When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses

Elisabeth A. Herniou, Elisabeth Huguet, Julien Thézé, Annie Bézier, Georges Periquet and Jean-Michel Drezen

Institut de Recherche sur la Biologie de l'Insecte, CNRS UMR 7261, Université François-Rabelais, Parc de Grandmont, 37200 Tours, France

1. Introduction

Paleovirolgy and the study of endogenous viral elements (EVEs), corresponding to ancient viral sequence insertions in eukaryotic genomes, are unveiling the long and rich interactions viruses have entertained with their hosts [1–4]. Although polydnaviruses (PDVs) might still be considered atypical, they represent the most compelling evolutionary success among EVEs. The virus ancestors of the Polydnaviridae family were integrated into the genomes of parasitoid wasps (Hymenoptera, Ichneumonoidea, Braconidae), where they were subjected to complex genomic rearrangements. Now, wasps produce and use for their own ends functional infective virus particles that enclose fragmented dsDNA genomes solely encoding virulence genes [5]. The benefit to the virus genome resides in its vertical transmission free of the mutation load generally incurred by non-functional EVEs [6]. Extraordinarily, EVE domestication (i.e. implicating genetic changes and regulated viral particle production) has occurred at least twice during wasp evolution from independent virus families (figure 1) [10,11], and we are probably just beginning to unravel the diversity of parasitoid wasp–virus associations.

Within the insect order Hymenoptera, ichneumonoid wasps encompass the Braconidae and Ichneumonidae, two highly diverse parasitoid families, both in terms of species richness and parasitic strategies [12]. Their larvae develop to the detriment of arthropod hosts, principally of the order Lepidoptera, Coleoptera, Hymenoptera and Hemiptera [13]. Both families have seen the evolution of numerous koinobiont endoparasitoids, in which wasp larvae grow inside...
developing hosts [12]. This particular lifestyle imposes relatively long and intimate relationships between hosts and parasites, giving rise to complex immune and physiological interactions [14]. To overcome host defences, wasps have evolved an arsenal of virulence factors present in their venoms and/or produced in their parasitized hosts by symbiotic PDV genes (figure 1).

PDVs are essentially chimeric viruses composed of viral particles enclosing DNA circles encoding virulence genes supposedly of wasp origin. PDVs have original infection cycles split between two hosts. PDV particles are produced only in wasps, but infect cells of the caterpillar host (figure 2). PDV genomes are stably integrated into the genomes of parasitoid wasps [15]. They are composed of (i) proviral segments used to produce the multiple dsDNA circles that encode virulence genes and that are packaged in infectious particles, and of (ii) genes, encoding the so-called viral machinery that produce the particles (figure 2). Expression of the structural genes as well as excision and packaging of PDV dsDNA circles occurs in specialized cells of the calyx, a particular region of the wasp ovaries located at the bases of the oviducts. During oviposition of parasitoid eggs, PDV particles are injected into the lepidopteran host and infect many lepidopteran cell types but do not replicate. Virulence gene expression leads to modifications in lepidopteran host physiology, such as inhibition of wasp egg encapsulation and developmental manipulations, allowing wasp development and emergence (figure 2) [16–19].

2. Polydnavirus origins

(a) Diversity of wasp–virus association

The Polydnaviridae family encompasses two genera: Bracovirus (BV) and Ichnovirus (IV), both associated with thousands of wasp species from six Braconidae subfamilies and the Ichneumonidae subfamily Campopleginae, respectively. All these wasps are koinobiont parasitoids of lepidopteran larvae [20]. Both BV and IV are symbiotically associated with wasps, and are produced in large amount as large fragmented dsDNA viruses in wasp ovaries, but their particles have different morphological features. Within Ichneumonidae, wasps from the Banchinae subfamily are also associated with viruses. Based on virus particle morphology and wasp phylogenetic relationships (figure 1), banchine PDV were proposed to form a third PDV group [21]. Only the characterization of the viral machinery producing banchine PDV would indicate whether they have a different or common origin with IVs. Furthermore, virus-like particles (VLPs; particles resembling viruses but devoid of nucleic acids) are produced by several wasps from the Figitidae (Hymenoptera, Cynipoidea) and Euphorinae (Braconidae) [22,23], and most notably by the campoplegine Venturia canescens (Ichneumonidae) [24] (figure 1). The origins of these diverse VLPs remain to be elucidated. Venturia canescens VLPs were the first immunosuppressive particles described in the ovaries of a parasitoid wasp, in a seminal paper [25]. Venturia canescens belongs to the Campopleginae, and thus is expected to harbour a regular ichnovirus, but it produces VLPs, in which no DNA is incorporated. One
hypothesis for the origin of *V. canescens* VLPs is that they might correspond to dysfunctional IVs having lost the ability to incorporate DNA, in which case it might be possible to identify the remnants of IVs in the wasp genome. Alternatively, they could be produced by as yet completely unknown cellular processes.

