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 [1]). 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 [2, 3]. 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 [4].

There are more than 18,000 braconid species described to date, but some estimates point to a total number exceeding 40,000 [5]. Braconid wasps are parasitoids of a wide range of insect orders, and are one of the most successful insects used in biological control programs [6]. An important species from this family is *Cotesia vestalis*, who has been found parasitizing Lepidoptera species in various regions of Asia, Europe, Africa and the Americas [7]. Also, *C. vestalis* is one of the main parasitoids of the diamondback moth *Plutella xylostella*, which is a pest of *Brassica* plants that costs annually US$4-5

52 billion worldwide in crop loss and management [8]. For those reasons, *C. vestalis* has been successfully
53 used in numerous biological control introductions [7].

54 The complete genome sequencing of *Cotesia vestalis* bracovirus (CvBV) revealed 157 ORFs
55 distributed in 35 encapsidated segments, with most of these genes also having homologs in other BVs
56 [9]. In the same work, the authors noted that a segment from CvBV (CvBV_c35) encodes a protein
57 displaying similarity with the human Pif1 helicase, which has no sequence homology with any other
58 gene described in BVs. The Pif1 family is a member of the superfamily 1B helicases and is involved in
59 many replication related functions in eukaryotes [10]. Because all PDVs analyzed to date only express
60 replication genes in the calyx cells of wasps, and do not replicate in host cells [1, 11], the function of this
61 helicase in a CvBV encapsidated segment is worth of investigation, as noted by Chen et al. [9].

62 In the present work, we found that segment CvBV_c35 is almost entirely made by a rolling-circle
63 Helitron transposon, which has fused to the CvBV genome. Although there has been reports of
64 transposable elements (TE) (including Helitrons) in BV genomes (e.g. [12–16]), we describe for the first
65 time a case where an autonomous TE took over a viral segment. Based on our results we propose a
66 scenario to explain the presence of this Helitron in the virus, and suggest two consecutive horizontal
67 transfer (HT) events involving this TE across species from three insect orders: Diptera (oriental
68 Drosophila), Hymenoptera (*C. vestalis*), and Lepidoptera (*Bombyx mori*). Finally, we discuss the
69 conservation of the Helitron as a PDV segment, the possibility of a co-option, and its implications for the
70 study of HT events.

71

72 Results and Discussion

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

74 The segment c35 (CvBV_c35) (gb: HQ009558.1) from *Cotesia vestalis* bracovirus (CvBV) has
75 5,667 bp and contains a single ORF, described as a DNA helicase that has never been reported on PDVs
76 [9]. Indeed, its annotated amino acid sequence (AEE09607.1) has a C-terminal Pif1-like helicase domain
77 which belongs to the P-loop containing Nucleoside Triphosphate Hydrolases (P-loop_NTPase)
78 superfamily. Because the Pif-like domain covers only a relatively small portion of the total amino acid
79 sequence (< 30%) and it is absent in all known PDVs, we further investigated the identity of this protein.

80 We found that the N-terminal half of the protein is a Rep domain which is typical of the Rep-

81 Helicases (Rep/Hel) from Helitron TEs [17]. Subsequent analysis revealed the presence of all three
82 conserved motifs present in the Rep catalytic core [18], confirming the identity of this protein as a
83 Rep/Hel (Fig 1). This group of proteins presents a basic structure consisting of a N-terminal HUH
84 endonuclease motif (Rep) and a C-terminal helicase motif (Hel) and is used by Helitrons during their
85 rolling-circle (RC) transposition (reviewed in [19]).

86 After we determined the Rep/Hel identity of this coding sequence, we verified if the rest of the
87 TE was present in segment c35. By comparing CvBV_c35 and Helitron sequences retrieved from Repbase
88 [20] we could determine the structure of the whole element, with its 5'- and 3'-termini flanking the
89 Rep/Hel ORF. This Helitron, hereafter named Hel_c35, has 5,294 bp, and covers most (~ 93.4%) of the
90 viral segment c35. Because Hel_c35 contains an intact ORF and all the structural hallmarks of Helitrons
91 (see Materials and Methods), we suggest that it should be classified as an autonomous TE. Only a small
92 373 bp region within segment CvBV_c35, but outside the Hel_c35 sequence, lacks significant homology
93 with Helitrons. Moreover, we found the same 373 bp sequence (100% identity) as part of a different
94 CvBV segment (CvBV_c19), indicating that this sequence is probably of viral nature. This 373 bp viral
95 sequence does not display significant similarity with any known transposable elements or insect genome
96 available to date.

97 The analysis performed with the CENSOR tool [21] against the whole Repbase repeat library,
98 using Hel_c35 as a query, indicated that this element has high sequence similarity (70-97%) with
99 Helitrons from Diptera, especially Drosophila species (S1 Fig). Although Hel_c35 displays high sequence
100 identity with these Helitrons, they share less than 80% similarity in their last 30 bp at the 3'-terminus.
101 Therefore Hel_c35 does not belong to any described family, according to the proposed classification of
102 Helitrons [19].

103 Helitron insertions in PDV genomes have been reported before, including in CvBV [13, 15, 16]. It
104 is worth mentioning that some of these studies describe Helitron insertions in the genomes of *Cotesia*
105 *plutellae* and *Cotesia plutellae* bracovirus (CpBV), which are junior synonyms of *C. vestalis* and CvBV,
106 respectively [22]. However, these elements are non-autonomous, and display low sequence identity (<
107 80%) with Hel_c35, including on their terminal sequences (data not shown), thus belonging to a distinct
108 Helitron family.

109 **Searching the CvBV_c35 proviral locus**

110 There are currently two lineages of *C. vestalis* with sequenced genomes: isolate ANU101

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

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

136 We found eight motifs that are conserved in many PDV-related sequences, including several BV
137 circles and proviral regions flanking segments (Fig 2 and S1 Table). Five of the eight conserved motifs are
138 also found within or nearby nudiviral-like genes from *Cotesia congregata* bracovirus (CcBV). Despite not
139 being encapsidated, these genes are involved in BV replication and production of structural viral
140 proteins [2, 3]. The fact that some of these motifs are found in both segment and nudiviral loci, and
141 others are exclusive to segment loci, might be related to their common or specific functions in BVs,

142 respectively. For example, common motifs could have a role in proviral replication, and specific motifs in
143 circle excision, encapsidation and transcription. All the conserved sequences analyzed here have an
144 average AT content of 71% (62.7-80%), and contain internal palindromic regions that could form hairpin
145 folds with variable sizes. Additionally, two of these motifs around the Hel_c35, one upstream and the
146 other downstream, contain long inverse repeats that could also form stem-loop structures (S2 Fig).
147 Similar features have been described in flanking and intermediate sequences of other BV proviral loci,
148 and are thought to be involved in viral replication [23, 24].

149 The presence of several motifs flanking this Hel_c35 in the *C. vestalis* genome suggests that this
150 copy was probably inserted on a proviral loci. If true, why this Hel_c35 copy is different from the one
151 present in the viral genome? This inconsistency might be explained if this Hel_c35 insertion does not
152 correspond to the main proviral CvBV_c35, but to a paralogous segment that could be either active (it
153 only produces a small part of c35 circles) or inactive (such as a 'pseudo-segment'). A large portion of
154 PDV genes found in viral segments belong to multigenic families, and segment reintegration, locus
155 duplication/deletion, and gene gain/loss, are common features during PDV evolution [4, 25].

