# Rep virulence genes have been recruited convergently by nudivirus-derived Virus-Like Particles and ichnoviruses of Campoplegine wasps

Alexandra Cerqueira de Araujo[a], Matthieu Leobold[b], Marie Cariou[c], Bernardo Ferreira dos Santos[d], Rustem Uzbekov[ef], Renato Ricciardi[g], Pierluigi Scaramozzino[g], Andrea Lucchi[g], Karine Musset[b], Jean-Michel Drezen[b], Thibaut Josse[b], Elisabeth Huguet[b]

[a]CBER, Department of Genetics, Evolution and Environment, University College London, London, UK, [b]Institut de Recherche sur la Biologie de l'Insecte (IRBI), UMR 7261, CNRS - Université de Tours, Tours, France, [c]Service Analyse de Données, Muséum National d'Histoire Naturelle, Centre National de la Recherche Scientifique, Paris, France, [d]Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Center for Integrative Biodiversity Discovery, Invalidenstraße 43, Berlin, 10115, Germany, [e]Université de Tours, Département des Microscopies, Tours, France, [f]Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia , [g]Department of Agriculture, Food and Environment (DAFE), University of Pisa, Italy.

Corresponding author: alexandra.cdea@gmail.com

17

18 ## Abstract

19 Viral endogenization events in the genomes of parasitoid wasps have led to the production
20 of immunosuppressive particles that are injected into the host and that either contain DNA
21 encoding virulence effectors or virulence proteins, essential for parasitism success. In some
22 Campopleginae (Ichneumonidae) wasps, viral endogenisation involves the production of
23 large dsDNA ichnoviruses encoding several dozens of rep virulence genes, to the exception
24 of *Venturia canescens*. Instead, *V. canescens* harbours an endogenous nudivirus producing
25 Virus-Like-Particles (VLPs) in which virulence proteins are wrapped into viral envelopes.
26 Here, we described that this nudiviral domestication event is more ancient and widespread
27 than previously thought as we identified homologous alphanudivirus sequences in the
28 genomes of three Campopleginae species belonging to the genera *Campoplex* and
29 *Rimphoctona*. We showed that *Campoplex capitator* produces morphologically similar VLPs
30 to those of *V. canescens* with almost the same viral protein content. Strikingly, the virulence
31 proteins are however different, suggesting that VLPs allow a high versatility in the molecular
32 strategies used by the wasps in response to host selection pressures. Furthermore, a marked
33 expansion of serpin and *rep* genes is present in *Campoplex capitator*, with one serpin and
34 several REP proteins identified as major VLP components using proteomics. REP proteins are
35 also well-described virulence factors initially shown to be delivered into hosts via the
36 expression of *rep* genes coded by ichnovirus DNA. In *Campoplex capitator*, these proteins
37 are packaged into nudivirus-derived VLPs indicating that convergent use of virulence factors
38 can imply endogenous viruses of different origin and with different delivery mechanisms.

Keywords: Nudivirus, Viral domestication, Campopleginae, Virus-Like-Particle, Parasitoid wasp, Ichnovirus

## Introduction

The endogenization of viral sequences within eukaryotic genomes is a common phenomenon in the evolutionary history of eukaryotic organisms (Holmes 2011; Feschotte and Gilbert 2012). These viral sequences, known as endogenous viral elements (EVEs), involve all virus types likely to integrate into any eukaryotic genome (Bejarano et al. 1996; Delaroque et al. 1999; Crochu et al. 2004; Maori et al. 2007; Belyi et al. 2010; Katzourakis and Gifford 2010; Liu et al. 2010; Feschotte and Gilbert 2012; Ballinger et al. 2014; Irwin et al. 2021). While a large proportion of EVEs are counter-selected and no longer functional (Tarlinton et al. 2006; Arbuckle et al. 2010; Katzourakis and Gifford 2010), there are, however, examples of viral domestications providing novel adaptive functions and a selective advantage to the eukaryotic host, such as EVEs involved in placentation in mammals (Mangeney et al. 2007; Lavialle et al. 2013).

In general, this type of domestication involves individual viral genes rather than entire viral genomes. However, in some parasitoid wasp species, entire portions of large double-stranded DNA (dsDNA) viruses have been integrated and perform crucial functions in the wasp's life cycle. Indeed, the wasps use the virus-derived particles produced in their ovaries and injected at the same time as wasp eggs in the insect host, to deliver virulence genes or virulence proteins that suppress the host immune response and spark other physiological

60 changes to ensure parasitism success (Bézier, Herbinière, et al. 2009; Volkoff et al. 2010;
61 Strand and Burke 2012; Burke et al. 2013; Strand and Burke 2014; Drezen et al. 2017).

62 To date, several viral domestication events have been identified in independent wasp
63 lineages, each involving large dsDNA viruses (Burke 2019; Di Giovanni et al. 2020; Burke et
64 al. 2021; Drezen et al. 2022; Guinet et al. 2023). These endogenization events can be
65 classified into different categories based on the nature of the virus that has been
66 endogenized, the wasp families involved, and the type of particles produced and their
67 content (dsDNA circles containing virulence genes or virulence proteins).

68 Approximately 100 million years ago, the integration of sequences from a deltanudivirus
69 (Petersen et al. 2022) in the ancestor of Braconidae Microgastrinae wasps – comprising at
70 least 46000 species – led this lineage to produce Bracovirus (BV) particles, which are
71 composed of a viral envelope enclosing capsids containing dsDNA circular molecules
72 harbouring virulence genes (Murphy et al. 2008; Bézier, Annaheim, et al. 2009; Thézé et al.
73 2011; Herniou et al. 2013).

74 Two independent endogenization events leading to the production of Ichnovirus (IV)
75 particles also containing circular dsDNA have involved a viral ancestor from a still
76 uncharacterized virus family in two distinct Ichneumonidae lineages (Banchinae and
77 Campopleginae) (Volkoff et al. 2010; Gundersen-Rindal et al. 2013; Strand and Burke 2014;
78 Béliveau et al. 2015; Volkoff and Cusson 2020). The production of these particles involves
79 the expression of several intronless genes tightly clustered into wasp genomic regions called
80 IVSPERs (Ichnovirus Structural Protein Encoding Regions) (Volkoff et al. 2010; Volkoff et al.
81 2012; Lorenzi et al. 2019). Although Bracoviruses and Ichnoviruses do not have the same
82 viral origin they share organisational and functional features, showcasing a remarkable

83 example of convergent evolution. Indeed, in both cases virus particles are produced by a
84 conserved set of endogenized viral genes of known (BV) or unknown (IV) origin. The particles
85 deliver virulence genes in the parasitized host that are harboured by dsDNA circles,
86 originating from linear segments in the wasp genome, that are circularized during BV or IV
87 production thanks to conserved direct repeat junctions (Pasquier-Barre et al. 2002;
88 Annaheim and Lanzrein 2007; Desjardins et al. 2008; Beck et al. 2011; Bézier et al. 2013).
89 These BV and IV circles also have the capacity to integrate into the DNA of the parasitized
90 insect host genomes, thanks to Host Integration Motifs present in some circles (Beck et al.
91 2011; Chevignon et al. 2018; Muller et al. 2021; Heisserer et al. 2023; Kim et al. 2023). The
92 integrative property of these ds DNA BV and IV circles is most probably in large part at the
93 origin of HGT events observed in the genomes of Lepidopteran species that are the main
94 order targeted by these endoparasitoid wasps (Gasmi et al. 2015; Heisserer et al. 2023). The
95 only similar virulence genes between BV and IV, possibly of wasp origin and likely resulting
96 from horizontal transfer between these viruses, correspond to genes encoding ANK proteins
97 involved in host immune suppression (Cerqueira de Araujo, Josse, et al. 2022). The other
98 virulence genes, probably also of eucaryote origin, are not shared between BV and IV.
99 Campopleginae-associated IV all contain a multigene virulence family having undergone
100 particularly high expansion corresponding to repeat element genes (rep) (Volkoff et al. 2002;
101 Tanaka et al. 2007; Legeai et al. 2020; Kim et al. 2023). Although the function of rep genes
102 has not yet been elucidated, their conservation and high expansion in these IV genomes
103 together with their reported expression in caterpillar hosts suggest that they play an
104 important role in the parasitoid-host interaction (Galibert et al. 2006; Rasoolizadeh et al.
105 2009; Etebari et al. 2011; Kim et al. 2023).

106 Remarkably, nudiviruses have been involved in other independent domestications, with the
107 endogenization of alphanudiviruses in the braconid *Fopius arisanus* (Burke et al. 2018) and
108 more surprisingly in the Campopleginae *Venturia canescens* (Pichon et al. 2015). In these
109 parasitoid wasps, the nudivirus endogenization event took a different evolutionary trajectory
110 and has led to the production of Virus-Like-Particles (VLPs) that do not contain DNA but
111 virulence proteins instead (Pichon et al. 2015; Burke et al. 2018; Burke 2019). Because *V.*
112 *canescens* belongs to the subfamily Campopleginae, known to harbour ichnoviruses, the
113 VLPs were initially thought to derive from IVSPER expression (Reineke et al. 2006). However,
114 sequencing of the wasp's genome identified only a few pseudogenised IVSPER genes, and it
115 was formally demonstrated that the VLPs are in fact of nudiviral origin (Pichon et al. 2015).
116 Venturia canescens Endogenous Nudivirus (VcENV), allows the production VLPs, which
117 consist of a viral envelope containing virulence proteins that protect the wasp eggs from the
118 lepidopteran host immune system after oviposition (Salt 1965; Rotheram 1967; Rotheram
119 1973; Bedwin 1979; Feddersen et al. 1986; Reineke et al. 2006; Pichon et al. 2015). Three
120 main virulence proteins have been described in VcVLPs: *vlp1*, *vlp2*, and *vlp3* which encode
121 respectively a glutathione peroxidase (PHGPx), a Rho GTPase-activating protein (RhoGAP),
122 and a metalloprotease (neprilysin) (Theopold et al. 1994; Hellers et al. 1996; Asgari et al.
123 2002; Reineke et al. 2002; Pichon et al. 2015).

124 Several lines of evidence had suggested that the nudivirus integration event in *V. canescens*
125 is more recent than the one described in the ancient braconid wasp 100 million years ago
126 (Murphy et al. 2008; Thézé et al. 2011; Pichon et al. 2015). Indeed, to date, an
127 alphanudivirus endogenization had only been described in a single campoplegine wasp
128 species. Also, the nudivirus genes are less dispersed in the *V. canescens* genome than in

129 braconid wasps (Bézier, Annaheim, et al. 2009; Burke et al. 2014; Pichon et al. 2015; Mao et
130 al. 2023), suggesting fewer genomic rearrangements had time to occur. It has also been
131 shown that in VcENV, losses of certain core nudivirus functions (*i.e.*: proteins involved in
132 capsid formation) are due to pseudogenization involving accumulation of mutations rather
133 than complete loss of genes or genomic regions as described for obligate symbionts of
134 bacterial origin (Leobold et al. 2018). The fact that pseudogenised nudiviral sequences are
135 not completely eroded could also be an indication that the nudivirus integration is more
136 recent in this species (Leobold et al. 2018).

137 Here, we describe that this nudiviral domestication event is in fact more ancient than
138 previously thought as we identified homologous alphanudivirus genomes in 2 species of the
139 sister genus *Campoplex* by genome sequencing, and in *Rimphoctona* by data mining. We
140 showed that *Campoplex capitator*, the main parasitoid of the vineyard pest *Lobesia botrana*,
141 produces morphologically similar VLPs to those of *V. canescens* with almost the same viral
142 protein content. Strikingly, the virulence proteins are completely different, suggesting that
143 VLPs allow a high versatility in the molecular strategies used by the wasps in response to
144 host selection pressures. Furthermore, a marked expansion of serpin and rep genes is
145 present in *Campoplex capitator* genome, with one serpin and several rep proteins identified
146 as major VLP components in proteomic analyses. Rep proteins are important virulence
147 factors in wasps carrying ichnoviruses. They are produced from *rep* genes located on
148 ichnovirus DNA circles once these circles have been delivered to the host. In *Campoplex*
149 *capitator*, these virulence proteins are packaged into nudivirus-derived VLPs indicating that
150 convergent use of virulence factors can imply endogenous viruses of different origin and
151 with different delivery mechanisms.

152 # Results

153 ## Viral integration Campopleginae wasps

154 In Campopleginae wasps, most species (of genera such as *Hyposoter*, *Campoletis*, *Diadegma*)
155 have endogenized an ichnovirus and produce ichnovirus particles, while *Venturia canescens*
156 was, so far, the only campoplegine wasp harboring an endogenized nudivirus and producing
157 VLPs lacking capsids and DNA. To investigate whether closely related species possess
158 endogenized viruses and determine their viral origin, we sequenced the genomes of two
159 *Campoplex* species, belonging to a sister group of *V. canescens*: *Campoplex capitator* and
160 *Campoplex nolae* (see Figure 1). Genome assemblies were obtained using hybrid sequencing
161 approaches, yielding a 261 Mb assembly for *C. capitator* (630 contigs, N50 = 7.8 Mb, 96%
162 BUSCO completeness, generated by PacBio sequencing) and a 218 Mb assembly for *C. nolae*
163 (4,593 contigs, N50 = 1.8 Mb, 94% BUSCO completeness, generated by Illumina and
164 Nanopore sequencing), with 11,288 and 13,929 predicted gene models, respectively (see
165 table S1 for details). We have also extended our analyses to *Rhimphoctona megacephalus*,
166 the most divergent species within the "Campoplex-Venturia-Dusona" clade after *Dusona sp.*,
167 a species that does not harbour any confirmed endogenized virus (Burke et al. 2021) nor
168 functional *rep* genes (see Figure 1) .

169

170 ## Identification of endogenized nudivirus genomes within other Campopleginae wasp species :
171 in *Campoplex spp.* and *Rimphoctona megacephalus* genomes

172 We identified 47 genes of viral origin in both *Campoplex* genomes and 62 viral genes in *R.*
173 *megacephalus* genome (Table1 and Table S2). In *C. capitator*, these genes are almost

174 exclusively expressed in adult wasp ovaries (Figure 2B and Table S3). Out of the 32 nudivirus
175 core genes, 22 have been found in the *Campoplex* genomes and 21 in *R. megacephalus*
176 (Table 1). Among these genes, some encode key viral functions in baculoviruses such as viral
177 transcription (*p47* and *lef* genes), DNA amplification (*helicase*), envelope
178 composition/infectiosity (*Ac81, p33* and *pif* genes) and virion morphogenesis (*vlf-1*). These
179 genes are grouped into 4 clusters in *Campoplex spp.* and *R. megacephalus* genomes (Figure
180 2A for *C. capitator* and Figure S3A for *C. nolae*), but genes are more physically distant from
181 each other in the *R. megacephalus* genome (see Figure S3B). In *V. canescens*, remnants of
182 some nudivirus genes, notably those involved in DNA amplification and capsid formation,
183 have been found as pseudogenes (Leobold et al. 2018). Therefore, for each missing nudivirus
184 gene, we searched for remnant sequences in *Campoplex* and *Rhimphoctona* genomes.
185 Several pseudogenes have been identified, some located inside or around the virus clusters
186 (Figure 2A and Figure S3), others located outside (Table S2). Genes that are typically involved
187 in DNA amplification (*integrase, FEN-1* and *DNApol*) and capsid formation (*p6.9, 38K* and
188 *vp39*) in baculoviruses have been found pseudogenized in both *Campoplex* species (Table 1),
189 except for *38K* which couldn't be found in *C. nolae*, probably due to the higher
190 fragmentation level of this genome. The same observation can be made in *R. megacephalus*,
191 although, unlike in *Campoplex* and *Venturia* wasps, *p6.9* and several *integrases* have been
192 found.

