Chelonus inanitus bracovirus encodes lineage-specific proteins and truncated immune IκB-like factors

Alexandra Cerqueira de Araujo, Thibaut Josse, Vonick Sibut, Mariko Urabe, Azam Asadullah, Valérie Barbe, Madoka Nakai, Elisabeth Huguet, Georges Periquet, Jean-Michel Drezen

Title: Chelonus inanitus bracovirus encodes lineage specific proteins and truncated immune IκB-like factors

Author names: Alexandra Cerqueira de Araujo[1], Thibaut Josse[1], Vonick Sibut[1], Mariko Urabe[2], Azam Asadullah[2], Valérie Barbe[3], Madoka Nakai[2], Elisabeth Huguet[1], Georges Periquet[1], Jean-Michel Drezen[1]

Affiliation:

[1]Institut de Recherche sur la Biologie de l'Insecte (IRBI), UMR 7261, CNRS - Université de Tours, Tours, France,

[2]Graduate School of Agriculture, Tokyo University of Agriculture and Technology, Tokyo 183-8509, Japan.

[3]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France.

Corresponding author: drezen@univ-tours.fr

Author Notes

3 supplementary tables and 5 supplementary figures are available with the online version of this article.

Keywords: polydnavirus, bracovirus, ichnovirus, cactus, IκB, parasitoid wasp, host-parasite arms race

Repositories: CiBV Sequences are available in Genbank: accession numbers ON351504 to ON351527

Abstract

Bracoviruses and Ichnoviruses are endogenous viruses of parasitic wasps that produce particles containing virulence genes expressed in host tissues and necessary for parasitism success. In the case of bracoviruses the particles are produced by conserved genes of nudiviral origin integrated permanently in the wasp genome, whereas the virulence genes can strikingly differ depending on the wasp lineage. To date most data obtained on bracoviruses concerned species from the braconid subfamily of Microgastrinae. To gain a broader view on the diversity of virulence genes we sequenced the genome packaged in the particles of Chelonus inanitus bracovirus (CiBV) produced by a wasp belonging to a different subfamily: the Cheloninae. These are egg-larval parasitoids which means that they oviposit into the host egg and the wasp larvae then develop within the larval stages of the host. We found that most of CiBV virulence genes belong to families that are specific to Cheloninae. As other bracoviruses and ichnoviruses however, CiBV encode *v-ank* genes encoding truncated versions of the immune cactus/IκB factor, which suggests these proteins might play a key role in host-parasite interactions involving domesticated endogenous viruses. We found that the structures of CiBV V-ANKs are different from those previously reported. Phylogenetic analysis supports the hypothesis that they may originate from a cactus/IκB immune gene from the wasp genome acquired by the bracovirus. However, their evolutionary history is different from that shared by other V-ANKs, whose common origin probably reflect horizontal gene transfer events of virus sequences between braconid and ichneumonid wasps.

Introduction

PolyDNAviruses (PDVs) are endogenous DNA viruses of parasitoid wasps that are used as tools to transfer genes into their host, in most cases a lepidopteran larva. Virus particles are produced in wasp ovaries from viral sequences integrated into wasp chromosomes [1, 2]. These particles contain a segmented genome made of multiple dsDNA circular molecules, a unique organization for viruses, which inspired their name “polyDNAvirus” [3]. Unlike most viruses that replicate in the cells after infection, these viruses have a life cycle split between two organisms: the wasp and the parasitized host. The replicative phase occurs in specialized cells in the wasp ovaries, where virus particles are produced. These particles consist of DNA circles packaged in nucleocapsids embedded in a protein matrix and surrounded by an envelope [4]. Depending on the wasp species, each particle contains one [5] or several nucleocapsids [6]). The infective phase begins by the introduction of the virus particles along with wasp eggs into

the parasitized caterpillar, followed by the entry of particles in host cells and release of viral DNA in their nuclei [7]. The genome packaged in the particles is then expressed by infected cells, which results in the production of virulence factors by the host cellular machinery [8, 9]. These factors induce a manipulation of parasitized host physiology, including an alteration of host immune defense, allowing successful development of the wasp progeny within the host body [10-12]. Unlike for most viruses, there is no particle replication in the infected tissues because replication genes are not present in the packaged genomes and reside permanently in the wasp genome. The virus is thus exclusively transmitted vertically as a part of the wasp genome. Obviously, polydnaviruses are very particular viruses and ICTV has therefore recently reclassified them as "viroforms" [13]. However, since they have retained many features of their virus ancestor, we recently proposed instead to classify them as "domesticated endogenous viruses" (DEVs) [14] to more clearly indicate their close relationship to viruses.

PDVs have been originally classified in two genera: ichnoviruses (IV) and bracoviruses (BV) [15] reflecting a convergent evolution after wasp high-jacking of viruses from different families. Bracoviruses are associated with over 46000 species [16] of 6 Braconidae subfamilies (Microgastrinae, Cardiochilinae, Miracinae, Mendeselinae, Khoikhoiinae, Cheloninae) that form a monophyletic group named the "Microgastroid complex" [17]. These viruses all originate from a unique event of integration of a virus in the last common ancestor of this group [18]. The virus originally integrated belonged to nudiviruses [19], a group of large DNA viruses infecting insects and crustaceans closely related to well-studied baculoviruses used in biological control against lepidopteran pests [13].

BV genomes have two components in wasp chromosomes. The first corresponds to genes of nudiviral origin coding for particle structural components, the products of which are necessary for particle production. These bracovirus genes named "nudiviral genes" because of their clear phylogenetic relationship with nudiviruses, reside permanently in the wasp genome. The second component corresponds to "proviral segments" which will allow production of the DNA circles packaged in the particles [20]. These segments do not contain genes involved in virus particle production but harbor genes expressed in the parasitized host. They are collectively named "virulence genes" because of their role in parasitism success, although the function of their products has been clearly determined for only some of them [21]. The size of circles packaged in particles may differ tremendously depending on bracoviruses as clearly visualized by the electrophoretic profile of virus DNA extracted from particles of wasps belonging to Microgastrinae and Cardiochilinae respectively [22], indeed the largest circles are 46 kb and 14 kb long respectively for Cotesia congregata BV [20] and Toxoneuron nigriceps

BV [23]. This suggests that the content of bracovirus packaged genomes is highly variable depending on the wasp lineage. It is thus likely that the arsenal of virulence genes that has already been described does not reflect the diversity of virulence genes from bracoviruses associated with parasitoid species belonging to different subfamilies and having different life-styles [16]. Indeed, the packaged genome sequences reported to date mainly stem from bracoviruses of a handful of species from Microgastrinae and one species of Cardiochilinae, all parasitizing their lepidopteran host at the larval stage (named "larval parasitoids").

The Cheloninae are egg-larval parasitoids which oviposit into eggs and develop within the developing caterpillar. The biology and physiology of the parasitoid-bracovirus-host interaction involving *Chelonus inanitus*-*Spodoptera littoralis* has been extensively studied (reviewed in [24]). All stages of host eggs can be successfully parasitized and depending on the stage, different strategies of host invasion are used [24, 25]. Analyses showed that up to 20 minutes before hatching of *Spodoptera littoralis*, *Chelonus inanitus* manages to successfully parasitize its host. The latter is then an almost fully grown first instar larva, and *Chelonus* places its egg into the haemocoel of the host. Presumably the host has then acquired an immune defense from which the parasitoid would have to defend itself. Whereas most studied species (from the genus *Cotesia*, *Glyptapanteles* or *Microplitis*) belong to Microgastrinae, a ~53 Millions years old group of larval parasitoids, *Chelonus inanitus* belong to Cheloninae a subfamily that diversified earlier, ~85 Million years ago (Mya) [26]. The hypothesis that Chelonus inanitus bracovirus (CiBV) circles have a unique gene content reflecting both life-style and evolutionary history of Cheloninae was sustained by the previous sequencing of 9 CiBV circles packaged in CiBV particles [27, 28]. Indeed, viral sequences were found to encode intron-rich specific genes sharing no similarities with available genes in data bank sequences (including virulence genes from other bracoviruses).

In the present study we used a high throughput 454 pyro-sequencing approach to more fully characterize the gene content of CiBV packaged genome. The annotation of novel sequences confirmed that CiBV encodes mostly lineage specific genes as recently described for CinsBV the polydnavirus of a related wasp *Chelonus insularis,* which was recently reported together with the whole genome sequence of the wasp [29]. We also similarly identified viral ankyrin genes (*v-ank*) sharing similarities to the conserved immune gene cactus of *Drosophila*. We provide here a detailed analysis of these genes focusing on their structure and evolution. Cactus is the homologue of the human IκB-alpha, a repressor of NfκB transcription factor that plays a key role in Toll pathways involved in immune acute phase response and apoptosis [30]. Previously sequenced PDV (BV and IV) genomes all encode V-ANK proteins, but we report

here that the structures of CiBV V-ANKs are different. Indeed, CiBV V-ANKs are truncated versions of Cactus corresponding either to all six repeats or to the first four ankyrin repeats of cactus/IκB Ankyrin Repeat Domain (ARD), whereas previously described PDV V-ANKs are composed of the last four repeats of Cactus/IκB ARD [31]. We also identified another CiBV *v-ank* gene (CcBV *v-ank6*), which is not clearly related to any reported ANK protein. We assessed the expression of CiBV *v-ank* genes during parasitism using RT-PCR analysis, as a first indication on whether they could play a role in *Chelonus inanitus* parasitism success. Finally, we performed phylogenetic analyses to determine whether CiBV *v-ank*s originated from a wasp cactus gene, as well as to characterize the relationships between PDV *v-ank*s. Overall, our results reveal that *v-anks* are shared by all polydnaviruses. They are the only virulence genes in this case, which suggests interaction of ankyrins with targeted host proteins might play an essential role in the molecular dialogue between PDVs associated with parasitoids and their hosts. However, the different structure of CiBV V-ANKs and the lack of presence of a previously reported PDV V-ANK conserved signature [31] indicate CiBV V-ANKs have followed a different evolutionary trajectory from the one shared by PDV V-ANKs from other braconid and ichneumonid lineages. We propose an evolutionary scenario that may explain the unexpected phylogenetic relationships observed among PDV V-ANKs.

Material and Methods

Insect rearing

C. inanitus (Braconidae) is a solitary egg-larval parasitoid which was reared on *S. littoralis* (Noctuidae) or *Spodoptera litura*. The biology and rearing of the parasitoid and the host have been described in [32, 33]. Virus DNA extraction used for CiBV packaged genome sequencing was performed in B. Lanzrein's laboratory (Bern, Switzerland), *v-ank* genes expression was studied in M. Nakai's laboratory (Tokyo, Japan) using the same *C. inanitus* strain in *Spodoptera litura* as their host and following the same rearing protocol.

DNA isolation and sequencing.

For CiBV packaged genome extraction calyx fluid was collected as described in Albrecht *et al.* [34]. Briefly the calyces from dissected ovaries were punctured with forceps and the calyx fluid was collected with a Gilson pipet. The collected material from 25-50 females, was centrifuged at 1000 g for 5 min to precipitate eggs and cellular debris. DNA from cleared calyx fluid was

extracted using QIAmp DNA midi kit (Qiagen). To produce the DNA quantity required for 454 sequencing purified DNA was amplified using the Illustra Templiphi kit (GE healthcare) for circular DNA using Phi29 phage DNA polymerase and 3 ng of viral DNA in 68 separate reactions. A total of 22 micrograms of amplified CiBV DNA was produced. The quality of DNA was then assessed using Pulse field electrophoresis (FIGE, Biorad) and PCR amplification of two previously characterized genes (CiBV17.7 and CiBV 15.8). Rolling circle amplification (RCA) was shown to induce a significant bias in the representation of the different segments of a multipartite virus genome [35] and of different bacterial genomes in metagenomic approaches [36]. However, this bias generally does not exceed 2 to 3 folds of over-representation and is less important when the amount of source DNA is not too low [37]. A 454 single read and mate pair libraries were prepared using viral DNA at "Genoscope" sequencing platform (Evry, France) and CiBV packaged genome sequences were obtained from 7 pyrosequencing runs using Roche 454 GS FLX+. This sequencing was performed before Pr B. Lanzrein retired and her laboratory was closed, at that time 454 pyrosequencing providing relatively long reads was commonly used to sequence viral genomes [38]. The study could be completed by V-ank genes expression analyses recently, thanks to the collaboration established with M. Nakay, who had maintained the rearing of the *Chelonus inanitus* strain.

Assembly and finishing

Assembly was performed using newbler version 2.3 (Roche/454 Life Sciences). 193 primary contigs were produced that were further reduced to 24 contigs among which 20 corresponded to circular molecules and 4 could not be circularized corresponding to incomplete circles (named linear contigs). We then compared primary contig sequences which had not been retained in the final assembly to the recently sequenced Chelonus insularis BV proviral sequences, to determine whether they could also be considered as parts of the bracovirus genome. Indeed, positions of bracovirus proviral loci were previously shown to be stably maintained in wasp genomes over 50 million years [2]). We thus estimated that 6 small primary contigs corresponding to a total length of 8 kb had not been incorporated in our final assembly. This allowed us to estimate that ~ 2.4 % of CiBV packaged genome sequence is probably lacking from our assembly.

To resolve ambiguities in the assembly regarding closely related CiBV8 and 9.7 circles, total DNA was extracted from females containing viral sequences and from males that do not

produce virus particles (as a negative control) using the Wizard Genomic DNA purification kit (Promega). Three adult wasps were homogenized in 600 µl lysis solution with a Polytron (Kinematica) and genomic DNA was extracted according to the protocol designed for animal tissues. PCR reactions were performed with GoTaq (Promega, France) in a final volume of 25 µl containing 50 ng of wasp genomic DNA, 1,25 U of GoTaq, 3 mM $MgCl_2$ and 20 pmole of each specific primer (see Fig. S1 for the sequence and position of primers) with the following cycling conditions: 4 minutes of initial denaturation at 94°C, followed by 35 cycles of denaturation at 94°C for 40 s, primer hybridization at 58°C for 40 s, extension at 72°C for 60 s, and final extension at 72°C for 10 minutes. The PCR products were cleaned with DyeEx columns (Qiagen) and analysed with an ABI PRISM 3700 Avant Genetic Analyzer (Applied Biosystems) using the BigDye sequencing kit (Applied Biosystems) according to the manufacturer's instructions.

Coverage

To assess the coverage of each assembled CiBV circle, sequenced raw reads were mapped on assembled CiBV circles with Bowtie2 (v2.3.4.2) (default parameters, unpaired reads). Reads that had been mapped on CiBV circles were then converted in read counts thanks to the Rsubread package (v2.2.6). In total, 29328 reads (86.7%) were successfully mapped on CiBV circles and 4501 reads (13.3%) could not be mapped. CiBV circle coverage was then assessed using the Lander-Waterman equation.

Gene annotation

CiBV packaged genome gene predictions were performed using FGENESH software from the SoftBerry platform with the *Nasonia vitripennis* training set (http:// softberry.com/all.htm). Genes coding for predicted proteins of at least 50 amino acids were retained, given that a gene coding for a protein of 50 amino acids was validated by expression data for CinsBV [29]. We verified that all mRNA sequences of CiBV previously reported in the literature corresponded to a gene predicted in this annotation. CinsBV genes sharing homology with CiBV genes (table S2) were identified by blastP analysis.

Duplicated regions and virus regulatory sequences analyses

Analyses of duplicated regions among CiBV circles were performed using MULTIALIN, MAFFT v. 7 (http://mafft.cbrc.jp/alignment/server/index.html)[39] and DIALIGN-TX (http://dialign-tx.gobics.de/submission?type=dna)[40], and BLAST tools available at NCBI. Results were displayed using the graphical tool WEBACT (http://www.webact.org/WebACT/home).

BV Proviral segments are terminated by direct repeats at both extremities, named DRJs [6, 27, 41] and bracovirus circles contain a unique sequence named "circle DRJ" produced by a recombination event between these DRJs [20, 42, 43]. CiBV circle DRJs were retrieved by a Blastn analysis (NCBI) using a CcBV circle DRJ. Alignments were performed on 130 bp containing these sequences using MULTIALIN (http://multalin.toulouse.inra.fr/multalin) [44]. Consensus motifs were generated using the MEME program suite [45] and visualized with WEBLOGO [46]. Circle DRJs clustering was performed using maximum likelihood on the Phylogeny platform (http://www.phylogeny.fr/version2_cgi/alacarte.cgi) with PhyML v. 3.0, SH-like test and the substitution model HYK85.

Analysis of CiBV *V-ank* genes expression

Spodoptera litura were parasitized at the egg stage. Depending on the larval stage of parasitized caterpillars, total RNAs were extracted either from whole body or different tissues (midgut, fat body, hemocytes) dissected in PBS using Isogen-II (Nippon Gene, Toyama, Japan). The samples were collected after molting the first day of each stage. All the samples (parasitized larvae) were dissected to verify that they were actually parasitized. After two treatments with DnaseI (Rnase free) (Takara) following supplier protocol, the quality of RNA samples was assessed by Nonovue (GE Healthcare) and by the amplification of the ß-actin genes (using primers indicated in table S3). Specific primers were designed for the 5 CiBV *V-anks* (supl. Mat. table 2) including *V-ank2* and *V-ank5,* the sequences of which were highly similar. For these genes the specificity of the amplification was verified by Sanger sequencing of a PCR product thus further confirming the presence of two closely related copies in CiBV packaged genome. Except *CiBV V-ank6,* CiBV *V-ank* genes contain introns and the primer pairs were chosen such as to encompass an intron, thereby allowing to distinguish amplification products resulting from cDNA or genomic DNA templates (primer sequences are shown in table S3). As bracovirus DNA circles can be difficult to eliminate during the process of RNA extraction, this

difference in amplicon size enabled us to firmly conclude that the genes were expressed in addition to the classical control sample corresponding to amplification of non-retrotranscribed RNA (RT-). During initial experiments it was also verified that no amplification was obtained using DNA extracted from unparasitized *S. litura* (data not shown). Reverse transcription was performed using Takara RNA PCR Kit (AMV) as described in the manufacturer protocol. PCR was performed by 5 minutes of initial denaturation at 94°C, followed by 30 cycles of denaturation at 94°C for 2 min, primer hybridization at 60°C for 60 s, extension at 72°C for 60 s, and final extension at 72°C for 5 minutes.

Phylogenetic analyses

To the exception of CiBV V-ANKs, cactus/IκB related proteins have been retrieved from Genbank (nucleotide and whole genome sequencing) by blastP or tblastn using *Drosophila* cactus and previously reported V-ANKs from domesticated viruses of wasps belonging to Microgastrinae, Banchinae and Campopleginae as queries. Alignment of V-ANK proteins have been performed using clustal omega at EBI (https://www.ebi.ac.uk/Tools/msa/clustalo) and conserved amino acids highlighted using boxshade version 3.21 (http://arete.ibb.waw.pl/PL/html/boxshade.html) written by Kay Hofmann and M. Baron. A first step of sequence alignment and curation was performed to ensure that different ANK repeats were correctly aligned. Then phylogenetic analyses of V-ANK proteins were performed at https://ngphylogeny.fr by maximum likelihood and 1000 replicates for bootstraping (using "advanced workflow Fast tree"). Conserved PDV V-ANK sites were characterized form the alignment of 50 proteins and conserved specific PDV V-ANK sites were identified by comparing PDV V-ANK conserved sites to insect cactus and vertebrate IκB-alpha, IκB-epsilon and NFκB p105 proteins.

Data availability

CiBV-annotated sequences from new circles and previously sequenced circles with the new annotation have been deposited at Genbank (ON351504 to ON351527).