156 Another possibility, which perhaps is more plausible, is that CvBV_c35 is polymorphic in the
157 different lineages used for the genome sequencing of *C. vestalis* (BioSample: SAMN03273265) and CvBV
158 (HQ009558.1). In addition to different CvBV_c35 segments in distinct lineages, this polymorphism could
159 also include lineages with no CvBV_c35 encapsidated circles. Before the work conducted by Chen et al.
160 [9], the CvBV (referred as CpBV) genome was partially sequenced from a wasp strain of South Korea
161 [26], and no circles containing a Rep/Hel gene were detected. Although this is a partial genome
162 sequence, the missing CvBV_c35 could be a consequence of its real absence in the CvBV particles from
163 this wasp strain. As we described above, the best candidate for a proviral CvBV_c35 sequence also
164 belongs to a wasp lineage from South Korea, in contrast to the CvBV_c35 sequence (HQ009558.1), which
165 is derived from a wasp strain of China. Hence, the differences between the wasp and BV genomes might
166 reflect a polymorphic CvBV_c35 which, depending on the lineage, exists as a proviral segment that
167 generates encapsidated viral circles, or as an 'ancestral' form with the Hel_c35 in a proviral region, but
168 not able to produce viral particles. Although the confirmation of this hypothesis of polymorphism would
169 require the genome sequencing of different *C. vestalis* strains and their respective BVs, it is worth
170 mentioning that there is evidence of significant variation between BVs from different wasp populations
171 of the same species [27, 28].

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

173 The data described here so far indicate a Helitron-BV fusion. In this context, it is important to
174 investigate when, during the evolution of BV, this event happened. Apart from CvBV, we did not find the
175 Hel_c35 Rep/Hel ORF in any other PDV sequenced genome available to date, including those from two
176 other bracoviruses present in two *Cotesia* species: *Cotesia congregata* bracovirus (CcBV) and *Cotesia*
177 *sesamiae* bracovirus (CsBV). Although there are no other *Cotesia* sequenced genomes available, Wu et
178 al. [29] described the influence of *Cotesia chilonis* on its lepidopteran host *Chilo suppressalis* during
179 parasitization, by analyzing the transcriptome of fat body and hemocytes in the larvae of this moth. In
180 addition to host genes, they identified 19 unique sequences associated with PDVs, which were classified
181 as *Cotesia chilonis* BV (CchBV) transcripts; so we further searched for sequences similar to Hel_c35 on
182 their Short Read Archive (SRA) (accession: SRR651040) and Transcriptome Shotgun Assembly (TSA) (gb:
183 GAJS00000000) databases in order to pinpoint the approximate phylogenetic position of the Hel_c35
184 integration into the ancestral BV genome.

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

194 Based on the distribution of the Rep/Hel in the phylogeny of *Cotesia* species (Fig 4) we might
195 assume that a homologous bracovirus segment with a Rep/Hel gene was independently lost twice
196 during *Cotesia* evolution, one in the *C. congregata* and the other in the *C. sesamiae* clade [30]; or that
197 two independent Helitron integration events occurred, one in CvBV and the other in CchBV. The
198 hypothesis of two integration events involving two different Helitron families would explain the disparity
199 of the Hel_c35 ORF sequence similarity with Unigene42046 (99%) and Unigene57509 (71%). In this case,
200 both Unigenes would be part of one single ORF from a Helitron family that is distinct from Hel_c35. This
201 hypothesis is also supported by our date estimate of when Hel_c35 was first inserted in the *C. vestalis*
202 genome (< 1 mya) (see topic "Tracing an evolutionary pathway of Hel_c35 in insect genomes"), which
203 appears to be too recent to have happened before the divergence of the two species. Nevertheless, the

204 genome sequencing of *C. chilonis* and CchBV will be necessary to test (i) the identity between Hel_c35
205 and Unigene42046, (ii) the presence of this TE within CchBV, and (iii) the homology of this insertion with
206 the one found in segment CvBV_c35.

207 **Horizontal transfer of Hel_c35 between three insect orders**

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

213 The resultant tree revealed some strikingly incongruences with the species' phylogeny (Fig 5).
214 Most notably, Hel_c35 from *Bombyx mori*, *Drosophila rhopaloa* and *Drosophila ficusphila* were allocated
215 in the same clade as the Hel_c35 from *Cotesia vestalis* bracovirus (CvBV). It is important to mention that
216 we did not find Hel_c35 in any other hymenopteran species, even though more than forty genomes
217 from this insect order were used in our search. Because the marked phylogenetic incongruences and the
218 extreme discontinuous or "patchy" distribution of Hel_c35 are indicative of horizontal transfer (HT)
219 events [31, 32], we decided to further investigate this possibility.

220 Another line of evidence used to evaluate the occurrence of HT is the high sequence similarity
221 between DNA sequences from distantly related taxa, when compared with more closely related ones
222 [31, 32]. We found that the nucleotide identity between Hel_c35 from CvBV, and Hel_c35 from *B. mori*,
223 *D. rhopaloa* and *D. ficusphila* sequences are ~ 99.7%, ~ 96% and ~ 94%, respectively.

224 For a HT event to be inferred, the species distributions also should be considered, as the
225 candidate taxa must overlap geographically in order to a HT event to occur [33, 34]. Thus, we classified
226 each taxon in the phylogeny to one of eight major geographical regions, based on their known
227 distributions (see Materials and Methods). The addition of this information revealed interesting aspects
228 that helped to explain some of the main inconsistencies found in the phylogeny (Fig 5). First, among the
229 ten closest taxa with Hel_c35, seven are native from eastern/southestern Asia, including the three
230 species that group immediately with the Hel_c35 from CvBV (*B. mori*, *D. rhopaloa* and *D. ficusphila*).
231 Second, the positioning of *Bactrocera dorsalis* (Tephritidae) within a well-supported clade of Drosophila
232 species, instead of a position closer to other tephritid fruit flies, also coincides with their common
233 distribution in southeastern Asia. Third, the incongruent grouping of *Calycopis cecrops* (Lepidoptera)

234 with *Drosophila willistoni* (Diptera), involve two species with geographic overlapping in southeastern
235 United States. Fourth, despite the fact that *Drosophila mojavensis* and the lepidopteran *Amyelois*
236 *transitella* belong to different insect orders, they are grouped together in a well-supported clade and
237 overlap geographically in southwestern United States. Thus, the geographical distribution of the
238 analyzed taxa supports the HT hypothesis as an explanation for at least some of the phylogenetic
239 incongruences, especially the one directly related to CvBV. Other incongruences outside the immediate
240 CvBV clade might also represent bona fide HTs; however, each of them requires a careful analysis that is
241 beyond the scope of the present work.

242 For HT events, the spatial overlapping of candidate species must be associated with at least
243 some degree of ecological overlapping [33, 34]. Accordingly, braconid wasps interact very closely with
244 several insect host orders, including lepidopteran and Drosophila species [5, 6] and as a rule, these
245 connections reach the cellular and even to the chromosomal level during wasp oviposition and
246 bracovirus infection of host cells [1]. It is noteworthy that these extremely direct interactions are not
247 occasional, but a fundamental part of the braconid wasps' life cycle.

248 In summary, our results suggest at least two HT events involving Hel_c35: one between a
249 Drosophila species and *C. vestalis* and another between *C. vestalis* and *B. mori*, which probably occurred
250 in this respective order, as indicated by the phylogeny. This hypothesis is supported by four lines of
251 evidence: (i) the marked incongruence between host and TE phylogeny, (ii) the patchy distribution of
252 this TE on the main taxa involved, (iii) the high sequence identity between Hel_c35 copies from different
253 insect orders, and (iv) the spatial/ecological overlap among the candidate species.