193

194 ## The endogenized nudiviruses to Alphanudiviruses

195 Viral sequences found in *Campoplex* and *Rhimphoctona* genomes share high similarity with
196 nudivirus sequences (Table S2). We therefore aimed to place these new endogenized viruses

197 in the context of the phylogeny of Naldaviricetes, a class of virus comprising Baculoviridae,
198 Nudiviridae, Hytrosaviridae and Nimaviridae. Using protein sequences from 34 genes (21
199 baculoviruses core genes and 13 nudivirus core genes, see Table S4 for more details), the
200 phylogenetic analysis places these endogenized viruses inside the Nudivirus family and as a
201 sister clade to VcENV, which belongs to the Alphanudivirus genus (Figure 3A). Due to their
202 placement in the Nudivirus family, we called these genomes CcapiENV, CnolaENV and
203 RmegaENV for Campoplex capitator Endogenous Nudivirus, Campoplex nolae Endogenous
204 Nudivirus and Rhimphoctona megacephalus Endogenous Nudivirus respectively. CcapiENV
205 and CnolaENV are from now on the closest relatives of VcENV.

206

207 ## Evidence for a single nudivirus endogenization event in Campopleginae wasps

208 Given the close phylogenetic position of these integrated alphanudiviruses in *Campoplex* and
209 *Rhimphoctona* (Figure 3A), we then investigated whether they resulted from the same
210 integration event as the one for VcENV. The same wasp genes around the virus clusters were
211 found in *C. capitator, R. megacephalus* and *V. canescens* species, indicating that these
212 regions are homologous. This observation confirms that these viruses come from the same
213 endogenization event that most probably occurred at the same place in a common ancestor
214 of these wasps (Figure 3B).

215 ## Impact of endogenization on nudivirus gene organisation

216 Integrated alphanudivirus genomes were compared to free virus genomes in order to
217 evaluate the impact of viral domestication after integration. Gene order among free
218 nudiviruses is fairly well conserved (Figure 3C), but gene order was globally lost following

219 viral endogenization (Figure 3D), suggesting that loss of certain syntenic regions might be
220 part of the viral domestication process.

221 We also compared the endogenized alphanudiviruses in Campopleginae wasps to other
222 endogenized alphanudiviruses such as FaENV (Fopius arisanus Endogenous Nudivirus) and
223 NlENV (Nilaparvata lugens Endogenous Nudivirus). First, as described before, the viral
224 genome structure in Campopleginae wasps is conserved after integration (Figure 3E).
225 However, for each independent alphanudivirus integration, the genomes seem to have been
226 reshuffled differently (Figure 3E). Viral genome shuffling after integration might be part of a
227 viral domestication process, as developed in Burke and collaborators, or might reflect
228 genome rearrangements dynamic (Burke et al. 2018; Leobold et al. 2018; see discussion).

229

230 No conclusive traces of Ichnovirus remnants in *Campoplex* genomes

231 As remnants of Ichnovirus were found in *V. canescens* (Pichon et al. 2015), we also searched
232 for sequences of IVSPER genes in the *Campopex spp.* and *R. megacephalus* genomes. In
233 *Campoplex* and *R. megacephalus* genomes, no trace of IVPSER genes could be found in using
234 either the IVSPER protein sequences described in campoplegine wasps associated to
235 ichnoviruses or pseudogenized IVSPER genes described in *V. canescens.* From genes
236 potentially related to ichnoviruses, we only have detected *rep-like* genes (see Table S2),
237 which are the most abundant genes in terms of copy number in IVs packaged genomes of
238 campoplegine wasps. In total, while only one to two sequences have been found in *R.*
239 *megacephalus* and *V. canescens*, 48 and 16 *rep-like* sequences have been found in *C.*
240 *capitator* and *C. nolae* respectively. In *C. capitator*, transcripts of these sequences can be

241 found specifically in the wasp ovaries, with some sequences showing high levels of
242 transcription (Table S3). Given the propensity for horizontal transfer in insects (see Figure 4),
243 we determined that the presence of *rep-like* sequences alone does not constitute evidence
244 of ichnoviral integration.

245

246 *Campoplex capitator* produces VLPs that are similar in morphology to those produced by
247 *Venturia canescens*

248 After identifying the endogenized nudiviruses in the *Campoplex* genomes, we investigated
249 whether *C. capitator* produces VLPs as described in *V. canescens*. Electron microscopy
250 revealed the presence of particles in the calyx region of *C. capitator* ovaries, a specialised
251 portion of the ovaries that produces virus-derived particles in braconid and Ichneumonid
252 wasps (Bézier, Annaheim, et al. 2009; Volkoff et al. 2010; Burke and Strand 2012; Pichon et
253 al. 2015). These particles in *C. capitator* are referred to as CcapiVLPs (Figure 5), and are
254 almost morphologically identical to VcVLPs (Pichon et al. 2015). CcapiVLPs are composed of
255 a membrane enclosing an electron-dense body (Figure 5E) and are massively present in the
256 lumen (Figure 5A and B) while only a few are visible in the cytoplasm and cell nuclei (Figure
257 5C). Virogenic stromas, electron-dense masses composed of proteins that form viral particles
258 in other wasps (Bézier, Herbinière, et al. 2009; Pichon et al. 2015), are visible in calyx cell
259 nuclei (Figure 5C), along with VLPs and empty envelopes (Figure 5F). As described for VcVLPs
260 (Pichon et al. 2015), CcapiVLPs appear to be be produced inside the nucleus by virogenic
261 stromas and transit through the nuclear membrane and then through the cytoplasmic
262 membrane by budding to finally end in the calyx lumen (Figure 5D,E and F).

263

264 ## *C. capitator* and *V. canescens* VLPs have a similar composition

265 We proceeded to determine whether the protein composition of the VLPs and their content
266 were the same between *C. capitator* and *V. canescens*. For CcapiVLPs, 28 proteins of
267 nudivirus origin were detected by mass spectrometry (see Table 1, Figure 6A and Table S5).
268 In most cases, proteins contained in VcVLPs deriving from VcENV genes were also found in
269 CcapiVLPs. For instance, all the VcVLP proteins composing the envelopes in baculoviruses
270 were found in purified CcapiVLPs. These include all the PIF proteins, P74, VP91 and P33
271 proteins (Table 1, Figure 6A and Table S5). However, OrNvorf46-like and a copy of PIF-5
272 proteins are present in VcVLPs but not in CcapiVLPs (Table 1, Figure 6A and Table S5).
273 Similarly, OrNVorf18-like, OrNVorf47-like, OrNVorf79-like, OrNVorf90-like and OrNVorf120-
274 like proteins appear to be present in CcapiVLPs but were not detected in VcVLPs.

275 Analysis of selection pressures on *C. capitator*, *C. nolae*, *V. canescens* and *R. megacephalus*
276 genes revealed that while 89.5% of 8,828 genes have dN/dS ratios below 1 (see Figure 6B
277 and Table S6), nudiviral genes putatively involved in infectivity and envelope protein
278 production have in average a higher dN/dS ratio (0.58) than genes putatively involved in
279 DNA amplification and viral transcription (0.16). Gauthier and colleagues had also found
280 equivalent results with nudiviral genes associated with Braconid wasps (Gauthier et al. 2018;
281 Gauthier et al. 2021). Furthermore, duplicated nudiviral genes have on average a higher
282 dN/dS ratio (by 1 to 2 times) than single copy nudiviral genes (Table S6).

283

284 ## The virulence protein contents of VcVLPs and CcapiVLPs are different

285 A large number of non-viral proteins, which should include virulence proteins, could be
286 identified by mass spectrometry within the VLPs (Table S7). Proteins described in VcVLPs
287 were first sought in the CcapiVLP protein dataset to determine whether the same proteins
288 are enveloped in VLPs in both wasp species. The VLP1 (PHGPx), VLP2 (RhoGAP) and VLP3
289 (Neprilysin) proteins of VcVLPs were not found in CcapiVLPs, indicating that VLPs produced
290 by *C. capitator* have a different content than those produced by *V. canescens*. Candidate VLP
291 virulence proteins were selected according to their peptide abundance (PSM for Peptide-
292 Spectrum Matches) value and only proteins with a PSM value higher than 50 were retained
293 (Table S8). Surprisingly, proteins encoded by eight *rep-like* genes have been found in the
294 CcapiVLPs (Table S8). Additionally, some proteins that have been described in parasitoid
295 wasp venoms (Poirié et al. 2014; Moreau and Asgari 2015; Inwood et al. 2023), such as a
296 serpin and a peroxiredoxin, have been found in *C. capitator* particles. Further investigation
297 on the serpin showed that the gene encoding this protein is a member of a multigenic
298 family, which comprises 15 members in *Campoplex spp*. and four in *V. canescens*. While the
299 four copies found in *V. canescens* have orthologous sequences in *Campoplex spp.*, the serpin
300 identified in CcapiVLPs is more distant, with only one ortholog found in *C. nolae* (Figure S5).
301 This serpin appears to be specific to *Campoplex* species and therefore represents an
302 interesting virulence protein candidate.

303

304 # Discussion

305 ## A single, relatively recent alphanudivirus endogenization event in Campopleginae wasps

306 The sequencing of *C. capitator* and *C. nolae* genomes, together with analyses of *V. canescens*
307 (Pichon et al. 2015; Mao et al. 2023) and *R. megacephalus* (Boparai and Broad 2024)
308 genomes, has provided the opportunity to characterize the early evolutionary events
309 following alphanudivirus domestication in Campopleginae wasps. Our results demonstrate
310 that CcapiENV, CnolaENV, and RmegaENV derive from the same endogenization event as
311 VcENV, as evidenced by strong synteny of flanking wasp genes around viral clusters (Figure
312 3B). This integration event is older than previously thought, having occurred in the common
313 ancestor of these species. However, as Campopleginae is a younger lineage (Murphy et al.
314 2008; Burke et al. 2021; Sharanowski et al. 2021), this event is still more recent than the
315 nudivirus endogenization in Microgastrinae braconids 100 million years ago (Murphy et al.
316 2008; Thézé et al. 2011).

317 Several lines of evidence support the recency of this integration. First, not all Campopleginae
318 harbor endogenized nudiviruses; genera such as *Hyposoter* and *Dusona* either possess
319 ichnoviruses or lack confirmed viral endogenization entirely (Burke et al. 2021), indicating
320 that nudivirus integration occurred within a specific Campopleginae lineage rather than
321 ancestrally. Second, viral genes in most Campopleginae wasps are organized into compact,
322 highly conserved clusters (Figures 2A, S3, and 3B), contrasting with the more dispersed
323 organization observed in braconid BVs (Bézier et al. 2009; Burke et al. 2014; Burke et al.
324 2018; Gauthier et al. 2021), suggesting fewer genomic rearrangements have occurred. Third,
325 we identified numerous pseudogenized viral genes in *Campoplex* spp. and *V. canescens*

326 genomes (Tables 1 and S2), whereas BV-associated wasps show complete loss of non-
327 functional genes (Bézier et al. 2009; Burke et al. 2014). The persistence of these degraded
328 sequences indicates insufficient time for complete erosion, consistent with previous
329 observations in VcENV (Leobold et al. 2018).

330

331 ## Genomic rearrangement as an early step in viral domestication

332 While gene order is highly conserved among free alphanudiviruses (Figure 3C), this synteny
333 was largely lost following endogenization in wasps (Figure 3D). Within the same
334 endogenization event, the endogenized viruses maintain strong synteny with each other
335 despite lacking synteny with their free-living and endogenized relatives (Figure 3E),
336 suggesting that major genome reshuffling occurs early in the domestication process.

337 This early genome shuffling may serve multiple functions in domestication. By scattering
338 viral genes across the host genome, rearrangement could disrupt coordinated viral
339 replication, effectively neutralizing viral pathogenicity while retaining genes useful for the
340 wasp (Drezen et al. 2017; Burke et al. 2018; Leobold et al. 2018). However, the subsequent
341 stabilization of these rearranged structures, as evidenced by conserved synteny across
342 Campopleginae wasp evolution, suggests that the clustered organization may provide
343 functional advantages for coordinated expression during VLP production. Similar conclusions
344 have been drawn from studies of FaENV (Burke et al. 2018) and other independent
345 endogenization events involving different virus families, indicating this may be a general
346 feature of viral domestication.

347

348 ## Conservation of essential functions and adaptive gene loss

349 Despite extensive genomic rearrangement, genes critical for VLP production have been
350 selectively retained across all Campopleginae species examined. The 22 nudivirus core genes
351 found in VcENV are present in CcapiENV and CnolaENV (Table 1), with particularly strong
352 conservation of genes involved in viral transcription (*p47*, *lef* genes) and DNA amplification
353 (*helicase*): these genes show significantly lower average dN/dS ratios (0.16) compared to
354 envelope and infectivity genes (0.58) (Figure 6B, Table S6), indicating strong purifying
355 selection. The functional importance of *lef* genes has been experimentally demonstrated, as
356 their silencing abolishes or reduces particle production in *V. canescens* and *M. demolitor*
357 (Burke et al. 2013; Cerqueira de Araujo et al. 2022).

358 In contrast, genes unnecessary for VLP production have been systematically lost or
359 pseudogenized. Genes involved in capsid formation (*vp39*, *38K*, *p6.9*) and DNA packaging
360 (*integrase*, *FEN-1*, *DNApol*) are pseudogenized in all VLP-producing Campopleginae wasps
361 (Table 1), consistent with the fact that these particles lack capsids and DNA, although *R.*
362 *megacephalus* retained one copy of *p6.9* and several copies of *integrase*, which could
363 indicate that this species produces a different kind of particle. This pattern contrasts with
364 FaENV and BVs, which retain functional copies of *38K* and *vp39* (Bézier et al. 2009; Burke et
365 al. 2014; Burke et al. 2018), reflecting their production of DNA-containing particles.

366 Interestingly, *vlf-1* is functional in *Campoplex* species but pseudogenized in *V. canescens* and
367 absent in *R. megacephalus* (Table 1). In baculoviruses, VLF-1 participates in late gene
368 transcription and virion maturation (McLachlin and Miller 1994; Yang and Miller 1999;
369 Vanarsdall et al. 2006). Although VLF-1 was not detected in CcapiVLPs, the gene is
370 specifically expressed in ovaries (Figure 2B, Table S3), suggesting it may regulate the

371 expression of some nudiviral genes involved in VLP production. Additionally, *OrNVof61-like* is
372 absent in RmegaENV (Table 1). In *V. canescens*, this gene is lately expressed in ovaries
373 (Cerqueira de Araujo et al. 2022) and its knockdown results in altered morphology in a
374 subset of VLPs and increases host-mediated egg encapsulation (Mao et al. 2025). Given
375 these differences in gene content between RmegaENV and the other endogenized viruses
376 examined here, it is plausible that RmegaVLPs differ structurally from those produced by *V.*
377 *canescens* and *Campoplex* species, which may reflect adaptation to distinct host taxa (*V.*
378 *canescens* and *Campoplex* species parasitize lepidopterans, whereas *R. megacephalus*
379 parasitizes wood-boring beetles, i.e., coleopterans). Structural divergence remains to be
380 explored experimentally in these wasps.

381

382 Diversification through gene duplication and divergent envelope evolution

383 We identified several duplicated nudivirus genes with distinct patterns across species: *pif-5*
384 is duplicated in VcENV, CcapiENV and CnolaENV, but not in RmegaENV; *p74*, *pif-1*, *pif-3*, *pif-*
385 *4*, *pif-6*, and *lef-8* are single-copy genes in *Campoplex* spp. and *V. canescens*, but duplicated
386 in *R. megacephalus*; VcENV has its own specific duplications: *Ac81* and *OrNVorf47-like*; and
387 *OrNVorf41-like* (an 11K family member) is duplicated in all studied species (Table 1). In
388 baculoviruses, 11K family genes (comprising *Ac145* and *Ac150*) mediate host-specific oral
389 infectivity (Lapointe et al. 2004; Beperet et al. 2015). Consistently, RNA interference
390 targeting *OrNVorf41*-like-1 and -2 in *V. canescens* results in the production of seemingly
391 empty virus-like particles (VLPs) with electron-dense membranes, and has furthermore
392 shown that these genes are essential to escape host immunity (Mao et al. 2025). The *pif*
393 genes encode envelope components involved in particle infectiosity, their proteins enable

394 particle entry in the host gut cells (Kikhno et al. 2002; Zheng et al. 2017; Boogaard et al.
395 2018). The functions of these genes suggest that their duplication may enable wasp species
396 to parasitize different host ranges, possibly by tailoring VLP specificity for host cell
397 membrane proteins and/or, hypothetically, modulating the recruitment of virulence factors.

