AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

The Complete Genome of *Chelonus insularis* Reveals Dynamic Arrangement of Genome Components in Parasitoid Wasps That Produce Bracoviruses

Meng Mao,[a] Michael R. Strand,[a] Gaelen R. Burke[a]

[a]Department of Entomology, University of Georgia, Athens, Georgia

ABSTRACT Bracoviruses (BVs) are endogenized nudiviruses in parasitoid wasps of the microgastroid complex (family Braconidae). Microgastroid wasps have coopted nudivirus genes to produce replication-defective virions that females use to transfer virulence genes to parasitized hosts. The microgastroid complex further consists of six subfamilies and ~50,000 species but current understanding of BV gene inventories and organization primarily derives from analysis of two wasp species in the subfamily Microgastrinae (*Microplitis demolitor* and *Cotesia congregata*) that produce M. demolitor BV (MdBV) and C. congregata BV (CcBV). Notably, several genomic features of MdBV and CcBV remain conserved since divergence of *M. demolitor* and *C. congregata* ~53 million years ago (MYA). However, it is unknown whether these conserved traits more broadly reflect BV evolution, because no complete genomes exist for any microgastroid wasps outside the Microgastrinae. In this regard, the subfamily Cheloninae is of greatest interest because it diverged earliest from the Microgastrinae (~85 MYA) after endogenization of the nudivirus ancestor. Here, we present the complete genome of *Chelonus insularis*, which is an egg-larval parasitoid in the Cheloninae that produces C. insularis BV (CinsBV). We report that the inventory of nudivirus genes in *C. insularis* is conserved but are dissimilarly organized compared to *M. demolitor* and *C. congregata*. Reciprocally, CinsBV proviral segments share organizational features with MdBV and CcBV but virulence gene inventories exhibit almost no overlap. Altogether, our results point to the functional importance of a conserved inventory of nudivirus genes and a dynamic set of virulence genes for the successful parasitism of hosts. Our results also suggest organizational features previously identified in MdBV and CcBV are likely not essential for BV virion formation.

IMPORTANCE Bracoviruses are a remarkable example of virus endogenization, because large sets of genes from a nudivirus ancestor continue to produce virions that thousands of wasp species rely upon to parasitize hosts. Understanding how these genes interact and have been coopted by wasps for novel functions is of broad interest in the study of virus evolution. This work characterizes bracovirus genome components in the parasitoid wasp *Chelonus insularis*, which together with existing wasp genomes captures a large portion of the diversity among wasp species that produce bracoviruses. Results provide new information about how bracovirus genome components are organized in different wasps while also providing additional insights on key features required for function.

KEYWORDS *Chelonus insularis*, bracovirus, endogenous virus element, parasitoid wasps, virus domestication

Integration of all or portions of a viral genome into the germ line of a eukaryotic host is referred to as endogenization (1, 2). Most endogenous virus elements (EVEs) rapidly decay but some have become fixed in the populations of hosts (1, 2). Some EVEs have also been coopted (domesticated) by hosts, in which their products provide beneficial functions (3). The most complex known example of endogenous virus domestication

Editor Colin R. Parrish, Cornell University

Address correspondence to Gaelen R. Burke, grburke@uga.edu.

The authors declare no conflict of interest.

Received 9 September 2021
Accepted 20 December 2021
Accepted manuscript posted online 5 January 2022
Published 9 March 2022

are the bracoviruses (BVs) found in parasitoid wasps of the family Braconidae (Hymenoptera) (4–6). BVs evolved from a virus in the family Nudiviridae (genus *Betanudivirus*) that integrated into the germ line of a braconid ancestor ~100 million years ago (MYA) (7–9). Speciation events have since resulted in ~50,000 BV-associated wasps in six subfamilies (Microgastrinae, Cardiochilinae, Miracinae, Mendeseliinae, Khoikhoiinae, Cheloninae) that form a monophyletic assemblage called the microgastroid complex (10). Most microgastroid wasps further appear to be specialists whose hosts are Lepidoptera (moths and butterflies) (11). BVs are much more complex than other EVEs because microgastroid wasps retain many genes from the nudivirus ancestor that still interact to produce virions, but virion function has also been repurposed for activities that enable females to parasitize their lepidopteran hosts.

The Nudiviridae is sister to the Baculoviridae with all known species in both families infecting insects or other arthropods (12, 13). Nudiviruses and baculoviruses also similarly have large, circular, double-stranded (ds) DNA genomes (80–230 kb), and share a partially overlapping set of core genes that primarily have functions in producing enveloped virions (12, 13). BV genome components are integrated into the germ line of all microgastroid wasps and also produce enveloped virions containing circular dsDNAs by replicating in ovary calyx cells of females (14). Calyx cell lysis releases virions into the lumen of the reproductive tract where they are stored with mature eggs, and females inject both into hosts. Virions rapidly infect host cells followed by the expression of BV genes that alter immune defenses and growth in ways that enable wasp offspring to develop (4, 15–17). However, the virions wasps inject into hosts cannot replicate, which results in BV genome components only being transmitted vertically (4).

Insights into BV genome evolution primarily derive from sequencing two wasp species in the subfamily Microgastrinae (*Microplitis demolitor* and *Cotesia congregata*) that produce M. demolitor bracovirus (MdBV) and C. congregata bracovirus (CcBV), respectively. BV genome components in both wasp species are extensively rearranged relative to nudiviruses. Nudivirus genes with known or predicted functions in producing virions are expressed in calyx cells but none are located in the DNA domains that are amplified, circularized, and packaged into virions (6, 18, 19). Almost half of these genes are also located in an ~100 kb region of the wasp genome called the nudivirus cluster (6, 18, 19). Gene content and order of the nudivirus cluster is almost identical in *M. demolitor* and *C. congregata*, which has been suggested to be both functionally important and a remnant of the nudivirus that integrated into the common ancestor of the microgastroid complex (6, 18, 19). However, other nudivirus genes with functions in virion formation are widely and disparately dispersed (18, 19). The DNA domains that are amplified and packaged into virions are called proviral segments, which contain the genes that are expressed in parasitized hosts. A majority of these proviral segments are tandemly arrayed in a large 'macrolocus' (6, 18, 19). Inferred orthology relationships further suggest the organization of proviral segments was already established in the common ancestor of *M. demolitor* and *C. congregata* which diverged ~53 MYA (7, 10, 18). The inventory of genes on proviral segments partially overlaps between *M. demolitor*, *C. congregata* and other species in the Microgastrinae (4, 16, 20). A majority of these genes also contain introns and share homology with genes from wasps, other insects, or other eukaryotes that suggest they have been coopted from diverse sources for functions in parasitizing hosts (4, 16, 20).

Altogether, the preceding results indicate the nudivirus ancestor of BVs was domesticated through genome rearrangements and regulatory alterations to produce replication-defective virions that wasps use to transfer virulence genes to hosts. Evolutionary constraints potentially maintain architectural features like the nudivirus cluster and macrolocus, but it is also possible gene content and organization differ in wasps outside the Microgastrinae. Among the subfamilies in the Microgastroid complex, the Cheloninae is of most interest because it diverged earliest from the Microgastrinae (~85 MYA) after endogenization of the nudivirus ancestor (7, 9). Unlike microgastrines that parasitize larval stage hosts, chelonines parasitize hosts during their egg stage, which could also select for differing traits. Some BV genome components were previously identified from Chelonus inanitus bracovirus (CiBV) from the wasp *Chelonus inanitus* (8, 21–23). However, no chelonine wasp genome has been sequenced, which prevents assessment of how chelonine and microgastrine BV genome

TABLE 1 Raw reads generated for the genome assembly

SRA accession	Library strategy	Sequence data (gbp)	Coverage
SRR11678241	Pacbio	18	136×
SRR11678242	Illumina	108.1	97×

components compare to one another. In this study, we sequenced the genome of *Chelonus insularis* (Braconidae: Cheloninae) that produces C. insularis bracovirus (CinsBV) (24). We report that the inventory of nudivirus genes in *C. insularis* significantly overlaps with *M. demolitor* and *C. congregata* but they are dissimilarly organized. Proviral segments in *C. insularis* share organizational features with *M. demolitor* and *C. congregata* but gene inventories exhibit almost no overlap. Altogether, our results reveal a more dynamic arrangement of nudivirus genes and proviral segments in the genomes of BV-carrying wasps than was previously known.

RESULTS

***De novo* sequencing generated a draft genome for *C. insularis* with high levels of completeness and contiguity.** Long-read PacBio sequencing together with short-read Illumina sequencing were used to generate a reference genome for *C. insularis* (Table 1). After removing likely bacterial contaminants with blobtools, the assembly yielded 455 scaffolds with an N50 of 1,162,728 bp, and an overall size of 135 Mb (average cumulative coverage = 233×; Table 1). This assembly was similar in quality to the *M. demolitor* genome assembly Mdem2 (25), but was more fragmented compared to the chromosome-level assembly available for *C. congregata* (18). The genome size was slightly larger than the estimate (122.5 Mb) from kmer analysis with KAT. An assessment of genome completeness using BUSCO indicated that 98.8% of the BUSCO "Insecta" protein-coding gene set was identified. The GC content of the genome assembly was 30.5%, which is similar to other parasitoid genomes (26). Genome annotation with the NCBI Eukaryotic Annotation Pipeline yielded 11,442 genes or pseudogenes, including 10,548 containing protein-coding regions, and 19,220 annotated mRNA transcripts (Table 2). Most predicted transcripts (N = 18,106) were fully supported by the RNA-Seq data sets we generated from different *C. insularis* life stages and parasitized *S. frugiperda* larvae that wasp larvae had been removed from but were CinsBV infected (Table 3). A total of 894 noncoding genes, 113 tRNAs, 848 lncRNAs and other genome components were also identified (Table 2).

