomofauna, 36, 121-176.

573 Gould SJ, Vrba ES. 1982. Exaptation—a missing term in the science of form. Paleobiology, 8(01), 4-15.

574 Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, et al. 2016. A Helitron transposon
575 reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nat. Commun, 7.

576 Graur D, Li W-H. 2000. Fundamentals of molecular evolution. Sinauer Associates, Sunderland, MA.

577 Gundersen-Rindal D, Dupuy C, Huguet E, Drezen JM. 2013. Parasitoid polydnaviruses: evolution,
578 pathology and applications: Dedicated to the memory of Nancy E. Beckage. Biocontrol Sci Technol,
579 23(1), 1-61.

580 Guo X, Gao J, Li F, Wang J. 2014. Evidence of horizontal transfer of non-autonomous Lep1 Helitrons
581 facilitated by host-parasite interactions. Sci. Rep, 4, 5119.

582 Gupta A, Das AK, Neog K, Verghese A. 2016. First report of *Cotesia dictyoplocae* (Hymenoptera:
583 Braconidae), a larval parasitoid of *Antheraea assamensis* (Lepidoptera: Saturniidae), from India. Fla.
584 Entomol, 99(3), 541-543.

585 Heimpel GE, Neuhauser C, Hoogendoorn M. 2003. Effects of parasitoid fecundity and host resistance on
586 indirect interactions among hosts sharing a parasitoid. Ecol. Lett, 6(6), 556-566.

587 Herniou EA, Huguet E, Thézé J, Bézier A, Periquet G, Drezen JM. 2013. When parasitic wasps hijacked
588 viruses: genomic and functional evolution of polydnaviruses. Phil. Trans. R. Soc. B, 368(1626), 20130051.

589 Ideo S, Watada M, Mitsui H, Kimura MT. 2008. Host range of *Asobara japonica* (Hymenoptera:
590 Braconidae), a larval parasitoid of drosophilid flies. Entomol. Sci, 11(1), 1-6.

591 Junier T, Pagni M. 2000. Dotlet: diagonal plots in a web browser. Bioinformatics, 16(2), 178-179.

592 Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a
593 database of eukaryotic repetitive elements. Cytogenet. Genome Res, 110(1-4), 462-467.

594 Kapitonov VV, Jurka J. 2001. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U.S.A, 98(15),
595 8714-8719.

596 Kapitonov VV, Jurka J. 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. Trends
597 Genet, 23(10), 521-529.

598 Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014a. Estimation of the spontaneous mutation rate per
599 nucleotide site in a *Drosophila melanogaster* full-sib family. Genetics, 196(1), 313-320.

600 Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, et al. 2014b. Estimation of
601 the spontaneous mutation rate in *Heliconius melpomene*. Mol Biol Evol 2015; 32 (1): 239-243. doi:

10.1093/molbev/msu302.

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: Repbase Submitter and Censor. BMC bioinformatics, 7(1), 474.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 2016; 33 (7): 1870-1874. doi: 10.1093/molbev/msw054.

Liu H, Jia Y, Sun X, Tian D, Hurst LD, Yang S. 2016. Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. Mol Biol Evol 2017; 34 (1): 119-130. doi: 10.1093/molbev/msw226.

Loreto ELS, Carareto CMA, Capy P. 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity, 100(6), 545-554.

Louis F, Bézier A, Periquet G, Ferras C, Drezen JM, 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(17), 9649-9660.

Maekawa H, Takada N, Mikitani K, Ogura T, Miyajima N, Fujiwara H, et al. 1988. Nucleolus organizers in the wild silkworm *Bombyx mandarina* and the domesticated silkworm *B. mori*. Chromosoma, 96(4), 263-269.

Malysh JM, Kazartsev IA, Frolov AN, Zverev AA, Tokarev YS. 2016. Molecular detection of *Cotesia vestalis* (Hymenoptera: Braconidae) in the beet webworm *Loxostege sticticalis* L. (Lepidoptera: Crambidae). J. Appl. Entomol., 140: 232–235.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res 2015; 43 (D1): D222-D226. doi: 10.1093/nar/gku1221.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature, 351(6328), 652.

Michel-Salzat A, Whitfield JB. 2004. Preliminary evolutionary relationships within the parasitoid wasp genus *Cotesia* (Hymenoptera: Braconidae: Microgastrinae): combined analysis of four genes. Syst. Entomol, 29(3), 371-382.