Results and Discussion

CiBV encapsidated genome contain mostly intron-rich specific genes

After high throughput sequencing, 29368 reads were obtained of which 86.7% were incorporated in the final assembly comprising 24 contigs ranging from 5419 bp to 24072 bp (Fig. 1, Genbank ON351504 to ON351527). The 9 previously sequenced CiBV circles [27, 28] were recovered. Interestingly 15 contigs corresponded to new CiBV sequences (those not labelled * in Fig.1): 11 complete circles and 4 linear molecules (incomplete circles labelled L in Fig.1). Further PCR and sequencing experiments were required to clearly differentiate CiBV 8 and CiBV 9.7 (that encode most *CiBV V-ank* genes) due to their high sequence identity (93% identity on 80% of their length) (Fig. S1). We measured the sequence coverage of each circle, which was found to vary from 6.56 (CiBV10.8) to over 202.81X (CiBV8) (Fig.1). These differences are not surprising since the abundance of different circles within bracovirus particles generally differs. Rolling circle amplification and NGS sequencing have been reported to induce a significant bias in the representation of the different segments of a multipartite virus genome. However, this bias did not exceed three folds over the estimated values [35]. Thus, the coverage of CiBV circles obtained, although it does not constitute an accurate measure as would be obtained by quantitative PCR, probably reflects the trend in the abundance of the different circles. In particular, the two *V-ank* containing circles were among the most highly covered circles, indicating that they are not minor components of the CiBV packaged genome. Accordingly, the unique homologous circle recently described in Chelonus insularis bracovirus is also among the most abundant [29]. The homology relationships that could be determined based on similarities between CiBV and CinsBV are reported in table S1 and Figure 1.

Our sequencing approach with an aggregated size of 350.771 Kb comprising the new molecules and those previously obtained provides a much more extensive view of CiBV packaged genome the total size being similar to that of CinsBV (341 kb) [29]. We estimated that only a low percentage of CiBV sequences were lacking (M&M) therefore having little impact on the predicted protein content, which was our main interest in this approach. Larger circles could result from incomplete resolution of smaller circles from molecules amplified during replication, a phenomenon called “nesting” reported for another bracovirus [20], most ichnoviruses [47, 48] and also specifically described in CiBV [27, 28]. The approach taken here allowed to obtain unique assemblies for each molecule and thus for smaller circles, and therefore did not allow to observe alternative organizations of viral sequences. Accordingly we did not obtain the largest circles (over 30 kilobases) previously reported from the measurement

of the sizes of circles released from CiBV particles using electronic microscopy [34] nor the large circle in which CiBV14 is predicted to be nested from a former Southern blot analysis [27].

Gene annotation performed on the whole set of sequences identified 71 genes coding for proteins of 50 to 760 amino acids. Unlike most viral genes, bracovirus genes contain introns, and CiBV genes are no exception as they also contain introns from our gene prediction. The intron abundance even exceeds that of previously annotated bracovirus packaged genomes. Indeed 57 genes (80%) have an intron compared to ~60% for CcBV, GiBV and GfBV [20, 49]. Half of these genes (26 genes) have a single intron. The genes for which expression profiles had been previously analyzed were all retrieved (Fig. 1). In comparison, 35 proteins have been predicted from the annotation of CinsBV packaged genome of comparable size using transcriptomic data [29], among which 30 proteins share homologous relationships with CiBV proteins (table S2). We used a threshold to annotate genes based on the smallest CinsBV protein predicted from expression data but this might overestimate the number of genes, in particular those encoding less than 100 amino acids should be verified by functional approaches.

Previously reported bracovirus packaged genomes [49-52] contain gene families common to microgastrinae encoding Protein Tyrosine Phosphatases (PTPs) [22, 53], Ben domain containing proteins [54] and viral ankyrins (V-ANKs)[55]) or proteins specific to particular lineages such as Early expressed Proteins (EP1, EP2, EP3) [56] or C-type lectins [57]. We found that CiBV packaged genome comprises 6 gene families, 5 of which are not found in other bracoviruses. The number of genes in a family does not exceed 6 (family 6), whereas they can comprise over 30 genes in other bracoviruses (for the PTP genes family) [49]). Most CiBV genes encode for proteins having no significant similarities in public data banks (except with homologous sequences from Cheloninae) and do not contain conserved domains, which makes it difficult to predict their function. However, CiBV encode *V-ank* genes, this gene family is the only one shared by all PDVs.

It should be remarked that the origin of most bracovirus packaged genes remains undetermined. Indeed, only a handful of PDV packaged genes were clearly shown to derive from wasp genes, and a few others were shown to derive from transposable elements inserted in bracovirus sequences (such as sola2 [2] and HAT [58] or a retroelement [59]). CiBV virulence proteins could have diverged to the extent that their phylogenetic relationship with

the original TE, nudivirus, wasp genes or other sources have changed to the point their origin is no longer recognizable. As an alternative hypothesis they may have evolved *de novo.* Indeed, orphan genes have recently been hypothesized to originate from non-coding sequences in primates, *Drosophila* [60] [61], yeast [62] and viruses [63]. A model predicts that recruitment of new genes might emerge from the expression of very short species-specific open reading frames (ORFs) located in non-genic sequences, which could sometimes provide adaptive potential and thus by selection might gradually become genes [62].

Sequences involved in circularization are conserved among bracoviruses

Evidence from CiBV and recently reported CinsBV indicate chelonine BVs encode a largely different inventory of virulence genes from microgastrine BVs, but nonetheless share motifs identified to have essential functions in segment circularization. It has been initially shown using a limited set of sequences that CiBV circles were circularized by a recombination mechanism using specific direct repeat sequences [28, 42] later named DRJs (Direct Repeat Junctions) [43] or WIMs (for Wasp excision/integration Motifs) [64]) located at the extremities of proviral segments. This was confirmed by extensive analyses for other bracoviruses (GfBV, GiBV, CcBV, CsBV and MdBV) [1, 20, 49]. During bracovirus replication large DNA molecules are produced from proviral segments. These amplified molecules contain the sequence of several circles that are almost contiguous in the wasp genome, separated by short spacer sequences [6, 28, 41, 65]. The amplified molecules are later resolved by site-specific recombination events resulting in the production of a single DRJ [42, 43] in each circle individually packaged in a nucleocapsid [34]. Analysis of the CiBV packaged genome confirmed that each CiBV circle and each linear contig (except CiBV 18.8) contained a circle DRJ (Fig. S2). In consequence we can assume that there are at least 23 CiBV circles produced. Alignment of these 23 CiBV sequences led to the identification of a full size conserved DRJ of ~120 bp. Within this DRJ we could more precisely identify the position corresponding to the highly conserved 5 bp direct sequence motif (AGCTT), which constitutes the DRJ core in other BVs [20]. However only the internal GCT is perfectly conserved among CiBV circles (Fig. 2), the DRJs of which display more divergence than among those of a microgastrine species, as also reported in CinsBV [29]. It was previously shown, by comparison between proviral segments and circular molecules for six CiBV circles, that the recombination between 5' and

3’ DRJs, allowing circle excision from amplified molecules, occurs within the DRJ core [28, 42]. Even if the proviral segment sequence is not available it is still possible to identify DRJs as 5’ or 3’ or as a circle junction (made of a left part originating from the 5’DRJ and a right part from the 3’ DRJ) because 5’ and 3’ DRJs although similar contain specific motifs. These conserved motifs could indeed be identified in all CiBV circle DRJs, thus confirming they actually correspond to a recombination event (sup Figure 1 and Fig.2). More precisely, in the alignment of CiBV DRJs we could identify both the highly conserved 80 bp motif upstream of the 3’DRJ core (gaAT in CiBV instead of TGAa/tT in the bracovirus of *Cotesia congregata*) (Fig.2) and a 5’DRJ motif following the core (ATnnAAnTAAngA(a/t)(t/c)AAT(a/t)), the latter is more divergent from bracoviruses of microgastinae than the two other motifs.

A sequence potentially involved in integration in host DNA is found in a single circle

Bracovirus circles may also integrate into the DNA of infected host cells. Depending on the wasp species, most circles (MdBV) [64] or only a subset of them, for the most part encoding *ptp* and *v-ank* genes (CcBV) [66-68] integrate into the DNA of infected host cells during parasitism. This integration occurs by a specific mechanism involving a bracovirus conserved regulatory signal named Host Integration Motif (HIM) distinct from the circle DRJ. We could identify a HIM-related sequence within the circle CiBV17.7, as in CinsBV10 circle as recently described [29] suggesting at least one CiBV circle may have the ability to integrate into the DNA of infected cells using a HIM mediated mechanism. As described for other HIM sites CiBV HIM is made of a palindromic structure (Fig. S3) having conserved extremities named J1 and J2 and a less conserved central region, which was shown previously to be deleted during integration [64]. The conservation of HIM sites between Cheloninae and Microgastrinae suggests the integration mechanism may have been inherited from the originally captured nudivirus. However, studies focusing on the integration of BV sequences in parasitized hosts *in vivo* such as those recently performed using various approaches [66-68] will be necessary to confirm that CiBV17.7 is actually able to integrate into the DNA of infected cells.

Chelonus inanitus bracovirus packaged genome contains large duplications

We performed comparisons of different CiBV circles and revealed that some of them share extensive similarities suggesting they have been produced following one or several duplication events.

The highest similarity between CiBV circles concerns CiBV8 and C9.6 (encoding *V-ank* genes) sharing 93% identity over their common sequence (Fig.3) suggesting they have been produced by a recent duplication. CiBV12, 14 and 10.9 share less extensive identities 70% between CiBV12 and 14 over their similar sequences (as previously reported [27]), 68% between CiBV10.9 and both 12 or 14- suggesting that they have been produced by two steps of duplication. CiBV21.4 and 22.5 share also 72% of similarity in aligned regions. All these duplicated circles share closely related DRJs (Fig.S1 and DRJs clustering not shown), suggesting that the duplicated regions have encompassed the whole sequences corresponding to these circles in the proviral form including 5' and 3' DRJs. This is consistent with previous observations in Microgastrinae that duplications of bracovirus sequences involve large regions containing several proviral segments, the boundaries of which do not correspond to conserved bracovirus regulatory sequences (DRJs or HIM) [20]. The viral mechanism resulting in integration of whole viral circles back into the wasp genome flanked by specific sequences (named J1 And J2), such as reported in *C. sesamiae* and *C. typhae* [68, 69] does not appear to be involved in these duplications, which are more likely produced by genomic rearrangements. Accordingly, some segments are partially duplicated, which occurs when the border of the duplicated region is localized within a segment. For example, in CiBV14.4 only the first part of CiBV24 is duplicated indicating the border of the duplicated region was localized within CiBV24. Moreover, CiBV15 and 15.9 share 78% similarity in aligned regions, but their DRJ circle junctions are not closely related, suggesting the duplicated region did not encompass the two DRJs of the ancestral segment. Comparisons of CiBV circle sequences with the *C. insularis* genome sustain the hypothesis that similarities between CiBV circles correspond to tandem duplications since we could identify by blastn analysis that homologues of CiBV segments sharing high similarities are localized in the same genomic regions in *Chelonus insularis*. In particular the proviral sequences homologous to C10.9, C12, C14, C15 and C15.9 are all localized in a *C.insularis* genomic region (contig 85) which appears to contain the largest number of circle sequences [29]. However, Chelonus insularis BV was reported to display extensive similarities between only two segments of the homologous locus (Locus 1) [29] suggesting that more duplications have occurred in *Chelonus inanitus* lineage.

Analyses of human mutations have attributed some complex duplicated regions in the genome to DNA replication errors. Replication fork stalling might cause the DNA polymerase to switch from one template to another and to go backwards and forwards sometimes several times [70]. The structure of proviral loci with tandem duplications (direct and/or inverted)

suggests that a similar mechanism operating at the level of the genome could be involved in the expansion of bracovirus proviral segments [20]. These tandem duplications are thought to provide new copies allowing the selection of particular beneficial alleles after accelerated mutation accumulation in duplicated genes, thus providing new weapons for the parasite [71].

The structure of CiBV viral ankyrins differs from that of other polydnaviruses V-ANKs

To study the relationship between CiBV V-ANKs and previously described proteins containing Ankyrin domains, we first performed BlastP and tblastN searches to identify proteins similar to CiBV V-ANKs in public data bank sequences. In addition to the expected cactus proteins from various insects, we identified ANK proteins from the annotated set of proteins derived from the *Chelonus insularis* genome, corresponding to bracovirus proteins encoded by the proviral form located in the wasp genome recently reported [29]. Surprisingly, we also retrieved closely related proteins encoded by the genome of the Lepidoptera *Chilo suppressalis*. This could be explained by promiscuous relationships of Lepidoptera with their parasites. More precisely the presence of a bracovirus gene in a Lepidoptera might reflect a contamination of the DNA used for genome sequencing [72]. According to this hypothesis *Chilo suppressalis* Ank genes might belong to the bracovirus *v-ank* genes of *Chelonus munakatae* a major parasitoid of the strip stem borer naturally present in southern Asia [73]. As an alternative hypothesis these *V-Ank* sequences could have been integrated in *Chilo suppressalis* genome by horizontal transfer. Bracovirus circles integrate into the DNA of infected cells [64, 66-68] as a part of the parasitoid/virus life cycle and the presence of bracovirus derived sequences in several genomes of Lepidoptera have been experimentally confirmed [74, 75], indicating that circle integration events in the germline of butterflies and moths do also occur. These sequences may reach fixation either because of the new function they provide to the Lepidoptera [75, 76] or by genetic drift in small populations. In any case *Chilo suppressalis* V-ANK sequences provide additional data on viral ankyrins from the Cheloninae and were therefore retained for phylogenetic analyses.

We performed an alignment between the predicted sequence of CiBV ankyrins and a set of similar proteins identified by blast analyses including human IκB, the crystal structure of which has been determined [77]. The well-characterized domains of IκB (ank repeats comprising alpha-helices and ß-sheets) were used to predict the localization of potential domains on homologous proteins.

Drosophila cactus contains an ankyrin repeat domain (ARD) that consists of six ankyrin repeats [78]. In addition, Cactus has a signal response region (SRR), N-terminal to the ARD, which is involved in ubiquitination targeting the protein for degradation by the proteasome [79], whereas the C-terminal domain contains a PEST domain (stretch of proline, glutamic acid, serine, and threonine residues) implicated in protein degradation by calpain protease [80]. PDV V-ANK proteins are much smaller because the ARD is reduced and they lack the N-terminal and C-terminal regulatory domains [31, 55, 81]. Based on the alignment between cactus and human IκB alpha we could deduce the structure of CiBV viral ankyrins. The longest proteins, CiBV ANK2 and CiBV ANK5 (differing by few AA residues in their N terminal part), contain an entire cactus ARD followed by a C-terminal stretch the length of which is comparable to those of hymenopteran cactus (Fig.4) but neither contain the typical PEST motif involved in IκB calpain mediated degradation [82], nor the SRR motif involved in ubiquitin mediated degradation. The shortest CiBV V-ANK proteins (CiBV V-ANK1, 3, 4) are composed of only the first four ankyrin repeats of the ARD (Fig.4, Fig.5). The structures of CiBV V-ANK proteins thus differ from that of previously sequenced bracovirus and ichnovirus V-ANKs, which are essentially composed of the last four ankyrin repeats of cactus ARD [31, 55, 81] (Fig.5). These shorter CiBV V-ANK 1,3,4 protein sequences are closely related to CiBV V-ANK 2/5 (Fig.4). Genes encoding these shorter V-ANK proteins likely correspond to duplicated versions of V-ANK2/5 that have been truncated and have diverged after duplication (accordingly they are encoded by duplicated circles CiBV 8 and CiBV 9.7). Such complete and truncated ARD are also found in CinsBV V-ANK proteins (Ank-CinsV1-3 ARD is complete while those of Ank-CinsV1-1 and Ank-CinsV1-2 are truncated and the three genes are all located on the same circle). Of note, unlike most PDV V-ank genes, these *CiBV V-ank* genes contain several introns (Fig. 3). Finally, an additional CiBV *v-ank* gene located on a different circle encodes a protein (CiBV V-ANK6, which is Ank-CinsV3 homologue) made of a three repeat ARD according to a pfam search but has diverged to such an extent that no clear phylogenetic relationship with a particular ANK protein can be detected. Because of this high divergence CiBV V-ANK6 was not included in the alignment. Interestingly a V-ANK protein of Cotesia congregata bracovirus (CcBV 26.5 gene product) was also shown to have similarly diverged to the extent that only a few residues allowed the identification of its relationship with other CcBV V-ANKs. We cannot exclude that CiBV V-ANK6 has further followed such a process of divergence rendering its relationship with CiBV V-ANKs no longer detectable, but it could as well derive from the capture of another wasp gene containing ankyrin repeats.

CiBV *V-ank* genes are expressed in parasitized host *Spodoptera littoralis* indicating a potential role of their products in the host/parasitoid interaction

We studied CiBV *V-ank* gene expression in *Spodoptera littoralis* during parasitism using RT-PCR and specific primers for each gene (see M&M). Given the small size of the larvae, whole larvae were used to extract RNA of the first three stages. From larval stage 4, we could dissect the midgut and fat body, and at stage 5 (feeding stage before the precocious onset of metamorphosis occurring during parasitism) we could also obtain the hemocytes. We did not detect any CiBV *V-ank* gene expression before in the host eggs (data not shown) whereas in first larval stages (L1, L2) a faint signal was sometimes detected, suggesting the possibility of an onset of expression at very low level before L3 (data not shown). In later stages the onset of expression differed between these genes: *CiBV V-ank1* and *V-ank3* were detected from L3 (Fig.6 A), *V-ank2, V-ank4* from the fourth instar (Fig.6 B), and *V-ank5* and *V-ank6* only in the fifth instar (Fig.6 C, D). An amplimer of the expected size corresponding to the cDNA was obtained for each CiBV *V-ank* gene in L5 stage indicating that all these genes were expressed during parasitism at this stage (Fig.6 C, D). Except for CiBV *V-ank6*, the expression of which was detected exclusively in the midgut, *CiBV ank* genes were expressed in the three tissues assessed, suggesting they might be ubiquitously expressed (Fig.6 C, D). Altogether these results indicate that *CiBV V-ank* genes were expressed during parasitism and readily detectable at late stages, indicating that corresponding V-ANK proteins are likely produced by parasitized host cells, and thus might play a role in the host-parasitoid interaction. Their biological functions will require further characterization by functional approaches inspired by the functions of V-ANK previously described, summarized below.

In *Drosophila*, the cactus protein regulates, as a repressor, several cellular responses triggered by NF-κB/Rel transcription factors, such as the dorso-ventral patterning during embryonic development, the release of antimicrobial peptides and apoptosis, involving Dif and Relish proteins. Whether they contain only half or the complete set of cactus ank repeats, some CiBV V-ANKs might have retained the ability to bind to NFκB-like transcription factors and interfere with cactus binding. Their biological functions might rely on their capacity to compete with and act as constitutive inhibitors of cactus, since they do not have N and C domains

mediating cactus degradation by the 26S proteasome and calpains respectively. Accordingly, it was shown previously by co-immunoprecipitation experiments that two V-ANK proteins from Microplitis demolitor bracovirus (MdBV) were actually able to bind to *Drosophila* Dif and Relish NF-κB proteins [55]. Moreover, MdBV V-ANKs were shown to reduce NF-κB-driven expression of reporter gene constructs in Hela cells according to their potential role as transcriptional inhibitors [55].