The *C. insularis* genome contains 43 homologs of known nudivirus genes. To characterize CinsBV genome components, we focused first on identifying nudivirus genes in the *C. insularis* genome. Our homology-based search identified 43 nudivirus genes on 21 scaffolds (Table 4). Twenty-four were dispersed in the genome while 19 were in small clusters of 2–5 genes on 7 scaffolds (Fig. 1; Table 4). The clustered nudivirus genes on five of these scaffolds (Cluster 1–3, 6, 7) were closely aligned and flanked by nonviral wasp genes, while pairs of nudivirus genes <75 kb apart were on two scaffolds (Clusters

TABLE 2 Gene annotation summary statistics

Feature	Count	Mean length (bp)	Median length (bp)	Min length (bp)	Max length (bp)
Genes	11,442	7,970	2,844	68	763,150
Protein-coding	10,548				
Non-coding	894				
All transcripts	20,679	2,954	2,225	68	56,124
mRNA	19,220	3,057	2,306	277	56,124
misc_RNA	400	2,944	2,331	244	15,328
tRNA	113	74	73	71	84
lncRNA	848	1,246	923	157	9,042
CDSs	19,220	2,170	1,521	105	54,753
Exons	77,264	399	218	2	15,371
Introns	63,902	1,406	120	30	176,056

TABLE 3 RNA-Seq reads from *C. insularis* samples used for annotation

Sample type	SRA accession	Raw reads sequenced
C. insularis		
Adult males (N = 4)	SRR11845185	16.7 million
Adult females (N = 4)	SRR11845186	15.5 million
Pupae (red eye stage, N = 3)	SRR11845184	11.9 million
Larvae (third instars, N = 3)	SRR11845190	14.3 million
Ovaries from adult females (N = 20)	SRR11845189	16.2 million
Parasitized *S. frugiperda* first instars (N = 10)		
Mapping to proviral segments	SRR11967921	1,863
Non-proviral segment reads	SRR11967920	13.5 million
Parasitized *S. frugiperda* fourth instars (N = 3)		
Mapping to proviral segments	SRR11967919	11,938
Non-proviral segment reads	SRR11967918	12.3 million

4, 5) with almost no wasp genes (Fig. 1). All sequenced nudiviruses share 21 core genes with baculoviruses while an additional 11 genes have been proposed to be nudivirus-specific, which yields a total nudivirus core gene set of 32 genes (13, 27). The Nudiviridae is also currently subdivided into two recognized genera (*Alphanudivirus*, *Betanudivirus*) (13) with certain genes potentially being genus specific. For the nudivirus core gene set, 21 were present in *C. insularis* but expansion of *11k* into a 3-member gene family resulted in a total of 23 homologs. Four of the nudivirus core genes (*p47*, *lef-4*, *lef-8*, *lef-9*) encode the subunits of a predicted DNA dependent RNA polymerase (RNApol), one (*lef-5*) encodes a predicted transcriptional initiation factor, while three (*helicase*, *vlf-1*, and *int-1*) have predicted replication functions (Table 4). Thirty-one nudivirus genes were classified as virion envelope (*odv-e66-1* through *−6*, *pif-0* through *−6*, *pif-8*, *p33*, *HzNVorf64-like*) or nucleocapsid components (*int-1*, *vlf-1*, *p33*, *38k*, *vp39*, *27b-like*, *HzNVorf9-1 and −2-like*, *HzNVorf93-like*, *HzNVorf118-like*, *HzNVorf106-like*, *HzNVorf128-like*, *HzNVorf140-like-1* and *−2*, *K425_459-like*) on the basis of a recent proteomic analysis of MdBV virions (Table 4) (28). None of the nudivirus genes in *C. insularis* had introns except *HzNVorf128-like* (see below). The most prominent absence from the nudivirus core gene set was a baculovirus/nudivirus-like DNA polymerase gene. We note that *odv-e66* is duplicated in some nudiviruses but in *C. insularis* formed a six-member family although two (*odv-e66-2* and *−5*) were truncated on their 3′ ends and thus likely pseudogenized (Table 4). We primarily used our unreplicated RNA-seq data sets for annotation purposes, but recognized they could provide qualitative information on expression patterns. Since BVs only produce virions in ovarian calyx cells, we expected the nudivirus genes to be primarily if not exclusively expressed in the adult female and ovary RNAseq samples we generated. Consistent with this expectation, FPKM values for nearly all of the nudivirus-like genes were higher in female wasps and ovaries than other wasp life stages while none were detected in host stages (Table S1). The only exceptions were *odv-e66-2* and *−5* that were largely not detected in any sample which further supported that both were pseudogenes (Table S1).

The *C. insularis* genome contains 20 proviral segments in 7 loci. We next identified the boundaries of proviral segments representing excision sites in the *C. insularis* genome by Illumina sequencing the DNAs in CinsBV virions, mapping the reads to the wasp genome, and identifying regions with marked differences in sequence read depth. Results identified 20 proviral segments (CinsV1-CinsV21) in 7 loci (Fig. 2A; Table 5) that were named to be consistent with naming of the partial inventory of segments identified in CiBV virions from *C. inanitus* (see below for information about segment homology between CinsBV and CiBV). Locus 1 contained 8 tandemly arrayed segments, loci 2–4 contained 2–4 segments, and loci 5–7 each consisted of a single segment (Fig. 2A). CinsV3 (locus 2) was at the end of scaffold NW_023276388.1 while CinsV8 (locus 7) was at the end of scaffold NW_023276780.1, which resulted in both being incomplete (Fig. 2A). The relative abundance of each viral segment in virions was estimated from the average sequence read depth of sites between segment boundaries. The abundance of the incomplete viral segments CinsV3 and CinsV8

TABLE 4 Nudivirus genes identified in the *Chelonus insularis* genome compared to baculoviruses, exogenous nudiviruses and three other BVs (CiBV, MdBV, CcBV)

Scaffold (length, bp)	Cluster	Start	End	Gene	Accession	Function	Baculovirus core	Nudivirus core	Alphanudivirus associated	Betanudivirus associated	CiBV (partial data)	MdBV	CcBV
NW_023276387.1 (1749501)		200477	202498	***pif-0***	KAG8148364.1	Envelope							
		830316	831975	*HzNVorf140-like-2*	KAG8148365.1	Nucleocapsid						&	&
		909501	910539	*HzNVorf93-like*	KAG8148366.1	Nucleocapsid							
	1	1356503	1357452	*p33*	KAG8148367.1	Nucleocapsid/ Envelope						&	&
	1	1357478	1358689	*HzNVorf64-like*	KAG8148368.1	Envelope							
	1	1358888	1360829	*OrNVorf18*	KAG8148369.1	Unknown							
	1	1361063	1362665	*DhNVorf67-like*	KAG8148370.1	Unknown							
	1	1363079	1367731	***helicase***	KAG8148371.1	Replication							
NW_023276792.1 (1272532)	2	508441	510691	*odv-e66-6*	KAG8148302.1	Envelope			*		&	&	&
	2	510677	512879	*odv-e66-1*	KAG8148303.1	Envelope			*		&	&	&
		900056	901731	***pif-1***	KAG8148304.1	Envelope							
		993103	994295	***pif-2***	KAG8148305.1	Envelope							
NW_023276469.1 (571634)	3	485835	486883	*int-1*	KAG8148347.1	Replication/ Nucleocapsid						&	&
	3	492695	493697	***38k***	KAG8148348.1	Nucleocapsid							
NW_023276571 (1274914)		73308	75126	***lef-4***	KAG8148338.1	Transcription							
		701252	703347	*odv-e66-4*	KAG8148339.1	Envelope			*			&	&
NW_023276637.1 (305325)	4	71211	72462	*27b*	KAG8148335.1	Nucleocapsid							
	4	130886	131956	***vp39***	KAG8148336.1	Nucleocapsid							
		292447	293517	*HzNVorf118-like*	KAG8148337.1	Nucleocapsid							
NW_023276654.1 (936017)		402340	403477	*HzNVorf140-like-1*	KAG8148334.1	Nucleocapsid						&	&
NW_023276741.1 (1424300)	5	1208611	1209746	*K425_459-like*	KAG8148330.1	Nucleocapsid							
	5	1255992	1257053	*HzNVorf106-like*	KAG8148331.1	Nucleocapsid							
		1403776	1405744	*HzNVorf9-1-like*	KAG8148332.1	Nucleocapsid					&	&	&
NW_023276782.1 (2432667)		2130891	2132412	***pif-5***	KAG8148307.1	Envelope						&	&
NW_023276563.1 (2102350)		732592	735049	***pif-8***	KAG8148342.1	Envelope							
NW_023276419.1 (619056)		197371	198210	*odv-e66-2, partial*	KAG8148350.1	Envelope			*			&	&
NW_023276451.1 (2804946)		2156207	2158313	***lef-9***	KAG8148349.1	Transcription							
NW_023276487.1 (932484)		154358	155069	***pif-4***	KAG8148346.1	Envelope							
NW_023276565.1 (763321)		284382	286455	*odv-e66-3*	KAG8148341.1	Envelope			*		&	&	&
NW_023276793.1 (1178256)	6	895702	897294	***p47***	KAG8148298.1	Transcription							
	6	897746	901393	***lef-8***	KAG8148299.1	Transcription							
	6	901603	903938	*HzNVorf128-like*	KAG8148300.1	Nucleocapsid							
NW_023276534.1 (344162)		311231	312305	*odv-e66-5, partial*	XM_035083910.1	Envelope			*			&	&
		312414	313024	*odv-e66-5, partial*	XM_035083910.1	Envelope			*			&	&
NW_023276344.1 (110044)		52079	52881	***pif-3***	KAG8148387.1	Envelope							
NW_023276514.1 (666923)		325543	326335	***pif-6***	KAG8148344.1	Envelope							
		559528	560504	*HzNVorf9-2-like*	KAG8148345.1	Nucleocapsid					&	&	&
NW_023276516.1 (347180)		260286	261312	***vlf-1***	KAG8148343.1	Replication/ Nucleocapsid							
NW_023276773.1 (1317781)		828205	828782	***lef-5***	KAG8148311.1	Transcription							
NW_023276392.1 (1600403)	7	1451077	1451582	*11k-1*	KAG8148352.1	Unknown		*	*	*	&	*	*
	7	1451858	1452707	*11k-2*	KAG8148353.1	Unknown		*	*	*	&	*	*
	7	1453425	1454017	*11k-3*	KAG8148355.1	Unknown		*	*	*	&	*	*
NW_023276685.1 (380495)		257410	258351	*ToNVorf29-like*	KAG8148333.1	Unknown							

Genes universally conserved in nudivirus and baculovirus genomes are indicated in bold type. Boxes shaded in gray are part of the core gene set, are present in all members of the *Alpha-* or *Betanudivirus*, or a given BV. Current ICTV nomenclature is used to categorize nudiviruses into two genera (*Alphanudivirus* and *Betanudivirus*) with Tipula oleracea nudivirus being most closely related to other members of the *Betanudivirus* genus (85). Boxes shaded in blue indicate gene products that were detected as virion components in baculoviruses (86), Tipula oleracea nudivirus (>2 unique peptides detected) (85), or a given BV (only partial data exists for CiBV and CcBV) (8, 23, 28). An exception is *11k*, which was detected as a structural component of Tipula oleracea nudivirus but has not been identified in BV virions to date. An asterisk (*) indicates the gene is single copy. The ampersand symbol (&) indicates the gene has duplicated into a multimember family.