Moretti S, Armougom F, Wallace IM, Higgins DG, Jongeneel CV, Notredame C. 2007. The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. Nucleic Acids Res, 35(suppl 2), W645-W648.

Nikam PK, Sathe TV. 1983. Life tables and intrinsic rate of natural increase of *Cotesia flavipes* (Cam.) (Hymen., Braconidae) population on *Chilo partellus* (Swin.) (Lep., Pyralidae). J Appl Entomol, 95(1-5), 171-175.

Nikam PK, Pawar CV. 1993. Life tables and intrinsic rate of natural increase of Bracon hebetor Say (Hym., Braconidae) population on *Corcyra cephalonica* Staint. (Lep., Pyralidae), a key parasitoid of *Helicoverpa armigera* Hbn. (Lep., Noctuidae). J. App. Entomol., 115(1-5), 210-213.

Pike JE, Henry RA, Burgers PM, Campbell JL, Bambara RA. 2010. An alternative pathway for okazaki

638 fragment processing resolution of fold-back flaps by Pif1 helicase. J. Biol. Chem., 285(53), 41712-41723.

639 Quicke DLJ. 2015. The Braconid and Ichneumonid Parasitoid Wasps: Biology, Systematics, Evolution and
640 Ecology. First Edition. John Wiley & Sons, Ltd.

641 Reddy KD, Abraham EG, Nagaraju J. 1999. Microsatellites in the silkworm, *Bombyx mori*: abundance,
642 polymorphism, and strain characterization. Genome, 42(6), 1057-1065.

643 Rincon C, Bordat D, Löhr B, Dupas S. 2006. Reproductive isolation and differentiation between five
644 populations of *Cotesia plutellae* (Hymenoptera: Braconidae), parasitoid of *Plutella xylostella*
645 (Lepidoptera: Plutellidae). Biol Control, 36(2), 171-182.

646 Rizki RM, Rizki TM. 1990. Parasitoid virus-like particles destroy *Drosophila* cellular immunity. Proc Natl
647 Acad Sci USA, 87(21), 8388-8392.

648 Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA,
649 Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a
650 large model space. Syst Biol, 61(3), 539-542.

651 Sathe T, Jadhav A. 2001. Host plant attractivity in a model, *Cotesia glomeratus- Bombyx mori* –
652 Mulberry. Sericologia, 41(3), 459-470.

653 Seetharam AS, Stuart GW. 2013. Whole genome phylogeny for 21 *Drosophila* species using predicted 2b-
654 RAD fragments. PeerJ, 1, e226.

655 Shaw, MR. 2003. Revised synonymy in the genus Cotesia (Hymenoptera: Braconidae: Microgastrinae):
656 the identity of Microgaster vestalis Haliday, 1834, as a senior synonym of Apanteles plutellae
657 Kurdjumov, 1912. Entomol Gaz., 54, 187-189.

658 Shaw MR, Huddleston T. 1991. Classification and biology of braconid wasps. Handbooks for the
659 identification of British insects, 7(11), 126.

660 Silva JC, Loreto EL, Clark JB. 2004. Factors that affect the horizontal transfer of transposable
661 elements. Curr Issues Mol Biol, 6, 57-71.

662 Strand MR, Burke GR. 2013. Polydnavirus-wasp associations: evolution, genome organization, and
663 function. Curr Opin Virol, 3(5), 587-594.

664 Suwito A, Ishida TA, Hattori K, Kimura MT. 2002. Environmental Adaptations of Two Flower Breeding
665 Species of *Drosophila* from Java, Indonesia (Behavior and Ecology). Entomol Sci, 5(4), 399-406.

666 Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-
667 transversion and G + C-content biases. Mol Biol Evol 9:678-687.

668 Thomas J, Schaack S, Pritham EJ. 2010. Pervasive horizontal transfer of rolling-circle transposons among
669 animals. Genome Biol Evol, 2, 656-664.

670 Thomas J, Pritham EJ. 2015. Helitrons, the eukaryotic rolling-circle transposable elements. Microbiol
671 Spectr, 3(4).

672 Tochen S, Dalton DT, Wiman N, Hamm C, Shearer PW, Walton VM. 2014. Temperature-related

673 development and population parameters for *Drosophila suzukii* (Diptera: Drosophilidae) on cherry and
674 blueberry. Environ Entomol, 43(2), 501-510.

675 Venner S, Miele V, Terzian C, Biémont C, Daubin V, Feschotte C, Pontier D. 2017. Ecological networks to
676 unravel the routes to horizontal transposon transfers. PLoS biology, 15(2), e2001536.