However, as more studies were performed the picture of V-ANK potential functions has become more complex. In particular, transcriptome analysis of *M. sexta* fat body and hemocytes did not show an inhibition of NF-kB immune peptide induction after bacterial challenge in parasitized larvae [83] as would be expected in the case of a constitutive repression of the NFκB pathway. Moreover, although sharing a similar structure PDV V-ANKs can be involved in various functions either inducing (TnBV1 [84]) or protecting from apoptosis (fat body expressed CsIV V-ANKs, [85]) or interfering with prothoracic gland signaling causing developmental arrest (TnBV1[86]). Such an impact on prothoracic gland could also apply to CiBV, as this BV was shown, in synergy with venom, to cause an inhibition of host prothoracic gland and reduction in ecdysteroids at a particular developmental stage (pupal cell formation) [87]. Some Ichnovirus V-ANKs have been shown to protect cells infected by baculovirus from apoptosis [85], a property which was used to increase protein production in cultured cells using recombinant baculoviruses containing these genes [88]. In addition, one might speculate that integration of bracovirus ANK gene-containing-circles in the DNA of infected hemocytes [66] might have an effect on the survival/death of infected cells as suggested by new findings on the inhibition of Toll pathways on the regulation of cell fitness during infection [89].

Besides the original interaction of cactus with NFκB, PDV ANKs might have acquired the ability to bind to new targets instead of the original NFκB by exploring mutational space. In particular co-immunoprecipitation experiments recently showed that TnBVANK1 binds to Alix, an interactor of apoptosis-linked gene protein 2 (ALG-2) resulting in induction of apoptosis in host haemocytes [90]. This interaction is probably also responsible for the impairment of the vesicular trafficking of the steroid precursor in the prothoracic gland causing host developmental arrest [86]. In line with this, we can speculate that the highly divergent CiBV ANK6 may have evolved to recognize a specific host protein target. Functional assays have first been performed using heterologous systems such as *Drosophila* which have given interesting clues but might be difficult to relate directly to actual interactions ongoing *in vivo* during host parasitoid interactions [91]. Yeast two hybrid experiments [92] might allow to

identify CiBV V-ANK targets among host proteins by using an approach without *a priori* assumptions. Moreover, the development of functional assays in parasitoid wasps such as gene knockout using CRISPER-cas9 [93] will probably allow us to characterize the role of V-ANKs, reflecting more accurately their function *in vivo* during host-parasitoid interactions, but this is the aim of future studies.

Phylogenetic analysis suggests that CiBV viral ankyrins originate from an insect cactus gene

We performed a phylogenetic analysis including CiBV VANK5, similar sequences from *Chelonus insularis* and *Chilo suppressalis*, cactus sequences from different insect orders, IκB-alpha, IκB-epsilon and NfκB p105 of vertebrates. CiBV V-ANK2 was not included because its sequence was almost identical to CiBV ANK5 and VANK1-3-4 were also excluded because these truncated proteins would have drastically reduced the length of the alignment used to build the tree. The grouping of CiBV V-ANK5 with insect cactus and vertebrate IκB-alpha, was well supported consistent with the hypothesis of a cactus wasp gene origin of the *V-ank* gene. This grouping is in accordance with results obtained using blastP analyses in which all 100 first retrieved sequences using CiBV V-ANK5 as a query belong to insects (data not shown). Surprisingly none of these sequences belong to those of other PDV V-ANKs. Of note in the phylogenetic tree within the group of Insect cactus and vertebrate IκB-alpha, CiBV-VANK5 appears to be closer to proteins from the hemiptera *Bemisia tabacci* than from Hymenoptera but the branches are not sufficiently supported to conclude on an actual closer relationship, which is unlikely, this grouping is rather probably a consequence of long branches attraction.

The structure of PDV V-ANKs composed almost exclusively of ANK repeats sharing similarities with cactus/IκB already suggested that these genes originated from a cactus gene of an ancestor wasp genome. However, we could not previously validate this hypothesis from a phylogenetic tree having well supported nodes because of the short length and high divergence of PDV V-ANK proteins, in contrast with the conservation of the ARD of cactus and IκB related genes. Indeed, PDV V-ANKs are as similar to insect cactus as to vertebrate related ankyrins except for a short stretch of amino acids (TYQLA in the 3' end of the protein, Fig.8) which is found in several other insects and very common in Hymenoptera. The high divergence with insect genes is a characteristic feature of insect-related BV genes packaged in the particles [94], and is thought to result from rapid evolution of virulence proteins [71, 95] interacting with host targets which themselves undergo rapid modifications in the context of the co-evolutionary

arms race between hosts and parasites. CiBV V-ANK5 is also very divergent but thanks to its larger size corresponding to the full cactus ARD group we could more clearly support a phylogenetic relationship with cactus/IκB.

A cactus gene could have been acquired by a bracovirus proviral segment following a wasp genome rearrangement or the integration of a cDNA after retro-transcription of a cactus mRNA using the retrotranscriptase of an endogenous retroviral element. This would have resulted in the presence of the gene in a circle and the possibility to express the corresponding protein in host cells during parasitism which could have been later selected, if adaptive. Such recent gene acquisition events of genes from the wasp genome have been already identified. For example, sugar transporter genes in bracoviruses of *Glyptapanteles* species [49] are not found in viruses of the sister genus *Cotesia*, suggesting a recent acquisition of these genes from the hymenoptera gene set. Moreover, in the case of bracovirus cystatin genes, the lack of the usually very conserved introns in these eukaryotic genes, suggests the mechanism of cystatin gene acquisition by the bracovirus has comprised a step of cDNA retrotransposition [96]. This type of mechanism was reported to be involved in the production of 8000 so-called "processed pseudogenes" in the human genome originating from the integration of reverse transcribed cDNAs by L1 retroelement retrotransposases [97].

CiBV viral ankyrins have a different evolutionary history than bracovirus and ichnovirus V-ANKs

We compared CiBV and other polydnavirus V-ANKs to determine whether CiBV V-ANK 2/5 protein sequence shares the previously identified specific signature that is common to many bracovirus and ichnovirus V-ANKs. This signature was originally reported to consist, in addition to the conserved motifs shared by all cactus related proteins, in a typical "WLC" motif together with five conserved amino acids present in different positions [31, 98] of the PDV proteins. This signature is not present in CiBV V-ANKs 1/3/4 since it is located in the part of the ARD that has been lost in these proteins. Surprisingly, however, we found that this motif is not present in CiBV V-ANK 2/5 either. In contrast, after retrieving V-ANK proteins from a much larger set of PDVs and cactus genes from insects not available at the time of our previous studies [31, 98] we confirmed the presence of the "WLC signature" in ~ 80% of PDV V-ANKs (Fig 8, Fig.9). This motif is specific to PDV V-ANKs and was not found in hymenopteran cactus proteins (having FLL or FIL in homologous positions) nor in the many other insect cactus proteins for which sequences are now available or can be deduced by tblastn

analysis of insect genomes. Moreover, we established a PDV V-ANKs consensus from 40 proteins (Fig. S4) and comparison of this sequence with cactus and related vertebrate proteins revealed that 19 sites were specific of PDV V-ANKs (Figure 8 and Fig. S5), thus providing a strong basis of their common grouping. Together with their common structure (ank3 to ank6 repeats), this confirms that V-ANK proteins from BVs and IVs share a common evolutionary history [31, 98], whereas CiBV V-ANK 5 does not belong to this group (Fig.9).

However, the fact that virulence genes of bracoviruses and ichnoviruses share a common history cannot be easily explained since the two PDV subfamilies evolved by convergence and are associated with different wasp lineages. Whereas bracoviruses are present in all subfamilies of the Microgastroid complex, members of the Ichnoviridae are associated with two subfamilies of Ichneumonidae -Campopleginae and Banchinae- separated in the phylogenetic tree by several subfamilies of wasps not associated with IVs [99]. Because of this patchy distribution, it is unclear whether Ichnoviruses originated from a single ancient viral capture event and were lost in some wasp subfamilies, or derive from independent captures of viruses in the two subfamilies [100, 101]. The gene set of ichnoviruses involved in particle production is conserved between Banchinae and Campopleginae [48, 100, 102], which indicates that if two virus captures have occurred, these viruses belonged to the same family. Unlike for bracoviruses however, these genes do not resemble those of any currently described pathogenic virus, suggesting that ichnovirus ancestor(s) might belong to a virus family of arthropods the free-living members of which have not yet been characterized or have become extinct [99, 100, 102]. Whatever the case, IVs and BVs clearly derive from viruses belonging to different families. Consequently, virulence gene families such as V-ANKs that would be shared between the two PDV families could not have been simply inherited from a common virus ancestor of PDVs.

Instead, we hypothesize that shared structure and signature between PDV V-ANKs might be explained by horizontal gene transfer events between PDVs during multi-parasitism events in common lepidopteran hosts (Fig.10). Once acquired by a receiver PDV, a virulence gene from the donor PDV can be readily used during parasitism since promoters of both PDV genera are expressed by the lepidopteran transcription machinery, thereby favoring this mode of acquisition.

Clues on how these horizontal events could take place are provided by several studies that have shown that bracoviruses and ichnoviruses use specific integration mechanisms to

insert packaged genome sequences into the DNA of infected cells [64, 66, 67, 103] [68]. A likely, although indirect consequence of this mechanism of integration is the occurrence on the one hand of horizontal gene transfer between wasps and Lepidoptera [75] and on the other hand of the reintegration of PDV circles back into the wasp genome [68, 69]. We hypothesize that rare events of integration of PDV circles in the genome of a wasp could also rarely occur in the context of multi-parasitism. Indeed, a lepidopteran species could be simultaneously parasitized both by IV and BV associated wasps. In consequence, an integration event occurring within the proviral sequences of an ancient ichneumonid wasp might have resulted in the acquisition by an IV of a bracovirus circle containing a gene encoding a V-ANK protein having the WLC signature, or conversely.

The PDV WLC signature is not found in any hymenopteran or insect cactus proteins (Hymenoptera having FIL or FLL in homologous positions) suggesting it was probably inherited from a virulence gene having already diverged from a cactus of wasp origin in the context of the host/parasite arms race. This signature has been lost in several PDV V-ANK proteins (alignment and tree) indicating its presence is not necessarily required for V-ANK function but could merely reflect that the ancestral sequence of the transferred gene is not yet completely eroded by divergence. Depending on whether Ichnoviruses have a single origin or derive from two independent captures of viruses from the same family, the widespread distribution of V-ANK genes could be explained by a single or at least two events of horizontal transfer that would have occurred between BVs and IVs (Fig.10).

In summary the phylogenetic analyses of CiBV V-ANK 5 sustain the hypothesis that this gene and related CiBV V-ANK genes (V-ANK1-2-3) originated from the acquisition of an insect cactus gene of probable wasp origin by proviral sequences leading to its incorporation in the particles allowing its expression during parasitism and thus adaptive selection. In addition, these analyses further highlight the common origin of other PDV ANKs which was unexpected given the different viral origins of BVs and IVs and suggest the occurrence of ancestral genetic exchanges of PDV circles between wasps parasitizing the same host species. Chelonus V-ANK genes have followed a different evolutionary history either from a unique ancestral acquisition of a cactus gene that occurred in a common ancestor of the microgastroid lineage or from an independent acquisition of wasp cactus gene in Cheloninae.

Conclusion

This study provides an extensive view of a packaged bracovirus genome of a Cheloninae, which are egg-larval parasitoids and is complementary to the recent analysis of *Chelonus insularis* genome and its endogenous bracovirus. An inventory of the genes packaged in CiBV virions shows that they share several features with the other Cheloninae BV analyzed (CinsBV) but differ from BVs associated with wasps Microgastrinae and Cardiochilinae. In contrast, results from this study lend further support that the regulatory sequences involved in circularization of the DNAs in virions are conserved among all BVs. These elements suggest a strong conservation of the viral functions probably inherited from the ancestral captured nudivirus, allowing these wasps to produce DNA containing particles; contrasting sharply with the highly variable gene content packaged in the particles. Indeed, most CiBV packaged genes are specific to Cheloninae to the notable exception of V-ANK genes, the products of which have a different structure from the one previously described for other PDV V-ANKs. Phylogenetic analyses lend some support to a cactus/IκB origin of CiBV V-ANKs despite the divergence of their sequences. They also indicated they had a different evolutive history than that shared by other bracovirus and ichnovirus V-ANKs, the latter potentially driven by horizontal gene transfer through multiparasitized hosts.

Acknowledgements

We thank Mohamed Amine Chebbi (Viroscan) for submission of data sequences to ENA. Elisabeth Herniou for having initially established the collaboration with Madoka Nakai. We thank Beatrice Lanzrein for inputs and critical reading of the manuscript.

Funding Information

The funding for sequencing and bioinformatic analyses was provided by Genoscope project AMW2010 “ADN particules virales” to J-M Drezen. Vonnick Sibut’s salary was funded by ANR Paratoxose (*ANR-09-BLAN-0243-01)*. The analyses were pursued using recurrent funding from CNRS and University of Tours.

Conflict of Interest

The authors declare they have no conflict of interest related to this study

Author contributions

A. C.deA.: Investigation, formal analysis, data curation, conceptualisation, writing-original draft preparation, visualisation, T.J.: ideas, investigation, visualisation writing-review and editing, supervision V.S.: investigation,validation, M.U.: investigation, A.A. : investigation, V.B.: resources, data curation, M.N.: resources writing-review and editing, E.H.: writing-review and editing, G.P.: ideas, visualisation, writing-review and editing, J.-M. D.: Conceptualisation, visualisation, writing-original draft preparation, supervision, funding

Figure Legends

Figure 1: Linearized map of CiBV DNA circles packaged in bracovirus particles and their genes represented as boxes (exon/introns structure is not shown). The sequencing coverage of these molecules is shown on the upper right panel. It is supposed to indirectly reflect the abundance of the different circles since a bias could be introduced by rolling circle amplification of CiBV DNA and high throughput sequencing [35]. CiBV packaged genome contains 6 gene families but only one (*V-ank* genes family) encodes proteins having a well-known conserved domain (Ankyrin repeat domain -ARD-). *V-ank* genes encode proteins almost completely made of a whole cactus/IκB-like ARD comprising 6 ankyrin repeats or of the first 4 ankyrin repeats only. The gene encoded by CiBV19.8 encodes a protein having a highly divergent ARD. Most CiBV genes are specific to the Cheloninae. The black stars in the boxes indicate that the annotated genes were found to be expressed during parasitism in this study (*V-anks*) or in previous analyses [8, 9, 100-102]. Green stars indicate CiBV circles for which sequences had been published previously by Pr B.Lanzrein's laboratory and the names of which have been conserved in this study. The names of the newly sequenced circles correspond to the size of their sequence to the first decimal. Probable homologous CinsBV circles are also indicated based on the similarities reported in table S. L: linear contigs corresponding to incomplete circles.

Figure 2: Conserved sequences involved in CiBV segments circularization obtained from the alignment of 23 CiBV circle DRJs and visualized using WEBLOGO (upper panel) compared to those of CcBV circle DRJs (bottom panel [20]). The height of the stack at a position indicates the sequence conservation, whereas the height of a base indicates the relative frequency of this base at this position. Circle DRJs are produced by a recombination between the direct repeats flanking the proviral form of a viral segment (5'DRJ and 3' DRJ) which are similar but have specific conserved motifs (5' DRJ motif and 3'DRJ motif). The left part of Circle DRJ comes from the 3'DRJ and their right part from the 5' DRJ.
The comparison highlights the conservation of these regulatory sequences, between Cheloninae and Microgastrinae with the 3' DRJ specific motif (GAAT in CiBV) then a central motif where circularization occurs (aGCT) [28, 38]). The 5' DRJ motif differs but is adenine-rich in both viruses. Based on data concerning genome packaging of other viruses, it was hypothesized that

these motifs correspond to binding sites of site-specific recombinases [60] thought to resolve amplified molecules produced during bracovirus DNA replication into circles individually packaged in nucleocapsids. The proteins of nudiviral origin Int-1 and Vlf1 have been shown to be involved in MdBV circles excision/circularization [103].

Figure 3: Dot plot comparison of CiBV8 and 9.6 sequences sharing 93% identity in aligned sequences (visualized by the slanted lines). The positions of *v-ank* genes encoded by these circles and their intron/exon organization are shown on the axes. Note that unlike most bracovirus *V-ank* genes those of CiBV contain introns.

Figure 4: Alignment of CiBV V-ANK proteins with closely related sequences in public data banks (named here VANK-like-Chilo, VANKX1chelonus for CinsV1-3, VANKX3chelonus for CinsV1-1) and with cactus proteins from a series of Hymenoptera species, *Drosophila* and Human IκB homologue (the accession numbers of which are indicated on the phylogenetic tree in figure 7). Above the alignment the secondary structure elements of Human IκB are shown with arrows for ß-strands, cylinders for alpha-helices, and H1, H2, loop and ß2 for the conserved structural elements of the ankyrin repeats. Amino acids highlighted in grey, ≥50% amino acid similarity; amino acids highlighted in black, ≥50% amino acid identity. Note that CiBV VANK2 and CiBV VANK5 ARD are identical; these proteins differ only in their N-terminal region.

Figure 5: Schematic representation of the proteins encoded by CiBV *V-ank* genes and closely related genes in public data banks, compared with the ankyrin repeats of human (Hum) IκB-alpha, *Drosophila* cactus (Droso) and previously reported V-ANKs, the structure of which is shared by both bracovirus and ichnovirus V-ANKs (represented here by 3 V-ANKs from Cotesia congregata bracovirus: CcBVank1, CcBVank4 and CcBVank6). For *Drosophila*, an insertion of 26 aa in the ank3 repeat, conserved across insects, is denoted with a double slash. The numbers under each protein representation indicate the amino acid positions delimiting the different ankyrin repeats as deduced from the alignment with IκB-alpha (Fig. 4 and [31]). Accession numbers of CcBVank1, CcBVank4 and CcBVank6 are AJ583542, AJ583545, AM180416 respectively). Human IκB-alpha regulatory regions: SRR for Signal Response Region which contains sites for phosphorylation by IKK (IKK: IκB kinase), for ubiquitination, and for nuclear export; PEST for carboxy-terminal PEST region composed of Proline, Glutamic acid, Serine and Threonine, responsible for protein turnover. SRR signals were not detected in CiBV V-ANKs the N terminal sequences of which are short and not conserved. The size of CiBV V-ANK2 and CiBV V-ANK5 is only a little shorter than that of cactus proteins from Hymenoptera but they do not contain typical amino acids of a PEST domain.

Figure 6: *CiBV V-ank* genes expression in 3rd (A) 4th (B) and 5th feeding stage (C and D) instar larvae. (A) *CiBV V-ank* gene expression in 3rd instar larvae: To test the expression of *CiBV V-ank* genes, RNA was extracted from 6 parasitized larvae. cDNA presence was tested by detecting ß-*actin* expression whereas viral genome DNA absence was indicated by no band detection in the minus RT control (-RT). *CiBV V-ank1* and *CiBV V-ank3* were expressed in 3rd instar larvae. *CiBV V-ank2* gene expression could not be detected in the 3rd instar larvae. No

expression was detected for *CiBV V-ank4, V-ank5 or v-ank6* in 3rd instar larvae (data not shown). (B) *CiBV ank* gene expression in 4th instar larvae: RNA was extracted from 5 parasitized larvae. The expression was assessed in two tissues: midgut and fat body. In each case, a control is shown to visualize the size of the amplimer obtained with *Chelonus inanitus* DNA (genome) corresponding to that of viral DNA, *CiBV V-ank* genes containing introns. In the upper part (+RT) results for *CiBV V-ank* gene expression is shown, and in the bottom part (-RT) is presented the control showing no viral DNA contamination. *CiBV V-ank1, 2* and *3 genes* are expressed in the midgut and fat body. For *CiBV V-ank4*, only very weak bands (white arrow) could be detected in midgut and fat body. No expression was detected fo *CiBV V-ank5 and 9* in 4th instar larvae (data not shown).
(C and D) *CiBV V-ank* gene expression in 5th instar larvae: RNA was extracted from 3 parasitized larvae. The expression was assessed in three tissues: midgut, fat body and hemocytes. In each case, the size of the amplimer obtained with *Chelonus inanitus* DNA (genome) is shown. In the upper part (+RT) results for CiBV *V-ank* genes expression is shown and controls are as indicated in B. (C): *CiBV V-ank1, 2, 3* and *4* genes are expressed in midgut, fat body and hemocytes. (D): *CiBV V-ank5* gene is expressed in midgut, fat body and hemocytes and *CiBV V-ank6* only in the midgut.