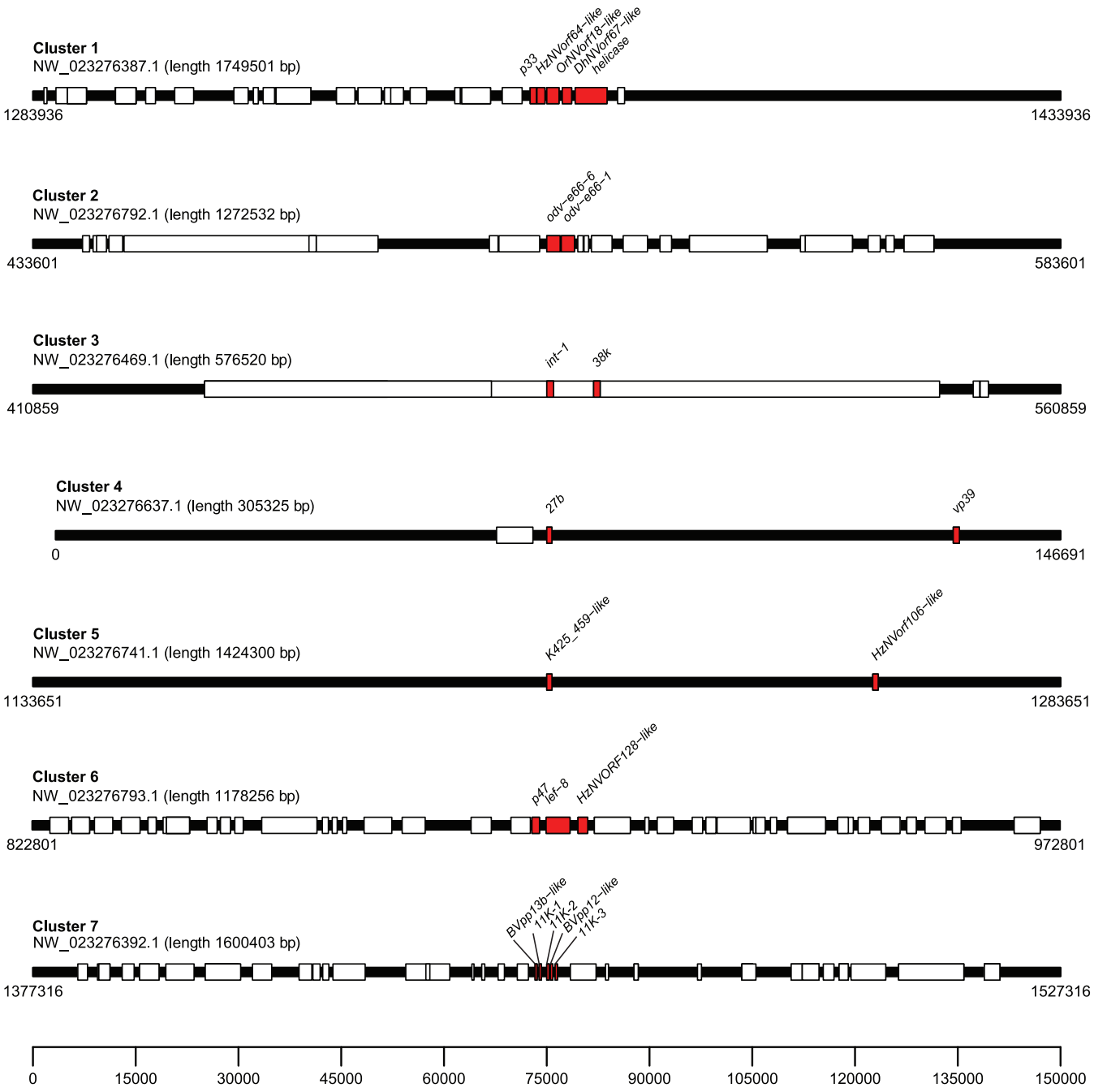

FIG 1 Locations of nudivirus gene clusters on *C. insularis* genome scaffolds. Boundaries of nudivirus genes or genes encoding putative structural components of CinsBV virions (*BVpp12-like* and *BVpp13b-like*) are shown as red boxes with names listed above. Nonviral wasp gene boundaries are shown as white boxes.

differed enough to make it unlikely they derive from a common proviral segment that was broken between two scaffolds (Table 5). Instead, our read depth data much more strongly support that each of these proviral segments are unique but incompletely assembled. Read depth data further suggested CinsV6.1 and CinsV14.1 contained nested segments, named CinsV6.2 and CinsV14.2, that were generated by a lesser-used alternative excision site extending further into the genome on one side of each segment (Table 5). Sequence homology searches indicated that CinsV10 and CinsV12 in locus 1 shared partial homology despite being separated by several other proviral segments. No appreciable intersegmental homology was identified among the other segments (Fig. S1).

Proviral segments in *M. demolitor* and *C. congregata* are initially amplified as replication units that can span one or more proviral segments in a given locus (29, 30). The

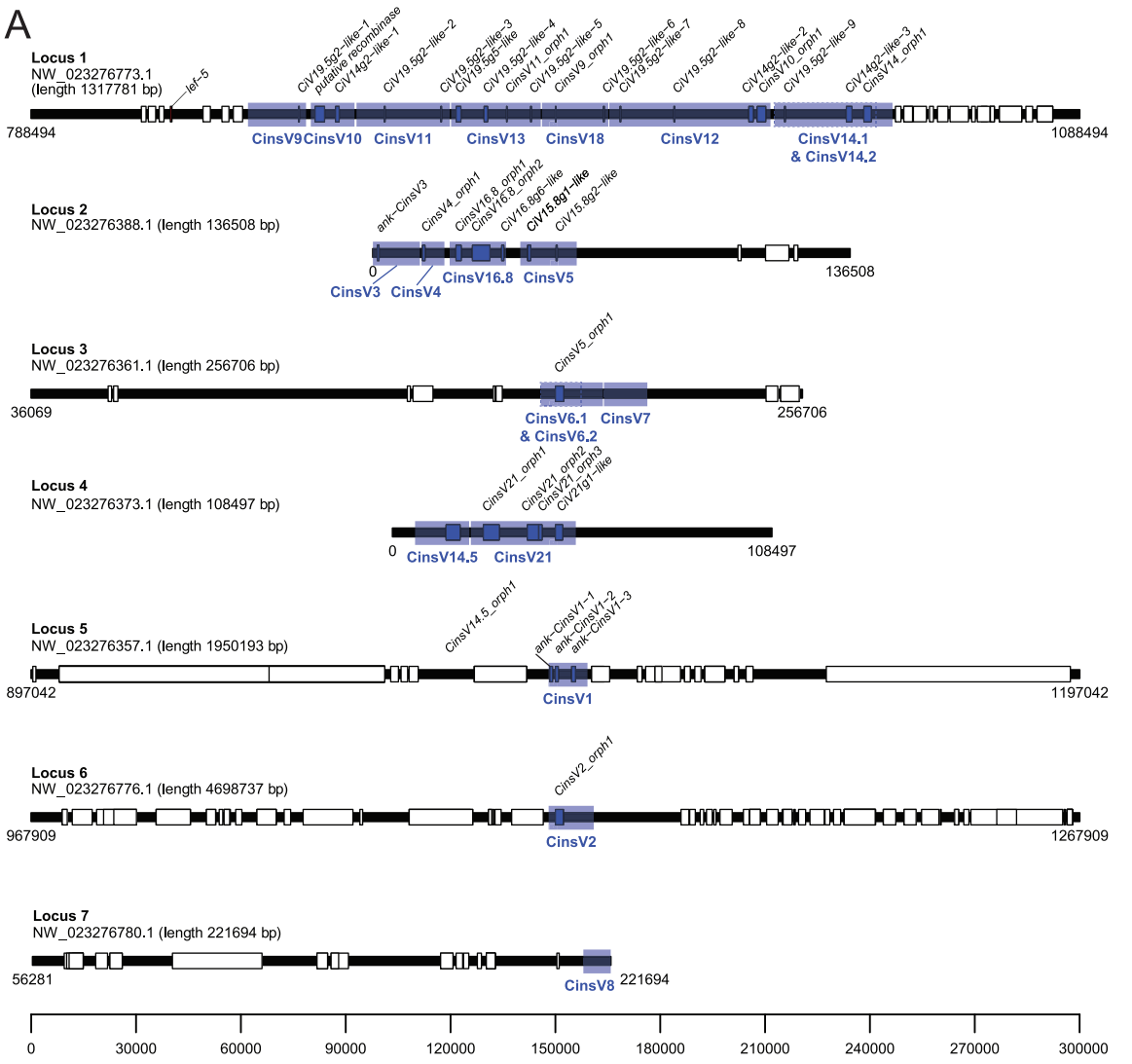

FIG 2 A. Locations of proviral segments in the *C. insularis* genome. Proviral segments reside in seven loci containing 1–8 segments at each locus. Proviral segment excision boundaries are shown with blue transparent boxes with names beneath, while proviral gene

(Continued on next page)

TABLE 5 Virulence genes identified in proviral segments of CinsBV

Segment					Gene			Accession
Location		Name	Coordinates	Relative coverage	Name	Start	End	
Scaffold	Locus							
NW_023276357.1	Locus 5	CinsV1	1045081-1056221	66x	*CinsV1_ncRNA-1*	1048267	1049983	CINS000011
					CinsV1_ncRNA-2	1052797	1056176	CINS000010
					ANK-CinsV1-1	1045411	1046896	KAG8148382.1
					ANK-CinsV1-2	1046915	1047942	KAG8148383.1
					ANK-CinsV1-3	1357478	1358689	KAG8148384.1
NW_023276776.1	Locus 6	CinsV2	1115970-1128883	58x	*CinsV2_orph1*	1116795	1120219	KAG8148308.1
NW_023276388.1	Locus 2	CinsV3	0-13786	50x	*CinsV3_ncRNA1*	9827	10722	CINS000034
					ANK-CinsV3	1143	1837	KAG8148356.1
		CinsV4	13849-20461	16x	*CinsV4_orph1*	13830	15872	KAG8148357.1
		CinsV16.8	21968-38113	63x	*CinsV16.8_orph1*	23763	26302	KAG8148358.1
					CinsV16.8_orph2	28170	33548	KAG8148359.1
					CiV16.8g6-like	34821	37796	KAG8148360.1
		CinsV5	42177-58326	24x	*CiV15.8g1-like, X1*	44221	48960	KAG8148361.1
					CiV15.8g1-like, X2	44221	51241	KAG8148362.1
					CiV15.8g2-like	52413	52886	KAG8148363.1
NW_023276361.1	Locus 3	CinsV6.1	181767-193504	14x	*CinsV6_orph1*	185094	188451	KAG8148378.1
		CinsV6.2	181767-199550	1x				
		CinsV7	199965-212379	14x	-			
NW_023276373.1	Locus 4	CinsV14.5	6435-21960	11x	*CinsV14.5_orph1*	15296	19323	KAG8148372.1
		CinsV21	22398-52428	5x	*CinsV21_orph1*	25926	31256	KAG8148373.1
					CinsV21_orph2	37461	41783	KAG8148374.1
					CinsV21_orph3	41808	42880	KAG8148375.1
					CiV21g1-like	44477	49882	KAG8148376.1
NW_023276780.1	Locus 7	CinsV8	213864-221694	24x	-			
NW_023276773.1	Locus 1	CinsV9	850527-867142	32x	*CiV19.5g2-like-1*	864629	867120	KAG8148312.1
		CinsV10	868495-881006	104x	*putative recombinase*	869662	873035	KAG8148313.1
					CiV14g2-like-1	875382	876788	KAG8148314.1
		CinsV11	881581-908291	15x	*CiV19.5g2-like-2*	888941	891378	KAG8148315.1
					CinsV11_ncRNA-1	890881	892148	CINS000080
					CinsV11_ncRNA-2	893601	894468	CINS000083
					CiV19.5g2-like-3	905189	908232	KAG8148316.1
		CinsV13	908648-934295	13x	*CiV19.5g5-like*	909999	914097	KAG8148317.1
					CiV19.5g2-like-4	918007	920474	KAG8148318.1
					CinsV13_orph1	923894	924859	KAG8148319.1
					CiV19.5g2-like-5	931008	933507	KAG8148320.1
		CinsV18	934652-953622	25x	*CinsV18_orph1*	937218	939087	KAG8148321.1
					CiV19.5g2-like-6	951353	952544	KAG8148322.1
		CinsV12	953917-1000033	13x	*CiV19.5g2-like-7*	956857	958320	KAG8148323.1
					CiV19.5g2-like-8	971100	973141	KAG8148324.1
					CiV14g2-like-2	993098	995439	KAG8148325.1
					CinsV12_orph1	996033	998726	KAG8148326.1
		CinsV14.1	1001207-1030340	18x	*CiV19.5g2-like-9*	1002464	1004575	KAG8148327.1
					CinsV14_ncRNA	1007744	1010755	CINS000084
					CiV14g2-like-3	1020669	1023588	KAG8148328.1
					CinsV14_orph1	1026749	1029110	KAG8148329.1
		CinsV14.2	1001207-1034965	6x	*CinsV14_ncRNA-2*	1031202	1032720	CINS000085