677 Wallau GL, Ortiz MF, Loreto ELS. 2012. Horizontal transposon transfer in eukarya: detection, bias, and
678 perspectives. Genome Biol Evol, 4(8), 689-699.

679 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2: A Multiple
680 Sequence Alignment and Analysis Workbench. Bioinformatics 25 (9) 1189-1191 doi:
681 10.1093/bioinformatics/btp033.

682 Wharton RA. 1993. Bionomics of the Braconidae. Ann Rev Entomol, 38(1), 121-143.

683 Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud
684 O, Paux E, SanMiguel P, Schulman AH. 2007. A unified classification system for eukaryotic transposable
685 elements. Nat Rev Genet, 8(12), 973-982.

686 Wu SF, Sun FD, Qi YX, Yao Y, Fang Q, Huang J, Stanley D, Ye G. 2013. Parasitization by *Cotesia chilonis*
687 influences gene expression in fatbody and hemocytes of *Chilo suppressalis*. PloS one, 8(9), e74309.

688 Yang S, Wang L, Huang J, Zhang X, Yuan Y, Chen JQ, Hurst LD, Tian D. 2015. Parent-progeny sequencing
689 indicates higher mutation rates in heterozygotes. Nature, 523(7561), 463-467.

690 Zalucki MP, Shabbir A, Silva R, Adamson D, Shu-Sheng L, Furlong MJ. 2012. Estimating the economic cost
691 of one of the world's major insect pests, *Plutella xylostella* (Lepidoptera: Plutellidae): just how long is a
692 piece of string? J Econ Entomol, 105(4), 1115-1129.

693 Zhang HH, Shen YH, Xiong XM, Han MJ, Qi DW, Zhang XG. 2016a. Evidence for horizontal transfer of a
694 recently active Academ transposon. Insect Mol Biol, 25(3), 338-346.

695 Zhang HH, Li GY, Xiong XM, Han MJ, Dai FY. 2016b. Horizontal transfer of a novel Helentron in
696 insects. Mol Genet Genom, 1-8.

697 Žikić V, Stanković SS, Ilić M. 2012. Checklist of the genus *Bracon* (Hymenoptera: Braconidae) in
698 Serbia. Biologica Nyssana, 3(1), 21-29.

699 Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids
700 Res., 31(13), 3406-3415.

701

702

703

704

705

Figure Captions

Figure 1. Schematic diagram of the Helitron-containing segment c35 from *Cotesia vestalis* Bracovirus (CvBV). (A) CvBV segments are encapsidated as double-strand DNA circles. (B) Segment c35 contains an ORF (light purple) flanked by sequences similar to Helitron TEs (dark purple). This segment also has a 373 bp viral sequence (green). (C) Structural and coding features of Hel_c35. The Rep catalytic core residues are depicted in black, and the Helitron consensus Rep domain (from Kapitonov and Jurka 2007) is shown below in grey.

Figure 2. Putative *C. vestalis* (strain ANU101) proviral locus which contains a sequence very similar (~ 99.9% identity) to Hel_c35. This region is marked by the presence of several sequences similar to other polydnaviruses (in blue), and the presence of direct and inverted repeats (black arrowheads). The asterisk denotes a PDV conserved region with a complex array of direct, inverted and tandem repeats.

Figure 3. Transcripts from the *Cotesia* parasitized host, *Chilo suppressalis* (Lepidoptera), displaying similarity with Hel_c35 regions.