254 **Tracing an evolutionary pathway of Hel_c35 in insect genomes**

255 Because our constructed phylogeny with Hel_c35 sequences suggests HT events between *D.*
256 *rhopaloa*, *C. vestalis* and *B. mori*, we estimated when these events occurred using the equation for
257 divergence time ($T = K/2r$) on non-coding sequences (~ 750 bp) from Hel_c35 copies of these species.
258 We also set minimum and maximum date thresholds, supposing different evolutionary constrains acting
259 on these sequences (see Materials and Methods). This analysis can also help to infer the order of the
260 two HT events which, according to the phylogeny, appear to have occurred from *D. rhopaloa* to *C.*
261 *vestalis* (represented by CvBV), and then from *C. vestalis* to *B. mori*. The results give an approximate
262 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
263 mya) for the *C. vestalis* – *B. mori* HT.

264 Based on our results, we suggest the following scenario to explain the observed distribution of
265 Hel_c35 copies among insect genomes. The transposable element Hel_c35 belongs to a Helitron family
266 originally present in southeastern Asian Drosophila species and, more specifically, within the oriental
267 subgroup of the *Drosophila melanogaster* species group. This taxon comprises the largest number of
268 species with Hel_c35 copies clustering with CvBV, and displays a topology roughly coherent with the
269 group's phylogeny [35]. Initially, a Hel_c35 element from a Drosophila species in the *rhopaloa* subgroup
270 was horizontally transferred into the ancestral genome of *C. vestalis* ~ 0.862 mya, probably during
271 parasitization. After genome integration, Hel_c35 transposed into the CvBV proviral genome in a locus
272 responsible for segment production, also becoming part of the encapsidated genome of this bracovirus.
273 Then, a second HT event occurred ~ 0.211 mya, this time involving the transfer of Hel_c35 from *C.*
274 *vestalis* to *B. mori*, through CvBV_c35 circles (Fig 6). If after entering the proviral locus, Hel_c35 'fused'
275 gradually with CvBV, the observed sequence inconsistencies between CvBV_c35 from different *C.*
276 *vestalis* lineages could represent incomplete stages of this fusion on different wasp strains. In that case,
277 the formation of segment CvBV_c35 would have been completed after geographical divergence of the
278 lineages used for wasp (S. Korea) and BV (China) genome sequencing.

279 To hold true, the above hypothesis might require at least two assumptions that should be
280 addressed. First, that *C. vestalis* is or at least was until recently capable to parasitize Drosophila species,
281 in addition to their typical lepidopteran hosts. This hypothesis seems improbable at a first sight, because
282 all hosts of Microgastrinae wasps described to date are Lepidoptera larvae [5]. Indeed, the major
283 radiation dates for both taxa seem to coincide, reinforcing the idea that Microgastrinae have evolved as
284 a group of specialized Lepidoptera parasitoids [36]. On the other hand, it is also important to note the
285 remarkable host diversity and variability within the Braconidae family. For instance, the subfamilies
286 Alysiinae and Opiinae are endoparasitoids of many cyclorrhaphous Diptera, including Drosophila and
287 tephritid fruit flies [6, 37], and several wasp species in the Braconinae subfamily can attack both
288 dipteran and lepidopteran hosts, depending on the location and time of the year [38, 39]. The
289 Exothecinae subfamily contains parasitoids of several insect orders: for instance, *Colastes braconius* can
290 parasitize dipteran, lepidopteran, coleopteran and even hymenopteran species [40]. Other ecological
291 aspects are also relevant if we consider the host range and specificity within the *Cotesia* genus. The first
292 is that, even though *Plutella xylostella* is commonly described as the main host of *C. vestalis* in the
293 literature, this wasp is capable to parasitize a wide range of Lepidoptera families and superfamilies,
294 which are extremely diverse in their ecology and morphology [41, 42].

295 It is also worth mentioning that, although to our knowledge there is no report of *C. vestalis*
296 parasitization on *B. mori*, the close related species *Cotesia glomerata* can use this moth as a host [43],
297 and *Cotesia dictyoplocae*, which is in the same species group of these two wasps, parasitizes moths of
298 the Bombycoidea superfamily [44]. Additionally, the present work is not the first to report a HT between
299 *C. vestalis* and *B. mori* (e.g. [15, 45, 46]), indicating that *C. vestalis* could indeed attack this lepidopteran
300 species, and might also parasitize a wider variety of hosts.

301 Secondly, the known hosts of a parasitoid wasp do not necessarily correspond to their actual
302 host range, and may simply represent the commonly attacked species of which successful parasitization
303 is more likely to ensue. It has been suggested that unusual conditions could induce attacks outside the
304 suitable host range, and even result in successful parasitization of unsuitable hosts, notwithstanding
305 their rarity [5, 47]. It is also noteworthy that some Drosophila species phylogenetically close to *D.*
306 *rhopaloa* can be resistant to attacks from the wasp *Asobara japonica*, probably because of a long-
307 continued interaction with braconid parasitoids [48, 49]. Additionally, on southeast Asia, two Drosophila
308 species closely related to *D. rhopaloa* breed on plants that are also commonly used as food by
309 Lepidoptera larvae [50], which could facilitate an encounter between Cotesia wasps and unusual
310 dipteran hosts. Hence, it is not difficult to conceive that *Cotesia* wasps attack, and sometimes can
311 successfully parasitize Drosophila larvae, even if these interactions may be rare.

312 **PDV-Helitron exaptation?**

313 There are several reports of TE integrations in PDV (e.g. [12, 14]). Moreover, Helitrons have
314 already been found in the genomes of Cotesia species and in their respective PDVs, and were also
315 involved in HT events [13, 15, 16]. However, all these insertions represent non-autonomous short or
316 fragmented elements. To our knowledge, we describe the first instance of a putative autonomous
317 Helitron that not only integrated the PDV genome, but effectively became a viral segment.

318 Exaptation refers to features originally evolved for some function but that were later co-opted
319 for a different role [51]. Repetitive DNAs like TEs can also take part on this process [52], as revealed by
320 many studies in eukaryotes (recently reviewed by [53]). Although there is evidence for the role of TEs on
321 PDV evolution, including the exaptation of a gene apparently derived from from a retroelement
322 (reviewed in [54]), we report the first intact TE that appears to have been co-opted by a bracovirus. The
323 main reasons for this suggestion are: (i) this TE appears to be autonomous; (ii) it has conserved an intact
324 structure in the genome of *C. vestalis* for the last ~ 1 my, despite the low number of copies; (iii) the
325 element occupies almost 94% of segment CvBV_c35, which is one of the most replicated circles of this

326 virus [9]; (iv) it contains the only ORF in CvBV_c35, and all known PDV segments have coding sequences,
327 with a few exceptions [55]; (v) the CvBV_c35 portion outside Hel_c35 has only 373 bp, and does not
328 contain any ORF or conserved sequences found in other PDVs (apart from segment CvBV_c19). Although
329 it is possible that this short sequence has a functional role (e.g. as an encapsidation signal), the whole
330 PDV segment CvBV_c35 is essentially a Helitron. Furthermore, we found a single full copy of Hel_c35 in
331 the *C. vestalis* genome, even though *D. rhopaloa* and *B. mori* contain several highly similar partial copies
332 (data not shown). This unique arrangement for a TE in a eukaryote genome reinforces the suggestion
333 that segment CvBV_c35 has been kept as an intact Helitron by selective constraints.