398 Most duplicated genes show elevated dN/dS ratios compared to single-copy genes (Table
399 S6), indicating relaxed constraint or positive selection. This pattern is particularly
400 pronounced for envelope and infectivity genes, which show 1- to 2-fold higher dN/dS ratios
401 than transcription machinery genes (Figure 6B). Similar patterns have been observed in BV-
402 associated braconids and filamentous virus-associated *Leptopilina* wasps (Di Giovanni et al.
403 2020; Gauthier et al. 2021), suggesting that envelope proteins evolve rapidly in response to
404 host-specific selection pressures while core replication machinery remains conserved.

405

406 Striking divergence in virulence protein content

407 Despite producing morphologically similar VLPs with nearly identical viral protein
408 compositions (Figures 5 and 6A, Table S5), *C. capitator* and *V. canescens* package completely
409 different virulence proteins. The three major VcVLP virulence proteins—VLP1 (PHGPx), VLP2
410 (RhoGAP), and VLP3 (neprilysin)—are absent from CcapiVLPs. Instead, CcapiVLPs contain a
411 distinct set of candidate virulence factors, including a serpin, peroxiredoxin, and remarkably,
412 proteins encoded by eight *rep*-like genes (Table S8).

413 The serpin identified in CcapiVLPs is particularly interesting as it belongs to an expanded
414 gene family with 15 members in *Campoplex* species versus only four in *V. canescens* (Figure
415 S5). This serpin is specific to *Campoplex* species and represents a prime candidate for

416 mediating immune suppression, as serpins in other parasitoid systems inhibit the
417 phenoloxidase cascade involved in melanisation (Beck and Strand 2007; Colinet et al. 2009).
418 The presence of other venom-associated proteins like peroxiredoxin, which protects against
419 oxidative stress during encapsulation (Perkin et al. 2015), could further suggest that
420 CcapiVLPs employ fundamentally different molecular strategies for host
421 immunosuppression.

422

423 ## Convergent recruitment of rep virulence factors via distinct delivery mechanisms

424 The most striking discovery is the packaging of REP proteins into CcapiVLPs. *Rep* (repeat
425 element) genes are the most abundant gene family in wasp-associated ichnoviruses, with 30
426 members in CsIV, 36 in HfIV, and 10 in TrIV (Galibert et al. 2006; Webb et al. 2006). While
427 their precise function remains unclear, their conservation, high expression in parasitized
428 hosts, and extensive expansion across IV genomes strongly suggest critical roles in parasitism
429 success (Volkoff et al. 2002; Tanaka et al. 2007; Kim et al. 2023).

430 We identified 48 and 16 *rep*-like sequences in *C. capitator* and *C. nolae*, respectively, versus
431 only one or two in *R. megacephalus* and *V. canescens* (Table S2), with several showing high
432 ovary-specific expression (Table S3). Critically, eight REP proteins were detected in purified
433 CcapiVLPs (Table S8), demonstrating that these virulence factors are packaged into VLPs.

434 This represents a remarkable example of convergent evolution: the same virulence protein
435 family is deployed by different Campopleginae lineages using entirely different delivery
436 mechanisms—DNA circles derived from ichnoviruses versus protein-filled VLPs derived from
437 nudiviruses. The expansion of *rep* genes specifically in *C. capitator* reflects a similar adaptive

438 strategy to venom protein diversification in parasitoids, whereby gene duplication enables
439 the generation of host-specific virulence arsenals in response to ongoing selection pressures.

440

441 ## No compelling evidence for virus replacement in Campoplex

442 Previous studies suggested *V. canescens* might have undergone "endogenous virus
443 replacement," losing an ancestral ichnovirus after acquiring the alphanudivirus (Pichon et al.
444 2015). However, we found no remnants of IVSPER genes in *Campoplex* and *Rhimphoctona*
445 genomes, despite thorough searches using both functional IVSPER sequences from
446 ichnovirus-associated wasps and pseudogenized IVSPERs from *V. canescens*. Similarly, Burke
447 and collaborators (Burke et al. 2021) found no IVSPER sequences in *Dusona* sp., a close
448 relative of *Campoplex*.

449 Current evidence suggests that ichnovirus integration occurred in the *Hyposoter-Campoletis*
450 clade, while the "*Campoplex-Venturia-Dusona*" clade (including *R. megacephalus*)
451 independently endogenized an alphanudivirus (present study; Burke et al. 2021;
452 Sharanowski et al. 2021). Given that *rep*-like sequences also appear in lepidopterans and
453 sawflies, we cannot definitively conclude that their presence alone indicates ancestral
454 ichnovirus integration in *Campoplex* species. These sequences could originate via horizontal
455 transfer from ichnovirus circles, as IV and BV integrations have been documented in
456 lepidopteran hosts (Gasmi et al. 2015; Heisserer et al. 2023). Under this scenario, *rep*-like
457 sequences in *Campoplex*, *V. canescens*, and *R. megacephalus* would reflect transfers from
458 parasitized lepidopterans that had themselves been parasitized by ichnovirus-carrying
459 wasps. However, no lepidopteran sequences clustered phylogenetically with these wasp REP

460 sequences (Figure 4A), and more genome assemblies of lepidopteran hosts are needed to
461 distinguish between horizontal transfer and ancestral ichnovirus integration.

462

463 ## Implications for viral domestication and host-parasite coevolution

464 The comparative analysis of endogenized alphanudiviruses in Campopleginae wasps reveals
465 key principles of viral domestication and highlights the remarkable versatility of virus-
466 derived particle systems in parasitoid wasps. Early genomic rearrangement appears to be a
467 hallmark of the domestication process, potentially disabling viral replication while preserving
468 beneficial functions. Following this initial disruption, strong purifying selection maintains
469 core particle production machinery, while envelope proteins diversify rapidly in response to
470 host-specific pressures. Most strikingly, the virulence cargo of morphologically and
471 structurally similar VLPs can diverge completely over evolutionary time, demonstrating
472 extraordinary flexibility in how endogenized viruses are weaponized against different hosts.
473 The VLP system in *C. capitator* exemplifies this versatility through its packaging of a
474 *Campoplex*-specific serpin and expanded REP protein families—the latter representing a
475 remarkable case of convergent evolution, as REP proteins are typically delivered via
476 ichnovirus DNA circles in other Campopleginae lineages.

477 This convergent deployment of similar virulence strategies through different viral systems
478 reflects a broader pattern across parasitoid wasps. Multiple viral families have been
479 independently domesticated to produce particles essential for successful parasitism:
480 deltanudiviruses generating BVs in braconids, ichnoviruses producing IVs in Banchinae and
481 some Campopleginae, alphanudiviruses producing VLPs independently in *F. arisanus* and in

482 the *Rhimphoctona-Venturia-Campoplex* clade, and hytrosavirus producing VLPs in
483 *Leptopilina*. The repeated evolution of REP protein deployment—whether packaged in
484 nudivirus-derived VLPs or delivered on ichnovirus circles—highlights how parasitoid wasps
485 repeatedly arrive at similar molecular solutions through distinct evolutionary paths, driven
486 by the relentless selection pressure of host immune responses.

487

488 To conclude, the genomes of four Campopleginae wasps harbour an alphanudivirus which
489 derive from the same endogenization event in the common wasp ancestor. The description
490 of these new endogenized nudiviruses sheds light on early evolutionary processes of viral
491 domestication. Comparison with the closely related *V. canescens* species shows that there
492 are mostly minute differences between the genomes of endogenized nudiviruses in
493 Campopleginae wasps. However, major differences can be seen in the virulence protein
494 content of VLPs which appears to be strikingly different between *C. capitator* and *V.*
495 *canescens*, showing early adaptation to wasp's hosts in the context of viral domestication.

496

# 497 Material and methods

## 498 Parasitoid model

499 *Campoplex capitator* wasps were obtained after emergence from the lepidopteran host
500 *Lobesia botrana*, a vineyard pest that can grow on *Daphne gnidium* (Loni et al. 2016).
501 Caterpillars of *L. botrana* were sampled from *D. gnidium* shrubs in the natural reserve of
502 Migliarino-San Rossore-Massaciuccoli (Tuscany, Italy). Samplings were carried out in early

503 July of 2019 and 2021, with wasps emerging between June and August (Scaramozzino et al.
504 2018). *L. botrana* nests were placed individually in test tubes and each tube was monitored
505 every day. Emerging *C. capitator* were used for experiments.

506 To produce enough wasps for our experiments, a *C. capitator* rearing was maintained as
507 described in (Lucchi et al. 2018; Benelli et al. 2020): *L. botrana* larvae were kept in plastic
508 boxes (20cm D x 15cm W x 10cm H) with a nutritious substrate (30g of agar, 60g of sugar,
509 50g of alfalfa, 36g of Brewer yeast, 25g of Wesson salt, 180g of wheat sprout, 80g of casein,
510 4g of sorbic acid, 15g of Wanderzahnt vitamins, 2.5g of cholesterol, 2.5g of tetracyclin, 5mL
511 of propionic acid (99,5%), 2mL of linoleic acid (95%), 5mL of olive oil and 1.5mL of distilled
512 water). After the emergence, adults were kept in the same box for reproduction where the
513 substrate was removed and droplets of honey were added. Once eggs were visible on the
514 box surface, adults were removed and a nutritious substrate was added to feed the future
515 larvae. *Campoplex capitator* females and males were meanwhile kept in plexiglass chambers
516 (40cm D x 25cm W x 30cm H) with honey to allow their reproduction. L2 to L4 host larvae
517 were put in the chambers for two hour parasitism periods and then individually reared into
518 small boxes (5cm Ø x 5cm H) with nutritious substrate. These boxes were monitored every
519 day to check for wasp emergence and honey was added when larvae were about to pupate.
520 All insects were kept at 25°C and 45% of humidity, with a 16:8 day/night photoperiod.

521 Meanwhile, a female specimen of *Campoplex nolae* was collected in the wild by active
522 sweeping at Powdermill Nature Reserve, Rector county, Pennsylvania, USA, and kept in
523 liquid nitrogen until the time of DNA isolation.

524

525 DNA extraction and sequencing

526 For *C. capitator*, DNA was extracted from one haploid male at emergence using the
527 MagAttract® HMW DNA Kit (Qiagen). The male wasp tissues were disrupted in a 2 mL tube
528 containing 200 µL of 1X DPBS. 20 µL of proteinase K, 4 µL of RNAse A and 150µL of buffer AL
529 were added to the tube. The tube was then mixed carefully and incubated at 56°C at 900
530 rpm for 2 hours. The manufacturer's protocol was followed from step 4 to step 16. Total
531 DNA was finally eluted in 100 µL of TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0) and
532 quantified using the Qubit™ DNA HS assay kit (Invitrogen) and the Qubit2.0 fluorometer
533 (Invitrogen). DNA quality and contamination were checked by measurement of 260/280 and
534 260/230 OD ratios using the Varian Cary® 50 Scan spectrophotometer. Total DNA was then
535 stored at 4°C. The library was prepared and sequenced using the PacBio Sequel II technology
536 by the Gentyane platform (France).

537 For *C. nolae*, extraction was performed using the MagAttract® HMW DNA Kit (Qiagen),
538 following the manufacturer's protocol with the whole specimen immersed in extraction
539 buffer and proteinase K overnight. An additional purification step using Agencourt AMPure
540 XP beads (Beckman Coulter, Brea, U.S.A.) at 0.5X was used to remove small fragments.
541 Genomic libraries were prepared for parallel sequencing in short-read (Illumina) and long-
542 read (Nanopore) platforms: the Illumina library was prepared using the Kapa HyperPrep kit
543 (Kapa Biosystems, Wilmington, MA). A-tailing, end-repair and ligation reactions were
544 performed at a quarter volume relative to the standard protocol, while library amplification
545 was performed at full volume. Custom, dual-indexing adapter-primers were used to allow for
546 in silico de-multiplexing of each sample (Glenn et al., 2016). Following stub ligation and PCR,
547 adapter-dimers were removed by performing a 0.8X bead cleaning using AMPure beads. The

548 library was pooled with other samples at equimolar concentrations and sequenced at 4 nM
549 as single lanes on Illumina NovaSeq 6000 S4 platform (2x150; Illumina Inc., San Diego, CA).
550 The library for long-read data was prepared using the Nanopore ligation kit, complemented
551 by reagents from the NEBNext Ultra II library prep kit (New England Biolabs, Ipswich,
552 Massachusetts, U.S.A.), and then sequenced at a MinION platform (Oxford Nanopore
553 Technologies, Oxford, U.K.) using the manufacturer's specifications.

554 For both wasp species, raw data are available in NCBI (accession number of
555 the BioProject: PRJNA936130).

556

557 RNA extraction and sequencing

558 Emerging *C. capitator* wasps were dissected in order to obtain ovaries where CcapiVLP are
559 produced. Venom glands and head-thorax of female adults and testis and head-thorax of
560 male adults were also dissected to investigate expression of CcapiENV genes outside of the
561 calyx region and in males. For all samples, two pools were analysed. Pools were composed of
562 5 individuals for head-thorax for both males and females, 21 female venom glands, 5 ovary
563 pairs and 15 male testis pairs. Total RNA was extracted from the 10 pools in total
564 using TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. Only the
565 isopropanol step was replaced by cold absolute ethanol (3 volumes per sample) and total
566 RNA was finally resuspended in 30µL of RNAse free water. RNA quantity of each pool was
567 quantified using the Qubit™ RNA HS assay kit (Invitrogen) designed for the Qubit2.0
568 fluorometer (Invitrogen). Samples of extracted RNA were then stored at -80°C.

569 For genome annotation, total RNA was extracted from two pools of four entire emerging
570 female wasps using the NucleoSpin®RNA kit (Macherey-Nagel). Total RNA was finally
571 resuspended in 40µL of RNAse free water. RNA quantity of the two pools was
572 quantified using the Qubit™ RNA HS assay kit (Invitrogen) designed for the Qubit2.0
573 fluorometer. Samples of extracted RNA were then stored at -80°C.

574 The sequencing of library preparations and the generation of paired-end reads were
575 performed on an Illumina platform by the Novogene company (UK). Raw data are available
576 in NCBI (accession number of the BioProject: PRJNA936130).

577

578 Genome assembly and annotation

579 For *C. capitator*, genome and K-mer sizes were estimated beforehand using KmerGenie
580 (Chikhi and Medvedev 2014). Raw reads were then corrected and trimmed using
581 respectively Canu–correct and Canu–trim modules (Canu software v1.8) (Koren et al. 2017)
582 in order to improve the base accuracy and to retrieve high quality sequence portions only.
583 The genome was then assembled using the Canu–assemble module. The newly assembled
584 genome was polished using Racon (v1.4.21) with the longest 80X raw reads to further
585 improve the base accuracy using the longest raw reads. In parallel, *C. nolae* raw reads from
586 Nanopore sequencing were filtered using Filtlong (v.0.2.1) before being assembled using Flye
587 (v.2.9-b1768) (Kolmogorov et al. 2020). The produced assembly was then polished with Hypo
588 (v.1.0.3.flye) using the Illumina reads. Scripts for *C. nolae* genome assembly are available at
589 https://gitlab.in2p3.fr/marie.cariou/campoplex_assembly/.