(b) Bracoviruses originate from an ancestral nudivirus

(i) Nudivirus genes are involved in *Bracovirus* production

The ovary transcriptomes of three braconid wasps, *Chelonus inanitus* (Cheloninae), *Cotesia congregata* (Microgastrinae) [10] and *Microplitis demolitor* [26], were analysed to identify genes involved in bracovirus production. A series of 29 nudivirus genes (nudiviruses are large dsDNA insect viruses related to baculoviruses) were expressed in the ovaries (table 1). Furthermore, one-third of these genes encoded BV particle components [10,27,28]. Last, 18 of these nudiviral genes, corresponding to baculovirus core genes [29], should perform essential functions of the virus cycle, based on functional characterization in baculoviruses [30] (table 1). Most viral functions [10,27] such as transcription, particle assembly and packaging and entry into host cells could be identified (table 1): (i) all RNA polymerase subunits involved in baculovirus transcription; (ii) genes encoding the equivalent of the major baculovirus capsid (VP39) [31,32] and a protein involved in nucleocapsid assembly (38K) [33]; (iii) all components of the baculovirus PIF complex involved in cell entry [34]. No transcripts involved in viral DNA replication could be identified, except for a nudiviral helicase gene in *Microplitis demolitor* [26].

Evolutionary rate analyses (GAbranch model [35]) performed on the whole nudiviral gene dataset showed they globally evolved under strong evolutionary constraints. Interestingly, the genetic algorithm showed the BV lineage had significantly higher evolutionary rates (0.06, $d_N/d_S$, 0.2) than the free virus lineages (0.004, $d_N/d_S$, 0.008). In particular, significant $d_N/d_S$ increase ($p < 0.01$, table 1) was observed for a number of genes mostly involved in cell entry (*pif* genes). There are two hypotheses to explain these relaxed selection pressures in domesticated versus free viruses: (i) domestication could have lowered functional constraints on viral genes, to globally evolve at the same rate as the host genome; or (ii) the genes producing BV particles could have been subject to diversifying selection episodes leading, for example, to changes in the number of cell types the virus could enter.

(ii) Genomic organization of nudiviral genes and proviral segments

The integrated form of most nudiviral genes has been identified by analyses of a *C. congregata* genomic bacterial artificial chromosome clone library (see [15] in this issue). Consistent with nudivirus genome integration in ancestral wasps, half of the nudiviral genes are organized in a cluster within the wasp genome (figure 2).

Figure 2. Bracovirus life cycle and genome organization. (a) The BV genome is integrated in the wasp genome (yellow). It is composed of (i) proviral segments (blue) used to produce the multiple dsDNA circles that encode virulence genes (coloured rectangles) and that are packaged in the particles, and of (ii) BV structural genes (nudiviral genes; black or grey rectangles) that are involved in particle production. (b) Nudiviral gene expression as well as amplification and excision of BV circles occurs in the calyx cells of the wasp ovaries. Direct repeat junctions (DRJs; red triangles) are involved in circularization. (c) DNA circles are packaged into BV particles. (d) BV particles are injected in the lepidopteran host during oviposition of parasitoid eggs and infect many lepidopteran cell types but do not replicate. (e) BV virulence gene expression leads to modifications in lepidopteran host physiology, such as inhibition of wasp egg encapsulation, allowing wasp development. (f) Emergence of adults carrying bracovirus genomes from wasp pupae. This figure is based on the life cycle of CcBV associated with *C. congregata* parasitoid wasp of *M. sexta*. (Photographs A. Bezier and A. Wild.)
### Table 1. Genes of nudiviral origin identified in braconid wasps.