replication units are amplified in one of two forms (head-tail or head-head/tail-tail), which are each associated with specific sequence motifs (18, 30). The circular dsDNA segments that are packaged into BV virions are then generated by excision and recombination, which occurs in association with conserved sequences named wasp integration motifs (WIMs) or direct repeat junctions (DRJs) containing the tetramer AGCT that identifies the site of recombination to produce circularized DNAs containing one WIM (29–33). There are additional, lesser-conserved motifs in the 10 bp immediately flanking the 5′ ends of WIMs and within proviral segments approximately 80 bp upstream of 3′ WIM sequences (TGAAT) (18, 19). Amplification is also non-equimolar, which results in some segments that are individu-

FIG 2 Legend (Continued)
boundaries are shown with blue opaque boxes with names indicated above. Nonviral wasp gene boundaries are shown as white boxes. Two segments have alternative excision sites as indicated by dashed boxes (CinsV6.1 and CinsV6.2, and CinsV14.1 and CinsV14.2). B. Wasp Integration Motifs (WIMs) for all CinsBV proviral segments. 200 nucleotides (nt) surrounding the 5′ and 3′ WIM sites of each segment are aligned, respectively. Similarity for each site is colored in shades of green. The tetramer AGCT and other conserved motifs are highlighted with black and gray boxes, respectively. Conservation of each site in the motifs is indicated in bits.

ally packaged into virions being more abundant than others (17, 19, 34). For *C. insularis*, read mapping data indicated proviral segments were differentially amplified with CinsV1, 2, 3, 16.8 and 10 being the most abundant (50–104× relative coverage) in CinsBV virions and the two nested segments (CinsV6.2 and 14.2) being the least (1–6× relative coverage) (Table 5). CinsV10 was much more abundant than the other segments (13–32×) in locus 1 suggesting it belonged to a different replication unit (Fig. 2A; Table 5). Abundance varied across all three segments (24–63×) in locus 2, but were less variable in loci 3 and 4 (Fig. 2A; Table 5). Loci 5–8 that consist of single segments also varied in abundance (Fig. 2A; Table 5). None of the specific sequence motifs associated with amplification units in *M. demolitor* or *C. congregata* were identified in *C. insularis*. However, all of the complete proviral segments in *C. insularis* were flanked by WIM sequences (Fig. 2B). We also identified two other conserved motifs in positions conserved with other bracovirus-producing wasps: one that was in the flanking region upstream of the 5′ AGCT sequence and another in the proviral domain upstream of the 3′ WIM sequence (GAAT) (Fig. 2B). Comparing proviral segments to the predicted circularized segments in virions indicated most were processed at several sites in a 'window' surrounding the conserved AGCT motif that ranged from 1–5 nucleotides (nt) at the 5′ end and 1–8 nt at the 3′ end. Within each window, most excision sites were used with relatively equal frequency (Fig. S2). Several MdBV and CcBV segments integrate into the genome of infected host cells with a conserved inverted repeat named the host integration motif (HIM) identifying the site of integration (17, 35). In contrast, only one CinsBV segment (CinsV10) contained a HIM domain (Fig. S3).

Gene coding densities on CinsBV segments are very low. Proviral segments in *C. insularis* ranged from 6.6–46.1 kb which summed to an overall size of 341 kb, but coding densities were very low with only 35 predicted protein-coding genes and 7 noncoding RNAs (Table 5). CinsV1 encoded five genes; CinsV11, CinsV14.1, and CinsV21 each encoded four, while the remaining proviral segments encoded three or fewer, including CinsV7 and CinsV8 with none (Fig. 3; Table 5). Half of the protein-coding genes further belonged to three expanded gene families named the *ank*, *CiV19.5g2-like*, and *CiV14g2-like* genes (Table 5). Unlike the nudivirus genes, 64% of the genes on proviral segments contained one or more introns. Most genes were named by the segment they resided on because they shared no significant homology with other genes outside Chelonus inanitis bracovirus (CiBV) (Fig. 3, Table 5) (see below). The exception was the four-member *ank* gene family on 2 segments (CinsV1, CinsV3) that was so named because of the presence of a similar ankyrin repeat domain (Fig. 3; Table 5).

We expected the genes on proviral segments would primarily be expressed in parasitized hosts. Our RNAseq data sets indicated this was the case for 11 CinsBV genes, including *CinsV16.8_orph1* and *ank-CinsV1-3* that had normalized expression values that were >500 in 4th instar parasitized hosts, and *ank-CinsV1-3*, *CinsV16.8_orph1*, *CinsV2_orph1*, *CinsV6_orph1*, and *CinsV16.8_orph2* that had high expression values in parasitized hosts compared to the samples we prepared from different wasp stages (Table S2). However, some genes on proviral segments were detected in both parasitized hosts and one or more wasp stages, others were preferentially detected in adult female wasps and/or ovaries, and a few were not detected in any of our RNAseq samples (Table S2).

Chelonine BV genome components partially overlap. As previously noted, comparative data from other chelonines are largely restricted to *C. inanitus* where ovary expressed sequence tags (ESTs) and proteomic analysis of virions identified some nudivirus genes while eight of the DNA segments in CiBV virions were cloned and sequenced (8, 21, 23, 36). A very early study also generated N-terminal sequence data (VGILDTVLSNTIQPH) for a 41 kDa virion protein from *Chelonus* sp. near *curvimaculatus* (37). *C. insularis* contained orthologs to all of the nudivirus genes identified in *C. inanitus* (Table 4) while the N-terminal sequence data for the 41 kDa protein from *Chelonus* sp. near *curvimaculatus* identified it as a *vp39* homolog. A second point of interest was that previously generated proteomic data for CiBV virions from *C. inanitus* identified several proteins and corresponding cDNAs that at the time had no similarity to known nudivirus genes (23). One of these genes, *27b*, was subsequently identified in a nudivirus (27) but not the others. Using these CiBV data, we identified orthologs to 23 genes in *C. insularis*, which included *27b* that we classified as a nudivirus gene (Table 4, Table 6). For the

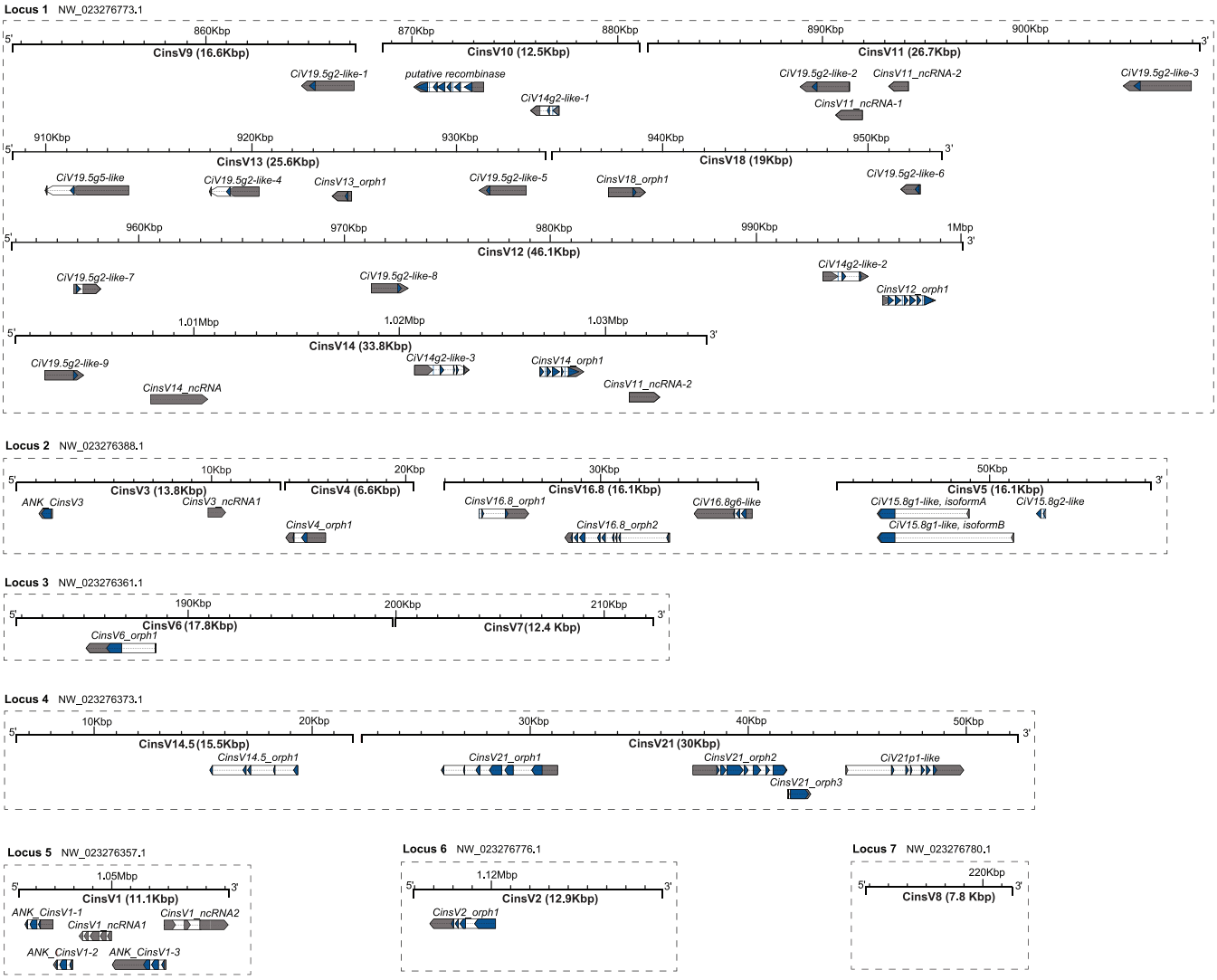

FIG 3 Locations of virulence genes in CinsBV proviral segments. Exons, introns, and transcripts are colored in dark blue, white, and gray, respectively. Distinct loci containing proviral segments are indicated by dashed boxes, while the edges of proviral segments are indicated by bars with ticks indicating proviral segment positions in genome scaffolds. Names of proviral segments are located under scale bars with sizes of segments in parentheses in kilobases (kb).

remaining, 10 were single copy genes and 12 formed 4 small families (Table 6). Seven of the single copy genes and all of the genes in families 1, 3 and 4 were intronless, while the other single copy genes plus all genes in family 2 contained multiple introns (Table 6). These genes are present in the *C. insularis* genome in six clusters of two genes, three of which appear to be generated by local duplications (Table 6). Our RNA-seq data sets further indicated each of these genes were preferentially expressed in female wasps and ovaries (Table S3).