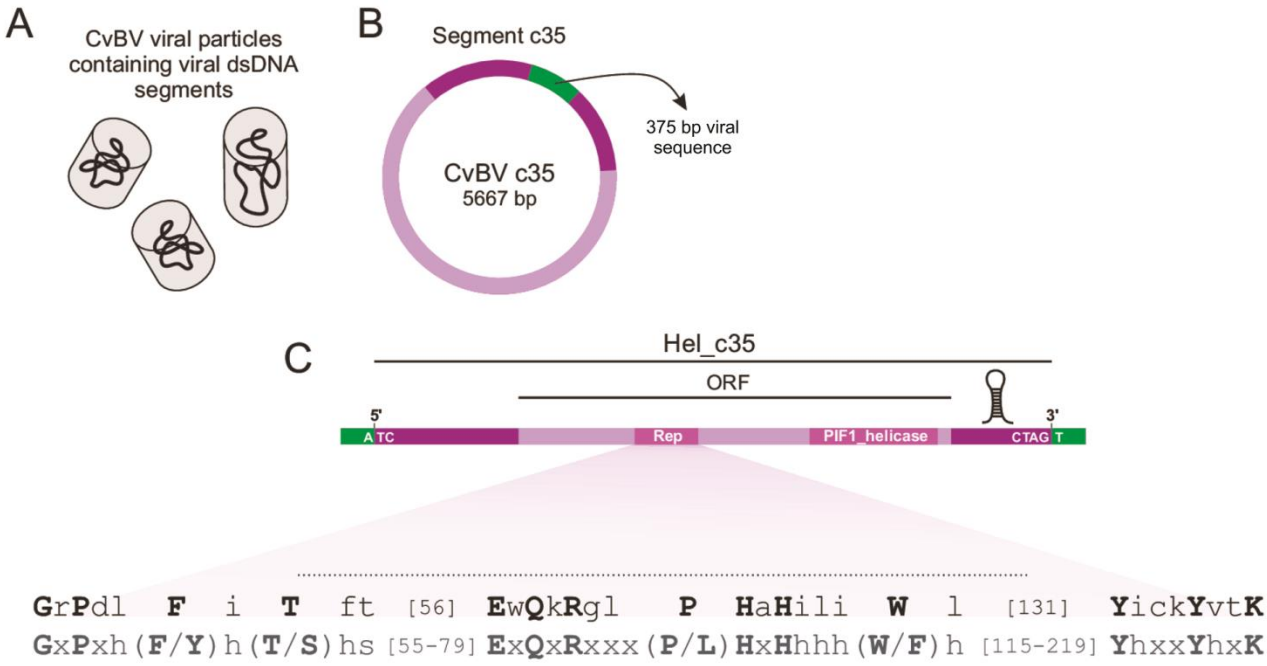
Figure 4. Cotesia phylogeny with four species (in bold) from which we investigated the presence of Hel_c35 in the respective BV genomes. Presence of Hel_c35 in C. vestalis, and putative presence of Hel_c35 in C. chilonis represented by green and orange stars, respectively. Adapted from Michel-Salzat and Whitfield (2004).

Figure 5. Phylogeny of Hel_c35 ORF related sequences in several insect genomes. Species name colors indicate the major geographical regions in which these species are found. *Cotesia vestalis*-CvBV clade is highlighted in yellow. A list of species and accession numbers is given in S2 Table.

Figure 6. Hypothetical evolutionary history of Hel_c35. This TE was already present in the ancestor of the oriental subgroup of the *melanogaster* species group. Less than one million years ago (mya), possibly during non-specific parasitation of *Drosophila rhopaloa*, Hel_c35 was horizontally transferred to the *Cotesia vestalis* lineage. Following transpositional activity in *C. vestalis* chromosomes, Hel_c35 eventually inserted into an active proviral locus and begun being replicated and encapsidated as the segment c35. Now replicated to high copy numbers and able to infect cells, it was horizontally transferred again (< 0.3 mya) to *Bombyx mori*, a likely Lepidopteran host of *Cotesia vestalis*.