<table>
<thead>
<tr>
<th>protein function</th>
<th>gene name</th>
<th>presence in bravoviruses</th>
<th>ORF number in nudiviruses and baculoviruses</th>
<th>variation in selection (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>nudivirus/baculovirus core genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>replication</td>
<td>helicase</td>
<td>+ n.d n.d.</td>
<td>104 88 34 95</td>
<td>0.45</td>
</tr>
<tr>
<td>transcription</td>
<td>RNA polymerase p47</td>
<td>+ + n.d.</td>
<td>75 69 20 40</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>RNA polymerase lef-8</td>
<td>+ + n.d.</td>
<td>90 49 64 50</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>RNA polymerase lef-4</td>
<td>+ n.d n.d.</td>
<td>90 96 42 90</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>RNA polymerase lef-9</td>
<td>+ n.d n.d.</td>
<td>75 24 96 62</td>
<td>0.07</td>
</tr>
<tr>
<td>initiation factor lef-5</td>
<td>+ + n.d.</td>
<td>101 85 52 99</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>packaging and assembly</td>
<td>very late factor vlf-1</td>
<td>+ n.d +</td>
<td>121 80 30 77</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>capsid protein vp91</td>
<td>+ n.d. +</td>
<td>46 2 106 83</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>capsid protein vp39c</td>
<td>+ + +</td>
<td>89 64 15 89</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>viral phosphatase 38K</td>
<td>+ + +</td>
<td>10 1 87 98</td>
<td>0.17</td>
</tr>
<tr>
<td>ODV envelope component (per os infectivity factor)</td>
<td>pif-0 (p74)</td>
<td>+ + +</td>
<td>11 45 126 138</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>pif-1</td>
<td>+ n.d. +</td>
<td>55 52 60 119</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>pif-2</td>
<td>+ n.d. +</td>
<td>123 66 17 22</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>pif-3</td>
<td>+ n.d. +</td>
<td>88 3 107 115</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>pif-4 (19 kDa)</td>
<td>+ + +</td>
<td>103 87 33 96</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>pif-5 (adv-e56)</td>
<td>+ + +</td>
<td>76 5 115 148</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>pif-6 (ac68)</td>
<td>+ + n.d.</td>
<td>74 55 72 68</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>nudivirus/baculovirus genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>envelope</td>
<td>odv-e66</td>
<td>+ + +</td>
<td>12 46</td>
<td>0.10</td>
</tr>
<tr>
<td><strong>nudivirus-specific genes</strong></td>
<td>DNA processing</td>
<td>integrate</td>
<td>+ n.d. +</td>
<td>144 57 75 –</td>
</tr>
<tr>
<td></td>
<td>flap endonucleasec</td>
<td>+ n.d. n.d.</td>
<td>68 65 16 –</td>
<td>–</td>
</tr>
<tr>
<td>particle component</td>
<td>HzNVorf9-likeabc</td>
<td>+ + +</td>
<td>9 –</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>HzNVorf10-like</td>
<td>+ + +</td>
<td>106 –</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>HzNVorf11-likeabc</td>
<td>+ + +</td>
<td>118 –</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>HzNVorf124-like</td>
<td>+ + +</td>
<td>124 95 41 –</td>
<td>–</td>
</tr>
<tr>
<td>unknown</td>
<td>HzNVorf54-like</td>
<td>+ + n.d.</td>
<td>64 –</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>HzNVorf94-like</td>
<td>+ + nd</td>
<td>94 –</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>HzNVorf128-like</td>
<td>+ + +</td>
<td>128 –</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>HzNVorf140-like</td>
<td>+ + +</td>
<td>140 –</td>
<td>–</td>
</tr>
<tr>
<td>bracovirus-specific genes</td>
<td>particle component</td>
<td>17a</td>
<td>+ + +</td>
<td>– – –</td>
</tr>
<tr>
<td></td>
<td>30ab</td>
<td>+ + +</td>
<td>– – –</td>
<td>– n.a.</td>
</tr>
<tr>
<td></td>
<td>35a</td>
<td>+ + +</td>
<td>– – –</td>
<td>– n.a.</td>
</tr>
<tr>
<td></td>
<td>97a</td>
<td>+ n.d. +</td>
<td>– – –</td>
<td>– n.a.</td>
</tr>
<tr>
<td></td>
<td>97b</td>
<td>+ n.d. +</td>
<td>– – –</td>
<td>– n.a.</td>
</tr>
<tr>
<td></td>
<td>27b</td>
<td>+ + +</td>
<td>– – –</td>
<td>– n.a.</td>
</tr>
<tr>
<td>unknown</td>
<td>Cs30C22.5</td>
<td>+ + n.d.</td>
<td>– – –</td>
<td>– n.a.</td>
</tr>
<tr>
<td></td>
<td>Cs30C22.6</td>
<td>+ + n.d.</td>
<td>– – –</td>
<td>– n.a.</td>
</tr>
</tbody>
</table>

*aProtein function assigned based on gene homology.

bBranch-site REL method was used to test whether the BV lineage was subjected to episodic diversifying selection compared with free-living viruses (significant p < 0.01).

cGene belonging to the nudiviral cluster identified within C. congregata wasp genome.

*dProtein sequence similarity identified using probabilistic methods (HMMER program; http://hmmer.janelia.org/search/phmmer) and confirmed by orthologous sequence alignment.

*eProtein identified as particle component in CcBV and/or GbBV.

fHzNVorf118-like is also referred to as PmV within bracovirus genomes and HzNVorf64-like as p51 within Heliothis zea nudivirus-1 genome.