590 The assembled genome completeness was assessed with BUSCO (Benchmarking Universal
591 Single-Copy Orthologs, v5.4.2) (Manni et al. 2021) using Insecta and Hymenoptera lineage
592 datasets (odb10) and genome contamination was checked with Blobtools (v1.1.1) (Laetsch
593 and Blaxter 2017) for both genomes.

594 For both genomes, repeated elements were annotated *de novo* using the TEdenovo pipeline
595 comprised in the REPET pipeline (v2.5) (Flutre et al. 2011). Two rounds of the TEannot
596 pipeline (contained in the REPER pipeline) were then used to annotate repeated elements
597 using existing databases. Repeated elements were then masked using Bedtools (v.2.29.1)
598 (Quinlan and Hall 2010) before gene prediction. Gene models were generated for *C.*
599 *capitator* using BRAKER-2 (v2.1.6) (Brůna et al. 2021) with assembled transcriptomes by
600 spades (v3.15.5) (Bushmanova et al. 2019). Models were then functionally annotated using
601 InterProScan (v5.53-87.0) (Blum et al. 2021) and BLASTp (NCBIblast+ v2.13.0) with the nr
602 database (downloaded the 14$^{th}$ of April 2023), before being polished by AGAT (v1.0.0)
603 (Dainat et al. 2023). Proteins were extracted from gene models using the
604 agat_sp_extract_sequences.pl tool from AGAT software. Finally, the genome completeness
605 was assessed again using the protein sequences with BUSCO.

606 In total, 44 CcapiENV genes were annotated with the automated annotation process. Using
607 the gene models predicted previously, additional viral genes were then annotated by
608 reciprocal blasts: translated sequences from predicted genes were aligned using BLAST tools
609 on the NCBI nr database filtered for sequences of endogenous and exogenous nudiviruses.
610 Aligned sequences were then extracted and realigned on the complete NCBI nr database.
611 The final BLAST outputs were filtered to keep only sequences which aligned on sequences of
612 endogenous and exogenous nudiviruses reported in the database.

613 To annotate nudivirus genes from *C. nolae* genome, open reading frames were predicted
614 using orfipy (v0.0.4) (Singh and Wurtele 2021). Nudivirus genes were then annotated by
615 reciprocal blasts using translated sequences from predicted ORFs. Miniprot (v0.9-0) (Li 2023)
616 was also used to align *C. capitator* proteins on *C. nolae* genome and validate previously
617 annotated nudivirus genes. Miniprot gene models were also used for gene clustering in the
618 dN/dS ratio analysis.

619 Genome assembly and gene predictions for *R. megacephalus* were obtained from the DToL
620 data portal (accession number: GCA_963989365.1, ). To annotate nudivirus genes in the *R.*
621 *megacephalus* genome, a complete list of nudiviral proteins was used. Nudiviral proteins
622 were searched against the wasp genome using tBLASTn (matrix: BLOSUM62; maximum E-
623 value: 1e-2). All hits were then analyzed individually, and open reading frames were
624 predicted using Geneious Prime® 2019.0.4.

625 Pseudogenized genes were identified by reciprocal blasts: tBLASTn was used first to map
626 nudivirus proteins on the wasp genomes. Hits were then verified by BLASTx to map potential
627 nudiviral nucleotide sequences on the NCBI nr database. Proteins corresponding to
628 nucleotide sequences mapping back on nudivirus were kept and incomplete and non-
629 transcribed sequences were considered as pseudogenes.

630 *Rep* genes and ichnovirus sequences were searched in *C. capitator*, *C. nolae*, *V. canescens*, *R.*
631 *megacephalus* and *Dusona sp.* Genomes. Sequences from ichnoviruses were first
632 downloaded from NCBI databases and manually completed. IVSPER sequences were
633 searched using a homemade pipeline: genomic IVSPER sequences were first aligned on the
634 genomes (the genome sequences were segmented into blocks of 500,000 bases) using
635 BLASTn and tBLASTx. Aligning chunks were then aligned with BLASTx on the NCBI nr

636 database. These hits were filtered to keep only hits aligning on IVSPER proteins. These hits
637 were then verified by aligning them back on the nr database using BLASTp. In parallel,
638 pseudogene sequences of IVSPER described in *V. canescens* were aligned on the genomes
639 using tBLASTn. Protein sequences of the matching regions were then extracted and blasted
640 against a portion of the NCBI nr database specific to Ichneumonidae (taxid:7408) and Viruses
641 (taxid:10239) using a PSI-BLAST with an e-value threshold and a word size set to 0.5 and 2
642 respectively. Finally, only hits aligning with IVSPER genes contained in the NCBI database
643 were retained. *Rep* genes and pseudogenes were all annotated manually using tBLASTn.

644 ### CO1 and nudivirus gene alignments

645 Because the endogenized nudivirus genomes were highly similar in both *Campoplex* species,
646 we collected evidence to confirm that the *Campoplex* wasps sequenced for this study are
647 part of different species. Methods and results for this part are available in the
648 supplementary materials.

649

650 ### Campopleginae and endogenized nudivirus evolution

651 To obtain a phylogenetic tree of ichneumonid wasps, complete BUSCO sequences from
652 RefSeq assemblies were extracted using BUSCO and the hymenoptera odb10 database.
653 Sequences were then aligned using MAFFT (v7.487) (Katoh et al. 2019) and the following
654 parameters: maxiterate=1000 (maximum number of iterations), genafpair (Altschul
655 algorithm (Altschul 1998)). Aligned sequences were then trimmed by TrimAL (v1.4.rev15)
656 (Capella-Gutierrez et al. 2009) with the gap threshold parameter fixed at 0.6. Trees were
657 then generated for each sequence using IQ-TREE 2 (v2.1.4) (Minh et al. 2020) and the

658 following parameters: m=TEST (use the best model found), B=1000 (number of bootstrap
659 iterations), alrt=1000 (number of SH-aLRT replicates). A species topology was then inferred
660 using ASTRAL (v 5.7.8) (Zhang et al. 2018). To build the final species phylogeny, sequences
661 were concatenated using goalign (v 0.3.7) (Lemoine and Gascuel 2021) and a partitioned
662 tree was built using IQ-TREE 2 with the constrained topology inferred by ASTRAL.

663 To perform the virus phylogeny, protein sequences of one nimavirus, three hytrosaviruses,
664 sixty-nine baculoviruses and twenty nudiviruses (including CcapiENV and *C. nolae* sequences)
665 were first aligned together to find orthologous genes. Results were then manually corrected
666 using the knowledge found in the literature (Bézier, Annaheim, et al. 2009; Bézier,
667 Herbinière, et al. 2009; Burand et al. 2012; Bézier et al. 2015; Burke 2019; Drezen et al.
668 2022). Protein sequences of orthologous genes were then aligned using MAFFT with the
669 previous parameters. Alignments were then trimmed by TrimAL (gap threshold = 0.6). For
670 each gene, a tree was finally generated with IQ-TREE 2 with the previous parameters.
671 Protein sequences of genes that were not supporting the accepted phylogeny (at the family
672 level) were removed from the further analysis. Finally, 21 baculovirus core genes and 13
673 alphanudivirus core genes were kept for the virus tree construction. Protein sequences of
674 these genes were aligned and trimmed as described previously. Sequences were then
675 concatenated using goalign and a partitioned tree was constructed with iqtree2 with the
676 same parameters as previously described. The tree image was then built using ape (v5.7-1),
677 treeio (v1.24.0) (Wang et al. 2020) and ggtree (v3.8.0) (Yu et al. 2017) packages with R
678 (v4.2.1).

679 A whole genome synteny analysis was performed using the R package GENESPACE (v1.2.3)
680 (Lovell et al. 2022). Parity plots (Figure 3C to E) showing syntenies among nudiviruses were
681 generated with Excel (Microsoft v16).

682 To estimate the evolutionary constraint occurring on groups of genes according to their
683 function, protein sequences of *C. capitator*, *C.nolae*, *V. canescens* and *R. megacephalus* were
684 first clustered using OrthoMCL (v2.0.9) (Li et al. 2003) and scripts written by Arun Seetharam
685 (https://github.com/ISUgenomics/common_scripts). Clustered protein sequences were then
686 aligned using ClustalOmega (v1.2.4) and aligned sequences were converted into codon
687 alignment using PAL2NAL (v14.1) (Suyama et al. 2006). Phylip trees were generated using IQ-
688 TREE 2 with the following parameters: -m JTT (Jones-Taylor-Thornton model (Jones et al.
689 1992)). Codeml (Paml v4.10.6) was finally used to calculate dN and dS values.

690 To construct the phylogenetic tree of *rep* genes, *rep* sequences from ichnovirus-associated
691 wasps and newly annotated *rep* were analysed altogether with *rep* sequences identified in
692 lepidopterans, sawflies, and phasmids - which were all annotated manually using tBLASTn.
693 Amino acid sequences were then aligned with MAFFT using parameters adapted for a
694 divergent gene family (--localpair, --maxiterate 1000, --bl 45). The phylogenetic tree was
695 subsequently generated using IQ-TREE 2 (v2.4.0) with the following parameters: -m MFP, -bb
696 1000, -alrt 1000. Sequences forming long branches and a group of REPs in *C. capitator* were
697 removed, and the entire analysis was repeated on the remaining sequences. The sequences
698 forming long branches were excluded to avoid artifacts in the tree topology. The group of
699 REPs in *C. capitator* corresponded to a set of sequences showing 100% similarity among
700 small genomic scaffolds, which were excluded due to a suspected assembly bias. To simplify
701 the final figure, only bootstrap values are shown (Figure 4).

702

703 ## Gene expression and regulatory sequences of CcapiENV genes

704 Results and methods are available in the supplementary materials.

705

706 ## CcapiVLP purification and protein sequencing

707 In order to purify Virus-Like-Particles, 61 adult female wasps of *C. capitator* were dissected 2
708 days after emergence. In total, 122 calyces were obtained and pooled in 1X DPBS. Calyces
709 were then disrupted thanks to a pellet pestle and CcapiVLPs were purified as described in
710 (Pichon et al. 2015): disrupted calyces were centrifuged at 770 g, at 4°C for 10 minutes in
711 order to separate CcapiVLPs and calyx fluid (supernatant) from cellular waste (pellet). The
712 supernatant was then retrieved and transferred into a new tube. The new tube was then
713 centrifuged at 15 400 g, at 4°C for 10 minutes. The supernatant was discarded and the pellet
714 containing CcapiVLPs was suspended in 20 µL of 1X DPBS and stored at -20°C.

715 CcapiVLP proteins were then separated by SDS-polyacrylamide gel electrophoresis (12.5%
716 polyacrylamide gel) under denaturing conditions. Separated bands were incubated for 1h
717 with a Coomassie blue solution (10% acetic acid, 50% ethanol, 0.1% R250 Coomassie blue,
718 q.s. to 1L of milliQ water) and 9 slices were cut from the gel. Slices were chopped into $1mm^3$
719 cubes before being stored at -20°C and analysed by the platform PIXANIM (France) using a
720 nanoUPLC_LTQ-VElos Pro Orbitrap Mass Spectrometer.

721 For each protein identified by the mass spectrometry, sites and domains were searched for
722 using InterProScan and gene ontology terms were attributed using sequence homology with

723 curated proteins from the Uniprot SWISSprot database (downloaded the 18th November
724 2022). Proteins with a peptide abundance higher than 50 were then retained as VLP
725 candidates.

726

727 ### Electronic microscopy of the calyx region

728 The calyx region of adult *C. capitator* wasp ovaries was observed under electronic
729 microscopy. Wasp ovaries were dissected two days after emergence and fixed in PrOx/EPON
730 epoxy resin. Samples were fixed, cut and observed with the same protocol as one used in
731 Cerqueira de Araujo and collaborators (Cerqueira de Araujo, Leobold, et al. 2022). The
732 protocol is available in Supplementary materials.

733

## 734 Acknowledgements

735 The authors would like to thank Francesca Cosci, Filippo Di Giovanni and Augusto Loni from
736 the Department of Agriculture, Food and Environment (DAFE) of the University of Pisa (Italy)
737 for the help provided during field wasp sampling. The authors would like to thank also Amine
738 Chebbi who helped the first author on the transposable element annotation pipeline and
739 Christophe Bressac who determined the sperm production period of male *C. capitator*. We
740 would like to thank the IBiSA Electron Microscopy Facility of the University of Tours and the
741 University Hospital of Tours for their assistance. Research funding was provided by an ADS
742 Research Grant from the Smithsonian Institution to BFS, who was also supported by a GGI
743 Peter Buck postdoctoral fellowship during early stages of this work.

744

745 ## Data availability

746 Raw reads and assemblies are deposited on the NCBI database under BioProject

747 PRJNA936130.

748 # Reference


749 Altschul SF. 1998. Generalized affine gap costs for protein sequence alignment. *Proteins*
750 *Struct. Funct. Genet.* 32:88–96.

751 Annaheim M, Lanzrein B. 2007. Genome organization of the Chelonus inanitus polydnavirus:
752 excision sites, spacers and abundance of proviral and excised segments. *J. Gen. Virol.* 88:450–
753 457.

754 Arbuckle JH, Medveczky MM, Luka J, Hadley SH, Luegmayr A, Ablashi D, Lund TC, Tolar J, De
755 Meirleir K, Montoya JG, et al. 2010. The latent human herpesvirus-6A genome specifically
756 integrates in telomeres of human chromosomes in vivo and in vitro. *Proc. Natl. Acad. Sci.*
757 107:5563–5568.

758 Asgari S, Reineke A, Beck M, Schmidt O. 2002. Isolation and characterization of a neprilysin-
759 like protein from *Venturia canescens* virus-like particles. *Insect Mol. Biol.* 11:477–485.

760 Ballinger MJ, Bruenn JA, Hay J, Czechowski D, Taylor DJ. 2014. Discovery and Evolution of
761 Bunyavirids in Arctic Phantom Midges and Ancient Bunyavirid-Like Sequences in Insect
762 Genomes.Doms RW, editor. *J. Virol.* 88:8783–8794.

763 Beck MH, Strand MR. 2007. A novel polydnavirus protein inhibits the insect
764 prophenoloxidase activation pathway. *Proc. Natl. Acad. Sci.* 104:19267–19272.

765 Beck MH, Zhang S, Bitra K, Burke GR, Strand MR. 2011. The Encapsidated Genome of
766 Microplitis demolitor Bracovirus Integrates into the Host Pseudoplusia includens. *J. Virol.*
767 85:11685–11696.

768 Bedwin O. 1979. An insect glycoprotein: a study of the particles responsible for the resistance
769 of a parasitoid's egg to the defence reactions of its insect host. *Proc. R. Soc. Lond. B Biol. Sci.*
770 205:271–286.

771 Bejarano ER, Khashoggi A, Witty M, Lichtenstein C. 1996. Integration of multiple repeats of
772 geminiviral DNA into the nuclear genome of tobacco during evolution. *Proc. Natl. Acad. Sci.*
773 93:759–764.

774 Béliveau C, Cohen A, Stewart D, Periquet G, Djoumad A, Kuhn L, Stoltz D, Boyle B, Volkoff A-
775 N, Herniou EA, et al. 2015. Genomic and Proteomic Analyses Indicate that Banchine and
776 Campoplegine Polydnaviruses Have Similar, if Not Identical, Viral Ancestors.McFadden G,
777 editor. *J. Virol.* 89:8909–8921.

778 Belyi VA, Levine AJ, Skalka AM. 2010. Sequences from Ancestral Single-Stranded DNA Viruses
779 in Vertebrate Genomes: the *Parvoviridae* and *Circoviridae* Are More than 40 to 50 Million
780 Years Old. *J. Virol.* 84:12458–12462.

781 Benelli G, Ricciardi R, Romano D, Cosci F, Stefanini C, Lucchi A. 2020. Wing-fanning frequency
782 as a releaser boosting male mating success—High-speed video analysis of courtship behavior
783 in *Campoplex capitator* , a parasitoid of *Lobesia botrana*. *Insect Sci.* 27:1298–1310.