Figure 7: Maximum likelihood tree of CiBV V-ANK5, closely related sequences available in data banks, insect cactus proteins and vertebrate IκB-alpha, IκB-epsilon and NfκB-p105. The Human Gankyrin was used as the outgroup. Note that the group of insect cactus including CiBV-ANK5 and closely related sequences, is well supported despite the divergence of bracovirus sequences and the high conservation of IκB and related genes from insects and vertebrates, which supports the hypothesis that CiBV V-ANKs originated from the gene set of an ancestral wasp genome.

Figure 8: Subset of the alignments performed to compare PDV V-ANKs (shared between bracoviruses and ichnoviruses) to CiBV V-ANK5, insect cactus and related vertebrate proteins (IκB-alpha: other vertebrate proteins are shown in Fig.S5 alignment). Conserved sites of V-ANK indicated in the penultimate line were identified using 40 proteins (Fig. S4). The 19 conserved PDV-VANK residues that are not found in corresponding sites of insect and vertebrate proteins are indicated below the alignment. WLC signature: previously described signature of PDV V-ANKs, note that this region of homology extends outside the signature (consensus EALEWLC—PGIDLE). Insect signature: consensus "GLTAYQLA" shared between PDV V-ANKs and some insect cactus proteins, in particular Hymenoptera suggesting that despite their divergence PDV V-ANKs originally derived from a gene of the hymenopteran gene set, possibly following an integration of a cactus gene copy into the proviral form of a virus circle.

Figure 9: Maximum likelihood tree of CiBV V-ANK5 and a set of PDV V-ANKs previously sequenced. Note bracovirus and ichnovirus V-ANKs form a well-supported monophyletic group to which CiBV V-ANK5 does not belong. The group of cactus/IκB proteins to which belongs CiBV V-ANK5 is not as well supported as in Fig.7 tree because of

the shorter length of sequences included in the alignment and the high divergence of CiBV V-ANK5 repeat 6 (Fig. 4).

Figure 10: Hypothetical scenario on the evolution of bracovirus and ichnovirus V-ANKs that may explain structural and sequence similarities between bracovirus and ichnovirus V-ANKs, whereas CiBV V-ANKs are different. 1°) The bracovirus might have acquired a copy of a wasp *cactus* gene in a common ancestor of the group of bracovirus associated wasps (note that cactus from braconid wasps have generally a stretch of amino acids "FLL" or "FIL" in the ANK5 repeat instead of the WLC motif specific of the other PDV V-ANKs). This acquisition might have occurred once at the basis of the group of bracovirus associated wasps (microgastroid complex), as shown in the figure, or twice independently at the basis of cheloninae and microgastrinae lineages respectively. 2°) This gene favoring parasitism success would have then diverged in the context of the host-parasite arms race. In the lineage common to the Microgastrinae and Cardiochilinae subfamilies the divergence would have led to a sequence having the specific phylogenetic signature of V-ANKs (made of dispersed shared amino acids and a "WLC" in the ANK5 repeat) and to the reduction of the V-ANK to the second half of the ARD. In Cheloninae the acquired cactus gene evolved differently since Chelonus inanitus bracovirus V-ANKs are made of the full ARD and do not display the V-ANK signature ("FIL" in ANK5 repeat) or are constituted of the first half of the ARD (ANK5 and ANK6 repeats were lost). 3°) and 4°) One horizontal transfer event of bracovirus *v-ank* gene to the common ancestor of Banchinae and Campopleginae lineages or two events to the ancestors of each lineage (as shown in the figure) would have resulted in Ichnoviruses having V-ANKs similar to those of bracoviruses. The scenario presented here is one of the most parsimonious, but alternative scenarii could be conceived involving the acquisition of bracovirus cactus from a virus having evolved the WLC signature or the transfer of a cactus gene having WLC from an ichnovirus to the other lineages. The phylogenetic tree is adapted from [96].

Table S1: CiBV Homologues of CinsBV genes

Table S2: Similarities and potential homologous relationships between CiBV and CinsBV circles

Homologues of CiBV circles were deduced from high similarity blocks of NCBI blastn graphic summary dispersed on the whole CiBV circle. Nd: not determined, i.e.: the similarities were insufficient to conclude or several CinsV circles shared similarities with the same CiBV circle.

Table S3: primers used for CiBV *V-ank* genes expression analyses

Figure S1: Sequence of primers used to determine the sequence of CiBV8 (7931 bp) and CiBV9.7 (9700 bp) from the assembly. Numbers in bold on top of the arrows are primary contig names, numbers in italics below the arrows correspond to primary contig length (base pair), primary contigs are not to scale.

Figure S2: Alignment of 23 circle DRJs identified in CiBV packaged genome

Figure S3: Alignment of the putative HIM motif from CiBV17.7 to the homologous sequence of CinsBV-10 and to functionally characterized HIM sites involved in integration into the DNA of parasitized host cells of MdBV and CcBV circles.

Figure S4: Alignment of the 40 proteins used to identify conserved sites of bracovirus and ichnovirus V-ANKs

Figure S5: Alignment comprising the different IκB related proteins of vertebrates highlighting the specific PDV V-ANK WLC and TAYQLA insect signatures indicating respectively their shared evolutionary history.

References

1. **Burke GR, Walden KK, Whitfield JB, Robertson HM, Strand MR**. Widespread genome reorganization of an obligate virus mutualist. *PLoS Genet* 2014;10(9):e1004660.
2. **Gauthier J, Boulain H, van Vugt J, Baudry L, Persyn E et al.** Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. *Commun Biol* 2021;4(1):104.
3. **Stoltz DB, Krell PJ, Summers MD, Vinson SB**. Polydnaviridae- a proposed family of insect viruses with segmented, double-stranded, circular DNA genomes. *Intervirology* 1984;21:1-4.
4. **De Buron I, Beckage NE**. Characterization of a polydnavirus (PDV) and virus-like filamentous particle (VLFP) in the braconid wasp *Cotesia congregata* (Hymenoptera : Braconidae). *J Invertebrate Pathol* 1992;59:315-327.
5. **Wyler T, Lanzrein B**. Ovary development and polydnavirus morphogenesis in the parasitic wasp *Chelonus inanitus*. II. Ultrastructural analysis of calyx cell development, virion formation and release. *J Gen Virol* 2003;84(Pt 5):1151-1163.
6. **Pasquier-Barre F, Dupuy C, Huguet E, Monteiro F, Moreau A et al.** Polydnavirus replication: the EP1 segment of the parasitoid wasp *Cotesia congregata* is amplified within a larger precursor molecule. *J Gen Virol* 2002;83(Pt 8):2035-2045.
7. **Stoltz DB, Guzo D, Belland ER, Lucarotti CJ, Mackinnon EA**. Venom promotes uncoating *in vitro* and persistence *in vivo* of DNA from a braconid polydnavirus. *J Gen Virol* 1988;69:903-907.
8. **Bonvin M, Kojic D, Blank F, Annaheim M, Wehrle I et al.** Stage-dependent expression of Chelonus inanitus polydnavirus genes in the host and the parasitoid. *J Insect Physiol* 2004;50(11):1015-1026.
9. **Weber B, Annaheim M, Lanzrein B**. Transcriptional analysis of polydnaviral genes in the course of parasitization reveals segment-specific patterns. *Arch Insect Biochem Physiol* 2007;66(1):9-22.
10. **Gundersen-Rindal D, Dupuy C, Huguet E, Drezen J-M**. Parasitoid Polydnaviruses: Evolution, Pathology and Applications. *Biocontrol Science and Technology* 2013;23(1):1-61.
11. **Strand MR, Burke GR**. Polydnaviruses: From discovery to current insights. *Virology* 2015;479-480:393-402.
12. **Gauthier J, Drezen JM, Herniou EA**. The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps. *Parasitology* 2018;145(6):713-723.
13. **Petersen JM, Bézier A, Drezen JM, van Oers MM**. The naked truth: An updated review on nudiviruses and their relationship to bracoviruses and baculoviruses. *J Invertebr Pathol* 2022;189:107718.
14. **Drezen JM, Bézier A, Burke GR, Strand MR**. Bracoviruses, ichnoviruses, and virus-like particles from parasitoid wasps retain many features of their virus ancestors. *Curr Opin Insect Sci* 2021.

15. **Strand MR, Drezen JM**. Polydnaviridae. In: king AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (editors). *Virus Taxonomy: 9th Report of the International Committee on Taxonomy of Viruses*: Elsevier; 2011. pp. 276-278.
16. **Rodriguez JJ, Fernandez-Triana JL, Smith MA, Janzen DH, Hallwachs W et al.** Extrapolations from field studies and known faunas converge on dramatically increased estimates of global microgastrine parasitoid wasp species richness (Hymenoptera: Braconidae). *Insect Conservation and Diversity* 2013;6:530–536.
17. **Whitfield JB**. Estimating the age of the polydnavirus/braconid wasp symbiosis. *Proc Natl Acad Sci U S A* 2002;21:7508-7513.
18. **Bézier A, Annaheim M, Herbinière J, Wetterwald C, Gyapay G et al.** Polydnaviruses of braconid wasps derive from an ancestral nudivirus. *Science* 2009;323(5916):926-930.
19. **Harrison RL, Herniou EA, Bézier A, Jehle JA, Burand JP et al.** ICTV Virus Taxonomy Profile: Nudiviridae. *J Gen Virol* 2020;101(1):3-4.
20. **Bézier A, Louis F, Jancek S, Periquet G, Thézé J et al.** Functional endogenous viral elements in the genome of the parasitoid wasp *Cotesia congregata*: insights into the evolutionary dynamics of bracoviruses. *Philos Trans R Soc Lond B Biol Sci* 2013;368(1626):20130047.
21. **Strand MR**. Polydnavirus gene products that interact with the host immune system. In: Beckage NE, Drezen J-M (editors). *Parasitoid viruses symbionts and pathogens*. San Diego: Elsevier; 2012. pp. 149-161.
22. **Provost B, Varricchio P, Arana E, Espagne E, Falabella P et al.** Bracoviruses contain a large multigene family coding for protein tyrosine phosphatases. *J Virol* 2004;78(23):13090-13103.
23. **Ibrahim AMA**. *Genome Sequencing and Annotation of Toxoneuron nigriceps Bracovirus*. Thesis of University of Naples « Federico II »; 2010.
24. **Lanzrein B, Pfister-Wilhelm R, Wespi G, Roth T**. The Orchestrated Manipulation of the Host by *Chelonus Inanitus* and its Polydnavirus. In: Beckage NE, Drezen JM (editors). *Parasitoid Viruses*: Academic Press; 2012. pp. 169-178.
25. **Kaeslin M, Wehrle I, Grossniklaus-Burgin C, Wyler T, Guggisberg U et al.** Stage-dependent strategies of host invasion in the egg-larval parasitoid *Chelonus inanitus*. *J Insect Physiol* 2005;51(3):287-296.
26. **Murphy N, Banks JC, Whitfield JB, Austin AD**. Phylogeny of the parasitic microgastroid subfamilies (Hymenoptera: Braconidae) based on sequence data from seven genes, with an improved time estimate of the origin of the lineage. *Mol Phylogenet Evol* 2008;47(1):378-395.
27. **Wyder S, Tschannen A, Hochuli A, Gruber A, Saladin V et al.** Characterization of Chelonus inanitus polydnavirus segments: sequences and analysis, excision site and demonstration of clustering. *J Gen Virol* 2002;83(Pt 1):247-256.
28. **Annaheim M, Lanzrein B**. Genome organization of the *Chelonus inanitus* polydnavirus: excision sites, spacers and abundance of proviral and excised segments. *J Gen Virol* 2007;88(Pt 2):450-457.
29. **Mao M, Strand MR, Burke GR**. The Complete Genome of *Chelonus insularis* Reveals Dynamic Arrangement of Genome Components in Parasitoid Wasps That Produce Bracoviruses. *J Virol* 2022;96(5):e0157321.
30. **Belvin MP, Anderson KV**. A conserved signaling pathway: the *Drosophila* toll-dorsal pathway. *Annu Rev Cell Dev Biol* 1996;12:393-416.
31. **Falabella P, Varricchio P, Provost B, Espagne E, Ferrarese R et al.** Characterization of the IκB-like gene family in polydnaviruses associated with wasps belonging to different Braconid subfamilies. *J Gen Virol* 2007;88(Pt 1):92-104.
32. **Grossniklaus-Bürgin C, Wyler T, Pfister-Wilheim R, Lanzrein B**. Biology and morphology of the parasitoid *Chelonus inanitus* (Braconidae, Hymenoptera) and effects on the development of its host *Spodoptera littoralis* (Noctuidae, Lepidoptera). *Invertebrate Reprod and Dev* 1994;25:143-158.
33. **Asadullah A, Kunimi Y, Inoue MN, M. N**. Effect of granulovirus infection of *Spodoptera litura* (Lepidoptera: Noctuidae) larvae on development of the endoparasitoid

Chelonus inanitus (Hymenoptera: Braconidae). *Applied entomology and zoology* 2016;51(3):479-488.
34. **Albrecht U, Wyler T, Pfister-Wilhelm R, Gruber A, Stettler P et al.** Polydnavirus of the parasitic wasp *Chelonus inanitus* (Braconidae): characterization, genome organization and time point of replication. *J Gen Virol* 1994;75 3353-3363.
35. **Gallet R, Fabre F, Michalakis Y, Blanc S**. The Number of Target Molecules of the Amplification Step Limits Accuracy and Sensitivity in Ultradeep-Sequencing Viral Population Studies. *J Virol* 2017;91(16).
36. **Wang J, Van Nostrand JD, Wu L, He Z, Li G et al.** Microarray-based evaluation of whole-community genome DNA amplification methods. *Appl Environ Microbiol* 2011;77(12):4241-4245.
37. **Ellegaard KM, Klasson L, Andersson SG**. Testing the reproducibility of multiple displacement amplification on genomes of clonal endosymbiont populations. *PLoS One* 2013;8(11):e82319.
38. **Chin-inmanu K, Suttitheptumrong A, Sangsrakru D, Tangphatsornruang S, Tragoonrung S et al.** Feasibility of using 454 pyrosequencing for studying quasispecies of the whole dengue viral genome. *BMC Genomics* 2012;13 Suppl 7:S7.
39. **Katoh K, Standley DM**. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30(4):772-780.
40. **Subramanian AR, Weyer-Menkhoff J, Kaufmann M, Morgenstern B**. DIALIGN-T: an improved algorithm for segment-based multiple sequence alignment. *BMC Bioinformatics* 2005;6:66.
41. **Louis F, Bézier A, Periquet G, Ferras C, Drezen JM et al.** The bracovirus genome of the parasitoid wasp *Cotesia congregata* is amplified within 13 replication units, including sequences not packaged in the particles. *J Virol* 2013;87(17):9649-9660.
42. **Gruber A, Stettler P, Heiniger P, Schumperli D, Lanzrein B**. Polydnavirus DNA of the braconid wasp *Chelonus inanitus* is integrated in the wasp genome and excised only in later pupal and adult stages of the female. *J GenVirol* 1996;77:2873-2879.
43. **Savary S, Beckage NE, Tan F, Periquet G, Drezen JM**. Excision of the polydnavirus chromosomal integrated EP1 sequence of the parasitoid wasp *Cotesia congregata* (Braconidae, Microgastrinae) at potential recombinase binding sites. *J Gen Virol* 1997;78:3125-3134.
44. **Corpet F**. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 1988;16(22):10881-10890.
45. **Bailey TL, Johnson J, Grant CE, Noble WS**. The MEME Suite. *Nucleic Acids Res* 2015;43(W1):W39-49.
46. **Crooks GE, Hon G, Chandonia JM, Brenner SE**. WebLogo: a sequence logo generator. *Genome Res* 2004;14(6):1188-1190.
47. **Cui L, Webb BA**. Homologous sequences in the Campoletis sonorensis polydnavirus genome are implicated in replication and nesting of the W segment family. *J Virol* 1997;71(11):8504-8513.
48. **Legeai F, Santos BF, Robin S, Bretaudeau A, Dikow RB et al.** Genomic architecture of endogenous ichnoviruses reveals distinct evolutionary pathways leading to virus domestication in parasitic wasps. *BMC Biol* 2020;18(1):89.
49. **Desjardins CA, Gundersen-Rindal DE, Hostetler JB, Tallon LJ, Fadrosh DW et al.** Comparative genomics of mutualistic viruses of *Glyptapanteles* parasitic wasps. *Genome Biol* 2008;9(12):R183.
50. **Espagne E, Dupuy C, Huguet E, Cattolico L, Provost B et al.** Genome sequence of a polydnavirus: insights into symbiotic virus evolution. *Science* 2004;306(5694):286-289.
51. **Webb BA, Strand MR, Dickey SE, Beck MH, Hilgarth RS et al.** Polydnavirus genomes reflect their dual roles as mutualists and pathogens. *Virology* 2006;347(1):160-174.
52. **Chen YF, Gao F, Ye XQ, Wei SJ, Shi M et al.** Deep sequencing of Cotesia vestalis bracovirus reveals the complexity of a polydnavirus genome. *Virology* 2011;414(1):42-50.