*gProtein identified as virion structural protein in HzNV-2 (11k).
3. Evolution of bracovirus genomes since integration

(a) Bracovirus genome evolution since integration within wasp genomes

Recent advances in wasp genomics combined with our knowledge of phylogenetics shed partial light on BV genome evolution. Transcriptomic analyses from the ovaries of the ichneumonid wasp *Hylespis plinius* allowed the identification of the genes involved in BV replication [11,48]. These genes are expressed within ichneumonid wasp ovaries and are not related to nudiviruses or to any other known viruses except for the *p12* and *p3* viral genes, which had previously been identified as structural proteins of Campodeitis sonorensis IV (CsiV) [49]. Similarly, as in BV, the genes encoding IV particle components are not present introns and are organized within gene-rich regions of the wasp chromosomes. These regions are referred to as ichnovirus structural protein encoding regions (IVSPERs). In *H. didymatirus*, three IVSPERs sharing related genes belonging to seven gene families were identified. They are thought to correspond to the remnants of a duplicated virus genome [11,48]. IVSPER genes are conserved among phylogenetically distant IV-associated parasitoid wasps such as *Transeoma rostrate* [11] and *Campodeis sonorensis* [48]. Altogether, the data suggest IVSPERs derive from a common viral ancestor. However, the lack of similarity between IVSPER genes and any known pathogenic viral genes implies the IV ancestor belonged to a virus family for which no present-day members are described [11].

(b) Endogenous viral element domestication

Convergence

At present, there is no doubt that IV and BV particles originated from distinct viruses [10,11]. Virus domestication occurred at least twice as two parasitoid wasp lineages independently integrated viruses into their genomes. In a remarkable example of convergent evolution, this resulted in the wasps delivering into their hosts pathogenic genes contained in viruses. Why have several mutualist associations with large DNA viruses been described in parasitoid wasps and not elsewhere in the tree of life? Altogether, there are a number of arguments to explain why parasitoid wasps might have had enhanced probability of association with viruses resulting in long-term domestication. Parasitoid wasps use diverse strategies for controlling host physiology. They generally consist of the injection of virulence proteins produced in the venom gland or the ovaries [50]. The association with viruses might allow the production of a larger set of virulence factors at a lower physiological cost for the wasp. This evolutionary benefit may explain why PDVs have been selected repeatedly. Another explanation might be linked with the life-history traits of the wasps themselves: endoparasitoid wasps insert their ovipositors in several individuals of a lepidopteran population either to probe the host quality before oviposition or to feed on the host. This behaviour may have favoured encounters with viruses, which can themselves use the wasps as vectors for horizontal transmission, as is the case for ascoviruses [51,52]. It should be noted that pathogenic virus infection induced during oviposition may benefit the development of parasitoid larvae by inhibiting host defences [53]. Moreover, nudiviruses might be particularly prone to domestication because they are able to infect gonads [54] and/or to integrate their genomes in host cells [55].

(c) Ichnoviruses originate from an uncharacterized virus family

Transcriptomic analyses from the ovaries of the ichneumonid wasp *Hypotoma dinutum* allowed the identification of the genes involved in IV particle production [11,48]. The genes expressed within ichneumonid wasp ovaries were not related to nudiviruses or to any other known viruses except for the *p12* and *p3* viral genes, which had previously been identified as structural proteins of Campodeitis sonorensis IV (CsiV) [49]. Similarly, as in BV, the genes encoding IV particle components do not possess introns and are organized within gene-rich regions of the wasp chromosomes. These regions are referred to as ichnovirus structural protein encoding regions (IVSPERs). In *H. didymatirus*, three IVSPERs sharing related genes belonging to seven gene families were identified. They are thought to correspond to the remnants of a duplicated virus genome [11,48]. IVSPER genes are conserved among phylogenetically distant IV-associated parasitoid wasps such as *Transeoma rostrate* [11] and *Campodeis sonorensis* [48]. Altogether, the data suggest IVSPERs derive from a common viral ancestor. However, the lack of similarity between IVSPER genes and any known pathogenic viral genes implies the IV ancestor belonged to a virus family for which no present-day members are described [11].
evolution. We can focus on two periods in the history of the microgastroid complex [15]. The first period concerns the events around the original viral genome integration over 100 Ma [7], and the second around 17 Ma relates to provirus divergence in Cotesia and Glyptapanteles (figures 1 and 3).

Bracoviruses evolved from integrated nudivirus genomes by several dramatic genomic transformations (figure 3). After integration of the ancestral nudivirus genome (figure 3a) into the wasp genome (figure 3b), the duplication of sequences, termed direct repeat junctions (DRJs; allowing virus genome excision and packaging), could have resulted in the packaging of wasp DNA instead of nudivirus genes in the particles (figure 3c,d). DRJs are specific DNA circularization sequence motifs flanking each segment and common to all bracoviruses [15]. This sequence motif could derive from the sequence of the ancestral nudivirus allowing the encapsidation of single genomes from concatemers produced during DNA replication [56]. Moreover, this period probably saw the first virulence gene translocations from the wasp genome towards proviral segments, resulting in the increase of viral particle capacity to counteract the immune system of the parasitized host (figure 3c). The acquisition of wasp genes by proviral segments and their subsequent diversification have produced a new entity that can be considered as an extended genome conferring to the wasp the ability to develop in their lepidopteran hosts [57], a notion akin to the extended phenotype [58]. It probably provided the braconid wasps with improved fitness, which favoured the diversification of both the microgastroid wasp complex and their symbiotic bracoviruses. Since genome integration 100 Ma, many events have occurred leading to diverse proviral genome organizations in different wasp lineages. However, comparative genomics on the Glyptapanteles and Cotesia genera revealed a surprising conservation in the localization of most proviral segments, which have remained stably integrated in braconid wasp genomes [15] since the separation of the wasps 17 Ma [7]. Proviral sequences were also subjected to specific gene transfers [37] and duplications [15] leading to unique PDV genome organizations in each wasp species.