784 Beperet I, Simón O, Williams T, López-Ferber M, Caballero P. 2015. The “11K” gene family
785 members sf68, sf95 and sf138 modulate transmissibility and insecticidal properties of
786 Spodoptera frugiperda multiple nucleopolyhedrovirus. *J. Invertebr. Pathol.* 127:101–109.

787 Bézier A, Annaheim M, Herbinière J, Wetterwald C, Gyapay G, Bernard-Samain S, Wincker P,
788 Roditi I, Heller M, Belghazi M, et al. 2009. Polydnaviruses of Braconid Wasps Derive from an
789 Ancestral Nudivirus. *Science* 323:926–930.

790 Bézier A, Herbinière J, Lanzrein B, Drezen J-M. 2009. Polydnavirus hidden face: The genes
791 producing virus particles of parasitic wasps. *J. Invertebr. Pathol.* 101:194–203.

792 Bézier A, Louis F, Jancek S, Periquet G, Thézé J, Gyapay G, Musset K, Lesobre J, Lenoble P,
793 Dupuy C, et al. 2013. Functional endogenous viral elements in the genome of the parasitoid
794 wasp *Cotesia congregata*: insights into the evolutionary dynamics of bracoviruses. *Philos.*
795 *Trans. R. Soc. B Biol. Sci.* 368:20130047.

796 Bézier A, Thézé J, Gavory F, Gaillard J, Poulain J, Drezen J-M, Herniou EA. 2015. The Genome
797 of the Nucleopolyhedrosis-Causing Virus from *Tipula oleracea* Sheds New Light on the
798 Nudiviridae Family.McFadden G, editor. *J. Virol.* 89:3008–3025.

799 Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-
800 Lafosse T, Qureshi M, Raj S, et al. 2021. The InterPro protein families and domains database:
801 20 years on. *Nucleic Acids Res.* 49:D344–D354.

802 Boogaard B, Van Oers M, Van Lent J. 2018. An Advanced View on Baculovirus per Os
803 Infectivity Factors. *Insects* 9:84.

804 Boparai J, Broad GR. 2024. The genome sequence of an ichneumonid wasp, Rhimphoctona (
805 Xylophylax) megacephalus (Gravenhorst, 1829). *Wellcome Open Res.* 9:472.

806 Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic
807 eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein
808 database. *NAR Genomics Bioinforma.* 3:lqaa108.

809 Burand JP, Kim W, Afonso CL, Tulman ER, Kutish GF, Lu Z, Rock DL. 2012. Analysis of the
810 Genome of the Sexually Transmitted Insect Virus *Helicoverpa zea* Nudivirus 2. *Viruses* 4:28–
811 61.

812 Burke GR. 2019. Common themes in three independently derived endogenous nudivirus
813 elements in parasitoid wasps. *Curr. Opin. Insect Sci.* 32:28–35.

814 Burke GR, Hines HM, Sharanowski BJ. 2021. The Presence of Ancient Core Genes Reveals
815 Endogenization from Diverse Viral Ancestors in Parasitoid Wasps. *Genome Biol. Evol.*
816 13:evab105.

817 Burke GR, Simmonds TJ, Sharanowski BJ, Geib SM. 2018. Rapid Viral Symbiogenesis via
818 Changes in Parasitoid Wasp Genome Architecture.Larracuente A, editor. *Mol. Biol. Evol.*
819 35:2463–2474.

820 Burke GR, Strand MR. 2012. Deep Sequencing Identifies Viral and Wasp Genes with Potential
821 Roles in Replication of *Microplitis demolitor* Bracovirus. *J. Virol.* 86:3293–3306.

822 Burke GR, Thomas SA, Eum JH, Strand MR. 2013. Mutualistic Polydnaviruses Share Essential
823 Replication Gene Functions with Pathogenic Ancestors.Schneider DS, editor. *PLoS Pathog.*
824 9:e1003348.

825 Burke GR, Walden KKO, Whitfield JB, Robertson HM, Strand MR. 2014. Widespread Genome
826 Reorganization of an Obligate Virus Mutualist.Malik HS, editor. *PLoS Genet.* 10:e1004660.

827 Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. 2019. rnaSPAdes: a de novo
828 transcriptome assembler and its application to RNA-Seq data. *GigaScience* 8:giz100.

829 Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated
830 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.

831 Cerqueira de Araujo A, Josse T, Sibut V, Urabe M, Asadullah A, Barbe V, Nakai M, Huguet E,
832 Periquet G, Drezen J-M. 2022. Chelonus inanitus bracovirus encodes lineage-specific proteins
833 and truncated immune IκB-like factors. *J. Gen. Virol.* [Internet] 103. Available from:
834 https://www.microbiologyresearch.org/content/journal/jgv/10.1099/jgv.0.001791

835 Cerqueira de Araujo A, Leobold M, Bézier A, Musset K, Uzbekov R, Volkoff A-N, Drezen J-M,
836 Huguet E, Josse T. 2022. Conserved Viral Transcription Plays a Key Role in Virus-Like Particle
837 Production of the Parasitoid Wasp *Venturia canescens*.Parrish CR, editor. *J. Virol.* 96:e00524-
838 22.

839 Chevignon G, Periquet G, Gyapay G, Vega-Czarny N, Musset K, Drezen J-M, Huguet E. 2018.
840 Cotesia congregata Bracovirus Circles Encoding *PTP* and Ankyrin Genes Integrate into the
841 DNA of Parasitized Manduca sexta Hemocytes.Jung JU, editor. *J. Virol.* 92:e00438-18.

842 Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome
843 assembly. *Bioinformatics* 30:31–37.

844 Colinet D, Dubuffet A, Cazes D, Moreau S, Drezen J-M, Poirié M. 2009. A serpin from the
845 parasitoid wasp Leptopilina boulardi targets the Drosophila phenoloxidase cascade. *Dev.*
846 *Comp. Immunol.* 33:681–689.

847 Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, Belhouchet M, Lemasson J-J, de Micco P,
848 de Lamballerie X. 2004. Sequences of flavivirus-related RNA viruses persist in DNA form
849 integrated in the genome of Aedes spp. mosquitoes. *J. Gen. Virol.* 85:1971–1980.

850 Dainat J, Hereñú D, Dr. K. D. Murray, Davis E, Crouch K, LucileSol, Agostinho N, Pascal-Git,
851 Zollman Z, Tayyrov. 2023. NBISweden/AGAT: AGAT-v1.2.0. Available from:
852 https://zenodo.org/record/8178877

853 Delaroque N, Maier I, Knippers R, M√°ller DG. 1999. Persistent virus integration into the
854 genome of its algal host, Ectocarpus siliculosus (Phaeophyceae). *J. Gen. Virol.* 80:1367–1370.

855 Desjardins CA, Gundersen-Rindal DE, Hostetler JB, Tallon LJ, Fadrosh DW, Fuester RW,
856 Pedroni MJ, Haas BJ, Schatz MC, Jones KM, et al. 2008. Comparative genomics of mutualistic
857 viruses of Glyptapantelesparasitic wasps. *Genome Biol.* 9:R183.

858 Di Giovanni D, Lepetit D, Guinet B, Bennetot B, Boulesteix M, Couté Y, Bouchez O, Ravallec
859 M, Varaldi J. 2020. A Behavior-Manipulating Virus Relative as a Source of Adaptive Genes for
860 *Drosophila* Parasitoids.Malik H, editor. *Mol. Biol. Evol.* 37:2791–2807.

861 Drezen J-M, Bézier A, Burke GR, Strand MR. 2022. Bracoviruses, ichnoviruses, and virus-like
862 particles from parasitoid wasps retain many features of their virus ancestors. *Curr. Opin.*
863 *Insect Sci.* 49:93–100.

864 Drezen J-M, Leobold M, Bézier A, Huguet E, Volkoff A-N, Herniou EA. 2017. Endogenous
865 viruses of parasitic wasps: variations on a common theme. *Curr. Opin. Virol.* 25:41–48.

866 Etebari K, Palfreyman RW, Schlipalius D, Nielsen LK, Glatz RV, Asgari S. 2011. Deep
867 sequencing-based transcriptome analysis of Plutella xylostella larvae parasitized by
868 Diadegma semiclausum. *BMC Genomics* 12:446.

869 Feddersen I, Sander K, Schmidt O. 1986. Virus-like particles with host protein-like antigenic
870 determinants protect an insect parasitoid from encapsulation. *Experientia* 42:1278–1281.

871 Feschotte C, Gilbert C. 2012. Endogenous viruses: insights into viral evolution and impact on
872 host biology. *Nat. Rev. Genet.* 13:283–296.

873 Flutre T, Duprat E, Feuillet C, Quesneville H. 2011. Considering Transposable Element
874 Diversification in De Novo Annotation Approaches.Xu Y, editor. *PLoS ONE* 6:e16526.

875 Galibert L, Devauchelle G, Cousserans F, Rocher J, Cérutti P, Barat-Houari M, Fournier P,
876 Volkoff A. 2006. Members of the Hyposoter didymator Ichnovirus repeat element gene
877 family are differentially expressed in Spodoptera frugiperda. *Virol. J.* 3:48.

878 Gasmi L, Boulain H, Gauthier J, Hua-Van A, Musset K, Jakubowska AK, Aury J-M, Volkoff A-N,
879 Huguet E, Herrero S, et al. 2015. Recurrent Domestication by Lepidoptera of Genes from
880 Their Parasites Mediated by Bracoviruses.Feschotte C, editor. *PLOS Genet.* 11:e1005470.

881 Gauthier J, Boulain H, van Vugt JJFA, Baudry L, Persyn E, Aury J-M, Noel B, Bretaudeau A,
882 Legeai F, Warris S, et al. 2021. Chromosomal scale assembly of parasitic wasp genome
883 reveals symbiotic virus colonization. *Commun. Biol.* 4:104.

884 Gauthier J, Gayral P, Le Ru BP, Jancek S, Dupas S, Kaiser L, Gyapay G, Herniou EA. 2018.
885 Genetic footprints of adaptive divergence in the bracovirus of *Cotesia sesamiae* identified by
886 targeted resequencing. *Mol. Ecol.* 27:2109–2123.

887 Guinet B, Lepetit D, Charlat S, Buhl PN, Notton DG, Cruaud A, Rasplus J-Y, Stigenberg J, De
888 Vienne DM, Boussau B, et al. 2023. Endoparasitoid lifestyle promotes endogenization and
889 domestication of dsDNA viruses. *eLife* 12:e85993.

890 Gundersen-Rindal D, Dupuy C, Huguet E, Drezen J-M. 2013. Parasitoid polydnaviruses:
891 evolution, pathology and applications: Dedicated to the memory of Nancy E. Beckage.
892 *Biocontrol Sci. Technol.* 23:1–61.

893 Heisserer C, Muller H, Jouan V, Musset K, Periquet G, Drezen J-M, Volkoff A-N, Gilbert C.
894 2023. Massive Somatic and Germline Chromosomal Integrations of Polydnaviruses in
895 Lepidopterans. *Mol. Biol. Evol.* 40:msad050.

896 Hellers M, Beck M, Theopoid U, Kamei M, Schmidt O. 1996. Multiple alleles encoding a virus-
897 like particle protein in the ichneumonid endoparasitoid *Venturia canescens*. *Insect Mol. Biol.*
898 5:239–249.

899 Herniou EA, Huguet E, Thézé J, Bézier A, Periquet G, Drezen J-M. 2013. When parasitic wasps
900 hijacked viruses: genomic and functional evolution of polydnaviruses. *Philos. Trans. R. Soc. B*
901 *Biol. Sci.* 368:20130051.

902 Holmes EC. 2011. The Evolution of Endogenous Viral Elements. *Cell Host Microbe* 10:368–
903 377.

904 Inwood SN, Harrop TWR, Dearden PK. 2023. The venom composition and parthenogenesis
905 mechanism of the parasitoid wasp Microctonus hyperodae, a declining biocontrol agent.
906 *Insect Biochem. Mol. Biol.* 153:103897.

907 Irwin NAT, Pittis AA, Richards TA, Keeling PJ. 2021. Systematic evaluation of horizontal gene
908 transfer between eukaryotes and viruses. *Nat. Microbiol.* 7:327–336.

909 Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices
910 from protein sequences. *Bioinformatics* 8:275–282.

911 Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment,
912 interactive sequence choice and visualization. *Brief. Bioinform.* 20:1160–1166.

913 Katzourakis A, Gifford RJ. 2010. Endogenous Viral Elements in Animal Genomes.Malik HS,
914 editor. *PLoS Genet.* 6:e1001191.

915 Kikhno I, Gutiérrez S, Croizier L, Croizier G, Ferber ML. 2002. Characterization of pif, a gene
916 required for the per os infectivity of Spodoptera littoralis nucleopolyhedrovirus. *J. Gen. Virol.*
917 83:3013–3022.

918 Kim J, Rahman M-M, Kim A-Y, Ramasamy S, Kwon M, Kim Y. 2023. Genome, host genome
919 integration, and gene expression in Diadegma fenestrale ichnovirus from the perspective of
920 coevolutionary hosts. *Front. Microbiol.* [Internet] 14. Available from:
921 https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1035669/ful
922 l

923 Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J,
924 Polevikov E, Smith TPL, et al. 2020. metaFlye: scalable long-read metagenome assembly using
925 repeat graphs. *Nat. Methods* 17:1103–1110.

926 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and
927 accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. *Genome*
928 *Res.* 27:722–736.

929 Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies.
930 *F1000Research* 6:1287.

931 Lapointe R, Popham HJR, Straschil U, Goulding D, O'Reilly DR, Olszewski JA. 2004.
932 Characterization of Two *Autographa californica* Nucleopolyhedrovirus Proteins, Ac145 and
933 Ac150, Which Affect Oral Infectivity in a Host-Dependent Manner. *J. Virol.* 78:6439–6448.

934 Lavialle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, Heidmann T. 2013.
935 Paleovirology of *'syncytins'*, retroviral env genes exapted for a role in placentation. *Philos.*
936 *Trans. R. Soc. B Biol. Sci.* 368:20120507.

Legeai F, Santos BF, Robin S, Bretaudeau A, Dikow RB, Lemaitre C, Jouan V, Ravallec M, Drezen J-M, Tagu D, et al. 2020. Genomic architecture of endogenous ichnoviruses reveals distinct evolutionary pathways leading to virus domestication in parasitic wasps. *BMC Biol.* 18:89.

Lemoine F, Gascuel O. 2021. Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. *NAR Genomics Bioinforma.* 3:lqab075.

Leobold M, Bézier A, Pichon A, Herniou EA, Volkoff A-N, Drezen J-M. 2018. The Domestication of a Large DNA Virus by the Wasp *Venturia canescens* Involves Targeted Genome Reduction through Pseudogenization.Abergel C, editor. *Genome Biol. Evol.* 10:1745–1764.

Li H. 2023. Protein-to-genome alignment with miniprot.Valencia A, editor. *Bioinformatics* 39:btad014.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13:2178–2189.

Liu H, Fu Y, Jiang D, Li G, Xie J, Cheng J, Peng Y, Ghabrial SA, Yi X. 2010. Widespread Horizontal Gene Transfer from Double-Stranded RNA Viruses to Eukaryotic Nuclear Genomes. *J. Virol.* 84:11876–11887.

Loni A, Samartsev KG, Scaramozzino PL, Belokobylskij SA, Lucchi A. 2016. Braconinae parasitoids (Hymenoptera, Braconidae) emerged from larvae of Lobesia botrana (Denis & Schiffermüller) (Lepidoptera, Tortricidae) feeding on Daphne gnidium L. *ZooKeys* 587:125–150.

Lorenzi A, Ravallec M, Eychenne M, Jouan V, Robin S, Darboux I, Legeai F, Gosselin-Grenet A-S, Sicard M, Stoltz D, et al. 2019. RNA interference identifies domesticated viral genes involved in assembly and trafficking of virus-derived particles in ichneumonid wasps.Strand MR, editor. *PLOS Pathog.* 15:e1008210.

Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. 2022. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes.Weigel D, editor. *eLife* 11:e78526.

Lucchi A, Ricciardi R, Loni A, Cosci F, Alvarez AR, Beeche M, Scaramozzino PL. 2018. Rearing Campoplex capitator Aubert in Italy and in Chile: preliminary achievements. *IOBCwprs Bull.* 139:96–99.

Mangeney M, Renard M, Schlecht-Louf G, Bouallaga I, Heidmann O, Letzelter C, Richaud A, Ducos B, Heidmann T. 2007. Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc. Natl. Acad. Sci.* 104:20534–20539.

Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: Assessing Genomic Data Quality and Beyond. *Curr. Protoc.* [Internet] 1. Available from: https://onlinelibrary.wiley.com/doi/10.1002/cpz1.323

Mao M, Simmonds TJ, Stouthamer CM, Kehoe TM, Geib SM, Burke GR. 2023. A chromosome scale assembly of the parasitoid wasp *Venturia canescens* provides insight into the process of virus domestication.Lawniczak M, editor. *G3 Genes Genomes Genet.* 13:jkad137.

978 Mao M, Stouthamer CM, Lorenzi A, Strand MR, Burke GR. 2025. Four nudivirus core genes
979 present in the genome of *Venturia canescens* are required for virus-like particle formation
980 and prevention of encapsulation of parasitoid wasp eggs.Van Oers MM, editor. *J.*
981 *Virol.*:e01305-25.

982 Maori E, Tanne E, Sela I. 2007. Reciprocal sequence exchange between non-retro viruses and
983 hosts leading to the appearance of new host phenotypes. *Virology* 362:342–349.

984 McLachlin JR, Miller LK. 1994. Identification and characterization of vlf-1, a baculovirus gene
985 involved in very late gene expression. *J. Virol.* 68:7746–7756.

986 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R.
987 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the
988 Genomic Era.Teeling E, editor. *Mol. Biol. Evol.* 37:1530–1534.

989 Moreau S, Asgari S. 2015. Venom Proteins from Parasitoid Wasps and Their Biological
990 Functions. *Toxins* 7:2385–2412.

991 Muller H, Chebbi MA, Bouzar C, Périquet G, Fortuna T, Calatayud P-A, Le Ru B, Obonyo J,
992 Kaiser L, Drezen J-M, et al. 2021. Genome-Wide Patterns of Bracovirus Chromosomal
993 Integration into Multiple Host Tissues during Parasitism.Parrish CR, editor. *J. Virol.*
994 95:e00684-21.

995 Murphy N, Banks JC, Whitfield JB, Austin AD. 2008. Phylogeny of the parasitic microgastroid
996 subfamilies (Hymenoptera: Braconidae) based on sequence data from seven genes, with an
997 improved time estimate of the origin of the lineage. *Mol. Phylogenet. Evol.* 47:378–395.

998 Pasquier-Barre F, Dupuy C, Huguet E, Monteiro F, Moreau A, Poirié M, Drezen J-M. 2002.
999 Polydnavirus replication: the EP1 segment of the parasitoid wasp Cotesia congregata is
1000 amplified within a larger precursor molecule. *J. Gen. Virol.* 83:2035–2045.

1001 Perkin LC, Friesen KS, Flinn PW, Oppert B. 2015. Venom gland components of the
1002 ectoparasitoid wasp, Anisopteromalus calandrae. *J. Venom Res.* 6:19–37.

1003 Petersen JM, Bézier A, Drezen J-M, van Oers MM. 2022. The naked truth: An updated review
1004 on nudiviruses and their relationship to bracoviruses and baculoviruses. *J. Invertebr. Pathol.*
1005 189:107718.

1006 Pichon A, Bézier A, Urbach S, Aury J-M, Jouan V, Ravallec M, Guy J, Cousserans F, Thézé J,
1007 Gauthier J, et al. 2015. Recurrent DNA virus domestication leading to different parasite
1008 virulence strategies. *Sci. Adv.* 1:e1501150.

1009 Poirié M, Colinet D, Gatti J-L. 2014. Insights into function and evolution of parasitoid wasp
1010 venoms. *Curr. Opin. Insect Sci.* 6:52–60.

1011 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
1012 features. *Bioinformatics* 26:841–842.

1013 Rasoolizadeh A, Béliveau C, Stewart D, Cloutier C, Cusson M. 2009. Tranosema rostrale
1014 ichnovirus repeat element genes display distinct transcriptional patterns in caterpillar and
1015 wasp hosts. *J. Gen. Virol.* 90:1505–1514.

1016 Reineke A, Asgari S, Ma G, Beck M, Schmidt O. 2002. Sequence analysis and expression of a
1017 virus-like particle protein, VLP2, from the parasitic wasp *Venturia canescens*. *Insect Mol. Biol.*
1018 11:233–239.

1019 Reineke A, Asgari S, Schmidt O. 2006. Evolutionary origin of *Venturia canescens* virus-like
1020 particles. *Arch. Insect Biochem. Physiol.* 61:123–133.

1021 Rotheram S. 1967. Immune Surface of Eggs of a Parasitic Insect. *Nature* 214:700–700.

1022 Rotheram S. 1973. The surface of the egg of a parasitic insect. I. The surface of the egg and
1023 first instar larva of *Nemeritis*. *Proc. R. Soc. Lond. B Biol. Sci.* 183:179–194.

1024 Salt G. 1965. Experimental studies in insect parasitism XIII. The haemocytic reaction of a
1025 caterpillar to eggs of its habitual parasite. *Proc. R. Soc. Lond. B Biol. Sci.* 162:303–318.

1026 Sayyari E, Mirarab S. 2016. Fast Coalescent-Based Computation of Local Branch Support from
1027 Quartet Frequencies. *Mol. Biol. Evol.* 33:1654–1668.

1028 Scaramozzino PL, Di Giovanni F, Loni A, Ricciardi R, Lucchi A. 2018. Updated list of the insect
1029 parasitoids (Insecta, Hymenoptera) associated with Lobesia botrana (Denis &
1030 Schiffermüller, 1775) (Lepidoptera, Tortricidae) in Italy. 2. Hymenoptera, Ichneumonidae,
1031 Anomaloninae and Campopleginae. *ZooKeys* 772:47–95.

1032 Sharanowski BJ, Ridenbaugh RD, Piekarski PK, Broad GR, Burke GR, Deans AR, Lemmon AR,
1033 Moriarty Lemmon EC, Diehl GJ, Whitfield JB, et al. 2021. Phylogenomics of Ichneumonoidea
1034 (Hymenoptera) and implications for evolution of mode of parasitism and viral
1035 endogenization. *Mol. Phylogenet. Evol.* 156:107023.

1036 Singh U, Wurtele ES. 2021. orfipy: a fast and flexible tool for extracting ORFs.Birol I, editor.
1037 *Bioinformatics* 37:3019–3020.

1038 Strand MR, Burke GR. 2012. Polydnaviruses as Symbionts and Gene Delivery Systems.Condit
1039 RC, editor. *PLoS Pathog.* 8:e1002757.

1040 Strand MR, Burke GR. 2014. Polydnaviruses: Nature's Genetic Engineers. *Annu. Rev. Virol.*
1041 1:333–354.

1042 Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence
1043 alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.

1044 Tanaka K, Lapointe R, Barney WE, Makkay AM, Stoltz D, Cusson M, Webb BA. 2007. Shared
1045 and species-specific features among ichnovirus genomes. *Virology* 363:26–35.

1046 Tarlinton RE, Meers J, Young PR. 2006. Retroviral invasion of the koala genome. *Nature*
1047 442:79–81.

1048 Theopold U, Krause E, Schmidt O. 1994. Cloning of a VLP-protein coding gene from a
1049 Parasitoid Wasp *Venturia canescens*. *Arch. Insect Biochem. Physiol.* 26:137–145.

1050 Thézé J, Bézier A, Periquet G, Drezen J-M, Herniou EA. 2011. Paleozoic origin of insect large
1051 dsDNA viruses. *Proc. Natl. Acad. Sci.* 108:15931–15935.

1052 Vanarsdall AL, Okano K, Rohrmann GF. 2006. Characterization of the Role of Very Late
1053 Expression Factor 1 in Baculovirus Capsid Structure and DNA Processing. *J. Virol.* 80:1724–
1054 1733.

1055 Volkoff A-N, Béliveau C, Rocher J, Hilgarth R, Levasseur A, Duonor-Cérutti M, Cusson M,
1056 Webb BA. 2002. Evidence for a Conserved Polydnavirus Gene Family: Ichnovirus Homologs of
1057 the CslV *Repeat Element* Genes. *Virology* 300:316–331.

1058 Volkoff A-N, Cusson M. 2020. The Unconventional Viruses of Ichneumonid Parasitoid Wasps.
1059 *Viruses* 12:1170.

1060 Volkoff A-N, Drezen J-M, Cusson M, Webb BA. 2012. The Organization of Genes Encoding
1061 Ichnovirus Structural Proteins. In: Parasitoid Viruses. Elsevier. p. 33–45. Available from:
1062 https://linkinghub.elsevier.com/retrieve/pii/B9780123848581000035

1063 Volkoff A-N, Jouan V, Urbach S, Samain S, Bergoin M, Wincker P, Demettre E, Cousserans F,
1064 Provost B, Coulibaly F, et al. 2010. Analysis of Virion Structural Components Reveals Vestiges
1065 of the Ancestral Ichnovirus Genome.Asgari S, editor. *PLoS Pathog.* 6:e1000923.

1066 Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, Guo P, Dunn CW, Jones BR, Bradley T, et al.
1067 2020. Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated
1068 and Associated Data.Kumar S, editor. *Mol. Biol. Evol.* 37:599–603.

1069 Webb BA, Strand MR, Dickey SE, Beck MH, Hilgarth RS, Barney WE, Kadash K, Kroemer JA,
1070 Lindstrom KG, Rattanadechakul W, et al. 2006. Polydnavirus genomes reflect their dual roles
1071 as mutualists and pathogens. *Virology* 347:160–174.

1072 Yang S, Miller LK. 1998. Expression and Mutational Analysis of the Baculovirus Very Late
1073 Factor 1 (vlf-1) Gene. *Virology* 245:99–109.

1074 Yang S, Miller LK. 1999. Activation of Baculovirus Very Late Promoters by Interaction with
1075 Very Late Factor 1. *J. Virol.* 73:3404–3409.

1076 Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an r package for visualization and
1077 annotation of phylogenetic trees with their covariates and other associated data. *Methods*
1078 *Ecol. Evol.* 8:28–36.

1079 Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree
1080 reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19:153.

1081 Zheng Q, Shen Y, Kon X, Zhang J, Feng M, Wu X. 2017. Protein–protein interactions of the
1082 baculovirus per os infectivity factors (PIFs) in the PIF complex. *J. Gen. Virol.* 98:853–861.

1083

1084

1085

1086

1087

## 1088 Tables and Figures

1089 Table 1: Nudivirus genes in the *C. capitator, C. nolae*, *V. canescens* and *R. megacephalus*
1090 genomes. Genes found in *C. capitator* (in orange), *C. nolae* (in blue) and *R. megacephalus* (in
1091 purple) were compared with those of VcENV (in green). Genes and pseudogenes are
1092 indicated with solid dots and crossed dots respectively. Copies are indicated by multiple
1093 dots. Encircled solid dots indicate genes for which proteins were found in VLPs (CcapiVLPs
1094 and VcVLPs). "na" indicates that no copy of the corresponding gene was found. Alternative
1095 gene names are given between parentheses. A more detailed table is given in Table S2.

1096

1097 Figure 1: History of virus integration in Campopleginae. The left tree represents the
1098 Campopleginae and Ophioninae phylogenetic tree, built using hymenopteran BUSCOs
1099 extracted from NCBI genome assemblies. Branch topology has been inferred by coalescent
1100 modelling (ASTRAL) while branch lengths were inferred using a maximum likelihood
1101 approach. Branch support is given by local posterior probabilities, i.e. the probability that a
1102 branch is the true branch given the set of trees built using BUSCOs (computed based on the
1103 quartet score (Sayyari and Mirarab 2016)). A more complete tree of ichneumonid wasps is
1104 available in Figure S1. The right tree represents a portion of the Ichneumonidae tree
1105 corresponding to Campopleginae wasps. Solid green dots indicate species for which *rep*
1106 genes were found, while empty green circles indicate that *rep* gene presence needs to be
1107 investigated. We can distinguish two clades: one with an integrated ichnovirus (bottom
1108 part, in green), and one with *Venturia canescens*, which bears an integrated nudivirus (top

1109 part, in darker orange). The ichnovirus endogenized in *Campopleginae sp.* genome can be
1110 found in Guinet et al. 2023. Species closely related to *V. canescens* for which we do not know
1111 whether they bear a virus in their genome are mentioned in orange. So far, no viral
1112 integration has been observed in *Dusona spp.*. Two scenarios explaining viral integrations in
1113 Campopleginae can be envisaged: 1- The ichnoviruses and nudiviruses were acquired in two
1114 different branches of the Campopleginae tree. In this scenario, the nudiviral and ichnoviral
1115 sequences should not be found together and the two viruses did not interact with each
1116 other after integration. 2- An initial integration of ichnovirus took place in the common
1117 ancestor of all these species, followed by the integration of a nudivirus and certain branches
1118 of the tree. In this scenario, the ichnovirus could have been replaced by the nudivirus as a
1119 result of selection, and it would be possible to find both ichnoviral and nudiviral sequences,
1120 provided that the non-conserved sequences can still be characterised.

1121

1122 Figure 2: The genome of CcapiENV. **A** - The nudiviral clusters of CcapiENV. Sequences of
1123 nudiviral origin are indicated by a red asterisk in the legend. Nudiviral genes were colourised
1124 according to their function described in baculoviruses. Cellular genes correspond to genes of
1125 non-nudiviral origin (with no nudiviral hit). The scale represents the cluster length in base
1126 pairs. **B** - Expression heatmap of CcapiENV genes. Expression levels are represented by log
1127 TPMs (transcripts per million) for each organ and each replicate. The wasp genes *vmp03*
1128 (venom metalloproteinase 3), *kh* (kynurenine 3-monooxygenase) and *otu* (ovarian tumor)
1129 were added as expression controls for venom glands, head-thorax and ovaries respectively.
1130 The Spearman correlation heatmap of the transcriptome samples is available in Figure S3. An

1131 Analysis of the potential regulatory sequences of CcapiENV genes is also available in
1132 supplementary materials.

1133

1134 Figure 3: Evolution of endogenized viruses. **A** - Phylogenetic tree of Naldaviricetes viruses. In
1135 total, 34 genes (21 baculovirus core genes and 13 alphanudivirus core genes) were used for
1136 the tree construction, using the maximum likelihood criterion. Branch supports are given by
1137 the SH-aLRT (Shimodaira-Hasegawa approximate likelihood ratio test) and bootstrap values.
1138 The scale bar represents the phylogenetic distance (percentage of genetic variation).
1139 Campoplex capitator Endogenous Nudivirus, Campoplex nolae Endogenous Nudivirus and
1140 Rhimphoctona megacephalus Endogenous Nudivirus are called respectively CcapiENV,
1141 CnolaENV and RmegaENV. Nudiviruses endogenized in Campopleginae wasps are mentioned
1142 in red. Virus and sequence information is given in supplementary data (Table S4). **B** -
1143 Syntenies between the CcapiENV clusters, VcENV clusters (from Mao et al. 2023) and
1144 RmegaENV clusters. Models were colourised according to their function. Light pink models
1145 correspond to genes annotated in Mao and collaborators (Mao et al. 2023) which have no
1146 match in the NCBI database, but have nevertheless been annotated as potential virus genes.
1147 The scale represents the length in base pairs. Blue tildes on RmegaENV clusters represent
1148 genome portions that are too big to be displayed. Genome macrosyntenies can be observed
1149 in Figure S4. **C** - Parity plot between the genomes of Oryctes rhinoceros nudivirus (OrNV) and
1150 other exogenous nudiviruses. The other nudiviruses are Gryllus bimaculatus nudivirus
1151 (GbNV), Kallithea virus (KV) Drosophila innubia nudivirus (DiNV), Esparto virus (ENV) and
1152 Tomelloso virus (TNV). **D** - Parity plot between the genomes of CcapiENV and the previous
1153 exogenous nudiviruses. **E** - Parity plot between the genomes of CcapiENV and other

1154 endogenized nudiviruses. The other endogenized nudiviruses are Nilaparvata lugens
1155 endogenous nudivirus (NlENV), Fopius arisanus endogenous nudivirus (FaENV), VcENV,
1156 CnolaENV and RmegaENV.