53. **Pruijssers AJ, Falabella P, Eum JH, Pennacchio F, Brown MR et al.** Infection by a symbiotic polydnavirus induces wasting and inhibits metamorphosis of the moth *Pseudoplusia includens*. *J Exp Biol* 2009;212(18):2998-3006.
54. **Ramjan Ali M, Kim Y**. A novel polydnaviral gene family, BEN, and its immunosuppressive function in larvae of *Plutella xylostella* parasitized by *Cotesia plutellae*. *J Invertebr Pathol* 2012;110(3):389-397.
55. **Thoetkiattikul H, Beck MH, Strand MR**. Inhibitor kappaB-like proteins from a polydnavirus inhibit NF-κB activation and suppress the insect immune response. *Proc Natl Acad Sci U S A* 2005;102(32):11426-11431.
56. **Harwood SH, Grosovsky AJ, Cowles EA, Davis JW, Beckage NE**. An abundantly expressed hemolymph glycoprotein isolated from newly parasitized *Manduca sexta* larvae is a polydnavirus gene product. *Virology* 1994; 205:381-392.
57. **Glatz R, Schmidt O, Asgari S**. Characterization of a novel protein with homology to C-type lectins expressed by the Cotesia rubecula bracovirus in larvae of the lepidopteran host, Pieris rapae. *J Biol Chem* 2003:M301396200.
58. **Zhang HH, Zhou QZ, Wang PL, Xiong XM, Luchetti A et al.** Unexpected invasion of miniature inverted-repeat transposable elements in viral genomes. *Mob DNA* 2018;9:19.
59. **Falabella P, Varricchio P, Gigliotti S, Tranfaglia A, Pennacchio F et al.** Toxoneuron nigriceps polydnavirus encodes a putative aspartyl protease highly expressed in parasitized host larvae. *Insect Mol Biol* 2003;12(1):9-17.
60. **Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ**. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc Natl Acad Sci U S A* 2006;103(26):9935-9939.
61. **Lange A, Patel PH, Heames B, Damry AM, Saenger T et al.** Structural and functional characterization of a putative de novo gene in *Drosophila*. *Nat Commun* 2021;12(1):1667.
62. **Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA et al.** Proto-genes and de novo gene birth. *Nature* 2012;487(7407):370-374.
63. **Legendre M, Fabre E, Poirot O, Jeudy S, Lartigue A et al.** Diversity and evolution of the emerging Pandoraviridae family. *Nat Commun* 2018;9(1):2285.
64. **Beck MH, Zhang S, Bitra K, Burke GR, Strand MR**. The encapsidated genome of Microplitis demolitor bracovirus integrates into the host *Pseudoplusia includens*. *J Virol* 2011;85(22):11685-11696.
65. **Desjardins CA, Gundersen-Rindal DE, Hostetler JB, Tallon LJ, Fuester RW et al.** Structure and evolution of a proviral locus of Glyptapanteles indiensis bracovirus. *BMC Microbiol* 2007;7:61.
66. **Chevignon G, Periquet G, Gyapay G, Vega-Czarny N, Musset K et al.** Cotesia congregata Bracovirus Circles Encoding PTP and Ankyrin Genes Integrate into the DNA of Parasitized *Manduca sexta* Hemocytes. *J Virol* 2018;92(15).
67. **Wang Z, Ye X, Zhou Y, Wu X, Hu R et al.** Bracoviruses recruit host integrases for their integration into caterpillar's genome. *PLoS Genet* 2021;17(9):e1009751.
68. **Muller H, Chebbi MA, Bouzar C, Periquet G, Fortuna T et al.** Genome-Wide Patterns of Bracovirus Chromosomal Integration into Multiple Host Tissues during Parasitism. *J Virol* 2021;95(22):e0068421.
69. **Serbielle C, Dupas S, Perdereau E, Hericourt F, Dupuy C et al.** Evolutionary mechanisms driving the evolution of a large polydnavirus gene family coding for protein tyrosine phosphatases. *BMC Evol Biol* 2012;12:253.
70. **Lee JA, Carvalho CM, Lupski JR**. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007;131(7):1235-1247.
71. **Jancek S, Bézier A, Gayral P, Paillusson C, Kaiser L et al.** Adaptive selection on bracovirus genomes drives the specialization of *Cotesia* parasitoid wasps. *PLoS One* 2013;8(5):e64432.
72. **Drezen JM, Gauthier J, Josse T, Bézier A, Herniou E et al.** Foreign DNA acquisition by invertebrate genomes. *J Invertebr Pathol* 2017;147:157-168.

73. **Quan W-L, Zheng XL, Li XX, Zhou XM, Ma WH et al.** Overwintering strategy of endoparasitoids in *Chilo suppressalis*: a perspective from the cold hardiness of a host *Entomologia Experimentalis and applicata* 2013;146:398–403.
74. **Schneider SE, Thomas JH**. Accidental genetic engineers: horizontal sequence transfer from parasitoid wasps to their lepidopteran hosts. *PLoS One* 2014;9(10):e109446.
75. **Gasmi L, Boulain H, Gauthier J, Hua-Van A, Musset K et al.** Recurrent domestication by Lepidoptera of genes from their parasites mediated by bracoviruses. *PLoS Genet* 2015;11(9):e1005470.
76. **Gasmi L, Ferre J, Herrero S**. High bacterial agglutination activity in a single-CRD C-type lectin from *Spodoptera exigua* (Lepidoptera: Noctuidae). *Biosensors (Basel)* 2017;7(1).
77. **Huxford T, Huang DB, Malek S, Ghosh G**. The crystal structure of the IκBα/NF-κB complex reveals mechanisms of NF-κB inactivation. *Cell* 1998;95(6):759-770.
78. **Gay NJ, Ntwasa M**. The *Drosophila* ankyrin repeat protein cactus has a predominantly alpha-helical secondary structure. *FEBS Lett* 1993;335(2):155-160.
79. **Chen Z, Hagler J, Palombella VJ, Melandri F, Scherer D et al.** Signal-induced site-specific phosphorylation targets IκB alpha to the ubiquitin-proteasome pathway. *Genes Dev* 1995;9(13):1586-1597.
80. **Belvin MP, Jin Y, Anderson KV**. Cactus protein degradation mediates *Drosophila* dorsal-ventral signaling. *Genes Dev* 1995;9(7):783-793.
81. **Kroemer JA, Webb BA**. IκB-related vankyrin genes in the Campoletis sonorensis ichnovirus: temporal and tissue-specific patterns of expression in parasitized *Heliothis virescens* lepidopteran hosts. *J Virol* 2005;79(12):7617-7628.
82. **Shumway SD, Maki M, Miyamoto S**. The PEST domain of IκBα is necessary and sufficient for *in vitro* degradation by mu-calpain. *J Biol Chem* 1999;274(43):30874-30881.
83. **Chevignon G, Thézé J, Cambier S, Poulain J, Da Silva C et al.** Functional annotation of Cotesia congregata bracovirus: identification of the viral genes expressed in parasitized host immune tissues. *J Virol* 2014;88:8795-8812.
84. **Lapointe R, Wilson R, Vilaplana L, O'Reilly DR, Falabella P et al.** Expression of a Toxoneuron nigriceps polydnavirus-encoded protein causes apoptosis-like programmed cell death in lepidopteran insect cells. *J Gen Virol* 2005;86(Pt 4):963-971.
85. **Fath-Goodin A, Kroemer JA, Webb BA**. The Campoletis sonorensis ichnovirus vankyrin protein P-vank-1 inhibits apoptosis in insect Sf9 cells. *Insect Mol Biol* 2009;18(4):497-506.
86. **Valzania L, Romani P, Tian L, Li S, Cavaliere V et al.** A polydnavirus ANK protein acts as virulence factor by disrupting the function of prothoracic gland steroidogenic cells. *PLoS One* 2014;9(4):e95104.
87. **Grossniklaus-Burgin C, Pfister-Wilhelm R, Meyer V, Treiblmayr K, Lanzrein B**. Physiological and endocrine changes associated with polydnavirus/venom in the parasitoid-host system *Chelonus inanitus-Spodoptera littoralis*. *J Insect Physiol* 1998;44(3-4):305-321.
88. **Steele KH, Stone BJ, Franklin KM, Fath-Goodin A, Zhang X et al.** Improving the baculovirus expression vector system with vankyrin-enhanced technology. *Biotechnol Prog* 2017;33(6):1496-1507.
89. **Germani F, Hain D, Sternlicht D, Moreno E, Basler K**. The Toll pathway inhibits tissue growth and regulates cell fitness in an infection-dependent manner. *Elife* 2018;7.
90. **Salvia R, Grossi G, Amoresano A, Scieuzo C, Nardiello M et al.** The multifunctional polydnavirus TnBVANK1 protein: impact on host apoptotic pathway. *Sci Rep* 2017;7(1):11775.
91. **Gueguen G, Kalamarz ME, Ramroop J, Uribe J, Govind S**. Polydnaviral ankyrin proteins aid parasitic wasp survival by coordinate and selective inhibition of hematopoietic and immune NFκB signaling in insect hosts. *PLoS Pathog* 2013;9(8):e1003580.
92. **Colinet D, Schmitz A, Depoix D, Crochard D, Poirie M**. Convergent use of RhoGAP toxins by eukaryotic parasites and bacterial pathogens. *PLoS Pathog* 2007;3(12):e203.
93. **Li M, Au LYC, Douglah D, Chong A, White BJ et al.** Generation of heritable germline mutations in the jewel wasp *Nasonia vitripennis* using CRISPR/Cas9. *Sci Rep* 2017;7(1):901.

94. **Bézier A, Herbinière J, Serbielle C, Lesobre J, Wincker P et al.** Bracovirus gene products are highly divergent from insect proteins. *Arch Insect Biochem Physiol* 2008;67(4):172-187.
95. **Serbielle C, Chowdhury S, Pichon S, Dupas S, Lesobre J et al.** Viral cystatin evolution and three-dimensional structure modelling: a case of directional selection acting on a viral protein involved in a host-parasitoid interaction. *BMC Biol* 2008;6:38.
96. **Espagne E, Douris V, Lalmanach G, Provost B, Cattolico L et al.** A virus essential for insect host-parasite interactions encodes cystatins. *J Virol* 2005;79(15):9765-9776.
97. **Hancks DC, Kazazian HH, Jr.** Roles for retrotransposon insertions in human disease. *Mob DNA* 2016;7:9.
98. **Huguet E, Serbielle C, Moreau SJM**. Evolution and origin of polydnavirus virulence genes. In: Beckage NE, Drezen J-M (editors). *Parasitoid viruses symbionts and pathogens*. San Diego: Elsevier; 2012. pp. 63-78.
99. **Sharanowski BJ, Ridenbaugh RD, Piekarski PK, Broad GR, Burke GR et al.** Phylogenomics of Ichneumonoidea (Hymenoptera) and implications for evolution of mode of parasitism and viral endogenization. *Mol Phylogenet Evol* 2021;156:107023.
100. **Beliveau C, Cohen A, Stewart D, Periquet G, Djoumad A et al.** Genomic and proteomic analyses indicate that banchine and campoplegine polydnaviruses have similar, if not identical, viral ancestors. *J Virol* 2015;89(17):8909-8921.
101. **Burke GR, Hines HM, Sharanowski BJ**. The presence of ancient core genes reveals endogenization from diverse viral ancestors in parasitoid wasps. *Genome Biol Evol* 2021;13(7).
102. **Volkoff AN, Jouan V, Urbach S, Samain S, Bergoin M et al.** Analysis of virion structural components reveals vestiges of the ancestral ichnovirus genome. *PLoS Pathog* 2010;6(5):e1000923.
103. **Wang ZH, Zhou YN, Yang J, Ye XQ, Shi M et al.** Genome-wide profiling of Diadegma semiclausum ichnovirus integration in parasitized *Plutella xylostella* hemocytes identifies host integration motifs and insertion sites. *Front Microbiol* 2020;11:608346.

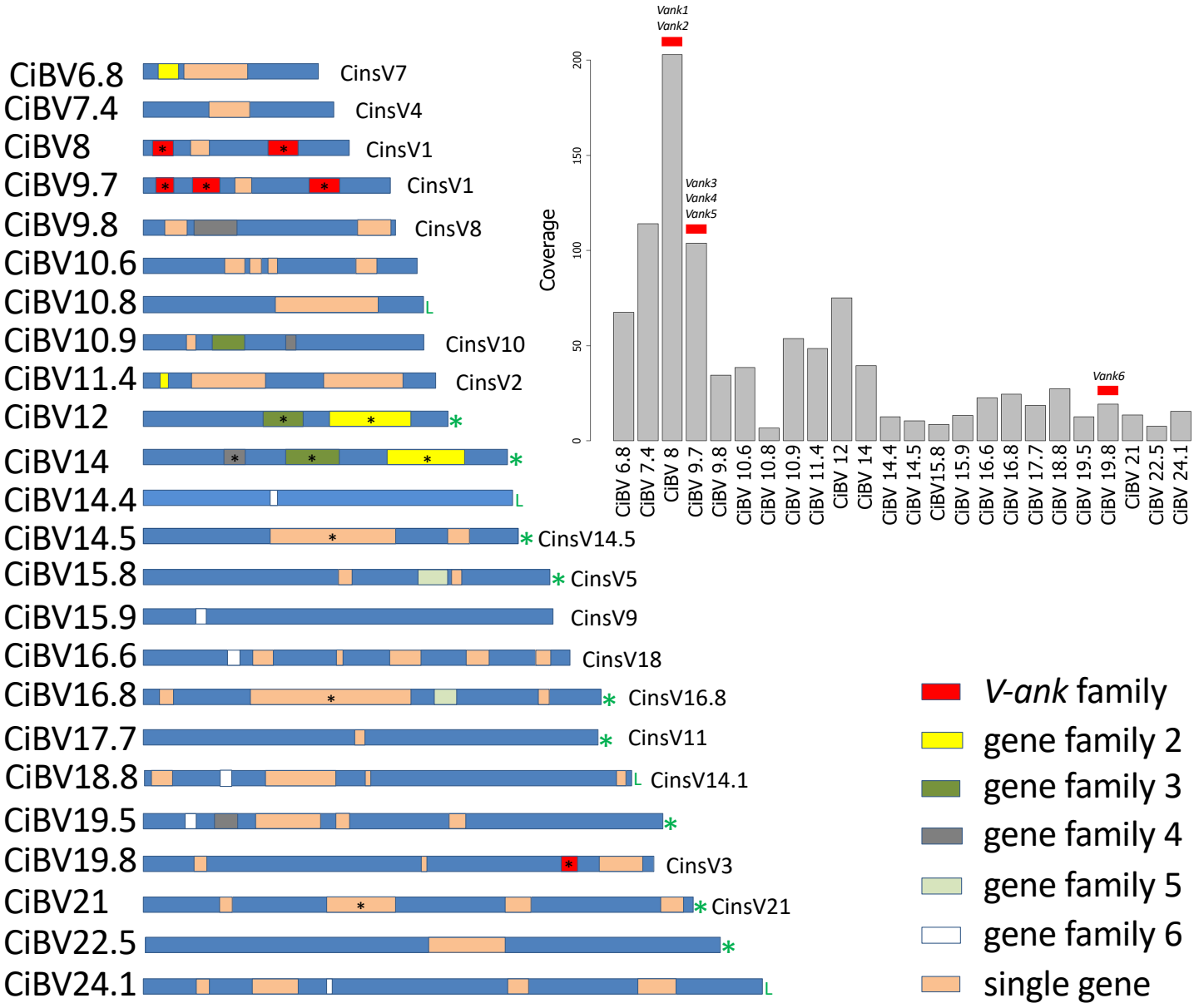
CiBV6.8
CinsV7
CiBV7.4
CinsV4
CiBV8
CinsV1
CiBV9.7
CinsV1
CiBV9.8
CinsV8
CiBV10.6
CiBV10.8
CiBV10.9
CinsV10
CiBV11.4
CinsV2
CiBV12
CiBV14
CiBV14.4
CiBV14.5
CinsV14.5
CiBV15.8
CinsV5
CiBV15.9
CinsV9
CiBV16.6
CinsV18
CiBV16.8
CinsV16.8
CiBV17.7
CinsV11
CiBV18.8
CinsV14.1
CiBV19.5
CiBV19.8
CinsV3
CiBV21
CinsV21
CiBV22.5
CiBV24.1
Vank1
Vank2
Vank3
Vank4
Vank5
Vank6
Coverage
0
50
100
150
200
CiBV 6.8
CiBV 7.4
CiBV 8
CiBV 9.7
CiBV 9.8
CiBV 10.6
CiBV 10.8
CiBV 10.9
CiBV 11.4
CiBV 12
CiBV 14
CiBV 14.4
CiBV 14.5
CiBV15.8
CiBV 15.9
CiBV 16.6
CiBV 16.8
CiBV 17.7
CiBV 18.8
CiBV 19.5
CiBV 19.8
CiBV 21
CiBV 22.5
CiBV 24.1
V-ank family
gene family 2
gene family 3
gene family 4
gene family 5
gene family 6
single gene

Figure 1

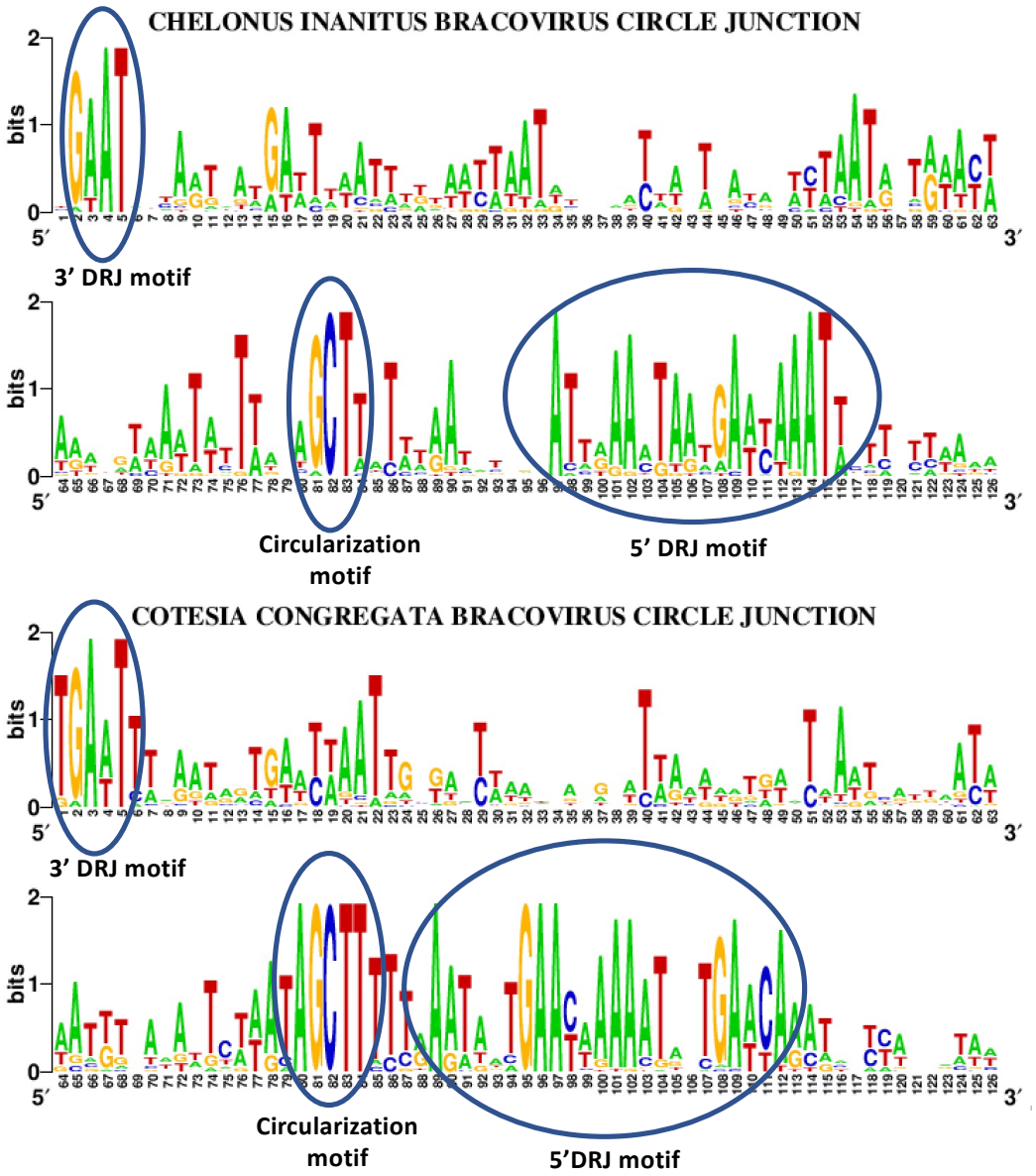
CHELONUS INANITUS BRACOVIRUS CIRCLE JUNCTION
bits
5′
3′
3’ DRJ motif
Circularization motif
5’ DRJ motif
COTESIA CONGREGATA BRACOVIRUS CIRCLE JUNCTION
3’ DRJ motif
Circularization motif
5’DRJ motif