(b) Gene transfers into bracoviruses
(i) Ancient wasp gene transfers into the provirus
Among the genes transferred into BVs, two families of virus gene transfers are particularly remarkable, the ankiryn and the ptp genes [59–62]. Apart from CiBV (Cheloninae), these genes are found in all BVs studied, including Toxoneuron nigriceps BV (Cardiochilinae) and MdBV, Glyptapanteles BV and Cotesia BV (Microgastrinae). Therefore, the insertion of the first ank and ptp genes must have happened before the separation of these different lineages 86 Ma (figure 1).

(ii) Recent gene transfers into the provirus
Although many PDV genes are similar to cellular genes, it has so far been difficult to formally demonstrate that they derived from gene transfer events between insect (wasp or lepidopteran) genomes and PDVs. High PDV gene divergence compared with insect homologues generally leads to loss of phylogenetic signal [63]. This divergence could be explained by the time (100 Myr) elapsed since integration in BV genomes but could also reflect high selection pressures imposed by the interaction with the lepidopteran host [60,63,64] (see §4). However, phylogenetic analyses of the sugar transporter gene family recently acquired by Glyptapanteles BVs showed the BV genes were more closely related to hymenopteran than lepidopteran genes [37]. To date, this represents the only robust example of gene transfer from wasps to PDVs. In addition, few packaged genes, such as a baculovirus pdh4 and an ascovirus gene, have most probably been acquired from other viruses by lateral transfer [41].

(c) Mechanisms involved in bracovirus gene evolution
Since integration of the nudivirus genome, different mechanisms, including duplications, gene mutations followed by selection and insertion of transposable elements (TEs), have been involved in shaping bracovirus genomes and driving the gene content diversity of contemporary BVs.

(i) Gene expansion is a hallmark of polydnaviridae genomes
A striking feature of PDV genomes is that over half of their genes belong to multigenic families. For example, the CcBV
Figure 4. Bracovirus sequences integrated into insect genomic DNA and encapsidated circles. (a) BV segments integrated in wasp genomic DNA. Within the C. sesamiae wasp genome (in yellow), (i) classical proviral BV segments containing virulence genes (coloured rectangles) delimited by DRJ sequences (red triangles) and (ii) a reintegrated segment can be identified. The reintegrated segment is not delimited by DRJ, but is bordered by left junction (LJ) and right junction (RJ) sequences. (b) BV segments reintegrated in lepidopteran host. Within lepidopteran genomes (in light blue), reintegrated BV segments bordered by LJ and RJ sequences can also be identified. (c) Encapsidated BV circles injected in lepidopteran host. Sequence comparison between circular and reintegrated viral sequences (in wasp or lepidopteran genomes) show that circle reintegration in both cases involves loss of a stretch of viral sequence (indicated by Δ), and is mediated by similar reintegration boundaries (LJ, RJ), suggesting that BV use a specific but unknown mechanism to reintegrate into genomic DNA. See table 2 for sequences involved in BV circle circularization and reintegration.

A re-emerging theme in the study of PDV evolution is the ability of viral circles to secondarily integrate into insect (lepidopteran or wasp) host genomic DNA [70]. Despite not replicating in lepidopteran host cells, chromosomal integration has been demonstrated both in vitro, in host and non-host derived cell cultures, and in vivo, in parasitized lepidopteran hosts for some PDV circles [70,71]. Viral circle reintegration (of ptp containing circles) in two different genomic locations of Cotesia sesamiae wasps have also been described [64]. Bracovirus DNA integration into lepidopteran host or secondary reintegration in wasp genomes might involve specific unknown mechanisms. Sequence comparison of circular and reintegrated viral sequences suggested that circle reintegration did not involve DRJs, but was mediated by specific sites in the circles, named left and right junction (figures 3 and 4) [64,70]. Furthermore, reintegration also resulted in the deletion of a 40–53 bp viral sequence (figures 3 and 4 and table 2) [64,70]. Similar viral reintegration motifs could be identified in 12 segments [70].
Functionally, PDV genome integration into lepidopteran host DNA could be important for wasp parasitism success, if the integration process was required to maintain PDV expression in the late stages of parasitism [74,75], or during prolonged interactions of wasps with their hosts (during diapause [76]).