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

337 The first hypothesis is that this Helitron Rep/Hel was exaptated by CvBV for its helicase function,
338 to aid the amplification of viral replication units or circles in wasp calyx cells. Most viral genes thought to
339 be essential for BV replication have not being identified, with a few exceptions, like a nudiviral helicase
340 and a fen-like flap endonuclease [4]. The Pif1 family of helicases (which include the Hel domain in
341 Rep/Hel proteins) and the flap endonucleases are important to process DNA secondary structures like
342 hairpin or fold-back substrates during replication [56, 57]. Interestingly, TA-rich inverted repeats and
343 other palindromic motifs capable of forming hairpins are found throughout the proviral genomes of BVs,
344 and probably serve as replication origins [23, 24]. Although possible, this hypothesis does not explain
345 why the Rep/Hel ORF is part of the CvBV encapsidated genome, as that would not be necessary for
346 replication in calyx cells.

347 The second hypothesis, is the use of the whole Helitron (with its non-coding sequences) as a
348 mechanism for specific segment CvBV_c35 amplification in calyx cells of wasps, which could explain why
349 CvBV_c35 constitutes one of the most abundant circles in CvBV, despite being a new segment. This
350 'auto-replication' would be useful if PDV capsids also play a role on virulence, independently of the
351 genes they carry. In that case, abundant segments could be selected for their ability to assemble more
352 capsids that are later injected into the host. This hypothesis is based on the observation that *Microplitis*
353 *demolitor* produces some BV encapsidated segments with no apparent coding sequences [55], and
354 several parasitoid wasp species, including braconids, use virus-like particles (VLP) with no detected PDV
355 sequences as virulence factors against their hosts [4, 49, 58]. In contrast to the first hypothesis, this
356 scenario also predicts the conservation of Hel_c35 non-coding sequences, as they would be at least

357 partially necessary for segment CvBV_c35 to take advantage of the Helitron rolling-circle replication
358 (RCR) mechanism. That is because Rep/Hel proteins recognize Helitron terminal sequence motifs to start
359 and finish RCR correctly [59]. Although the sequences used in our analysis display slightly higher
360 nucleotide conservation on coding regions in comparison with non-coding regions (data not shown), the
361 exact level of sequence stability necessary for Rep/Hel recognition and RCR viability of Helitrons is not
362 well understood.

363 The third hypothesis is that CvBV_c35 ORF could be used to replicate its own circle and other
364 CvBV circles, once within host cells. Like the second hypothesis, this explanation considers the fact that
365 Hel_c35 Rep/Hel is part of the BV encapsidated genome, but in this case, the Rep/Hel protein would
366 further amplify other CvBV genes after their injection in the host, in addition to CvBV_c35 circles. This
367 mechanism of circle replication in host cells could further increase the number of CvBV genes, and
368 therefore enhance virulence. This scenario would be particularly interesting to investigate, as all PDVs
369 analyzed to date only express replication genes in the calyx cells of wasps, and apparently do not
370 replicate after encapsidation and injection in host cells [1, 11]. Thus, the hypothesis of Hel_c35
371 exaptation by CvBV for some replication-related function is plausible, even though the replication of this
372 group of viruses is not fully understood.

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

379

380 Conclusion

381 There are several reports of TE integrations in PDVs. However, all these cases involve classical TE
382 insertions within PDV genomes, mostly of non-autonomous and fragmented elements. In contrast, our
383 study revealed an autonomous Helitron within a bracovirus segment, which is not only an integrated
384 copy, but a TE covering almost entirely a viral circle. The segment CvBV_c35 is effectively an intact
385 Helitron, and our results suggest that this TE might have been recently co-opted by CvBV, probably for
386 its helicase and/or rolling-circle replication function. This PDV-Helitron fusion points to a new type of

387 relationship between TEs and viruses. In addition, our data reinforces the idea of PDVs as effective
388 agents of horizontal transfer (HT), and of Helitrons as one of the TEs most commonly involved in those
389 events. Specifically, we reported two consecutive transfers of Hel_c35 across three insect orders.

390 The number of reported HT events in eukaryotes has been growing recently, although the
391 probability of their occurrence must be very low, as they require the presence of several biological
392 features and conditions. Notably, the tripartite parasitoid system composed by braconid wasps, PDVs
393 and hosts fulfill most of those requirements [60], being exceptional candidates for the study of HT in
394 eukaryotes. Indeed, there are several reports of HT events involving insect species associated with these
395 parasitoid systems [61]. As pointed out by Quicke [5], because the number of estimated Microgastrinae
396 species ranges between 16,000-40,000 and apparently, all of them have associated PDVs, this could be
397 the biggest known group of viruses. The future genome sequencing of more PDVs and wasp species will
398 help to understand the real importance of HT for the evolution of parasitoids and their hosts.
399 Furthermore, the investigation of events like this can contribute to ecological analyses, by revealing
400 potential hosts and cryptic interactions not detected in field studies.

401

402 ## Materials and Methods

403 ### Rep/Hel amino acid analysis

404 To confirm that the CvBV_c35 codes a Rep/Hel we first analyzed its amino acid sequence (AEE09607.1)
405 using the NCBI's conserved domain database (CDD) [62]. Rep/Hel ORF sequences were retrieved from
406 Repbase [20], either directly or using the CENSOR tool [21] with CvBV_c35 ORF as a query. Sequences
407 were then aligned using MEGA7 [63]. We determined the residues that characterize Rep domains by
408 visually inspecting the CvBV_c35 amino acid sequence, based on the consensus from Kapitonov and
409 Jurka [18] (Fig 1C).

410 ### Hel_c35 structure hallmarks

411 To determine the whole Helitron structure we first used the CvBV_c35 sequence as a query in a search
412 against the Repbase reference collection [21] (S1 Fig). After visually establishing the element's putative
413 boundaries, we determined its precise termini by aligning CvBV_c35 with the best results, and verified if
414 the sequence contained the hallmarks of Helitrons, like the insertion between AT nucleotides, the
415 hairpin structure close to the 3'-end, and the conserved terminal nucleotides. To further validate the TE
416 limits, we used the putative Helitron sequence (named Hel_c35) as a query for a Blastn search [64]

417 against the genome of *Cotesia vestalis* available on GenBank [65], and then used the best results,
418 together with their flanking sequences as a query to a second Blastn search. Both searches were
419 conducted using default parameters. Most results with multiple hits only display similarity with Hel_c35
420 sequences within the terminal nucleotides, suggesting that our defined limits for the element
421 encompass the whole TE.

422 Search for CvBV_c35 proviral locus and analysis of PDV conserved sequences

423 To Blast-search [64] the CvBV_c35 proviral locus, we used its sequence (HQ009558.1) as a query against
424 the *C. vestalis* genome on GenBank [65] using default parameters and selecting the best result based on
425 score, query cover and identity. We then scanned the selected contig (gb: JZSA01007369.1) searching
426 for sequences displaying similarity with PDV genomes and other PDV-related sequences available on
427 Genbank [65]. To detect structural features on the PDV-like sequences found in the selected contig, we
428 first used dotplots from the software Dotlet [66], followed by manual curation of sequences. Finally, the
429 analysis of palindromic or inverted repeat motifs, and their predicted secondary structures, were made
430 using the softwares mfold [67] and Palindrome analyser [68].

431 Transcriptome analysis of *Chilo suppressalis*

432 We used the Rep/Hel coding DNA sequence (CDS) of CvBV_c35 (HQ009558.1) as a query for Blastn
433 searches against the Short Read Archive (SRA) (SRR651040) and Transcriptome Shotgun Assembly (TSA)
434 (GAJS00000000) of *Chilo suppressalis* generated by Wu et al. [29]. Reads with identity > 90%, and
435 Unigenes with score ≥ 200 were selected. The sequences of selected Unigenes were then used as
436 queries for Blastn searches against genomes from all insect orders for comparison. All searches were
437 conducted using default parameters. The sequences from Unigene42046 and the corresponding region
438 from the CvBV_c35 CDS were aligned using M-coffee [69] (S3 Fig).