Figure 2

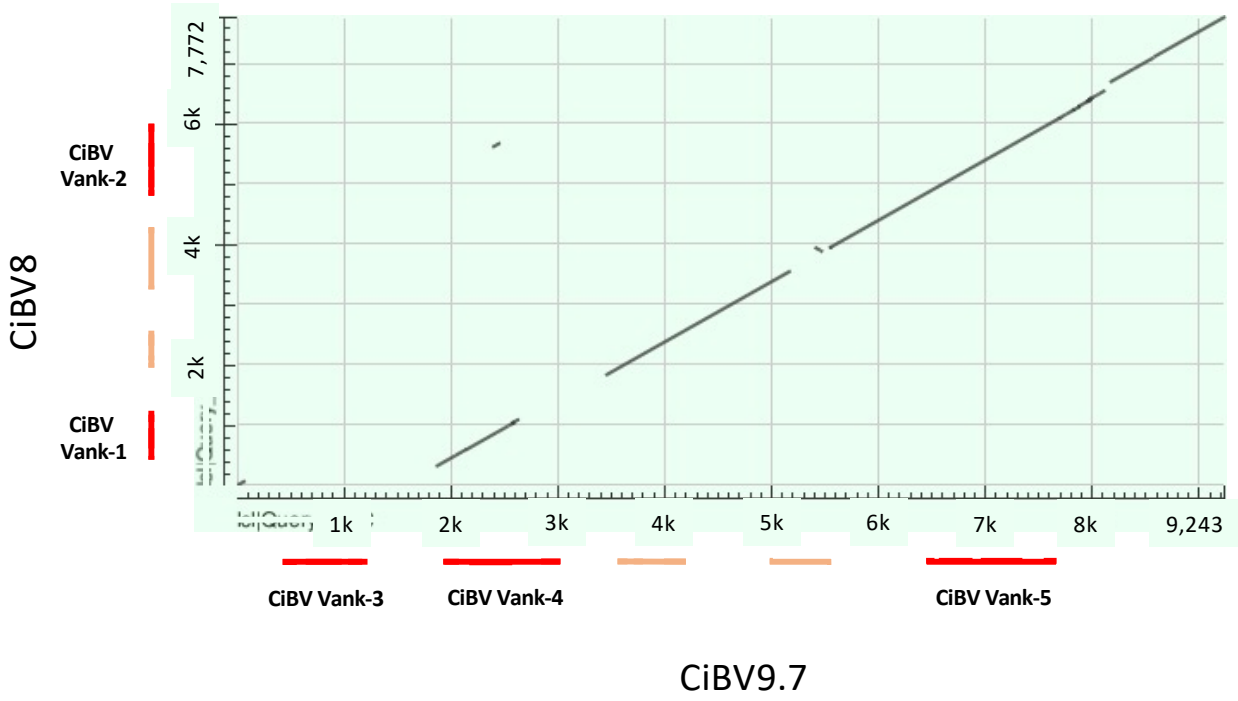
CiBV8
7,772
6k
4k
2k
CiBV Vank-2
CiBV Vank-1
1k
2k
3k
4k
5k
6k
7k
8k
9,243
CiBV Vank-3
CiBV Vank-4
CiBV Vank-5
CiBV9.7

Figure 3

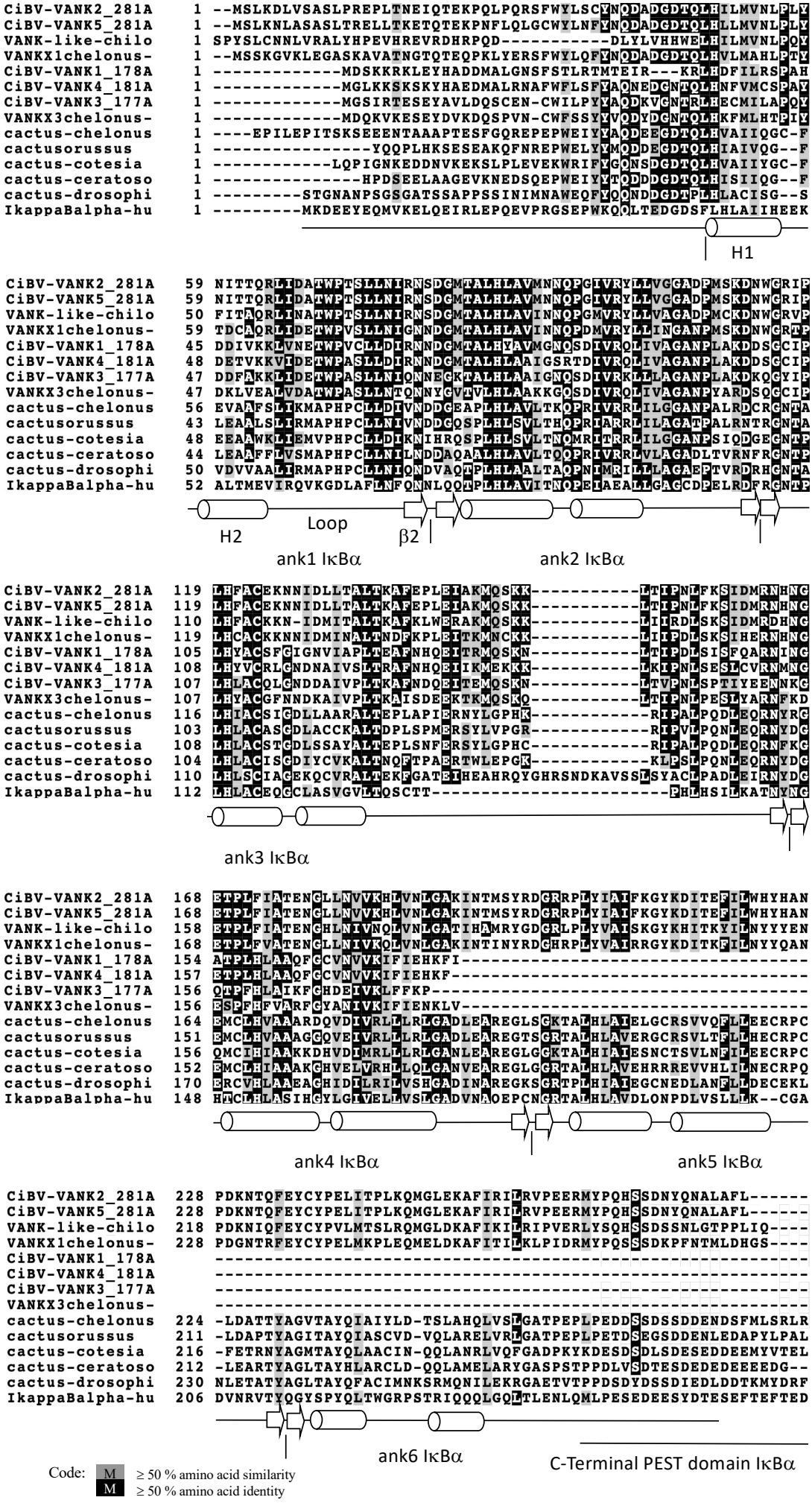
CiBV-VANK2_281A
CiBV-VANK5_281A
VANK-like-chilo
VANKX1chelonus-
CiBV-VANK1_178A
CiBV-VANK4_181A
CiBV-VANK3_177A
VANKX3chelonus-
cactus-chelonus
cactusorussus
cactus-cotesia
cactus-ceratoso
cactus-drosophi
IkappaBalpha-hu
H1
H2
Loop
β2
ank1 IκBα
ank2 IκBα
ank3 IκBα
ank4 IκBα
ank5 IκBα
ank6 IκBα
C-Terminal PEST domain IκBα
Code:
≥ 50 % amino acid similarity
≥ 50 % amino acid identity

Figure 4

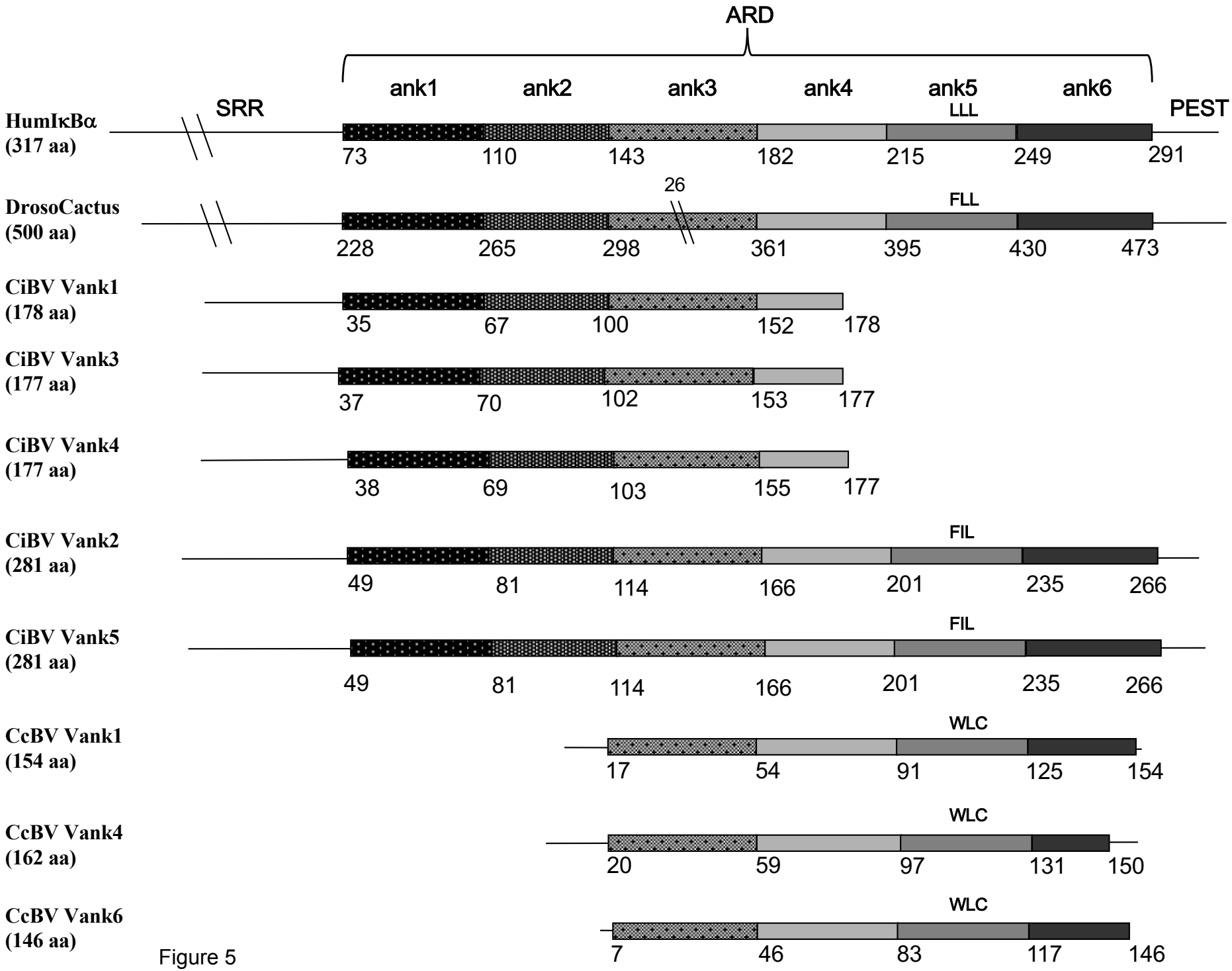
ARD
ank1
ank2
ank3
ank4
ank5
ank6
SRR
LLL
PEST
HumIκBα
(317 aa)
73
110
143
182
215
249
291
26
FLL
DrosoCactus
(500 aa)
228
265
298
361
395
430
473
CiBV Vank1
(178 aa)
35
67
100
152
178
CiBV Vank3
(177 aa)
37
70
102
153
177
CiBV Vank4
(177 aa)
38
69
103
155
177
FIL
CiBV Vank2
(281 aa)
49
81
114
166
201
235
266
FIL
CiBV Vank5
(281 aa)
49
81
114
166
201
235
266
WLC
CcBV Vank1
(154 aa)
17
54
91
125
154
WLC
CcBV Vank4
(162 aa)
20
59
97
131
150
WLC
CcBV Vank6
(146 aa)
7
46
83
117
146

Figure 5

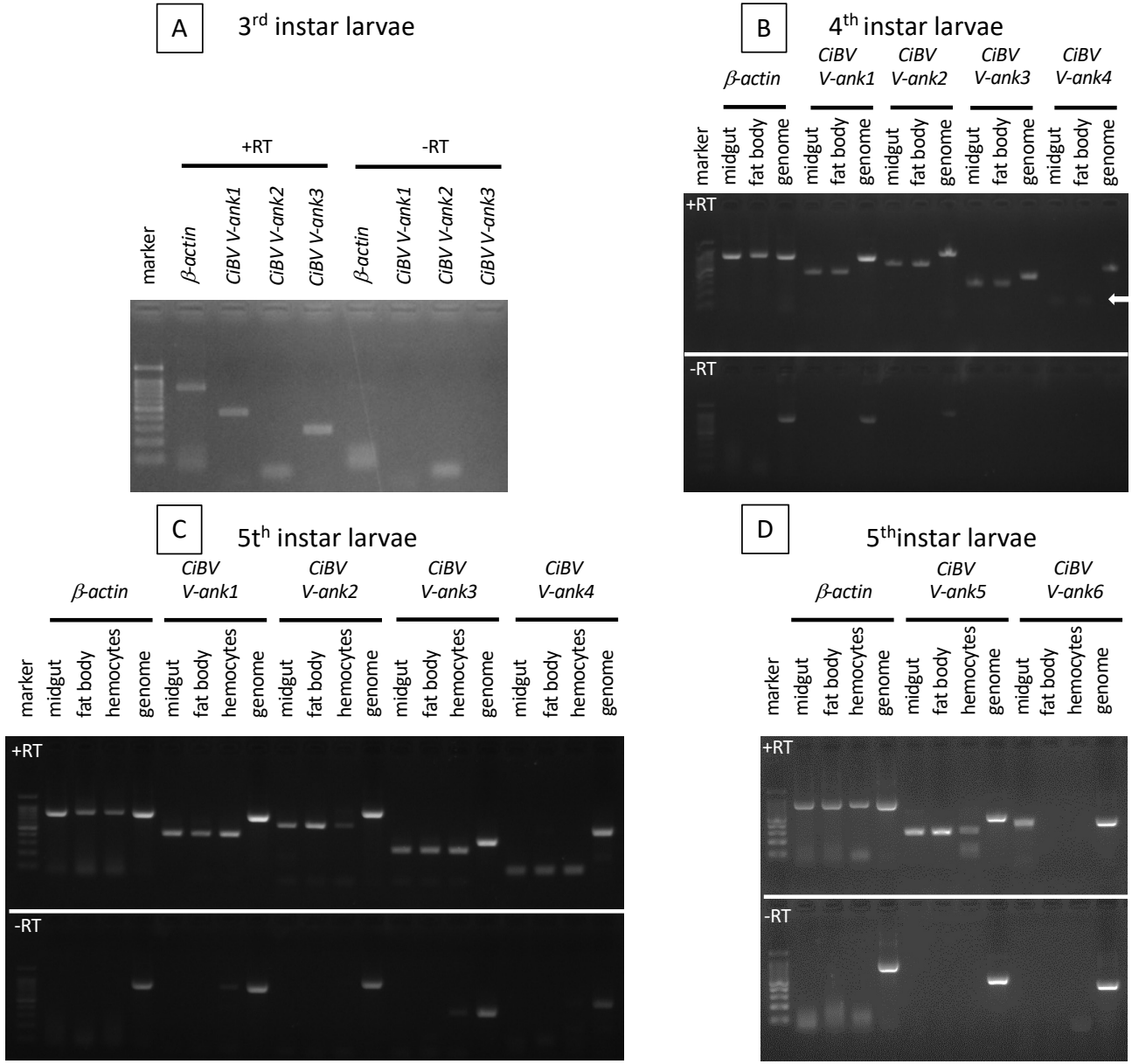
A
3rd instar larvae
+RT
-RT
marker
β-actin
CiBV V-ank1
CiBV V-ank2
CiBV V-ank3
B
4th instar larvae
CiBV V-ank4
midgut
fat body
genome
C
5th instar larvae
hemocytes
D
5thinstar larvae
CiBV V-ank5
CiBV V-ank6

Figure 6


0.20 0.40 0.60 0.80 1.0 1.2 1.4 1.6 1.8 2.0
XP_021703819_Aedes_aegypti_Cactus
AAA85908_Drosophila_melanogaster_Cactus
XP_037814227_Lucilia_sericata_Cactus
XP_026478858_Ctenocephalides_felis_Cactus
D910_04204_Dendroctonus_ponderosae_Cactus
NP_001157183_Tribolium_castaneum_Cactus
CAD6208429_Cotesia_congregata_Cactus
XP_014299284_Microplitis_demolitor_Cactus
XP_012283790_Orussus_abietinus_Cactus
KOX75348_Melipona_quadrifasciata_Cactus
XP_017755869_Eufriesea_mexicana_Cactus
XP_035732908_Vespa_mandarinia_Cactus
XP_015122357_Diachasma_alloeum_Cactus
XP_011304926_Fopius_arisanus_Cactus
GFG39294_Coptotermes_formosanus_Cactus
XP_021919362_Zootermopsis_nevadensis_Cactus
PSN55380_Blattella_germanica_Cactus
AWJ58417_Locusta_migratoria_Cactus
NP_001166191_Bombyx_mori_Cactus
XP_030020657_Manduca_sexta_Cactus
XP_031762805_Galleria_mellonella_Cactus
XP_022129983_Pieris_rapae_Cactus
XP_034934298_VANKX1_Chelonus_insularis
RVE47639_VANKlike_Chilo_suppressalis
CiBV_VANK5_Chelous_inanitus
XP_018900556_Bemisia_tabaci_Cactus
XP_007436766_Python_bivittatus_IkB_alpha
XP_007436766_Corvus_cornix_IkB_alpha
XP_004270018_Orcinus_orca_IkB_alpha
NP_065390_Homo_sapiens_IkB_alpha
XP_007429861_Python_bivittatus_NF_kB_p105
XP_039406723_Corvus_cornix_NF_kB_p105
XP_033258008_Orcinus_orca_NF_kB_p105
NP_001369557_Homo_sapiens_NF_kB_p105
XP_007429544_Python_bivittatus_IkB_epsilon
XP_039424387_Corvus_cornix_IkB_epsilon
XP_004267574_Orcinus_orca_IkB_epsilon
NP_004547_Homo_sapiens_IkB_epsilon
NP_002805_Homo_sapiens_Gankyrin
Insect
Cactus
Bootstrap support: *>0.6 ** >0.8