The fact that BV circles can reintegrate into the genomes of both wasp and lepidopteran hosts suggests BVs could mediate horizontal gene transfer between these insects. Initial experimental data indicate reintegrations do happen during natural interactions and could be more frequent than initially expected [64–70]. For PDV reintegration into wasp DNA, PDV circles would need to enter the wasp germline and to stably integrate into the wasp genome during the phase when parasitoid wasp eggs and larvae are exposed to virus circles in the haemolymph of the lepidopteran host. Concerning integration and transmission of PDV sequences in Lepidoptera, this implies wasp oviposition in the late stages of parasitism [74,75], or during prolonged integration process was required to maintain PDV expression [81]. Integrated approaches combining wasp ecological and life-history traits, phylogenetic analyses, comparative BV genomic data and controlled parasitism experiments should help determine whether BV gene sequences are genuinely linked to given host ranges.

We shall discuss the mechanisms that could be implicated in shaping BV genomic features involved in determining wasp host range and describe an ecological and phylogenomic framework where this work has been initiated.

(a) Mechanisms involved in shaping bracovirus genomes that could be involved in wasp adaptation

(i) Bracovirus gene content links with wasp life-history traits

Each packaged BV genome has a unique gene content, with most genes organized into gene families [15,82]. Among the 37 gene families so far identified, some are present in practically all BVs (i.e. ank, ptp), whereas others are specific to particular wasp lineages [5]. Specific genes or gene families could reflect how physiological interactions with different hosts have modelled these genomes. For example, despite originating from the same ancestral nudivirus, no common virulence genes have been identified to date between CiBV originating from the same ancestral nudivirus, no common virulence genes have been identified to date between CiBV and other BV genomes [83]. Like all Cheloninae, C. inanitus oviposits into the eggs of lepidopteran hosts, whereas the other BV-associated wasps inject their eggs inside immune competent larvae (figure 1) [14]. Parasitism strategies involving the oviposition into host eggs versus larvae, exposing the wasp larvae to different physiological contexts, could explain the different panel of virulence proteins found in chelonines [83,84] compared with other wasps. Other differences in life-history strategies of the wasps could impact BV gene sets, for example, solitary versus gregarious development [5].

(ii) Bracovirus gene evolution: duplications and diversification

The remarkable expansion of many PDV genes into multigene families is a feature that could also be associated with wasp adaptive capacities. Gene duplications and their subsequent divergence play an important role in the evolution of novel

### Table 2. Bracovirus circle motifs involved in genomic DNA integration. LJ, left junction, RJ, right junction.

<table>
<thead>
<tr>
<th>bracovirus</th>
<th>wasp host</th>
<th>DRJ</th>
<th>no. of segments$^a$</th>
<th>host DNA</th>
<th>LJ</th>
<th>RJ</th>
<th>deletion (bp)</th>
<th>no. of segments$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GiBV</td>
<td>G. indiensis</td>
<td>AGCTT</td>
<td>24 [72]</td>
<td>Lepidopteran</td>
<td>CATGGT</td>
<td>n.d.</td>
<td>n.d.</td>
<td>1 [71]</td>
</tr>
<tr>
<td>MdBV</td>
<td>M. demolitor</td>
<td>AGCTT</td>
<td>13 [72]</td>
<td>Lepidopteran</td>
<td>ACCA</td>
<td>TAGT</td>
<td>50–51</td>
<td>12 [70]</td>
</tr>
<tr>
<td>CsBV</td>
<td>C. sesamiae</td>
<td>AGCTT</td>
<td>16 [73]</td>
<td>Hymenopteran</td>
<td>ACCA</td>
<td>TGGT</td>
<td>40–53</td>
<td>3 [64]</td>
</tr>
</tbody>
</table>

$^a$Number of segments on which the studies are based.
gene functions [85,86]. In host–parasitoid interactions where endoparasitoid wasps develop within caterpillars and have to face the arsenal of a functional immune system, BV virulence genes are expected to be under strong evolutionary constraints. Selection pressures are imposed on BV genes to overcome new host resistances either in the process of a coevolutionary arms race with the host species in which the wasp is established, or in the process of host switch or host range expansion. Duplications and subsequent divergence of PDV genes could therefore be involved in wasp adaptation to new or evolving hosts.

In BVs, molecular evolutionary approaches gave insights into the evolutionary processes involved in the expansion and diversification of particular BV multigene families [64,73,87]. To assess whether BV genes are coevolving with their lepidopteran hosts, rates of non-synonymous versus synonymous substitutions were initially measured at the interspecific level. Higher rates of amino acid changes, leading to innovation of protein function [88,89], are expected in BV genes involved in wasp adaptation to hosts. High positive selection was identified in cystatins, which are inhibitors of C1 cysteine proteases, and the positively selected residues were located in the vicinity of active sites assumed to directly interact with host proteases [87]. In the case of different ptp genes, the largest BV gene family, some gene copies evolved under relaxed selection pressure, whereas others underwent positive selection episodically [64,73]. For example, we detected selected amino acids within PTPE and PTPX in regions predicted to be involved in PTP substrate specificity suggesting recent protein target shifts in the host [64]. The evolution of the ptp gene family shows (i) evidence for classical gene duplication models assuming fixation of the duplicated copy is a neutral process, as well as (ii) evidence for alternative models proposing the gene duplication process is itself under positive natural selection [86]. Evidence for gene loss implying a ‘birth and death’ model (iii) could also be observed [90]. According to the ‘birth and death’ model, genes arise continuously by duplication and are lost by deletion or by mutational events. An ongoing process of pseudogenization was also observed for copies corresponding to different ptp genes in different species [60]. As shown in primates, Drosophila or large DNA viruses such as poxviruses, gene expansion and contraction could explain important adaptive traits allowing physiological adaptations of their host species [91–93]. The expansion of the ptp gene family could therefore be an important source of evolutionary innovations conferring new adaptive traits to the wasps. In accordance, different ptp expression patterns and different functions have been described in the context of host–parasite interactions [16].