1157

1158 Figure 4: Phylogenetic tree of REP sequences across insects. The tree has been built using
1159 229 proteins using the maximum likelihood approach. Bootstrap support values are
1160 indicated by node color. For each type of campoplegine wasp associated with an
1161 endogenous nudivirus or ichnovirus, tip label colours indicate in which species REP proteins
1162 have been identified. The pseudogenized REP identified in *Dusona sp.* is shown in red and is
1163 not associated with an active endogenous virus. Silhouettes are showing horizontal transfer
1164 of rep genes in lepidopterans, sawflies and phasmids. Red dots at the leaves indicate genes
1165 for which corresponding proteins have been found in CcapiVLPs.

1166

1167 Figure 5: The calyx region of a *C. capitator* ovaries of adult wasp observed using light and
1168 electron microscopy. **A** - light microscopic photograph of a semi-thin section of the calyx
1169 region, stained with toluidine blue. **B**, **D** - electron microscopy photographs of the calyx
1170 region shown in A. **C**, **E** - enlarged areas (white frames) of photographs B and D, respectively.
1171 Virogenic stromas (VS) and VLPs (dark arrow) are visible in cell nuclei (N). While virogenic
1172 stromas can be seen only in the nuclei, VLPs appear to travel from the nucleus to the lumen
1173 (L) where they accumulate, passing through the cytoplasm (C) and the membrane microvilli
1174 (MV). **F** - electron microscopy photograph showing the nucleus-cytoplasm-lumen interface

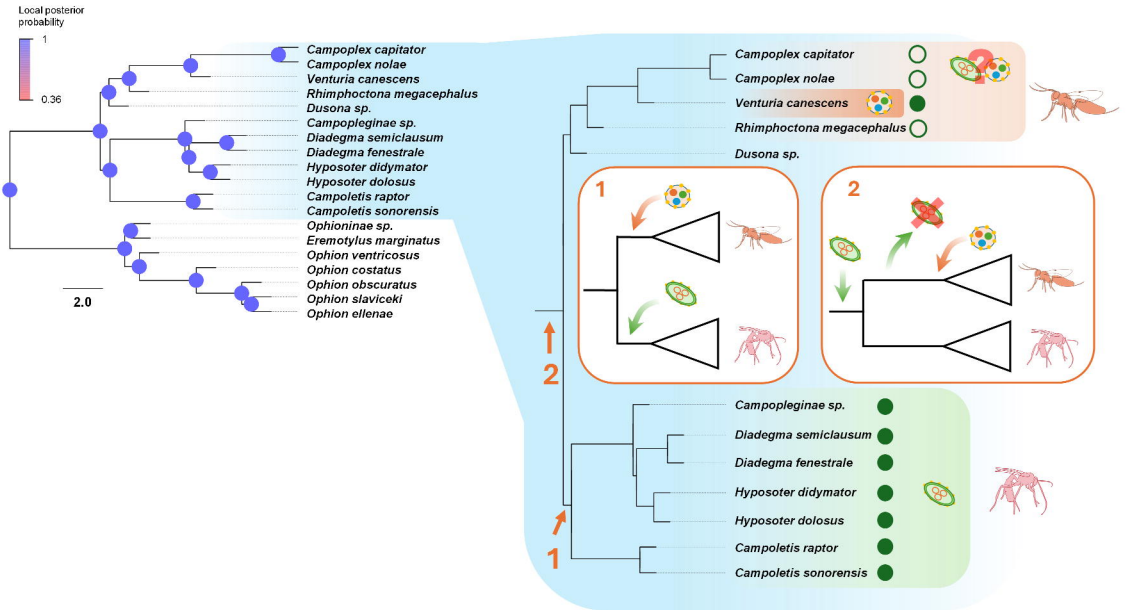
1175 where empty VLP envelopes can be seen in the nucleus (white circle), as described in *V.*
1176 *canescens* (Pichon et al. 2015).

1177

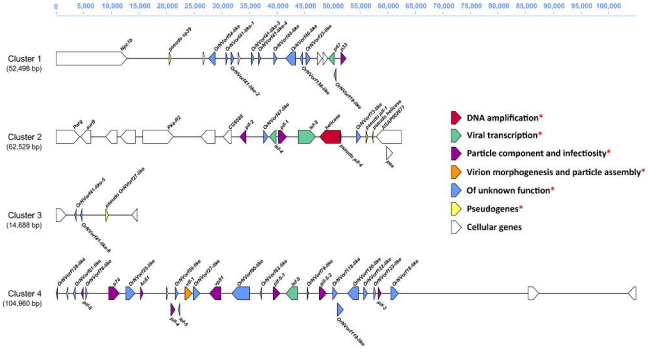
1178 Figure 6: Evolution of endogenized nudivirual genes in Campopleginae. **A** - Protein
1179 composition of VLP found using mass spectrometry. The protein content of VLPs is given in
1180 the blue panels. In green are mentioned the nudiviral genes involved, or likely to be
1181 involved, in VLP production whose proteins are not contained in VLPs. Main differences
1182 between *V. canescens* and *C. capitator* are indicated in bold. Result details of mass
1183 spectrometry are given in Table S5. **B** - Estimation of selective pressures acting on wasp and
1184 endogenized nudivirus genes. dN/dS ratio values were calculated from wasp (in grey) and
1185 nudivirus genes of *C. capitator*, *C. nolae*, *V. canescens* and *R. megacephalus*. Numbers below
1186 each boxplot represent the number of sequences used for dN/dS ratio calculation for each
1187 category. Numbers above each boxplot represent the number of values higher than 2, which
1188 could not be plotted. In red are the dN/dS ratio values for *helicase*, *p47* and *lef* genes, in
1189 green are the values for *pif* (*per-os-infectivity factor*) genes, in pink the values for *Ac81*, *vp91*
1190 and *p33* genes, and in blue the values for *OrNVorf-like* genes. See Table S6 for details.

1191

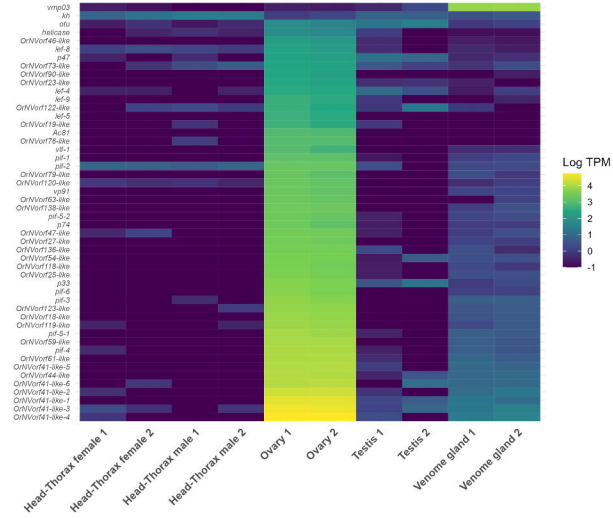
| Function in baculoviruses | Gene name | Presence in | | | |
|---|---|---|---|---|---|
| | | CcapiENV | CnolaENV | VcENV | RmegaENV |
| DNA amplification | DNApol | ⊗ | ⊗ | ⊗ | ⊗ |
| | FEN-1 | (⊗) | ⊗ | (⊗) | na |
| | helicase | •(•) | • | • | •(•) |
| | helicase2 | na | na | na | na |
| | Integrase | ⊗ | ⊗ | ⊗ | •••• |
| Viral transcription | lef-4 | • | • | •⊗ | • |
| | lef-5 | • | • | • | • |
| | lef-8 | •(•) | • | • | ••⊗ |
| | lef-9 | • | • | • | • |
| | p47 | • | • | • | • |
| Envelope components, infectiosity | ac81 | • | • | ••• | • |
| | **p33 (ac92)** | (•) | • | (•) | • |
| | **p74 (pif-0)** | (•) | • | (•) | ••• |
| | **pif-1** | (•)⊗ | • | (•)⊗ | ••••⊗⊗ |
| | **pif-2** | (•) | • | (•) | • |
| | **pif-3** | (•) | • | (•) | •• |
| | **pif-4 (19 Kda)** | (•)⊗ | •⊗ | (•) | •• |
| | **pif-5 (odv-e56)** | (•)(•) | •• | (•)(•)(•) | • |
| | **pif-6 (ac68)** | (•) | • | (•) | •• |
| | **vp91(pif-8)** | (•) | • | (•) | •(•) |
| Morphogenesis, particle assembly | 38K | (•) | na | (•) | na |
| | vlf-1 | • | • | (•) | na |
| | vp39 | (•) | ⊗ | (•) | na |
| | OrNVorf22-like (p6.9) | (•) | ⊗ | (•) | • |
| Nucleotid metabolism | tk1 | na | na | na | na |
| | tk2 | na | na | na | na |
| | tk3 | na | na | na | na |
| Unknown | OrNVorf2-like | na | na | ⊗ | na |
| | OrNVorf8-like | na | na | na | • |
| | OrNVorf9-like | na | ⊗ | ⊗ | na |
| | OrNVorf12-like (odv-e66) | na | na | na | • |
| | **OrNVorf18-like** | (•) | • | • | ••(•) |
| | OrNVorf19-like | • | • | • | • |
| | **OrNVorf23-like** | (•) | • | (•) | • |
| | OrNVorf24-like | na | na | na | • |
| | **OrNVorf25-like** | (•) | • | (•) | • |
| | **OrNVorf27-like** | (•)(•) | •(•) | (•) | • |
| | **OrNVorf41-like (11K)** | (•)(•)(•)••• | •••••• | (•)(•)(•)••• | ••• |
| | **OrNVorf44-like** | (•) | • | (•) | • |
| | **OrNVorf46-like** | • | • | (•) | • |
| | **OrNVorf47-like** | (•) | • | (•)•• | •⊗ |
| | OrNVorf53-like | na | na | na | •• |
| | **OrNVorf54-like** | (•) | • | (•) | • |
| | **OrNVorf59-like** | (•) | •⊗ | (•) | • |
| | **OrNVorf61-like** | (•) | • | (•) | na |
| | OrNVorf62-like | ⊗ | ⊗ | ⊗ | na |
| | OrNVorf63-like | • | • | • | • |
| | OrNVorf73-like | • | • | • | •(•) |
| | **OrNVorf76-like** | (•) | • | • | • |
| | **OrNVorf79-like** | (•) | • | • | • |
| | **OrNVorf90-like** | (•) | • | • | • |
| | OrNVorf95-like | na | na | na | (•)(•) |
| | OrNVorf99-like | (•) | ⊗ | (•) | na |
| | OrNVorf117-like | (•) | ⊗ | (•) | na |
| | **OrNVorf118-like** | (•) | • | (•) | • |
| | **OrNVorf119-like** | (•) | • | (•) | • |
| | **OrNVorf120-like** | (•) | • | • | • |
| | OrNVorf121-like | na | na | na | na |
| | OrNVorf122-like | • | • | • | • |
| | **OrNVorf123-like** | (•) | • | (•) | • |
| | OrNVorf124-like | na | na | na | ⊗ |
| | OrNVorf128-like | ⊗ | ⊗ | ⊗ | na |
| | OrNVorf130-like | na | na | ⊗ | na |
| | OrNVorf136-like | • | • | • | • |
| | **OrNVorf138-like** | (•) | • | • | • |
| | OrNVorf137-like | na | na | na | • |
| | OrNVorf139-like | ⊗ | ⊗ | ⊗ | na |

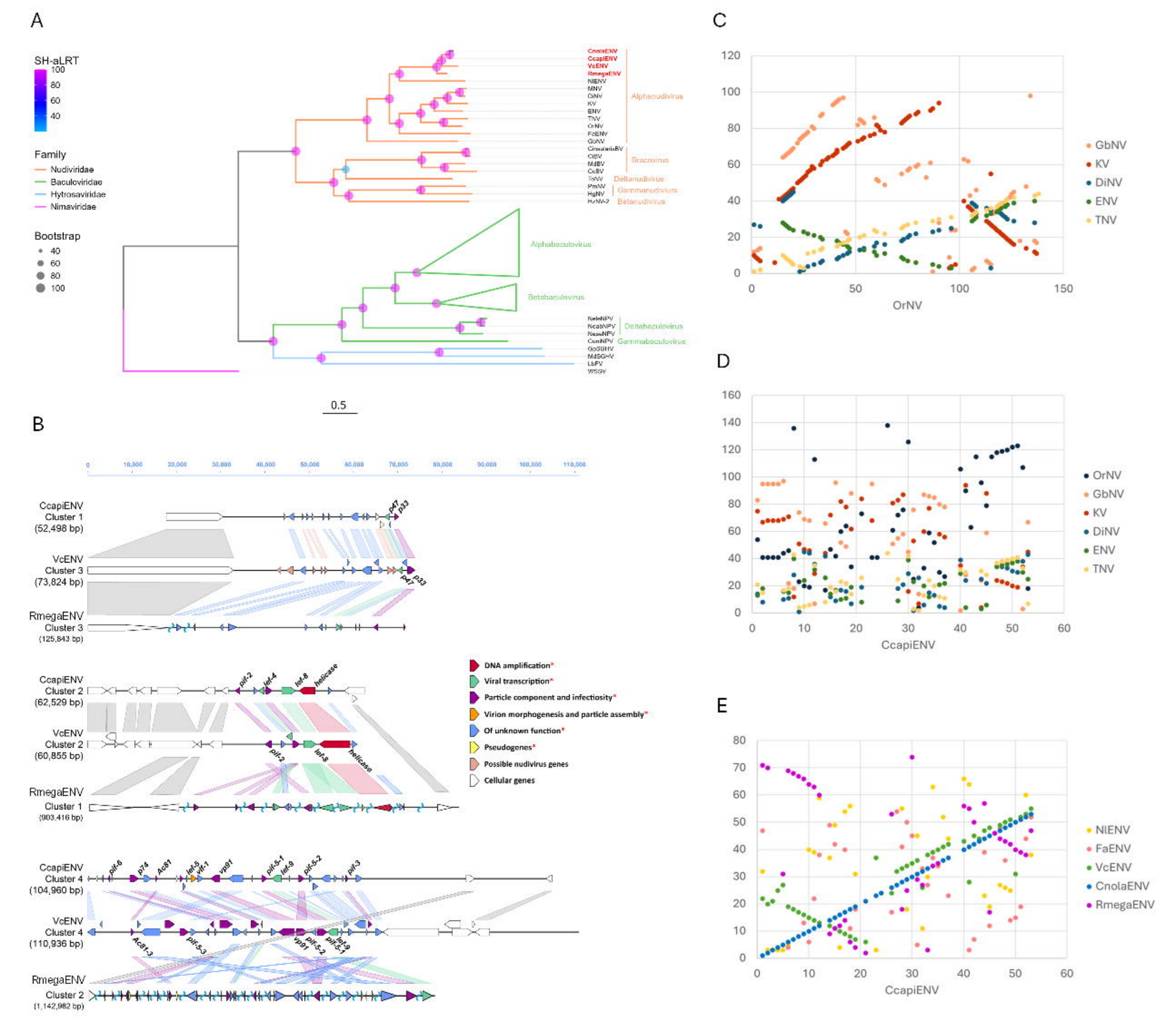

Local posterior probability
1
0.36
Campoplex capitator
Campoplex nolae
Venturia canescens
Rhimphoctona megacephalus
Dusona sp.
Campopleginae sp.
Diadegma semiclausum
Diadegma fenestrale
Hyposoter didymator
Hyposoter dolosus
Campoletis raptor
Campoletis sonorensis
Ophioninae sp.
Eremotylus marginatus
Ophion ventricosus
Ophion costatus
Ophion obscuratus
Ophion slaviceki
Ophion ellenae
2.0
Campoplex capitator
Campoplex nolae
Venturia canescens
Rhimphoctona megacephalus
Dusona sp.
2
1
2
Campopleginae sp.
Diadegma semiclausum
Diadegma fenestrale
Hyposoter didymator
Hyposoter dolosus
Campoletis raptor
Campoletis sonorensis
1