Figure 7

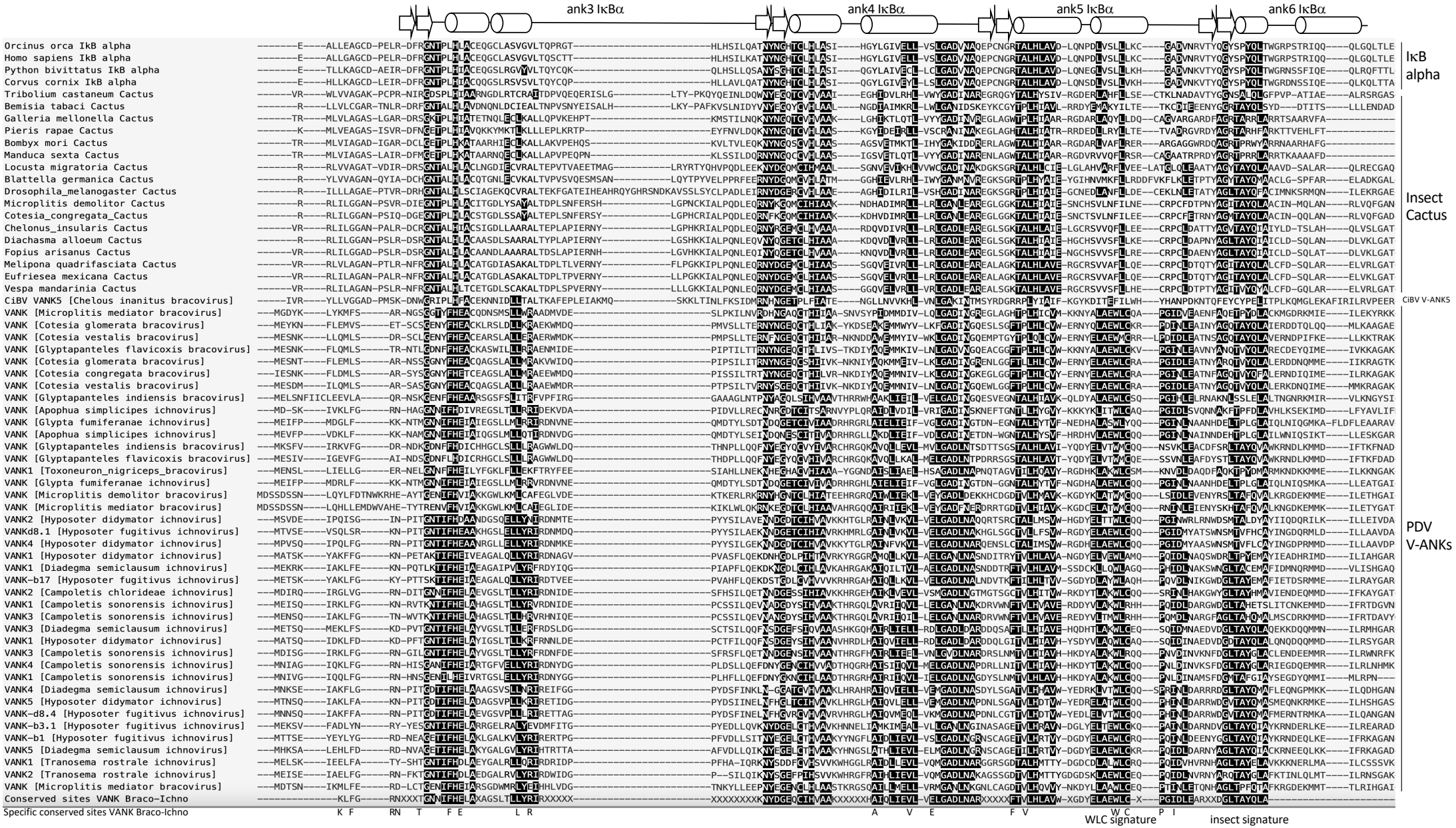
ank3 IκBα
ank4 IκBα
ank5 IκBα
ank6 IκBα
Orcinus orca IkB alpha
Homo sapiens IkB alpha
Python bivittatus IkB alpha
Corvus cornix IkB alpha
Tribolium castaneum Cactus
Bemisia tabaci Cactus
Galleria mellonella Cactus
Pieris rapae Cactus
Bombyx mori Cactus
Manduca sexta Cactus
Locusta migratoria Cactus
Blattella germanica Cactus
Drosophila_melanogaster Cactus
Microplitis demolitor Cactus
Cotesia_congregata_Cactus
Chelonus_insularis Cactus
Diachasma alloeum Cactus
Fopius arisanus Cactus
Melipona quadrifasciata Cactus
Eufriesea mexicana Cactus
Vespa mandarinia Cactus
CiBV VANK5 [Chelous inanitus bracovirus]
VANK [Microplitis mediator bracovirus]
VANK [Cotesia glomerata bracovirus]
VANK [Cotesia vestalis bracovirus]
VANK [Glyptapanteles flavicoxis bracovirus]
VANK [Cotesia glomerata bracovirus]
VANK [Cotesia congregata bracovirus]
VANK [Cotesia vestalis bracovirus]
VANK [Glyptapanteles indiensis bracovirus]
VANK [Apophua simplicipes ichnovirus]
VANK [Glypta fumiferanae ichnovirus]
VANK [Apophua simplicipes ichnovirus]
VANK [Glyptapanteles indiensis bracovirus]
VANK [Glyptapanteles flavicoxis bracovirus]
VANK1 [Toxoneuron_nigriceps_bracovirus]
VANK [Glypta fumiferanae ichnovirus]
VANK [Microplitis demolitor bracovirus]
VANK [Microplitis mediator bracovirus]
VANK2 [Hyposoter didymator ichnovirus]
VANKd8.1 [Hyposoter fugitivus ichnovirus]
VANK4 [Hyposoter didymator ichnovirus]
VANK1 [Hyposoter didymator ichnovirus]
VANK1 [Diadegma semiclausum ichnovirus]
VANK-b17 [Hyposoter fugitivus ichnovirus]
VANK2 [Campoletis chlorideae ichnovirus]
VANK1 [Campoletis sonorensis ichnovirus]
VANK3 [Campoletis sonorensis ichnovirus]
VANK3 [Diadegma semiclausum ichnovirus]
VANK1 [Hyposoter didymator ichnovirus]
VANK3 [Campoletis sonorensis ichnovirus]
VANK4 [Campoletis sonorensis ichnovirus]
VANK1 [Campoletis sonorensis ichnovirus]
VANK4 [Diadegma semiclausum ichnovirus]
VANK5 [Hyposoter didymator ichnovirus]
VANK-d8.4 [Hyposoter fugitivus ichnovirus]
VANK-b3.1 [Hyposoter fugitivus ichnovirus]
VANK-b1 [Hyposoter fugitivus ichnovirus]
VANK5 [Diadegma semiclausum ichnovirus]
VANK1 [Tranosema rostrale ichnovirus]
VANK2 [Tranosema rostrale ichnovirus]
VANK [Microplitis mediator bracovirus]
Conserved sites VANK Braco-Ichno
Specific conserved sites VANK Braco-Ichno
IκB alpha
Insect Cactus
CiBV V-ANK5
PDV V-ANKs
WLC signature
insect signature

Figure 8

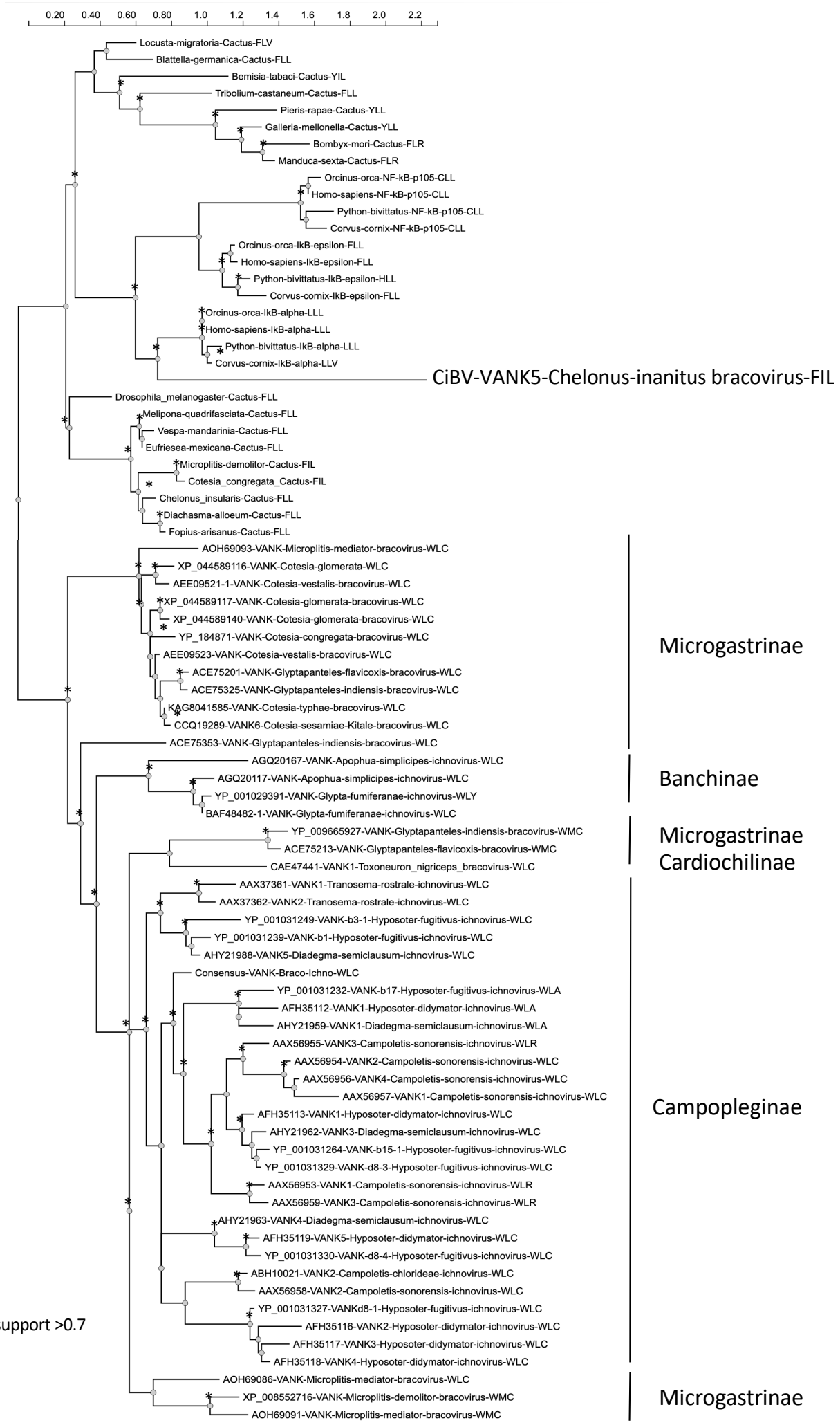
0.20 0.40 0.60 0.80 1.0 1.2 1.4 1.6 1.8 2.0 2.2
Locusta-migratoria-Cactus-FLV
Blattella-germanica-Cactus-FLL
Bemisia-tabaci-Cactus-YIL
Tribolium-castaneum-Cactus-FLL
Pieris-rapae-Cactus-YLL
Galleria-mellonella-Cactus-YLL
Bombyx-mori-Cactus-FLR
Manduca-sexta-Cactus-FLR
Orcinus-orca-NF-kB-p105-CLL
Homo-sapiens-NF-kB-p105-CLL
Python-bivittatus-NF-kB-p105-CLL
Corvus-cornix-NF-kB-p105-CLL
Orcinus-orca-IkB-epsilon-FLL
Homo-sapiens-IkB-epsilon-FLL
Python-bivittatus-IkB-epsilon-HLL
Corvus-cornix-IkB-epsilon-FLL
Orcinus-orca-IkB-alpha-LLL
Homo-sapiens-IkB-alpha-LLL
Python-bivittatus-IkB-alpha-LLL
Corvus-cornix-IkB-alpha-LLV
CiBV-VANK5-Chelonus-inanitus bracovirus-FIL
Drosophila_melanogaster-Cactus-FLL
Melipona-quadrifasciata-Cactus-FLL
Vespa-mandarinia-Cactus-FLL
Eufriesea-mexicana-Cactus-FLL
Microplitis-demolitor-Cactus-FIL
Cotesia_congregata_Cactus-FIL
Chelonus_insularis-Cactus-FLL
Diachasma-alloeum-Cactus-FLL
Fopius-arisanus-Cactus-FLL
AOH69093-VANK-Microplitis-mediator-bracovirus-WLC
XP_044589116-VANK-Cotesia-glomerata-WLC
AEE09521-1-VANK-Cotesia-vestalis-bracovirus-WLC
XP_044589117-VANK-Cotesia-glomerata-bracovirus-WLC
XP_044589140-VANK-Cotesia-glomerata-bracovirus-WLC
YP_184871-VANK-Cotesia-congregata-bracovirus-WLC
AEE09523-VANK-Cotesia-vestalis-bracovirus-WLC
ACE75201-VANK-Glyptapanteles-flavicoxis-bracovirus-WLC
ACE75325-VANK-Glyptapanteles-indiensis-bracovirus-WLC
KAG8041585-VANK-Cotesia-typhae-bracovirus-WLC
CCQ19289-VANK6-Cotesia-sesamiae-Kitale-bracovirus-WLC
Microgastrinae
ACE75353-VANK-Glyptapanteles-indiensis-bracovirus-WLC
AGQ20167-VANK-Apophua-simplicipes-ichnovirus-WLC
AGQ20117-VANK-Apophua-simplicipes-ichnovirus-WLC
YP_001029391-VANK-Glypta-fumiferanae-ichnovirus-WLY
BAF48482-1-VANK-Glypta-fumiferanae-ichnovirus-WLC
Banchinae
YP_009665927-VANK-Glyptapanteles-indiensis-bracovirus-WMC
ACE75213-VANK-Glyptapanteles-flavicoxis-bracovirus-WMC
CAE47441-VANK1-Toxoneuron_nigriceps_bracovirus-WLC
Microgastrinae
Cardiochilinae
AAX37361-VANK1-Tranosema-rostrale-ichnovirus-WLC
AAX37362-VANK2-Tranosema-rostrale-ichnovirus-WLC
YP_001031249-VANK-b3-1-Hyposoter-fugitivus-ichnovirus-WLC
YP_001031239-VANK-b1-Hyposoter-fugitivus-ichnovirus-WLC
AHY21988-VANK5-Diadegma-semiclausum-ichnovirus-WLC
Consensus-VANK-Braco-Ichno-WLC
YP_001031232-VANK-b17-Hyposoter-fugitivus-ichnovirus-WLA
AFH35112-VANK1-Hyposoter-didymator-ichnovirus-WLA
AHY21959-VANK1-Diadegma-semiclausum-ichnovirus-WLA
AAX56955-VANK3-Campoletis-sonorensis-ichnovirus-WLR
AAX56954-VANK2-Campoletis-sonorensis-ichnovirus-WLC
AAX56956-VANK4-Campoletis-sonorensis-ichnovirus-WLC
AAX56957-VANK1-Campoletis-sonorensis-ichnovirus-WLC
AFH35113-VANK1-Hyposoter-didymator-ichnovirus-WLC
AHY21962-VANK3-Diadegma-semiclausum-ichnovirus-WLC
YP_001031264-VANK-b15-1-Hyposoter-fugitivus-ichnovirus-WLC
YP_001031329-VANK-d8-3-Hyposoter-fugitivus-ichnovirus-WLC
AAX56953-VANK1-Campoletis-sonorensis-ichnovirus-WLR
AAX56959-VANK3-Campoletis-sonorensis-ichnovirus-WLR
AHY21963-VANK4-Diadegma-semiclausum-ichnovirus-WLC
AFH35119-VANK5-Hyposoter-didymator-ichnovirus-WLC
YP_001031330-VANK-d8-4-Hyposoter-fugitivus-ichnovirus-WLC
ABH10021-VANK2-Campoletis-chlorideae-ichnovirus-WLC
AAX56958-VANK2-Campoletis-sonorensis-ichnovirus-WLC
YP_001031327-VANKd8-1-Hyposoter-fugitivus-ichnovirus-WLC
AFH35116-VANK2-Hyposoter-didymator-ichnovirus-WLC
AFH35117-VANK3-Hyposoter-didymator-ichnovirus-WLC
AFH35118-VANK4-Hyposoter-didymator-ichnovirus-WLC
Campopleginae
AOH69086-VANK-Microplitis-mediator-bracovirus-WLC
XP_008552716-VANK-Microplitis-demolitor-bracovirus-WMC
AOH69091-VANK-Microplitis-mediator-bracovirus-WMC
Microgastrinae
* :Bootstrap support >0.7

Figure 9

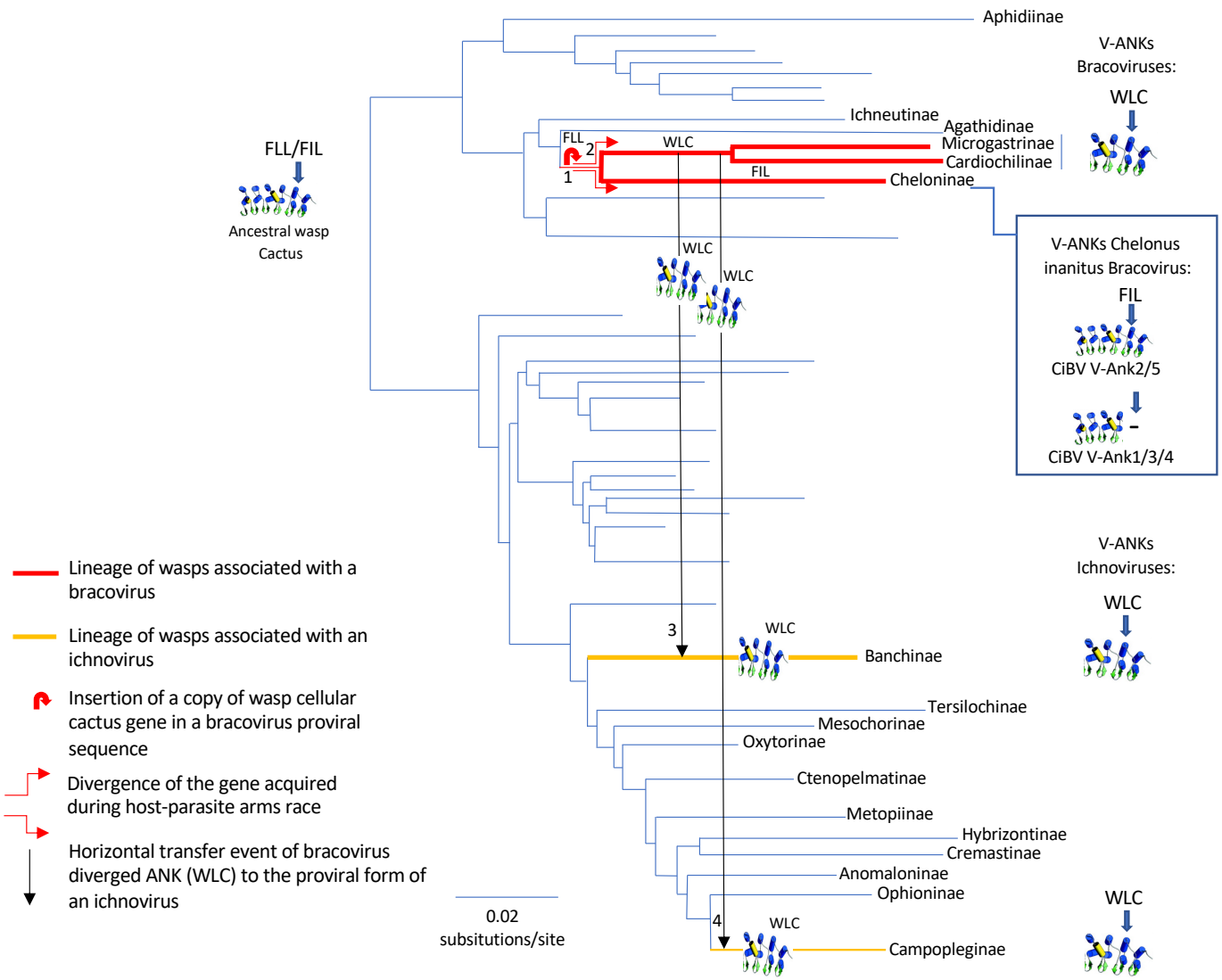
Aphidiinae
V-ANKs
Bracoviruses:
WLC
Ichneutinae
Agathidinae
Microgastrinae
Cardiochilinae
Cheloninae
FLL/FIL
Ancestral wasp
Cactus
FLL
2
1
WLC
FIL
WLC
WLC
V-ANKs Chelonus
inanitus Bracovirus:
FIL
CiBV V-Ank2/5
-
CiBV V-Ank1/3/4
V-ANKs
Ichnoviruses:
WLC
3
WLC
Banchinae
Tersilochinae
Mesochorinae
Oxytorinae
Ctenopelmatinae
Metopiinae
Hybrizontinae
Cremastinae
Anomaloninae
Ophioninae
WLC
4
WLC
Campopleginae
Lineage of wasps associated with a bracovirus
Lineage of wasps associated with an ichnovirus
Insertion of a copy of wasp cellular cactus gene in a bracovirus proviral sequence
Divergence of the gene acquired during host-parasite arms race
Horizontal transfer event of bracovirus diverged ANK (WLC) to the proviral form of an ichnovirus
0.02
subsitutions/site