439 Search on Arthropoda genomes, alignment and phylogeny

440 We conducted Blastn searches [64] on all Arthropoda genomes available on GenBank [65], using a query
441 of 1,675 bp from the Hel_c35 Rep/Hel ORF located in the 'Rep region' and applying default parameters.
442 This region contains the Rep motif that is exclusive from Helitrons (S1 Data), in contrast to the 'Hel
443 region' and its Hel motif belonging the Pif1 family of helicases, which are involved in many cellular
444 processes and pervasive in eukaryotes [10]. Therefore, to avoid false positives, and because most of the
445 initial sampling searches using the whole ORF (4,538bp) retrieved only fragmented hits, we decided to
446 use a shorter query covering only the Rep region. This selected nucleotide stretch correspond to the

447 “Helitron_like_N” domain annotated on the NCBI's CDD [62], the rest of the catalytic core that we
448 manually selected using the Helitron consensus (see ‘Rep/Hel amino acid analysis’ above), and an
449 intermediate region between Rep and Hel (S1 Data). In order to include in our study only TEs from the
450 same family (see [70] for a hierarchical proposal for transposable elements classification), results with >
451 70% cover and > 80% identity were selected based on their Max score, and then aligned using MUSCLE
452 on MEGA7 [63]. The final alignment was formatted on Jalview [71] to be better visualized (S4 Fig).
453 Although we recognize that Helitrons families should be classified in a manner distinct from most DNA
454 TEs, due to their peculiar structure and evolutionary dynamics [19], our interest was to retrieve coding
455 sequences, including the ones from elements lacking their terminal 30 bp used in the mentioned
456 classification. For the phylogeny of aligned Rep sequences, we used MrBayes 3.2 [72] to run a mixed
457 analysis, which samples across the different nucleotide substitution models. The results were compared
458 with the phylogeny estimated using MEGA7 [63], with the Maximum Likelihood method, and using the
459 Tamura 3-parameter model [73], which resulted a similar branch topology (data not shown). The main
460 geographical distributions of the taxa included in the phylogeny were retrieved from various Web
461 sources.

462 Date estimation of HT events

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

465 $T = K/2r$

466 where T is the number of generations, K is the number of substitutions per site, and r is the rate of
467 nucleotide substitution, which is equal to the mutation rate (μ) for neutral mutations [74]. The
468 sequences used comprised the 5’ and 3’ non-coding flanking regions of the Hel_c35 ORF, which together
469 sum ~ 750 bp. The copy from CvBV was used to represent *C. vestalis*, and as a query for Blast searches in
470 the genomes of *D. rhopaloa* and *B. mori*. Although we used non-coding sequences flanking the Hel_c35
471 ORF for the analysis, assuming they evolve neutrally, the importance of these regions for Helitron
472 transposition are not fully understood, apart from the terminal 40 bp on each end [59]. As the selective
473 pressure on these sequences cannot be discarded, the assumption of neutrality should be taken
474 cautiously. For that reason, we used one equation ($T = K/2\mu$) for a minimum date estimation, with both
475 sequences evolving neutrally, and another ($T = K/\mu$) for a maximum date estimation, with only one
476 sequence evolving neutrally. This second equation (maximum threshold) gives the divergence time for a
477 single branch in the phylogeny [75], and represents the hypothetical scenario where the whole Helitron

478 sequence has been conserved (or 'static') since the HT events. We considered an μ value of 3.0 x 10^{-9}
479 mutations per generation per site per haploid genome for *D. rhopaloa*, *C. vestalis* and *B. mori*, based on
480 the direct measures conducted on four insect species from three different orders [76–79], and on the
481 possibility that insects might have the same mutation rate [79]. For alignment and estimation of
482 evolutionary divergence between sequences (*K*) we used MEGA7 [63]. As the *T* values are given in
483 number of generations, and we were interested in estimate the HT dates in millions of years (*mya*), two
484 averages for the number of generations per year (gen/y) were used: one for the *D. rhopaloa – C. vestalis*
485 (14 gen/y) and another for the *C. vestalis – B. mori* (11 gen/y) HT event. They were based on the known
486 estimates for each taxon: ~ 10 gen/y for the *D. melanogaster* species group [75, 80, 81, 82], ~ 18 gen/y
487 for Braconidae [83, 84], and ~ 4 gen/y for *B. mori* [85, 86].

488

489

490

491 **References**

492

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

496 2. Strand MR, Burke GR. Polydnavirus-wasp associations: evolution, genome organization, and
497 function. Curr Opin Virol. 2013;3(5): 587–594.

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

500 4. Herniou EA, Huguet E, Thézé J, Bézier A, Periquet G, Drezen JM. When parasitic wasps hijacked
501 viruses: genomic and functional evolution of polydnaviruses. Phil Trans R Soc B. 2013;368(1626):
502 20130051.

503 5. Quicke DLJ. The Braconid and Ichneumonid Parasitoid Wasps: Biology, Systematics, Evolution
504 and Ecology. First Edition. John Wiley & Sons, Ltd.; 2015.

505 6. Wharton RA. Bionomics of the Braconidae. Ann Rev Entomol. 1993;38(1): 121–143.

506 7. Furlong MJ, Wright DJ, Dosdall LM. Diamondback moth ecology and management: problems,
507 progress, and prospects. Annu. Rev. Entomol. 2013;58: 517–541.

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

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

513 10. Bochman ML, Sabouri N, Zakian VA. Unwinding the functions of the Pif1 family helicases. DNA
514 repair. 2010;9(3): 237–249.

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

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

519 13. Thomas J, Schaack S, Pritham EJ. Pervasive horizontal transfer of rolling-circle transposons
520 among animals. Genome Biol Evol. 2010;2: 656–664.

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

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

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

527 17. Kapitonov VV, Jurka J. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U.S.A.
528 2001;98(15): 8714–8719.

529 18. Kapitonov VV, Jurka J. Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet.
530 2007;23(10): 521–529.

531 19. Thomas J, Pritham EJ. Helitrons, the eukaryotic rolling-circle transposable elements. Microbiol
532 Spectr. 2015;3(4): MDNA3-0049-2014

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

535 21. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive
536 elements in Repbase: Repbase Submitter and Censor. BMC bioinformatics. 2006;7(1): 474.
537 Available from: http://www.girinst.org/repbase/.

538 22. Shaw, MR. Revised synonymy in the genus Cotesia (Hymenoptera: Braconidae: Microgastrinae):
539 the identity of Microgaster vestalis Haliday, 1834, as a senior synonym of Apanteles plutellae
540 Kurdjumov, 1912. Entomol Gaz. 2003;54: 187–189.

541 23. Louis F, Bézier A, Periquet G, Ferras C, Drezen JM, Dupuy, C. The bracovirus genome of the
542 parasitoid wasp *Cotesia congregata* is amplified within 13 replication units, including sequences
543 not packaged in the particles. J. Virol. 2013;87(17): 9649–9660.

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

547 25. Burke GR, Strand MR. Deep sequencing identifies viral and wasp genes with potential roles in
548 replication of *Microplitis demolitor* bracovirus. J Virol. 2012;86(6): 3293–3306.

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

551 27. Rincon C, Bordat D, Löhr B, Dupas S. Reproductive isolation and differentiation between five
552 populations of *Cotesia plutellae* (Hymenoptera: Braconidae), parasitoid of *Plutella xylostella*

(Lepidoptera: Plutellidae). Biol Control. 2006;36(2): 171–182.