The challenge is now to link mutations, diversifying selection and gene gain and loss observed in BV genomes with potential wasp host shifts, or wasp counter-adaptations to resistant hosts. In order to do this, it is now important to be able to perform similar studies on wasps for which reliable ecological data is available. Only a complete dataset, combining wasp phylogeny, wasp ecology (i.e. host range) and selection analyses, could enable us to understand the consequences of BV gene content and evolution on wasp specialization.

(b) Case study: Cotesia sesamiae
To understand the potential adaptive role of bracoviruses, one needs to conduct detailed population studies of the wasps in an ecological context that includes accurate knowledge of their host range. However, there are currently few models linking BV genes to the ecological adaptation of their carrier wasps. One of the best examples, the wasp C. sesamiae, parasitizes over 20 species of African stem borers belonging to the Noctuidae and Crambidae families [94]. As its hosts include major cereal pests, such as Bussola fusca and Sesamia calaminis, detailed host preference studies have been done for biological control improvement. Bussola fusca was found to be either susceptible or resistant to parasitization by different C. sesamiae populations [95]. These wasp populations were found to carry different alleles of the CrV1 baculovirus gene. These alleles, labelled virulent and avirulent, are involved in the success or failure to parasitize B. fusca [96,97]. Only the wasps of overlapping geographical range with B. fusca were found to carry the virulent CrV1 BV allele, suggesting wasp adaptation is linked with this BV gene [96]. Comprehensive population studies further showed that host range was the main factor explaining C. sesamiae population structure based on eight CrV1 alleles [94]. This suggested that, even though most wasp populations were generalists parasitizing several host species, cryptic specialization could occur [94]. However, based on partial genomic data, C. sesamiae BVs encode over 130 genes, mostly involved in physiological and immune interactions [73]. Genome-wide molecular evolution studies revealed at least 17 genes, including CrV1, histone H4, ep1, ep2 and lectin, were under positive selection within the Cotesia genus and likely implicated in wasp ecological adaptation [73]. A comparative study of three positively selected genes, namely CrV1, histone-H4 and ep2, in 21 wild C. sesamiae populations associated with different host species, identified several haplotypes for the three BV genes. Different allelic combinations of these genes, implicated in several immune pathways, were found in different populations. In addition, the signature of positive selection was detected in the three genes, but in branches leading to different wasp groups. This suggested the three genes carry different adaptive potential, depending on local host–wasp interactions [73]. As they can parasitize many lepidopteran species, C. sesamiae wasps are involved in a complex landscape comprising multiple coevolutionary interactions, which can take them through different directions to different adaptive peaks. Cotesia sesamiae relies on several BV mediated molecular pathways to overcome multiple host resistance. There is a need for further integrative studies to detail the role BVs play in the ecological adaptation of the wasps.

(c) Evolutionary model for polydnavirus genome extension
In the context of parasitoid evolution, the multiple EVE domestications appear to have been particularly beneficial to the diversification of the wasps, based on species richness [12,13,96,99]. However, there might be a cost to the wasp for the replication of PDVs. Some wasps do not use PDVs at all, and the case of V. canescens [25,100] suggests reversal from PDV use could be possible. In this light, why are PDV genomes so large? Serial knockout experiments could help determine whether all PDV genes are absolutely required to foster wasp development. However, working in non-model systems has so far been problematic for this kind of approach. Furthermore, they would be unlikely to
Figure 5. Polydnavirus mediated host parasite coevolution. (a) PDV adaptation to a single host species harbouring either a single locus susceptibility (SLS), a single locus resistance (SLR) or multilocus resistance (MLR); (b) PDV adaptation to two hosts species harbouring either a multilocus susceptibility (MLS), or multilocus resistance (MLR). Pie and semicircle shapes indicate PDV genes; squares and circles host resistance genes. Corresponding shapes and colours indicate an efficient PDV effector targeting of host factor allowing the wasp larvae to develop and the transmission of the PDV gene in their chromosomes. Lightning shapes depict mutations and virulence or resistance gene acquisitions.

reflect the complete picture of the conditions a wasp could encounter in the wild. The maintenance of large PDV genomes, at least in the case of BVs, might be explained by the way viral DNA circles are packaged in the virus particles. There is no real constraint on the size of the DNA circles encapsidated in the virus particles, and the segmented nature of the genome allows the presence of many genes. If the main cost encountered by the wasp is on the production of the viral particles, the encapsidation of many versus few virulence genes might not be as costly as it seems at first sight. Most of the physiological cost is endorsed by the parasitized host, which express the viral genes using its own cellular machinery.