741 **Figure 1**

742



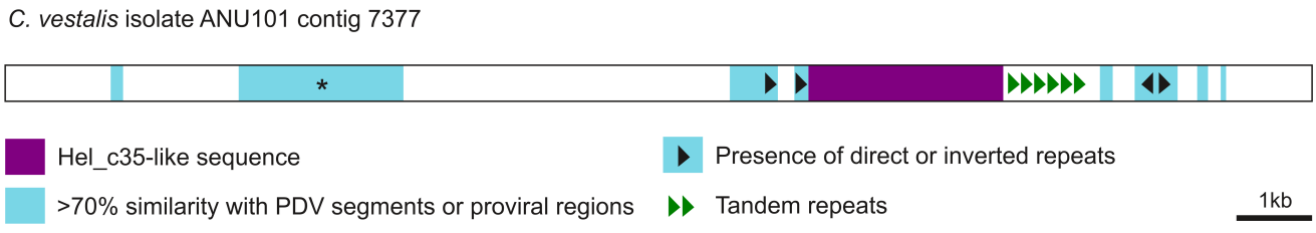
743

744

745

746 **Figure 2**

747



748

749

Figure 3

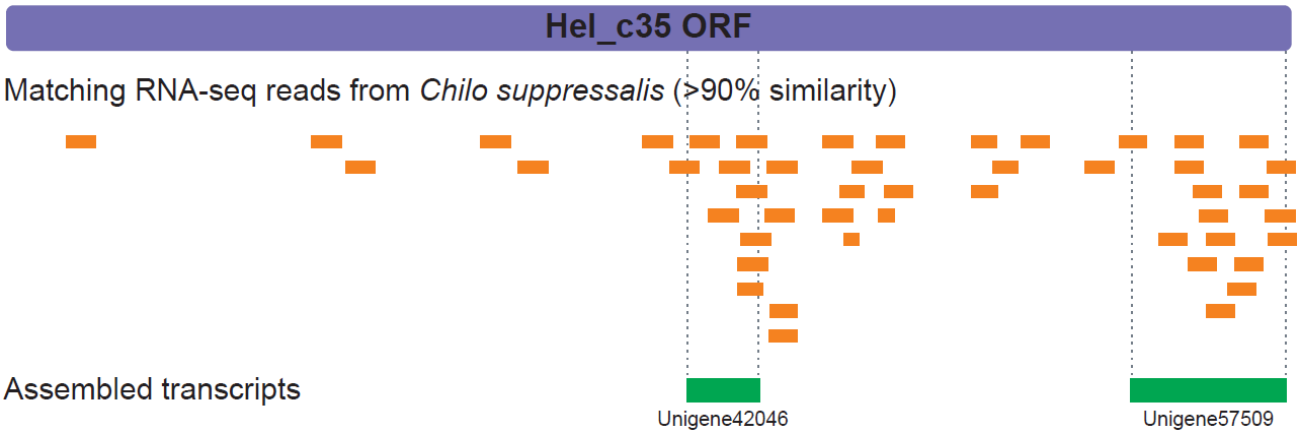