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

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

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

31. Silva JC, Loreto EL, Clark JB. Factors that affect the horizontal transfer of transposable elements. Curr Issues Mol Biol. 2004;6: 57–71.

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

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

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

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

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

37. Carton Y, Boulétreau M, van Alphen JJM, van Lenteren JC. 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

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

39. Gadallah NS, Ghahari H. An annotated catalogue of the Iranian Braconinae (Hymenoptera: Braconidae). Entomofauna. 2015;36: 121–176.

40. Shaw MR, Huddleston T. Classification and biology of Braconid wasps. Handbooks for the Identification of British Insects, vol. 7 (Part 11). Royal Entomological Society of London, London, UK; 1991. 126 p.

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

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

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

594 44. Gupta A, Das AK, Neog K, Verghese A. First report of *Cotesia dictyoplocae* (Hymenoptera:
595 Braconidae), a larval parasitoid of *Antheraea assamensis* (Lepidoptera: Saturniidae), from
596 India. Fla Entomol. 2016;99(3): 541–543.

597 45. Zhang HH, Shen YH, Xiong XM, Han MJ, Qi DW, Zhang XG. Evidence for horizontal transfer of a
598 recently active Academ transposon. Insect Mol Biol. 2016;25(3): 338–346.

599 46. Zhang HH, Li GY, Xiong XM, Han MJ, Dai FY. Horizontal transfer of a novel Helentron in
600 insects. Mol Genet Genomics. 2016;292(1): 243–250.

601 47. Heimpel GE, Neuhauser C, Hoogendoorn M. Effects of parasitoid fecundity and host resistance
602 on indirect interactions among hosts sharing a parasitoid. Ecol Lett. 2003;6(6): 556–566.

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

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

608 50. Suwito A, Ishida TA, Hattori K, Kimura MT. Environmental Adaptations of Two Flower Breeding
609 Species of *Drosophila* from Java, Indonesia (Behavior and Ecology). Entomol Sci. 2002;5(4): 399–
610 406.

611 51. Gould SJ, Vrba ES. Exaptation—a missing term in the science of form. Paleobiology. 1982;8(1):
612 4–15.

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

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

617 54. Burke GR, Strand MR. Polydnaviruses of parasitic wasps: domestication of viruses to act as gene
618 delivery vectors. Insects. 2012;3(1): 91–119.

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

621 56. Pike JE, Henry RA, Burgers PM, Campbell JL, Bambara RA. An alternative pathway for okazaki
622 fragment processing resolution of fold-back flaps by Pif1 helicase. J Biol Chem. 2010;285(53):
623 41712–41723.

624 57. Balakrishnan L, Bambara RA. Flap endonuclease 1. Annu Rev Biochem. 2013;82: 119–138.

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

627 59. Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, et al. A Helitron transposon
628 reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nat
629 Commun. 2016;7: 10716.

630 60. Venner S, Miele V, Terzian C, Biémont C, Daubin V, Feschotte C, Pontier D. Ecological networks
631 to unravel the routes to horizontal transposon transfers. PLoS Biol. 2017;15(2): e2001536.

632 61. Drezen J-M, Gauthier J, Josse T, Bézier A, Herniou E, Huguet E. Foreign DNA acquisition by
633 invertebrate genomes. J Invertebr Pathol. 2016; https://doi.org/10.1016/j.jip.2016.09.004.

634 62. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's
635 conserved domain database. Nucleic Acids Res. 2015;43(D1): D222–D226. Available from:
636 https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml.

637 63. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for
638 bigger datasets. Mol Biol Evol. 2016;33(7): 1870–1874.

639 64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol
640 Biol. 1990;215(3): 403–410.

641 65. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW.
642 GenBank. Nucleic Acids Res. 2013;41(D1): D36–D42. Available from:
643 https://www.ncbi.nlm.nih.gov/genbank/.

644 66. Junier T, Pagni M. Dotlet: diagonal plots in a web browser. Bioinformatics. 2000;16(2): 178–179.
645 Available from: http://dotlet.vital-it.ch/.

646 67. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids
647 Res. 2003;31(13): 3406–3415. Available from: http://unafold.rna.albany.edu/?q=mfold.

648 68. Brázda V, Kolomazník J, Lýsek J, Hároníková L, Coufal J, Šť'astný J. Palindrome analyser – A new
649 web-based server for predicting and evaluating inverted repeats in nucleotide
650 sequences. Biochem Biophys Res Commun. 2016;478(4): 1739–1745. Available from:
651 http://bioinformatics.ibp.cz/#/en/palindrome.

652 69. Moretti S, Armougom F, Wallace IM, Higgins DG, Jongeneel CV, Notredame C. The M-Coffee web
653 server: a meta-method for computing multiple sequence alignments by combining alternative
654 alignment methods. Nucleic Acids Res. 2007;35(suppl 2): W645–W648. Available from:
655 http://tcoffee.crg.cat/apps/tcoffee/do:mcoffee.

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

659 71. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2: A Multiple
660 Sequence Alignment and Analysis Workbench. Bioinformatics. 2009;25(9): 1189–1191.

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

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

666 74. Graur D, Li W-H. Fundamentals of molecular evolution. 2nd ed. Sunderland (MA): Sinauer
667 Associates; 2000.

668 75. Cutter AD. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of
669 the neutral mutation rate. Mol Biol Evol. 2008;25(4): 778–786.

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

672 77. Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, et al. Estimation of
673 the spontaneous mutation rate in *Heliconius melpomene*. Mol Biol Evol. 2015;32(1): 239–243.

674 78. Yang S, Wang L, Huang J, Zhang X, Yuan Y, Chen JQ, Hurst LD, Tian D. Parent-progeny sequencing

675 indicates higher mutation rates in heterozygotes. Nature. 2015;523(7561): 463–467.

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

679 80. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature.
680 1991;351(6328): 652.

681 81. Tochen S, Dalton DT, Wiman N, Hamm C, Shearer PW, Walton VM. Temperature-related
682 development and population parameters for *Drosophila suzukii* (Diptera: Drosophilidae) on
683 cherry and blueberry. Environ Entomol. 2014;43(2): 501–510.

684 82. Asplen MK, Anfora G, Biondi A, Choi DS, Chu D, Daane KM, et al. Invasion biology of spotted
685 wing Drosophila (*Drosophila suzukii*): a global perspective and future priorities. J Pest Sci.
686 2015;88(3): 469–494.

687 83. Nikam PK, Sathe TV. Life tables and intrinsic rate of natural increase of *Cotesia flavipes* (Cam.)
688 (Hymen., Braconidae) population on *Chilo partellus* (Swin.) (Lep., Pyralidae). J Appl Entomol.
689 1983;95(1-5): 171–175.

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

693 85. Maekawa H, Takada N, Mikitani K, Ogura T, Miyajima N, Fujiwara H, et al. Nucleolus organizers
694 in the wild silkworm *Bombyx mandarina* and the domesticated silkworm *B. mori*. Chromosoma.
695 1988;96(4): 263–269.

696 86. Reddy KD, Abraham EG, Nagaraju J. Microsatellites in the silkworm, *Bombyx mori*: abundance,
697 polymorphism, and strain characterization. Genome. 1999;42(6): 1057–1065.