We would like to propose a model, based on the gene for gene coevolution model [101], to explain how ecological constraints could foster PDV genome extension (figure 5). The model assumes that the only issue to a parasitism challenge is the death of one of the partners, either the wasp or the caterpillar, as is the case in host–parasitoid interactions. In the gene for gene model, a single host mutation can counteract the virulence gene, but is in turn overcome by a single parasite gene mutation (SLR in figure 5a). Taking on a new locus or even a new resistance metabolic pathway to overcome virulence gives an immediate advantage to the host (MLR in figure 5a). In the case of PDVs, we can postulate the transfer of new virulence genes or duplication and diversification of existing virulence genes in PDV circles could overturn this kind of multilocus resistance (figure 5a). If the cost of encapsidation is low, several virulence genes could be assayed at once, but only the gene conferring a genuine selective advantage should be retained through time; non-adaptive genes could be lost through pseudogenization [15]. So far, we have placed this arms race in the context where the wasps parasitize a single host species. However, we know from C. sesaniae that the situation is far more complex in the field where the wasps can develop on several hosts within a locality [94]. We therefore extended our model to two hosts, both harbouring multilocus resistance (figure 5b). In this situation, the PDV has to produce twice as many virulence genes to overcome distinct resistances. Even if both hosts can rely on the same pathways (and even on orthologous genes) to overcome parasitism (MLS in figure 5b), point mutations in each host gene could be sufficient to confer resistance to a particular virulence protein (MLR in figure 5b). The acquisition by the PDV of new virulence genes targeting each of these resistance mutations would be necessary to overturn resistance in both host species (MLR in figure 5b). The model therefore proposes an explanation for the large gene family extension observed in PDV genomes in the ecological adaptive landscape in which parasitoid wasps evolve. Only detailed population genomics studies, in multiple host species, could validate these hypotheses for global genome extension.
5. Conclusion

Nudiviral genes in braconid wasp genomes can be considered as EVEs because they originated from a nudivirus integrated into the genome of wasp ancestor species [10]. The same holds true for ichneumonid IVSPER genes, even if the corresponding pathogenic virus has yet to be characterized [11]. Contrary to most EVEs studied [4], bracoviruses are not fossil genomic remnants but active viruses. Indeed, most viral functions described for pathogenic viruses have been conserved and BVs undergo all the steps of a classical virus cycle. The main difference between PDVs and conventional viruses resides in the separation of the virus cycle between two cell types and hosts: calyx cells in the wasp ovaries produce infectious particles, and parasitized lepidopteran host cells are infected. This conforms to the definition of viruses as infectious agents with nucleic acid genomes, which replicate inside living host cells to produce particles transferring their genome to other cells [102].

In viruses, however, the genome packaged in the particles is usually supposed to contain all the information required for replication, which is not the case for PDVs. As obligatory symbionts, PDVs contain a genome involved in wasp adaptation to its host, and the information required for particle production now resides permanently in the wasp genome. PDVs are not the only viruses transferring genetic material other than their own. In prokaryotes, several phases of independent origins have been identified in a wide range of species. They were called gene transfer agents (GTAs) [103]. GTAs mediate genetic exchanges between bacteria of the same species in particular environmental conditions by transmitting random bacterial DNA instead of their own. To account for this functional diversity in the viral world, Stoltz & Krell [104] recently proposed establishing three categories: classical viruses, GTAs and endogenous viruses.

Genomic advances in all branches of the tree of life should reveal the diversity of viruses and EVEs associated with cellular organisms. This would give clues as to whether symbiotic viruses are a particular feature of parasitoid wasps or are fairly common, such as GTA in prokaryotes, and have been ignored in less studied taxa. Within parasitoid wasps, deep sequencing of closely related wasps with contrasting life-histories and host ranges would give new insights as to whether particular PDV genes or gene sets are required for wasp adaptation. The PDV genome extension model predicts generalist wasps should harbour more PDV genes than specialists.

Acknowledgements. We thank G. Rohrmann for discussion on gene homology and P. Gayral and F. Dedene for discussion.

Funding statement. This work was supported by the ERC project no. 205206 ‘GENOVIR’ and the ANR project ‘Paratoxose’.

References


64. Serbieille C, Dupuy S, Perdereau E, Héricourt F, Dupuy C, Huguet E, Drezen J-M. 2012 Evolutionary mechanisms driving the evolution of a large polydnavirus gene family coding for protein tyrosine