Figure 10

Supplementary Materials for

Chelonus inanitus bracovirus encodes lineage specific proteins and truncated immune IκB-like factors

Alexandra Cerqueira de Araujo, Thibaut Josse, Vonick Sibut, Mariko Urabe, Azam Asadullah, Valérie Barbe, Madoka Nakai, Elisabeth Huguet, Georges Periquet, Jean-Michel Drezen*

*corresponding author: drezen@univ-tours.fr

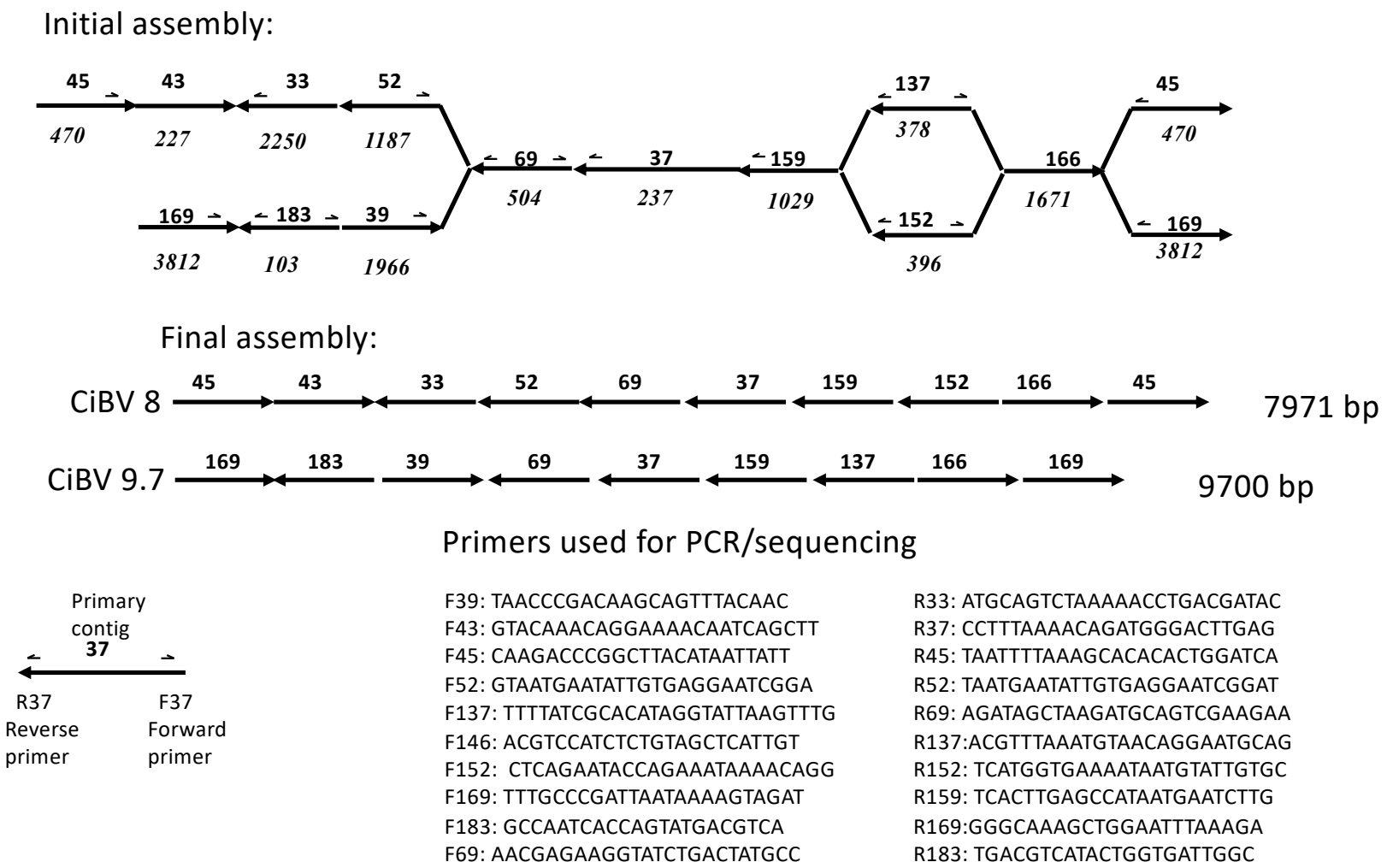

Figure S1: Sequence of primers used to determine the sequence of CiBV8 (7931 bp) and CiBV9.7 (9700 bp) from the assembly. Numbers in bold on top of the arrows are primary contig names, numbers in italics below the arrows correspond to primary contig length (base pair), primary contigs are not to scale.

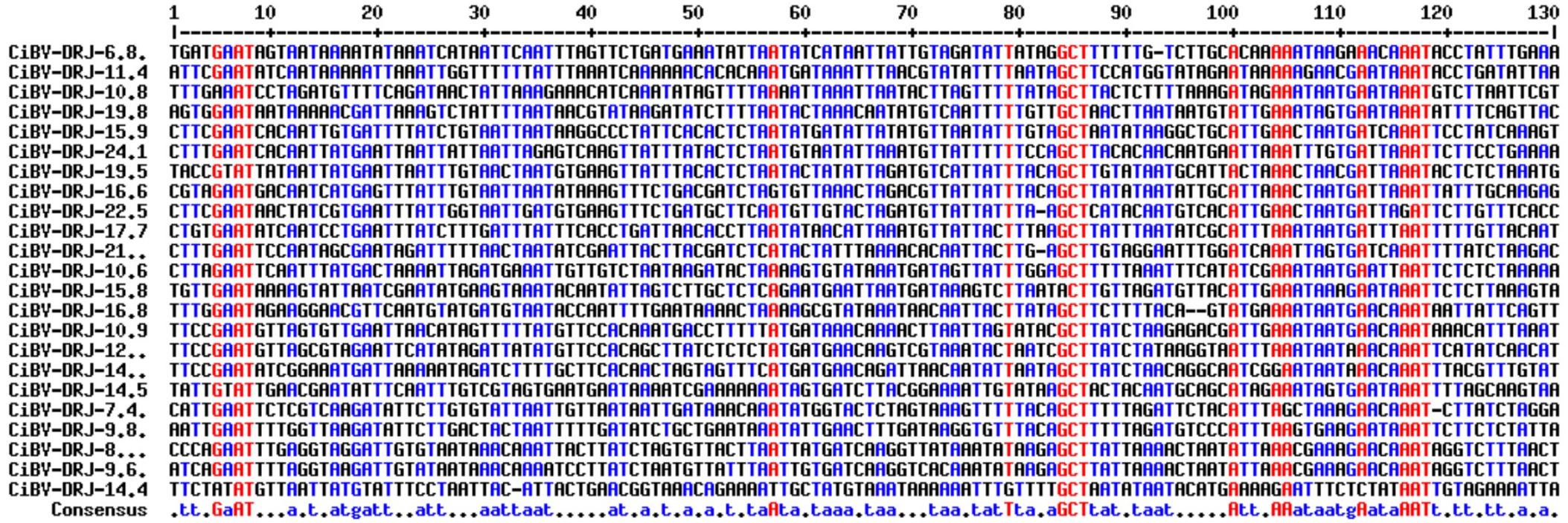

Figure S2: Alignment of 23 circle DRJs identified in CiBV packaged genome

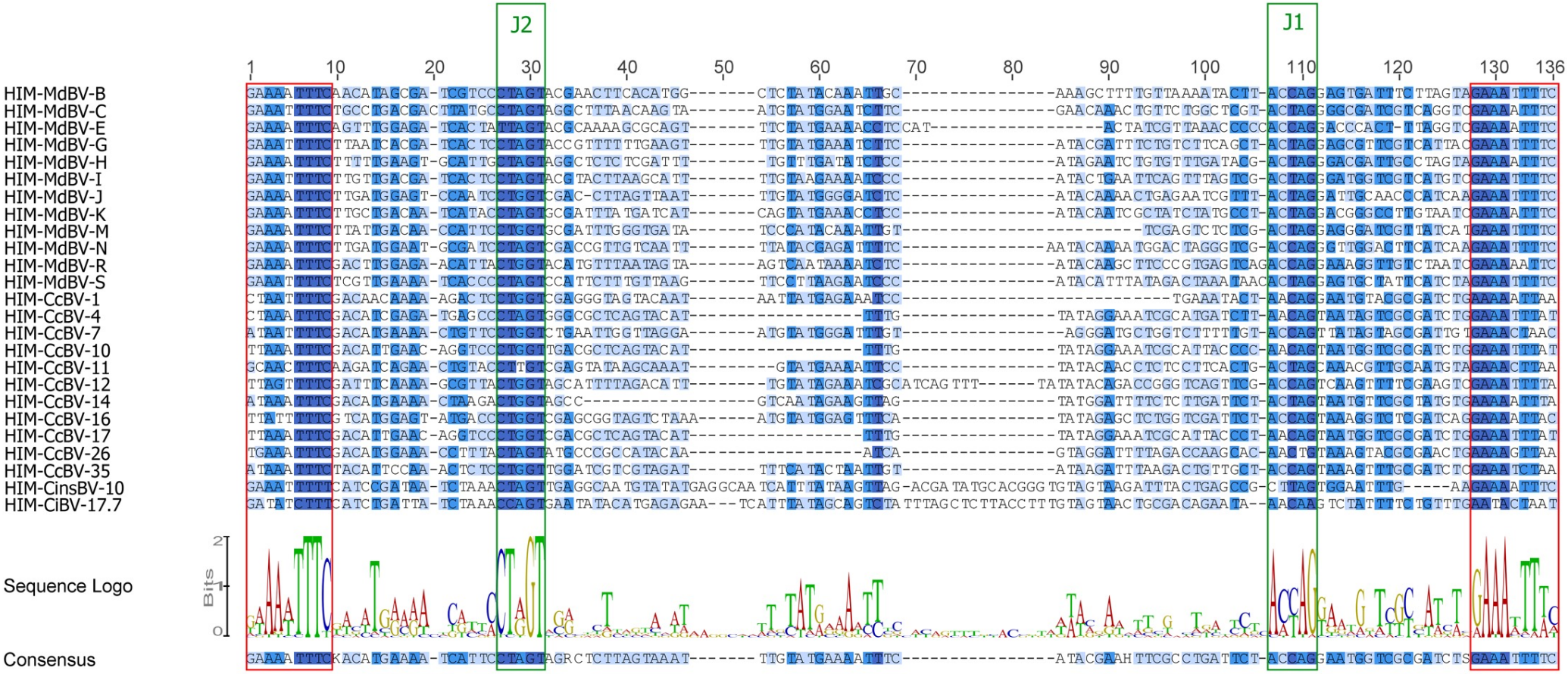

Figure S3: Alignment of the putative HIM motif from CiBV17.7 and CinsBV-10 and to functionally characterized HIM sites involved in integration into the DNA of parasitized host cells of MdBV and CcBV circles.

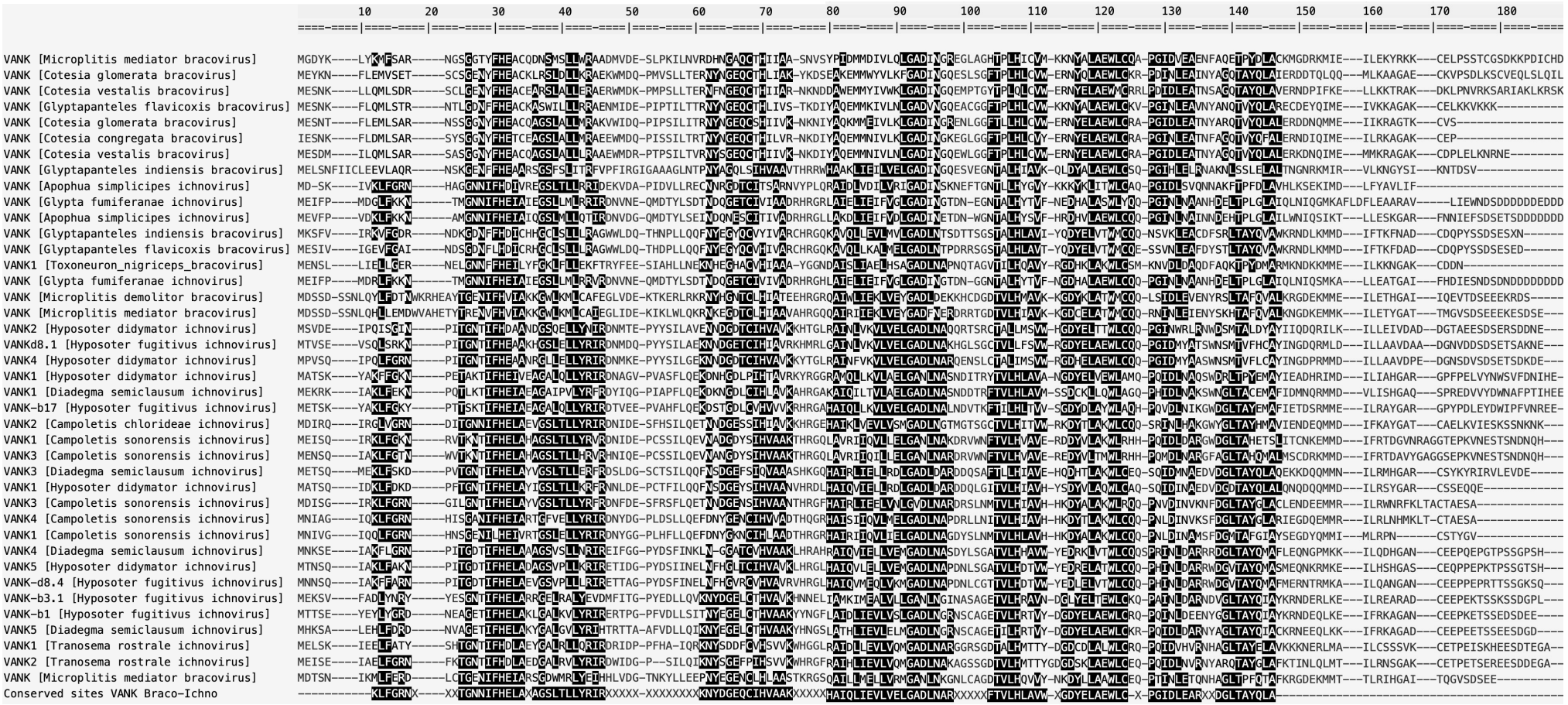

Figure S4: Alignment of the 40 proteins used to identify conserved sites of bracovirus and ichnovirus V-ANKs

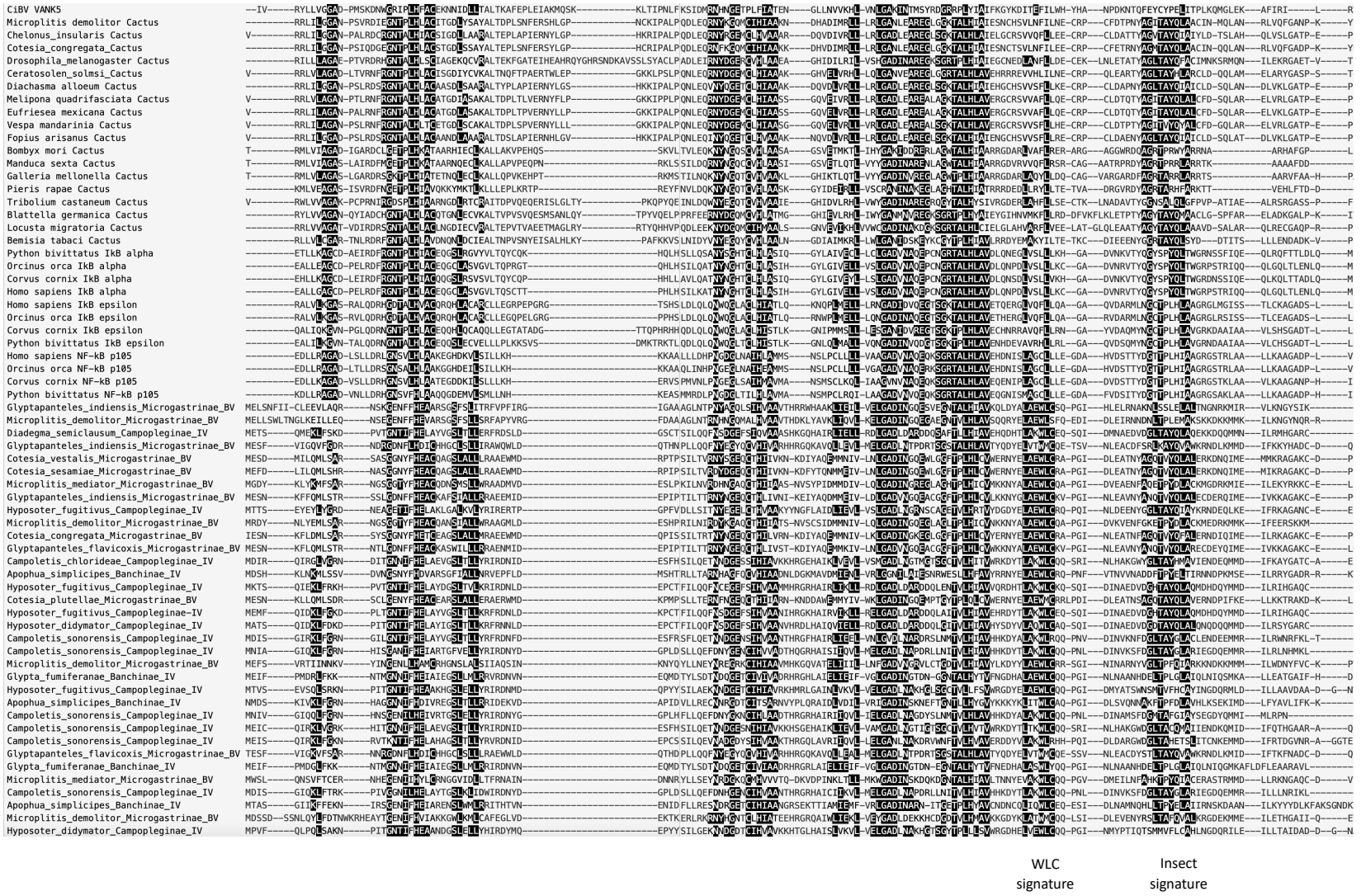

Figure S5: Alignment comprising the different IκB related proteins of vertebrates highlighting the specific PDV V-ANK WLC and TAYQLA insect signatures indicating respectively their shared evolutionary history.

Primer name	Primer sequence (5’-3’)	Product length cDNA (bp)	Product length gDNA (bp)
vank-1-F	CGCTCAGAACCATGACGGAA	438	752
vank-1-R	ACAACGTTAACGCATCCGAA		
vank-2-F	GCGTCTCTTCCTAGAGAGCCA	589	896
vank-2-R	GTGGTCGACGTCCATCTCTG		
vank-3-F	CGCTCAGGACAAAGTCGGAA	270	360
vank-3-R	TCGTTGCCCAATTGACATGC		
vank-4-F	TACGCTCAGAACGAAGACGG	106	483
vank-4-R	ATGCTGGCCAGGTTTCATCT		
vank-5-F	CTCAGACGGAAAAACCTAAC	294	493
vank-5-R	GAATACGACCCCAGTTGTC		
vank-6-F	CATCGTGCCCGTGTGTTAAT	380	380
vank-6-R	TAGTGGCGACCATCCACGAT		
β-actin-F	GAGATCCACATCTGCTGGAAGGTGGACAG	842	842
β–actin-R	AACTGGGATGACATGGAGAAGATCTGGC		

Table S3: primers used for CiBV *V-ank* genes expression analyses