Figure 4

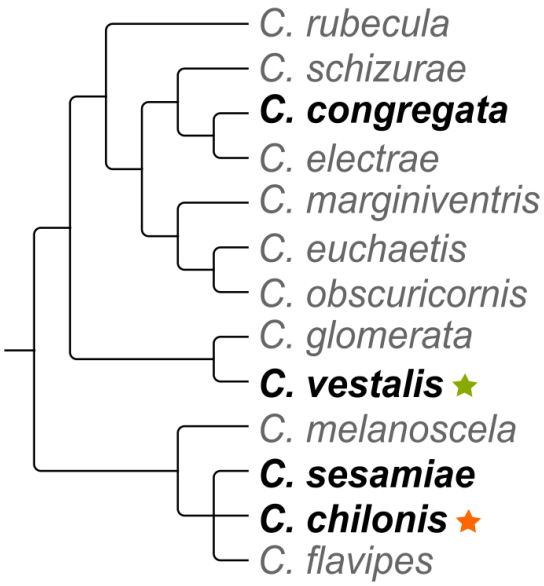
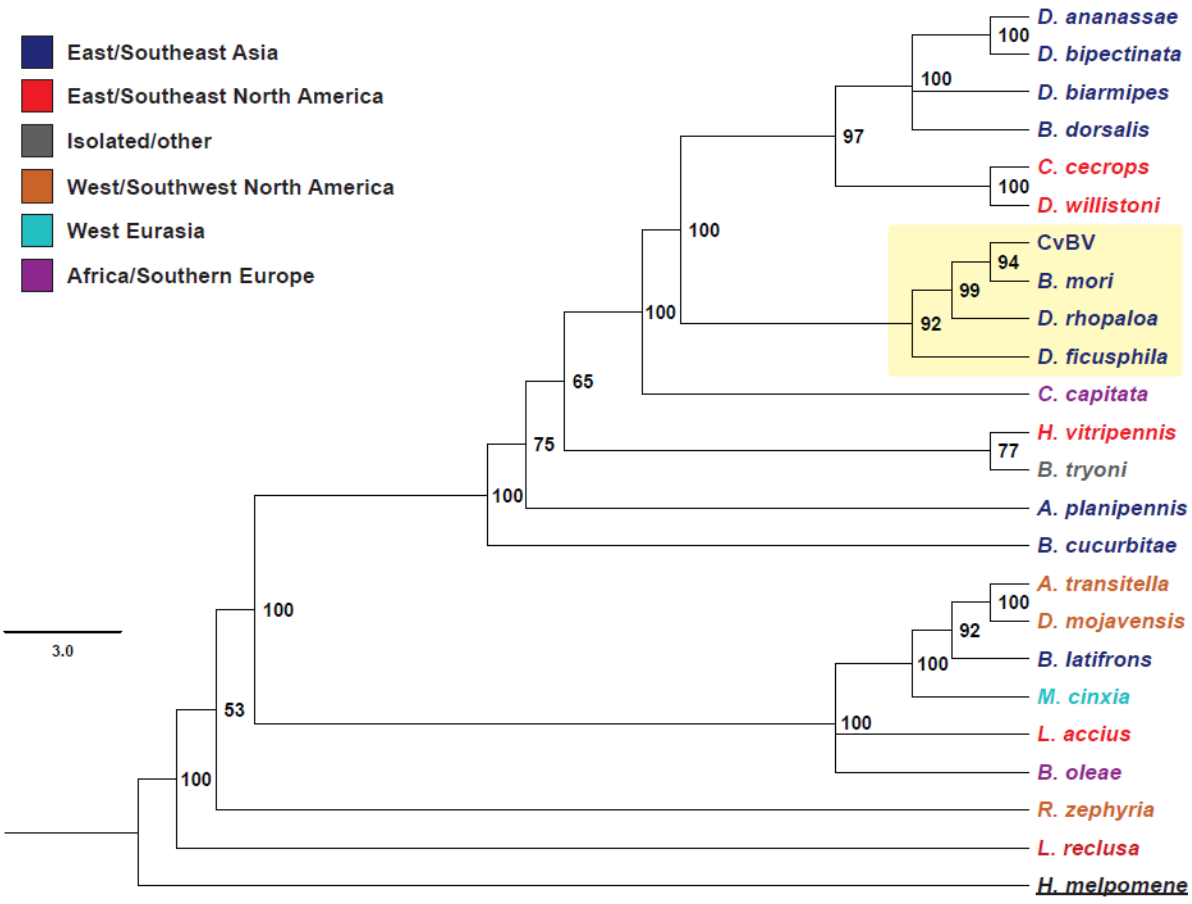
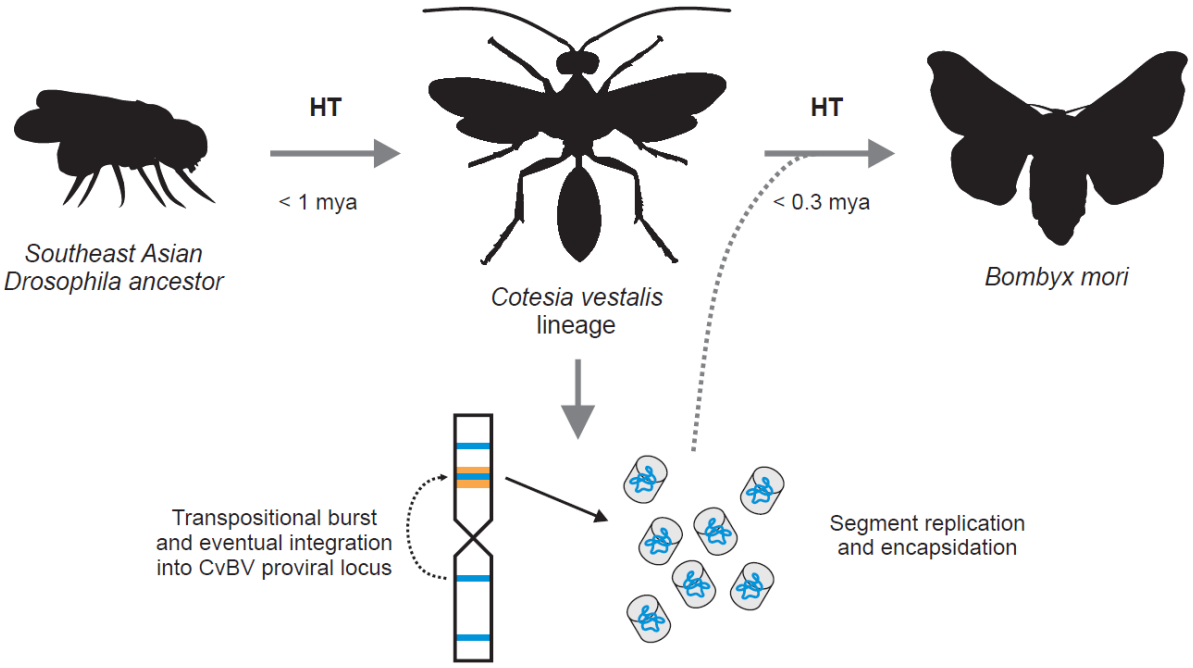


Figure 5



767 **Figure 6**

768



769