698

699

700

701

702

703

704

705

706

707

708

709

710

711 ## Figure Captions

712

713 **Figure 1.** Schematic diagram of the Helitron-containing segment c35 from *Cotesia vestalis* Bracovirus
714 (CvBV). (A) CvBV segments are encapsidated as double-strand DNA circles. (B) Segment c35 contains an
715 ORF (light purple) flanked by sequences similar to Helitron TEs (dark purple). This segment also has a
716 373 bp viral sequence (green). (C) Structural and coding features of Hel_c35. The Rep catalytic core
717 residues are depicted in black, and the Helitron consensus Rep domain (from [18]) is shown below in
718 grey.

719 **Figure 2.** Putative *C. vestalis* (strain ANU101) proviral locus which contains a sequence very similar (~
720 99.9% identity) to Hel_c35. This region is marked by the presence of several sequences similar to other
721 polydnaviruses (in blue), and the presence of direct and inverted repeats (black arrowheads). The
722 asterisk denotes a PDV conserved region with a complex array of direct, inverted and tandem repeats.

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

725 **Figure 4**. Cotesia phylogeny with four species (in bold) from which we investigated the presence of
726 Hel_c35 in the respective BV genomes. Presence of Hel_c35 in *C. vestalis*, and putative presence of
727 Hel_c35 in *C. chilonis* represented by green and orange stars, respectively. Adapted from [30].

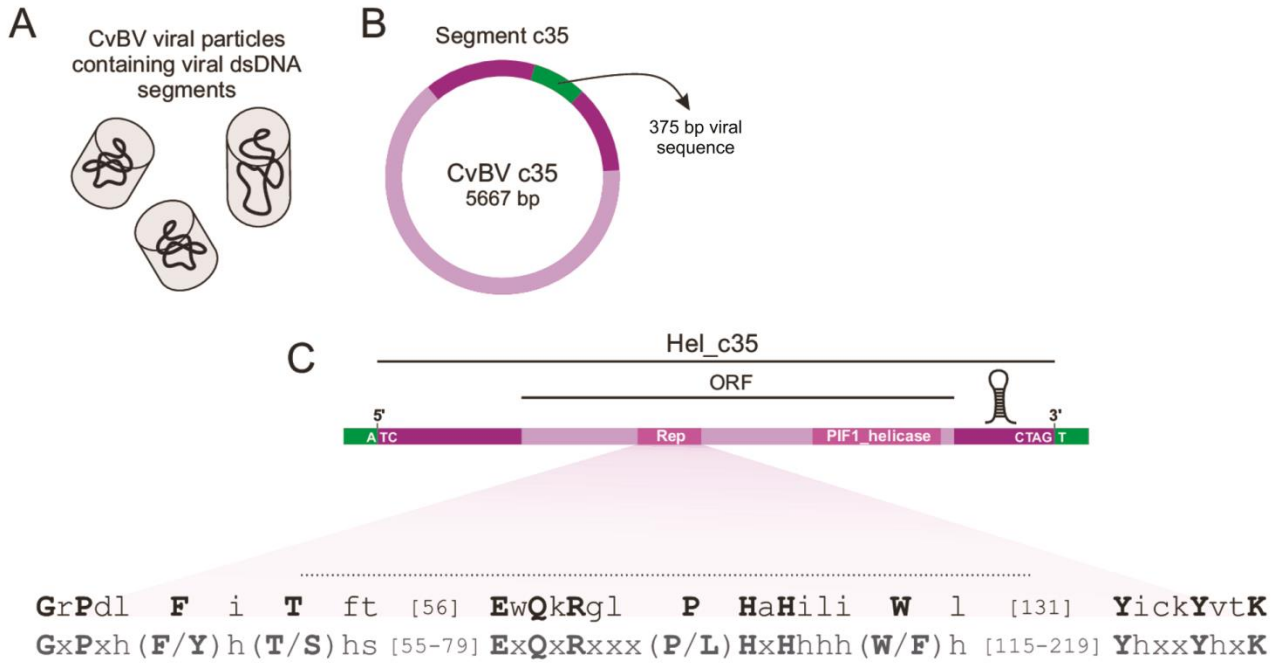
728 **Figure 5**. Phylogeny of Hel_c35 Rep domain sequences in several insect genomes. Species name colors
729 indicate the major geographical regions in which these species are found. CvBV clade is highlighted in
730 yellow. A list of species and accession numbers is given in S2 Table.