Also as noted above, we named CinsBV proviral segments in a manner consistent with how CiBV segments had been earlier named. This was because our analysis indicated five CinsBV and CiBV segments shared significant similarity, which suggested they were homologous (Fig. S4). We thus named the homologous CinsBV segments CinsV12, CinsV14.2, CinsV14.5, CinsV16.8 and CinsV21 to correspond to CiBV segments CiBV12, CiBV14.2, CiBV14.5, CiBV16.8 and CiBV21, while the other, nonhomologous, CinsBV proviral segments were named numerically starting with CinsV1 and extending through CinsV18 (Fig. 3; Table 5). Five single copy genes on the CinsBV segments that were homologous to CiBV segments were orthologs of single copy CiBV genes (*CiV15.8g1-like*, *CiV15.8g2-like*, *CiV16.8g6-like*, *CiV19.5g1-like*, *CiV21g1-like*) of which three also resided in similar locations on homologous CiBV and CinsBV segments (*CiV14.2-like-3*, *CiV16.8g6-like*, and *CiV21g1-like*). In contrast, the other two single copy homologous genes were on different CinsBV and CiBV segments (Fig. 3; Table 5). We further

TABLE 6 Genes encoding putative structural components of CinsBV virions

Scaffold	Start	End	Gene	Accession	Function	CiBV	CcBV	MdBV	Introns	Gene Family
NW_023276357.1	812961	813992	*BVpp26-like-1*	KAG8148380.1	Unknown	*BVpp26*			None	1
NW_023276357.1	814015	815182	*BVpp26-like-2*	KAG8148381.1	Unknown	*BVpp26*			None	1
NW_023276357.1	1568085	1570820	*BVpp48b-41b-like*	KAG8148385.1	Unknown	*BVpp48b, BVpp41b*			12	2
NW_023276363.1	34059	36059	*BVpp69a-like*	KAG8148377.1	Unknown	*BVpp69a*			None	
NW_023276392.1	1449999	1450956	*BVpp13b-like*	KAG8148351.1	Unknown	*BVpp13b*			None	
NW_023276392.1	1452864	1453202	*BVpp12-like*	KAG8148354.1	Unknown	*BVpp12*			None	
NW_023276571.1	726748	728393	*BVpp48a-like*	KAG8148340.1	Unknown	*BVpp48a*			None	
NW_023276776.1	2407974	2408837	*BVpp28d_like*	KAG8148309.1	Unknown	*BVpp28d*			None	3
NW_023276776.1	2408849	2409881	*BVpp35a_like*	KAG8148310.1	Nucleocapsid/ Envelope	*BVpp35a*	*35a-1, -2*	*35a-2, 4-14*	None	3
NW_023276782.1	825725	826713	*BVpp19-like*	KAG8148306.1	Unknown	*BVpp19*			None	
NW_023276795.1	1971579	1973867	*BVpp66b_like*	KAG8148296.1	Unknown	*BVpp95a, BVpp97a, BVpp66b, BVpp66c*			None	4
NW_023276795.1	1973943	1975776	*BVpp85_like*	KAG8148297.1	Unknown	*BVpp85, BVpp97a, BVpp66c*			None	4
NW_023276609.1	647340	649332	*BVpp66b-66c_like*	XM_035088262.1	Unknown	*BVpp66b, BVpp66c*			None	4
NW_023276783.1	284812	285693	*BVpp12b-like*	XM_035093253.1	Unknown	*BVpp12b*			None	3
NW_023276451.1	564790	567563	*BVpp48b-41b-97b-69b_like-1*	XM_035079749.1	Unknown	*BVpp48b, BVpp41b, BVpp97b, BVpp69b*			11	2
NW_023276451.1	567924	570526	*BVpp48b-41b-97b-69b_like-2*	XM_035079793.1	Unknown	*BVpp48b, BVpp41b, BVpp97b, BVpp69b*			12	2
NW_023276776.1	3185775	3189384	*BVpp69b-97b_like-1*	XM_035091807.1	Unknown	*BVpp69b, BVpp97b*		*LOC103570572*	13	2
NW_023276522.1	430841	432336	*BVpp13c-like, X1*	XM_035083422.1	Unknown	*BVpp13c*			3	
NW_023276522.1	430841	432336	*BVpp13c-like, X2*	XM_035083423.1	Unknown	*BVpp13c*			2	
NW_023276419.1	247125	249434	*BVpp58b-like*	XM_035079006.1	Unknown	*BVpp58b*	*58b*	*LOC103574890, LOC103575872*	10	
NW_023276352.1	168968	170143	*30b (K425_475_like)*	XM_035087411.1	Unknown	*BVpp30b*	*30b*	*K425_475*	None	
NW_023276357.1	776651	777212	*17a*	KAG8148379.1	Envelope	*BVpp17a*	*17a-1*	*17a*	None	

Boxes shaded in gray indicate homologous genes are present in *C. congregata* that produces CcBV or *M. demolitor* that produces MdBV. Boxes shaded in blue indicate the gene product is also detected in MdBV virions.

determined that genes forming 3- (*CiV14g2-like-1 through −3*) and 9-member (*CiV19.5g2-1 through −9-like*) families on multiple CinsBV segments shared homology with two genes identified on CiBV segments (Fig. 3; Table 5). Semiquantitative data previously indicated CiBV segments CiV12 and CiV16.8 were approximately five times more abundant than CiV14, CiV14.5, and CiV21 (36). Our read mapping data likewise classified CinsV16.8 as a higher abundance segment (relative coverage 53×), but suggested CinsV12 (13×) and the other CinsBV segments that were homologous to CiBV segments were lower abundance (5–18× relative coverage) (Table 5).

Chelonine and microgastrine wasps encode overlapping inventories of nudivirus but not proviral segment genes. A complete genome for *C. insularis* also allowed us to compare BV gene content and organizational features to *M. demolitor* and *C. congregata*. The inventory of nudivirus genes substantially overlapped with homologs of all of the 43 nudivirus genes in *C. insularis* being present in *M. demolitor* and *C. congregata,* respectively (Table 4). *DhNVorf067-like* in *C. insularis*, which shared 25.6% identity with its homolog in Dikerogammarus haemobaphes nudivirus (38) was absent in the nudivirus gene inventories of *M. demolitor* and *C. congregata;* however, two uncharacterized genes in *M. demolitor* (LOC106693848) and *C. congregata* (CAJNRD030001114.1, position 2079873 to 2081417), both with 47% identity with *DhNVorf067-like*, were identified using blastp or tblastN. Eight candidate nudivirus genes in *M. demolitor* and *C. congregata* (*HzNVorf94-like*, *K425_438*, *K425_456*, *K425_461*, *int-2*, *fen-1*, *PmNVorf87-like*, and *ToNVorf54-like*) were absent in *C. insularis* (Table 4). Other differences included *odv-e66*, which as noted above was modestly expanded into a 6-member family in *C. insularis* but is greatly expanded into 21 and 36 member families in *M. demolitor* and

C. congregata, respectively (6, 18, 19). Three genes orthologous with *C. insularis*, *odv-e66-1*, *−3*, *and −6*, were identified in partial data from *C. inanitus*, *odv-e66-3* (CBA62617.1), *odv-e66-1* (CBA62593.1), *and odv-e66* (CAR40197.1), respectively. Unlike *C. insularis*, the *int* genes have expanded into a multimember family in *M. demolitor* and *C. congregata* (18, 19). A maximum likelihood tree placed *int-1* from *C. insularis*, *C. inanitus*, *M. demolitor* and *C. congregata* in one clade that was separate from *int-2*, which was consistent with duplication of these genes in the microgastrines occurring after divergence from chelonines (Fig. S5). In contrast, this analysis suggested *HzNVorf140-like* had duplicated prior to divergence of the Cheloninae and Microgastrinae (Fig. S5). Like *C. insularis*, all of the nudivirus genes in *M. demolitor* and *C. congregata* were intronless except *HzNVorf128-like*. However, the introns in *HzNVorf128-like* were not in the same place in *C. insularis* (in the 5′ untranslated region and the coding sequence) and *M. demolitor* (in the 5′ untranslated region) which suggested they evolved independently. For the *C. insularis* genes that were first identified as CiBV virion components but are unknown from nudiviruses, *M. demolitor* and *C. congregata* also encoded *17a*, *35a*, *30b* and *58b*, while *M. demolitor* but not *C. congregata* encoded a *BVpp69-like* gene (Table 6).

The number of proviral segments (N = 20) and genes (N = 42) on proviral segments in *C. insularis* were both much lower than in *M. demolitor* (26 proviral segments, 95 genes) or *C. congregata* (38 proviral segments, 222 genes) (6, 18, 19). Most genes on CinsBV segments also shared no homology with MdBV or CcBV. The one possible exception was the presence of ankyrin domain-containing genes that also exist on proviral segments in all studied microgastrine BVs and ichnoviruses (IVs) from parasitoids in the family Ichneumonidae (16, 39). Genes with ankyrin domains also exist in the genomes of *M. demolitor* and other microgastroids, other wasp species, as well as other insects. A phylogenetic analysis of wasp-associated *ank* genes identified a clade containing the CinsBV *ank* genes plus several other wasp *ank* genes with a high support value that was outside the clade containing *ank* genes on microgastrine BV and IV proviral segments (Fig. S6). This finding suggested independent origins for the chelonine and microgastrine BV *ank* genes although uncertainty also existed because most clades were supported by low bootstrap values due to high sequence divergence.

Organizationally, chelonine and microgastrine proviral segments were similar in regard to some being tandemly arrayed and all being flanked by WIM domains (18, 19, 21, 36, 40, 41). However, the process of proviral segment amplification may differ in microgastrine and chelonine species, given the differential abundance of segments derived from the same loci and the lack of replication unit associated motifs in *C. insularis*. However, as earlier noted recombination occurs variably within a 1–8 bp window surrounding the AGCT motif of CinsBV proviral segments which was also found for several CiBV proviral segments (21, 36), whereas it is restricted to the 4 bp AGCT motif in microgastrine proviral segments (17–19, 29, 30). Thus, the conserved AGCT motif may be important for WIM recognition by presumptive integrases in chelonine wasps but sequence specificity may be less essential for recombination. The greatly reduced prevalence of HIM domains in CinsBV proviral segments relative to MdBV and CcBV further suggested fewer chelonine DNA segments integrate into the genome of infected host cells.