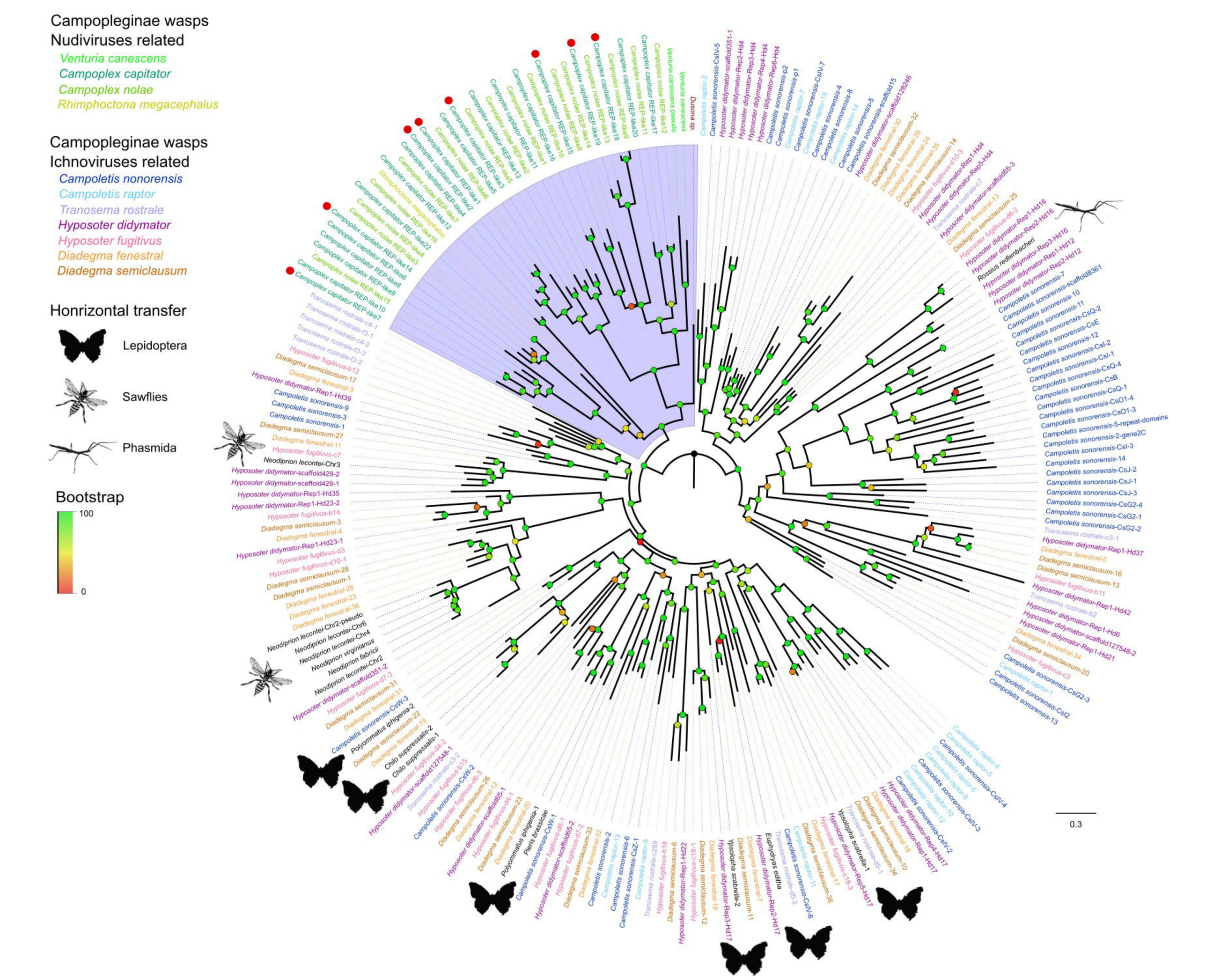
A

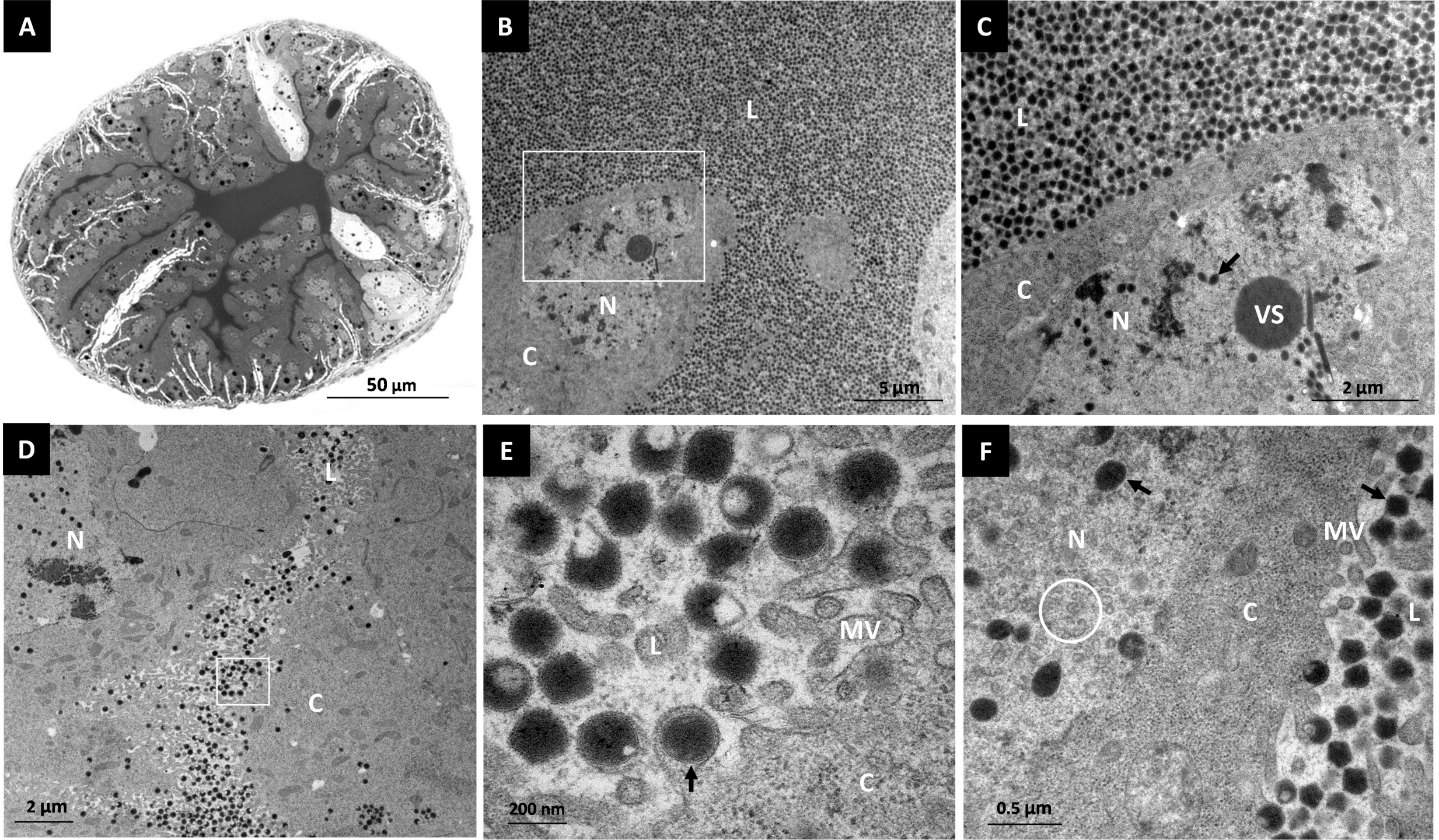

0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
45,000
50,000
55,000
60,000
65,000
70,000
75,000
80,000
85,000
90,000
95,000
100,000
Cluster 1
(52,496 bp)
Cluster 2
(62,529 bp)
Cluster 3
(14,688 bp)
Cluster 4
(104,960 bp)
DNA amplification*
Viral transcription*
Particle component and infectiosity*
Virion morphogenesis and particle assembly*
Of unknown function*
Pseudogenes*
Cellular genes

B


vmp03
kh
otu
helicase
OrNVorf46-like
lef-8
p47
OrNVorf73-like
OrNVorf90-like
OrNVorf23-like
lef-4
lef-9
OrNVorf122-like
lef-5
OrNVorf19-like
Ac81
OrNVorf76-like
vlf-1
pif-1
pif-2
OrNVorf179-like
OrNVorf120-like
vp91
OrNVorf63-like
OrNVorf138-like
pif-5-2
p74
OrNVorf47-like
OrNVorf27-like
OrNVorf136-like
OrNVorf54-like
OrNVorf118-like
OrNVorf25-like
p33
pif-6
pif-3
OrNVorf123-like
OrNVorf18-like
OrNVorf119-like
pif-5-1
OrNVorf59-like
pif-4
OrNVorf61-like
OrNVorf41-like-5
OrNVorf44-like
OrNVorf41-like-6
OrNVorf41-like-2
OrNVorf41-like-1
OrNVorf41-like-3
OrNVorf41-like-4
Head-Thorax female 1
Head-Thorax female 2
Head-Thorax male 1
Head-Thorax male 2
Ovary 1
Ovary 2
Testis 1
Testis 2
Venome gland 1
Venome gland 2
Log TPM
4
3
2
1
0
-1


A
SH-aLRT
100
80
60
40
Family
Nudiviridae
Baculoviridae
Hytrosaviridae
Nimaviridae
Bootstrap
40
60
80
100
CnolaENV
CcapiENV
VcENV
RmegaENV
NlENV
MNV
DiNV
KV
ENV
TNV
OrNV
FaENV
GbNV
CinsularisBV
CfBV
MdBV
CcBV
ToNV
PmNV
HgNV
HzNV-2
Alphanudivirus
Bracovirus
Deltanudivirus
Gammanudivirus
Betanudivirus
Alphabaculovirus
Betabaculovirus
NeleNPV
NcabNPV
NaseNPV
CunNPV
GpSGHV
MdSGHV
LbFV
WSSV
Deltabaculovirus
Gammabaculovirus
0.5
B
0
10,000
20,000
30,000
40,000
50,000
60,000
70,000
80,000
90,000
100,000
110,000
CcapiENV
Cluster 1
(52,498 bp)
p47
p33
VcENV
Cluster 3
(73,824 bp)
RmegaENV
Cluster 3
(125,843 bp)
CcapiENV
Cluster 2
(62,529 bp)
pif-2
lef-4
lef-8
helicase
VcENV
Cluster 2
(60,855 bp)
RmegaENV
Cluster 1
(903,416 bp)
CcapiENV
Cluster 4
(104,960 bp)
pif-6
p74
Ac81
lef-5
vif-1
vp91
pif-5-1
lef-9
pif-5-2
pif-3
VcENV
Cluster 4
(110,936 bp)
Ac81-3
pif-5-3
RmegaENV
Cluster 2
(1,142,982 bp)
DNA amplification*
Viral transcription*
Particle component and infectiosity*
Virion morphogenesis and particle assembly*
Of unknown function*
Pseudogenes*
Possible nudivirus genes
Cellular genes
C
OrNV
GbNV
KV
DiNV
ENV
TNV
D
CcapiENV
OrNV
GbNV
KV
DiNV
ENV
TNV
E
CcapiENV
NlENV
FaENV
VcENV
CnolaENV
RmegaENV


Campopleginae wasps
Nudiviruses related
Venturia canescens
Campoplex capitator
Campoplex nolae
Rhimphoctona megacephalus
Campopleginae wasps
Ichnoviruses related
Campoletis nonorensis
Campoletis raptor
Tranosema rostrale
Hyposoter didymator
Hyposoter fugitivus
Diadegma fenestral
Diadegma semiclausum
Honrizontal transfer
Lepidoptera
Sawflies
Phasmida
Bootstrap
100
0
0.3


A
50 µm
B
L
N
C
5 µm
C
L
C
N
VS
2 µm
D
L
N
C
2 µm
E
L
MV
C
200 nm
F
N
MV
C
L
0.5 µm

A

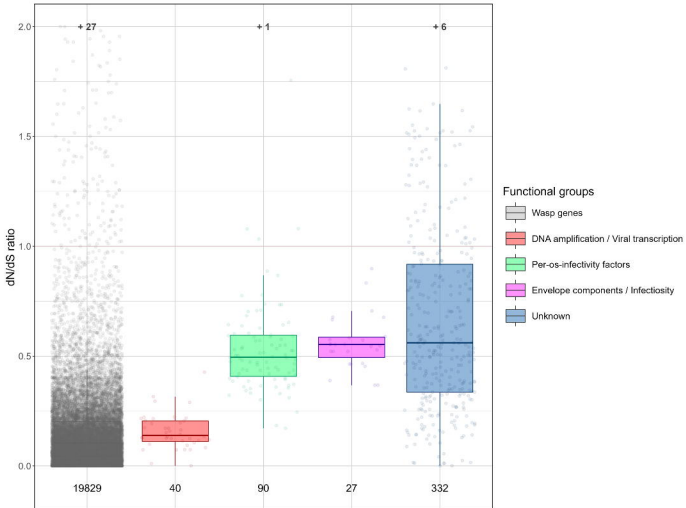
*Venturia canescens*

| DNA amplification | Viral transcription |
|---|---|
| *helicase* | *tef-4* |
| | *tef-5* |
| **Particle morphogenesis and envelope components** | *tef-8* |
| | *tef-9* |
| | *p47* |
| *ac81* | |

| Envelop components | Unknown OrNVorf-like |
|---|---|
| PIF-1 | ORF23 |
| PIF-2 | ORF25 |
| PIF-3 | ORF27 |
| PIF-4 | ORF41 (x3) |
| PIF-5 (x3) | ORF44 |
| PIF-6 | **ORF46** |
| VP91 | ORF47 |
| P33 | ORF54 |
| | ORF49 |
| | ORF61 |
| | ORF118 |
| | ORF119 |
| | ORF123 |

*Campoplex capitator*

| DNA amplification | Viral transcription |
|---|---|
| *helicase* | *tef-4* |
| | *tef-5* |
| | *tef-8* |
| **Particle morphogenesis and envelope components** | *tef-9* |
| | *p47* |
| ***vlf-1*** | |
| *ac81* | |

| Envelop components | Unknown OrNVorf-like |
|---|---|
| PIF-1 | **ORF18** |
| PIF-2 | ORF23 |
| PIF-3 | ORF25 |
| PIF-4 | ORF27 |
| PIF-5 (x2) | ORF41 (x4) |
| PIF-6 | ORF44 |
| VP91 | ORF47 |
| P33 | ORF54 |
| | ORF49 |
| | ORF61 |
| | **ORF76** |
| | **ORF79** |
| | **ORF90** |
| | ORF116 |
| | ORF119 |
| | **ORF120** |
| | ORF123 |
| | **ORF138** |

B