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

738

739 **Figure 1**

740



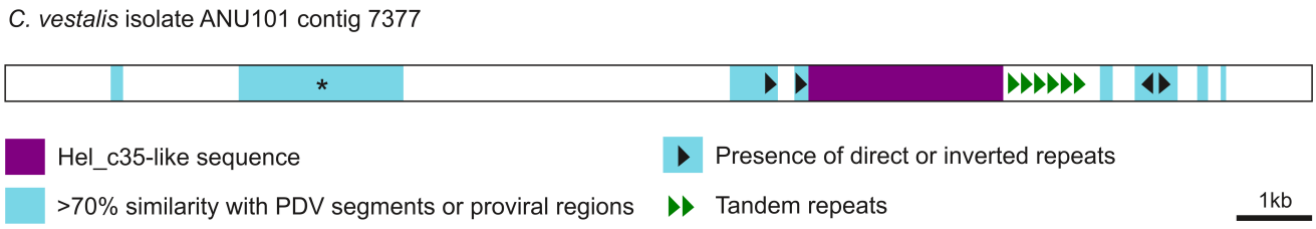
741

742

743

744 **Figure 2**

745

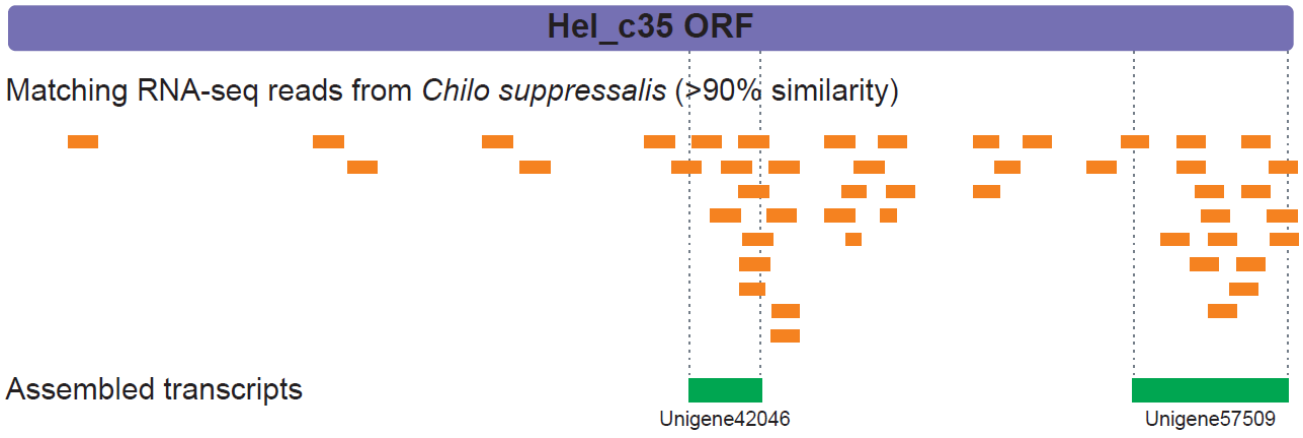


746

747

748 **Figure 3**

749



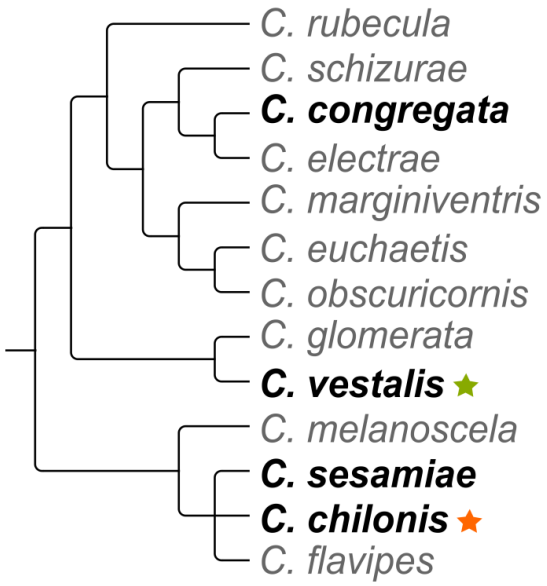
750

751

752

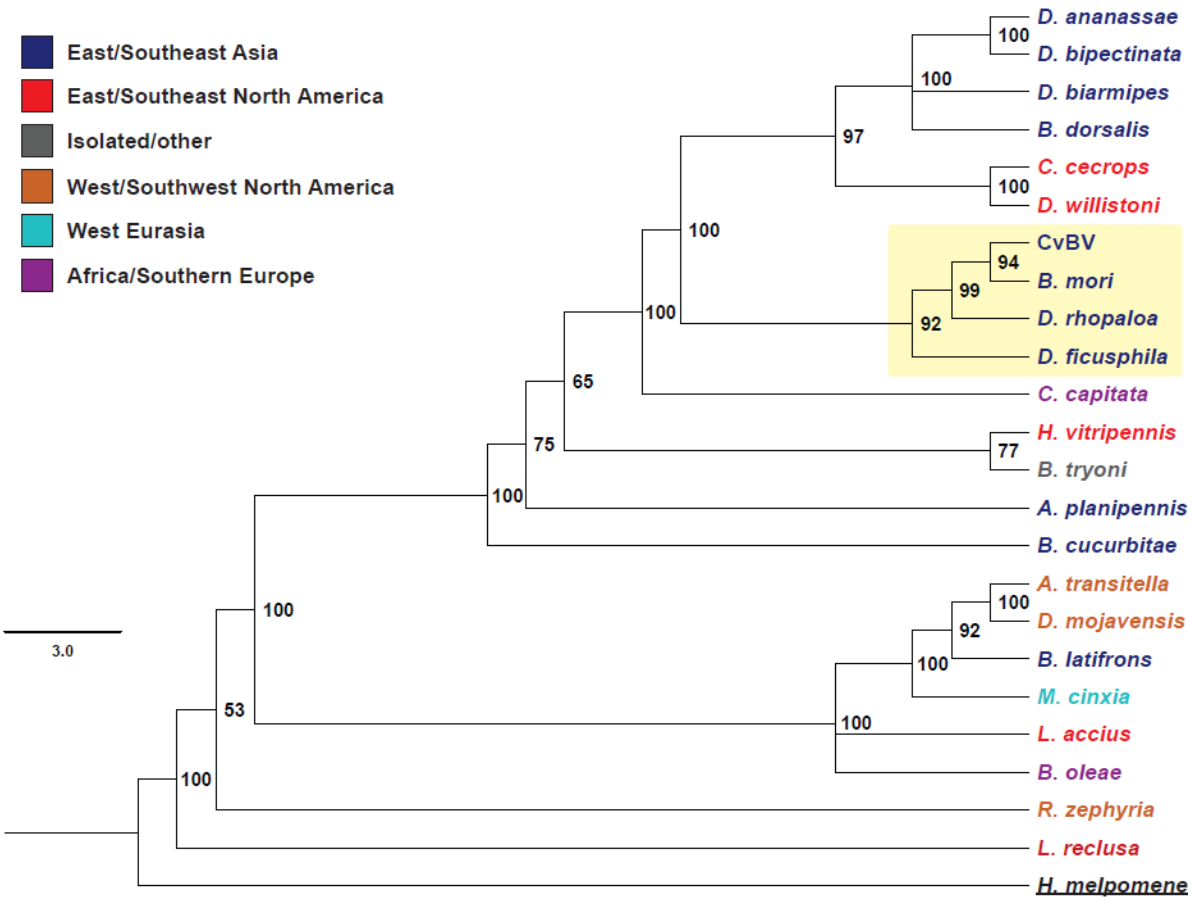
753 **Figure 4**

754



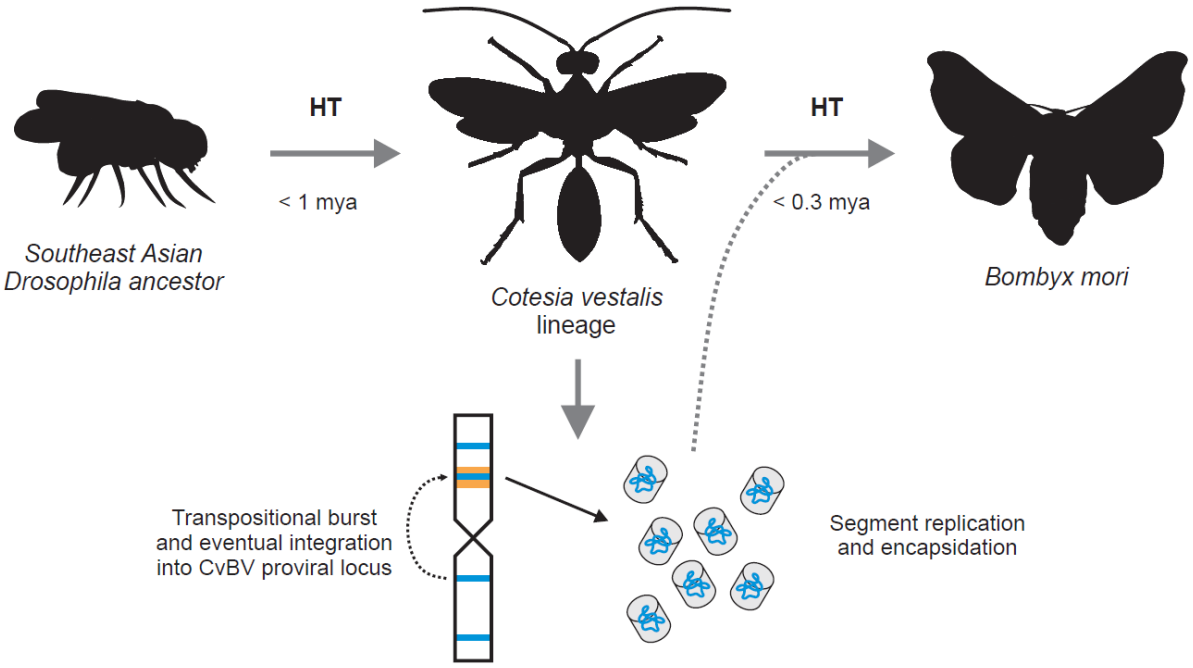
755

Figure 5


East/Southeast Asia
East/Southeast North America
Isolated/other
West/Southwest North America
West Eurasia
Africa/Southern Europe
3.0
D. ananassae
D. bipectinata
D. biarmipes
B. dorsalis
C. cecrops
D. willistoni
CvBV
B. mori
D. rhopaloa
D. ficusphila
C. capitata
H. vitripennis
B. tryoni
A. planipennis
B. cucurbitae
A. transitella
D. mojavensis
B. latifrons
M. cinxia
L. accius
B. oleae
R. zephyria
L. reclusa
H. melpomene
100
100
97
100
100
94
99
92
100
65
75
77
100
100
100
92
100
100
53
100

765 **Figure 6**

766



767