***C. insularis* and *M. demolitor* exhibit little synteny.** Comparing the *C. insularis* and *M. demolitor* genomes identified only 493 syntenous blocks containing 5 or more genes with average block size being only 10 genes. None of the blocks in *C. insularis* corresponded to the nudivirus cluster in *M. demolitor* but 3 small clusters of nudivirus genes in *C. insularis* shared syntenic features. Cluster 4 contained *27b*, *vp39*, and *HzNVorf118-like* with a wasp gene located between *vp39* and *HzNVorf118-like*, which was similar to *27b*, *vp39*, *pif-3*, and *HzNVorf118-like* in the same order in the *M. demolitor* nudivirus cluster (Fig. 4A). *Int-1* and *38k* in cluster 3 and *K425_459-like*, *HzNVorf106-like*, and *HzNVorf9-1-like* in cluster 5 also exhibited syntenic order with the *M. demolitor* nudivirus cluster (Fig. 4A), which overall suggested a nudivirus gene cluster existed in the common ancestor of microgastroid wasps that has persisted in microgastrines like *M. demolitor* and *C. congregata* but has largely dispersed in chelonines like *C. insularis*. Four other small clusters of nudivirus genes were also

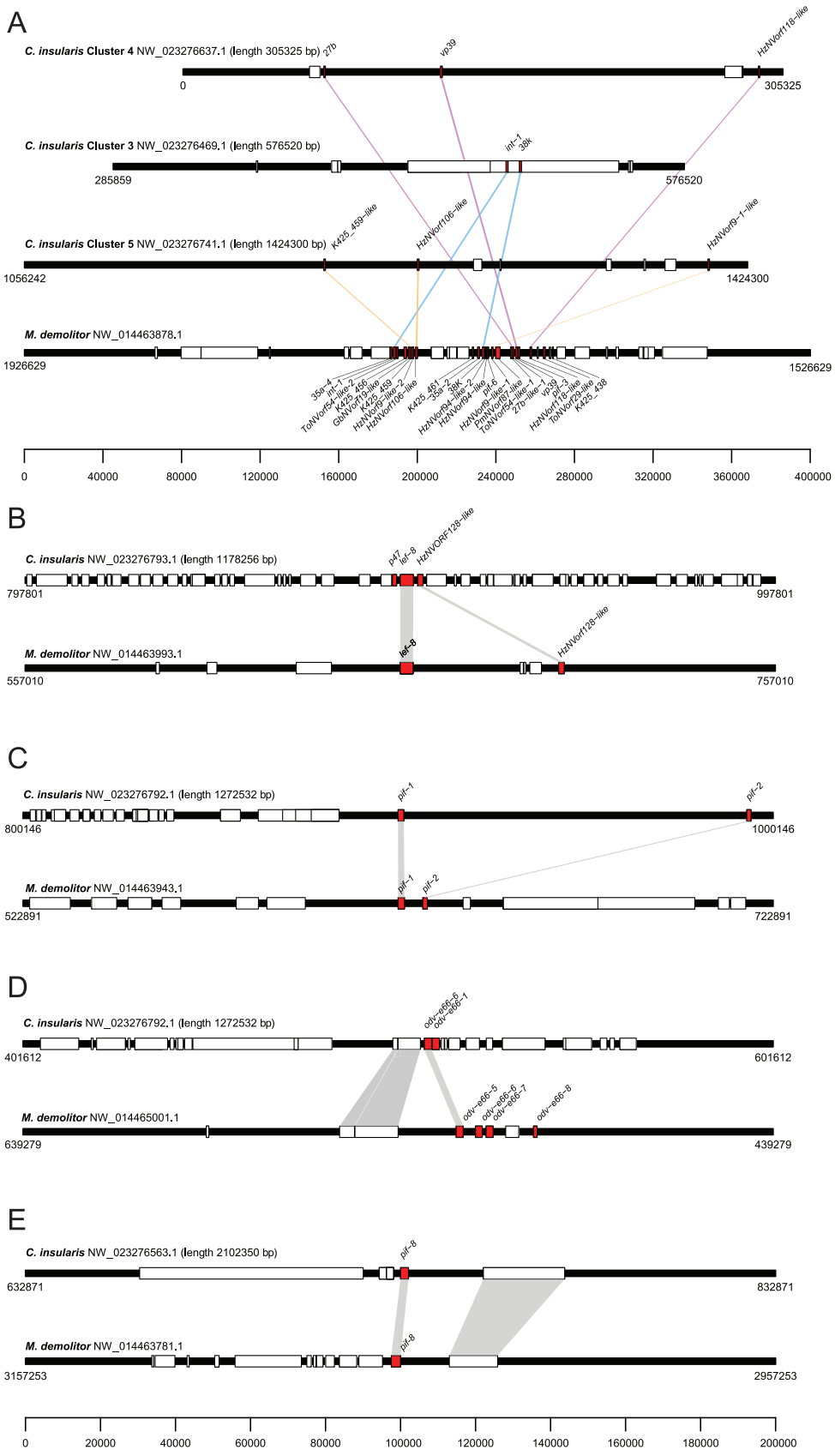

FIG 4 Regions of nudivirus gene microsynteny between *C. insularis* and *M. demolitor*. Boxes identify genes as defined in Fig. 1. Shading between scaffolds indicate orthologous genes.

identified in *C. insularis* that exhibited syntenic features with *M. demolitor*. In *C. insularis*, *p47*, *lef-8*, and *HzNVorf128-like* clustered together while in *M. demolitor lef-8* and *HzNVorf128-like* reside on the same scaffold with intervening wasp genes gained and *p47* moving to another location in the genome (Fig. 4B). We found that *pif-1* and *pif-2* were located together in both genomes albeit in different locations as evidenced by different neighboring wasp genes (Fig. 4C). Paralogs of *odv-e66* in both species independently underwent localized duplications next to the same wasp gene (nucleolar preribosomal associated protein), while *pif-8* in both species was located next to a conserved-position wasp gene (*zwei Ig domain protein zig-8-like*) (Fig. 4D and E).

Microsynteny of genes surrounding proviral segments has been used to determine homologous segment locations in genomes as well as segment gain and loss events between species in the Microgastrinae (18, 19). However, we identified no syntenous blocks of greater than three genes between any of the proviral segments in *C. insularis* and *M. demolitor*. While most nudivirus-like genes reside distantly from proviral segments, the small number near proviral segments in *M. demolitor* and *C. congregata* have been suggested as evidence that certain nudivirus genes and proviral segments reflect an ancestral association (18, 19). For example, *pif-0* resides within 50 kb of locus 2 that contains five tandemly arrayed proviral segments in *M. demolitor*, while *lef-5* is distant from any proviral segment but resides next to *vlf-1* (19). Yet in *C. insularis lef-5* is distant from *vlf-1* but resides within 40 kb of locus 1 that contains 8 tandemly arrayed proviral segments (Fig. 2), while *pif-0* is located on scaffold with no proviral segments (Table 4). Rather than reflecting an ancestral association, these results may simply be a product of different genome rearrangements that have occurred in *C. insularis* and *M. demolitor* due to movement of proviral segments or erosion of synteny surrounding proviral segments over time.

DISCUSSION

BVs retain many genes from their *Betanudivirus* ancestor, but function as transducing agents that female wasps use to transfer genes to hosts for the benefit of their offspring. Dispersal of nudivirus genes to locations outside proviral segments prevents BVs from being able to replicate outside wasps. However, the nudivirus cluster in *M. demolitor* and *C. congregata* has been suggested to be important for virion formation, while the clustering of proviral segments could be important for amplification, processing, or the evolution of virulence gene variants in molecular arms races with hosts (18). Our analysis of the *C. insularis* genome indicates several nudivirus genes and organizational features of proviral segments are likely conserved across all wasps in the microgastroid complex, while other features like the nudivirus cluster are not.

Twenty-one core genes are conserved in all baculovirus and nudivirus genomes: *dnapol* and *helicase* (DNA replication); *lef-4*, *lef-5*, *lef-8*, *lef-9*, *p47* (transcription of viral genes); *38k*, *p33*, *p6.9*, *vlf-1*, *vp39* (DNA packaging, virion production and assembly); *pif-0 through pif-6*, *pif-8* (infectivity); and *ac81* (unknown function) (12, 13). Nudivirus gene inventories in *C. insularis*, *M. demolitor* and *C. congregata* indicate all except *dnapol* and *ac81* are present (*p6.9* is difficult to detect but may be present in all three species [18]) (Table 4). The absence of a baculovirus/nudivirus *dnapol* in all three of these species underscores this gene was lost while an unknown wasp DNA polymerase was likely recruited early in BV evolution to replicate proviral segments. Eleven additional genes present in all sequenced nudiviruses have been suggested as nudivirus-specific core genes: *helicase-2*, *integrase* (*int*), *fen-1*, three thymidine kinases (*tk1-3*), *11k*, *HzNVorf143-like*, *OrNVorf18-like*, *PmNVorf62-like*, and *GbNVorf19-like*. Among these genes, only the *int*, *OrNVorf18-like*, and *11k* genes remain recognizable in *C. insularis*, *M. demolitor* and *C. congregata* (Table 4). Like other large DNA viruses, baculoviruses and nudiviruses encode a number of other genes that are present in some but not all lineages or that are unique to particular species (12, 13). *C. insularis*, *M. demolitor* and *C. congregata* all encode *odv-e66* genes that are present in many but not all baculoviruses and nudiviruses as well as several genes that are present in beta- but not alphanudiviruses (Table 4). Altogether, the genes shared by *C. insularis*, *M. demolitor* and *C. congregata* may comprise a BV conserved gene

set, which consists of most of the core genes shared by baculoviruses and nudiviruses, three of the proposed nudivirus-specific core genes (*int*, *OrNVorf18-like*, and *11k*), *odv-e66*, and certain genes (*HzNVorf9*, *HzNVorf64*, *HzNVorf93*, *HzNVorf106*, *HzNVorf118*, and *HzNVorf128*) known only from betanudiviruses.

It is possible other genes from the nudivirus ancestor are also shared among all microgastroid wasps but cannot be conclusively categorized as such because of their absence from currently sequenced nudivirus genomes. Due to their presence in the *M. demolitor* and *C. congregata* nudivirus cluster, four intronless genes are counted as nudivirus genes (*K425_438*, *K425_456*, *K425_459*, *K425_461*) (18, 19), but results from this study indicate only *K425_459* is present in *C. insularis*. Additionally, 15 intronless genes were identified as encoding structural components of CiBV virions that are also present in *C. insularis* (Table 6), with only *17a*, *35a*, and *30b* also present in *M. demolitor* and *C. congregata*. *C. insularis* further lacks nudivirus genes like *HzNVorf94-like*, *int-2*, *fen-1*, *PmNVorf87-like*, and *ToNVorf54-like* that are present in *M. demolitor* and *C. congregata*, suggesting some genes from the nudivirus ancestor have persisted in certain lineages of microgastroid wasps but not others, which in turn could result in differing replication or structural properties of BV virions among lineages.

Our results overall support that the single large nudivirus cluster present in *Microplitis* and *Cotesia* spp. derives from a syntenic domain from the ancestral virus that has largely been lost in *C. insularis*. However, this finding also indicates the large nudivirus cluster present in *M. demolitor* and *C. congregata* is likely not essential for high-level production of virions in calyx cells, which does not appear to differ among *Chelonus*, *Microplitis* and *Cotesia* spp. (14, 42, 43). The dispersal of BV genomes also does not functionally impede high level replication of proviral segments or virion formation, which in the case of MdBV exceeds replication rates of baculoviruses (44). We thus conclude that three other previously noted features (19) conserved between chelonine and microgastrine wasps are likely more important for virion formation. The first is the baculovirus/nudivirus-like RNA polymerase from the *p47*, *lef-4*, *lef-8*, and *lef-9* genes. Similar to baculoviruses, the viral RNA polymerase in *M. demolitor* specifically transcribes nudivirus structural genes through promoter recognition which likely enables it to transcribe these genes regardless of their location in different microgastroid species (45). The second is the still unknown DNA polymerase(s) that amplifies proviral segments by also recognizing conserved motifs, while the third is the nudivirus-like integrase/recombinase genes (*vlf-1*, *int*) that experimental data from *M. demolitor* implicate in recognizing the WIM domains that likely flank all proviral segments in microgastroid wasps to produce the circularized segments in virions (45).

The similarities in proviral flanking sequences between *C. insularis*, *M. demolitor* and *C. congregata* support shared ancestry across the microgastroid complex although the origin of these motifs remains unclear as no sequenced nudiviruses encode similar domains. However, it is also possible integrase/recombinase gene function has potentially diverged between chelonine and microgastrine wasps given the greater variability in the sites of circularization for segments produced by *C. inanitus* and *C. insularis* versus *M. demolitor* and *C. congregata*. In contrast, duplication of genes into families appears to be a key means of generating novelty on proviral segments in all microgastroid wasps given the prevalence of multimember gene families across all BVs that have been examined to date (4, 5). Previous studies of *M. demolitor*, *C. congregata* and other species in the Microgastrinae indicate some genes and gene families on proviral segments are recent acquisitions from wasps (46–48), while others show evidence of acquisition from other insects or are very ancient and thus could have originated from the nudivirus ancestor (20, 44, 48). Most striking from sequencing *C. insularis* is that CinsBV proviral segments and associated genes share some homology with CiBV but gene inventories greatly differ from MdBV and CcBV, indicating gene gain and loss occurs frequently. Some of the differences in viral segment gene content and activity between microgastrine and chelonine species could be a consequence of the greater time of divergence compared to within-microgastrine comparisons. However, it seems likely that the inventory of genes on proviral segments is also strongly shaped by the specific interactions that have evolved between different lineages of microgastroid wasps and the host stages and/or species they parasitize. In the case of *C. insularis*, differences in proviral

segment gene inventories may reflect differing host usage strategies given that chelonine braconids are egg-larval parasitoids that cause hosts to precociously initiate metamorphosis, while most microgastrine braconids are larval parasitoids that usually prevent hosts from pupating (4, 49, 50).

Like all insects, the lepidopteran hosts of chelonine braconids grow as larvae by molting which is followed after the last instar by a metamorphic molt to the pupal stage (51). Larval-larval versus larval-pupal molts are also primarily regulated by the titers of two key hormones: 20-hydroxyecdysone (20E) which stimulates insects to molt when its titer rises, and juvenile hormone (JH) that at high titer when 20E is released results in molting to another instar (larva) but at low titer results in molting to a pupa (52, 53). Chelonine braconids are referred to as solitary, egg-larval parasitoids because females oviposit a single egg into the egg stage of their hosts which is followed by the host hatching and developing as a larva. In turn, the wasp egg inside the host hatches into a larva that develops by feeding on hemolymph in the host larva (54). Hosts parasitized by chelonine braconids undergo 'precocious' metamorphosis, which refers to the host larva initiating behavioral and developmental processes associated with molting to the pupal stage one instar earlier than normally occurs (54). Precocious metamorphosis of the host is further associated with the mature wasp larva emerging from the host's body to pupate while the host itself is unable to complete a pupal molt. Prior studies of *C. inanitus* establish that precocious metamorphosis correlates with host JH titers declining to undetectable levels one instar earlier than occurs in nonparasitized hosts. Concurrently, 20E titers do not increase normally, which is associated with the host larva being unable to complete a pupal molt (55, 56). Results from *C. curvimaculatus* and *C. inanitus* implicate both the wasp larva and BVs in regulating precocious metamorphosis but the specific genes involved are unknown (57–59).

Results from this study do not identify the gene products that cause precocious metamorphosis but they do suggest that similar to CiBV (22, 60–62) certain CinsBV genes are preferentially expressed in early or late instar parasitized hosts while others are expressed at similar levels. We hypothesize that CinsBV genes expressed similarly between host stages could have functions in processes like protection against host immune defenses that wasp progeny would be expected to require over the course of development. In contrast, genes that are preferentially expressed in early versus late instar hosts are potential candidates for having functions in regulating the endocrine processes that control host growth, including precocious metamorphosis. Lastly, we were surprised that some genes on *C. insularis* proviral segments were only detected in wasps or both wasps and hosts because few genes on MdBV and CcBV proviral segments are expressed in any wasp stage (62–64). We currently are uncertain about the function of these genes but speculate that they could be recently introduced into proviral segments from other locations in the wasp genome and as a result may still be responsive to regulatory factors and have functions that result in continued expression in wasps. It is also possible differences in methodology underestimate expression levels of some genes on proviral segments in species like *M. demolitor* and *C. congregata*, which would indicate functions in wasp biology may not be restricted to BVs associated with wasps in the subfamily Cheloninae.

MATERIALS AND METHODS

Insects. *Chelonus insularis* was collected at the University of Florida Everglades Research and Education Center in Belle Glade, Florida and then maintained at the University of Georgia on its host, *Spodoptera frugiperda*. *S. frugiperda* larvae were fed corn leaves while adult moths were provided 10% sucrose in plastic containers covered with paper towels which is where females laid eggs. *C. insularis* was reared by allowing adult females to oviposit into *S. frugiperda* eggs. After hatching, parasitized host larvae were provided corn leaves until wasp offspring emerged to pupate. Adult wasps were then kept in cages and fed a 10% sucrose solution in water. All cultures were maintained at 23°C, 40–50% humidity, with a 12 h light: 12 h dark photoperiod.

Genome Sequencing, Assembly and Validation. The Blood and Cell Culture DNA minikit (Qiagen) was used to isolate high molecular weight (HMW) DNA from adult male wasps. PacBio sequencing was performed using the PacBio Sequel System at the Georgia Genomics and Bioinformatics Core (GGBC). Genomic DNA for Illumina sequencing was extracted from one male wasp by homogenizing in lysis buffer (500 μl of 1xPBS, 2% sarkosyl and 0.5 mg/ml proteinase K) at 62°C for 1 h. After phenol:chloroform extraction, DNA was precipitated with 0.3 M sodium acetate (pH 5.2), 25 μg glycogen, and 100% isopropanol. The DNA

pellet was then resuspended in 20 μl Nuclease-Free water. The sample library was prepared and sequenced (2 $\times$ 75 bp reads) on a NextSeq 500 system at GGBC.

Pacbio raw reads were assembled using Flye V2.7.1 with a genome size setting of 200 Mb (65). The assembly was polished using Pilon V1.22 with Illumina paired-end reads (66). First, reads were adapter trimmed and quality filtered with Trimmomatic V0.36 (program settings: ILLUMINACLIP:2:20:10:1 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:36) (67). Processed reads were then mapped to the assembly using BWA V0.7.17 with the MEM option (68). Finally, Pilon was performed to improve assembly quality with default settings. Blobtools V1.1.1 was used to remove potential contaminants (69). The completeness of the genome assembly was assessed by BUSCO V4.0.5. with the "Insecta" data set (70) and assembly statistics were generated with QUAST V5.0.2 (71). We further performed a kmer analysis using KAT V2.4.1 to assess the completeness and heterozygosity of the assembly (72).

RNA purification and sequencing. RNA was extracted from different wasp stages (larvae, pupae, adults), ovaries from adult females, and parasitized first and fourth instar hosts from which wasp larvae were removed by dissection (Table 3). Samples were first extracted using the RNeasy minikit with an on-column DNase step (Qiagen), followed by re-extraction with acid phenol:chloroform and ethanol precipitation in the presence of 0.5 M NaCl. Samples were then treated with the Ambion TURBO DNA-free kit (Invitrogen). Standard strand-specific Illumina-compatible libraries were then constructed from more than 1 μg of total RNA starting material for each sample, and sequenced (2 $\times$ 150 bp reads) by GGBC using the NextSeq system.

DNA extraction and sequencing from virions. Ovaries were dissected from 50 adult female wasps and homogenized in 200 μl of TURBO 1$\times$ DNase buffer. The tissue was gently spun down at 200 $\times$ *g* and the supernatant containing CinsBV virions was passed through a 0.45 μm filter. DNase was added to the supernatant and incubated at 37°C for 1 h to remove exogenous DNA outside virions. After adding RNase A (Qiagen) to 0.8 mg/ml for 2 min, the DNase was inactivated by adding EDTA to 10 mM. DNA was then extracted from virions using proteinase K, sarkosyl and phenol-chloroform extraction method (44) followed by sequencing on a NextSeq system as described above.

Genome annotation. Annotation of the *C. insularis* genome assembly was performed using the NCBI Eukaryotic Genome Annotation Pipeline (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/). RNASeq data for *C. insularis* were compared to NCBI RefSeq protein sets for *Diachasma alloeum*, *Microplitis demolitor*, *Fopius arisanus*, *Nasonia vitripennis*, *Apis mellifera*, *Chelonus inanitus*, 39,059 other Insecta RefSeq proteins, and 106,743 GenBank Insecta proteins for gene prediction. Statistics and the evidence used for annotation are available at https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Chelonus_insularis/100/.

Nudivirus gene annotation. Nudivirus homologs in the *C. insularis* genome were identified using two strategies. First, ORFs from the *C. insularis* genome assembly and annotations generated by NCBI were searched against a database of all nudivirus-like proteins from *C. inanitus*, *M. demolitor*, and *C. congregata* using BLASTP (E value $<$ 0.01) (8, 18, 19). Second, the set of genome ORFs or annotations was searched against a custom protein database containing all viral protein sequences from the NCBI nr database (downloaded July 2020). All of the identified ORFs or gene annotations were then manually converted into annotated gene models with the *C. insularis* jBrowse/Apollo instance on the i5k workspace (https://i5k.nal.usda.gov/available-genome-browsers). Manual annotations were merged with NCBI annotations to generate a *C. insularis* Official Gene Set OGSv1.0 (DOI 10.15482/USDA.ADC/1523023). Bigwig coverage blots generated from the transcriptome data sets generated were then used to delineate nudivirus gene transcription boundaries.

Proviral segment analysis. Paired-end Illumina sequence reads from DNA that was isolated from CinsBV virions were mapped to the *C. insularis* genome using BWA V0.7.17 with the MEM option. A BAM file was imported to Tablet to view the mapping result (73). Regions of scaffolds with high read coverage and clear coverage boundaries indicating the presence of segments were then selected. Relative coverage was calculated for each proviral segment by normalizing coverage values to set the segment with the lowest coverage to 1$\times$. Segments previously cloned and sequenced from CiBV virions (21) were used to search the proviral segments identified in *C. insularis* using BLASTN (E value $<$ 1e-20). Coordinates of each alignment were plotted as line segments on grids representing pairwise comparisons between segments from *C. inanitus* and *C. insularis* with ggplot2 in R. The genes on segments were predicted by NCBI and supplemented with manual annotations from tblastn results of CiBV proteins against the *C. insularis* genome (E value $<$ 0.01) and the FGENESH gene-finder (74), which were then searched against the nr database and manually annotated with Apollo. Wasp Integration Motifs (WIMs) were identified using read mapping data as described above. A conserved tetramer AGCT in WIMs was used as segment start and end coordinates (31, 32). A 100 bp flanking sequence at both ends of each segment was extracted and aligned using MAFFT (75) to identify segment direction and conserved motifs. Host integration motifs (HIMs) from *M. demolitor* and *C. congregata* were used as queries for blastn searches (E value $<$ 0.01), and were also used to construct a Hidden Markov Model (HMM) to search against the *C. insularis* proviral segments using hmmsearch (18, 19). Candidate *C. insularis* HIMs were then further aligned with the *M. demolitor* and *C. congregata* HIMs using MAFFT. Sequence logos of the conserved motifs were generated using the WebLogo server (76).

Expression analysis of CinsBV genes. Raw reads from samples listed in Table 3 were adapter-trimmed and quality filtered with Trimmomatic v0.36 using the above program settings. Expressed genes in different wasp or host stages were determined by read mapping to the assembled genome with HISAT2 (77). Transcript assembly and abundance estimation were performed using StringTie and read counts per gene were obtained using Ballgown (78).

Phylogenetic analysis of CinsBV genes. Integrase/recombinase family members, *int-1* (integrase-1), *vlf-1* (very late expression factor-1) and *HzNVorf140-like*, were identified in *C. insularis* using select sequences from other BVs, nudiviruses and baculoviruses that were used previously (44) as queries in a blastp search (E value $<$0.01). Proteins containing ankyrin domains from *C. insularis* and *M. demolitor*

were identified using hmmsearch with the PFAM Hidden Markov Models Ank, Ank_2, Ank_3, Ank_4, and Ank_5. Ankyrin domain-containing sequences from other BVs and ichnoviruses (IVs) were retrieved using accession numbers reported in earlier studies (79, 80). Amino acid sequences of each gene were aligned with MAFFT using the L-INS-i model (75). Poorly aligned positions were excluded by trimAl V1.2 (81). Substitution models for each gene were determined with jModelTest2 (82). Maximum Likelihood trees were inferred with RAxML-HPC2 via the CIPRES Science Gateway portal (83, 84).

Data availability. RNA-seq and Pacbio reads from *Chelonus insularis* samples are available in the Sequence Read Archive under accession numbers SRR11845184 to SRR11845190, SRR11967918 to SRR11967921, and SRR11678241 and SRR11678242.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE1, PDF file, 2.7 MB.

ACKNOWLEDGMENTS

We thank Hannah Boomgarden and Jena Johnson for assistance with insect rearing, and Amy Rowley and Robert Meagher for initial *C. insularis* collections and the gift of their wasp colony along with rearing advice. This work was supported by the US National Science Foundation (DEB-1916788).

REFERENCES

1. Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. PLoS Genet 6:e1001191. https://doi.org/10.1371/journal.pgen.1001191.
2. Holmes EC. 2011. The evolution of endogenous viral elements. Cell Host Microbe 10:368–377. https://doi.org/10.1016/j.chom.2011.09.002.
3. Feschotte C, Gilbert C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet 13:283–296. https://doi.org/10.1038/nrg3199.
4. Strand MR, Burke GR. 2014. Polydnaviruses: nature's genetic engineers. Annu Rev Virol 1:333–354. https://doi.org/10.1146/annurev-virology-031413-085451.
5. Gauthier J, Drezen J-M, Herniou EA. 2018. The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps. Parasitology 145:713–723. https://doi.org/10.1017/S0031182017000725.
6. Burke GR. 2019. Common themes in three independently derived endogenous nudivirus elements in parasitoid wasps. Curr Opin Insect Sci 32:28–35. https://doi.org/10.1016/j.cois.2018.10.005.
7. Whitfield JB. 2002. Estimating the age of the polydnavirus/braconid wasp symbiosis. Proc Natl Acad Sci U S A 99:7508–7513. https://doi.org/10.1073/pnas.112067199.
8. Bézier A, Annaheim M, Herbinière J, Wetterwald C, Gyapay G, Bernard-Samain S, Wincker P, Roditi I, Heller M, Belghazi M, Pfister-Wilhem R, Periquet G, Dupuy C, Huguet E, Volkoff A-N, Lanzrein B, Drezen J-M. 2009. Polydnaviruses of braconid wasps derive from an ancestral nudivirus. Science 323:926–930. https://doi.org/10.1126/science.1166788.
9. Thézé J, Bézier A, Periquet G, Drezen J-M, Herniou EA. 2011. Paleozoic origin of insect large dsDNA viruses. Proc Natl Acad Sci U S A 108:15931–15935. https://doi.org/10.1073/pnas.1105580108.
10. Murphy N, Banks JC, Whitfield JB, Austin AD. 2008. 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 47:378–395. https://doi.org/10.1016/j.ympev.2008.01.022.
11. Smith MA, Rodriguez JJ, Whitfield JB, Deans AR, Janzen DH, Hallwachs W, Hebert PD. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc Natl Acad Sci U S A 105:12359–12364. https://doi.org/10.1073/pnas.0805319105.
12. Rohrmann GF. 2019. Baculovirus molecular biology, Bethesda (MD): National Center for Biotechnology Information (US) ed
13. Harrison RL, Herniou EA, Bézier A, Jehle JA, Burand JP, Theilmann DA, Krell PJ, van Oers MM, Nakai M, Consortium IR. 2020. ICTV virus taxonomy profile: *nudiviridae*. J Gen Virol 101:3–4. https://doi.org/10.1099/jgv.0.001381.
14. Stoltz DB, Vinson SB. 1979. Viruses and parasitism in insects. Adv Virus Res 24:125–171. https://doi.org/10.1016/s0065-3527(08)60393-0.
15. Webb BA, Strand MR. 2005. The biology and genomics of polydnaviruses. Comprehensive Molecular Insect Science 6:323–360.
16. Strand MR. 2012. Polydnavirus gene products that interact with the host immune system, p 149–161, Parasitoid Viruses. Elsevier.
17. Beck MH, Inman RB, Strand MR. 2007. Microplitis demolitor bracovirus genome segments vary in abundance and are individually packaged in virions. Virology 359:179–189. https://doi.org/10.1016/j.virol.2006.09.002.
18. Gauthier J, Boulain H, van Vugt JJ, Baudry L, Persyn E, Aury J-M, Noel B, Bretaudeau A, Legeai F, Warris S, Chebbi MA, Dubreuil G, Duvic B, Kremer N, Gayral P, Musset K, Josse T, Bigot D, Bressac C, Moreau S, Periquet G, Harry M, Montagné N, Boulogne I, Sabeti-Azad M, Maïbèche M, Chertemps T, Hilliou F, Siaussat D, Amselem J, Luyten I, Capdevielle-Dulac C, Labadie K, Merlin BL, Barbe V, de Boer JG, Marbouty M, Cônsoli FL, Dupas S, Hua-Van A, Le Goff G, Bézier A, Jacquin-Joly E, Whitfield JB, Vet LE, Smid HM, Kaiser L, Koszul R, Huguet E, Herniou EA, et al. 2021. Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. Commun Biol 4:1–15. https://doi.org/10.1038/s42003-021-02480-9.
19. Burke GR, Walden KK, Whitfield JB, Robertson HM, Strand MR. 2014. Widespread genome reorganization of an obligate virus mutualist. PLoS Genet 10:e1004660. https://doi.org/10.1371/journal.pgen.1004660.
20. Huguet E, Serbielle C, Moreau SJ. 2012. Evolution and origin of polydnavirus virulence genes, p 63–78, Parasitoid Viruses. Elsevier.
21. Wyder S, Tschannen A, Hochuli A, Gruber A, Saladin V, Zumbach S, Lanzrein B. 2002. Characterization of Chelonus inanitus polydnavirus segments: sequences and analysis, excision site and demonstration of clustering. J Gen Virol 83:247–256. https://doi.org/10.1099/0022-1317-83-1-247.
22. Weber B, Annaheim M, Lanzrein B. 2007. Transcriptional analysis of polydnaviral genes in the course of parasitization reveals segment-specific patterns. Arch Insect Biochem Physiol 66:9–22. https://doi.org/10.1002/arch.20190.
23. Wetterwald C, Roth T, Kaeslin M, Annaheim M, Wespi G, Heller M, Mäser P, Roditi I, Pfister-Wilhelm R, Bézier A, Gyapay G, Drezen J-M, Lanzrein B. 2010. Identification of bracovirus particle proteins and analysis of their transcript levels at the stage of virion formation. J Gen Virol 91:2610–2619. https://doi.org/10.1099/vir.0.022699-0.
24. Jourdie V, Alvarez N, Turlings TC. 2008. Identification of seven species of hymenopteran parasitoids of *Spodoptera frugiperda*, using polymerase chain reaction amplification and restriction enzyme digestion. Agric Forest Ent 10:129–136. https://doi.org/10.1111/j.1461-9563.2008.00362.x.
25. Burke GR, Walden KK, Whitfield JB, Robertson HM, Strand MR. 2018. Whole genome sequence of the parasitoid wasp *Microplitis demolitor* that harbors an endogenous virus mutualist. G3 (Bethesda) 8:2875–2880. https://doi.org/10.1534/g3.118.200308.
26. Geib SM, Liang GH, Murphy TD, Sim SB. 2017. Whole genome sequencing of the braconid parasitoid wasp *Fopius arisanus*, an important biocontrol agent of pest tepritid fruit flies. G3 (Bethesda) 7:2407–2411. https://doi.org/10.1534/g3.117.040741.
27. Bézier A, Thézé J, Gavory F, Gaillard J, Poulain J, Drezen J-M, Herniou EA. 2015. The genome of the nucleopolyhedrosis-causing virus from *Tipula oleracea* sheds new light on the Nudiviridae family. J Virol 89:3008–3025. https://doi.org/10.1128/JVI.02884-14.

28. Arvin MJ, Lorenzi A, Burke GR, Strand MR. 2021. MdBVe46 is an envelope protein that is required for virion formation by Microplitis demolitor bracovirus. J Gen Virol 102:e001565.
29. Louis F, Bézier A, Periquet G, Ferras C, Drezen J-M, Dupuy C. 2013. The bracovirus genome of the parasitoid wasp *Cotesia congregata* is amplified within 13 replication units, including sequences not packaged in the particles. J Virol 87:9649–9660. https://doi.org/10.1128/JVI.00886-13.
30. Burke GR, Simmonds TJ, Thomas SA, Strand MR. 2015. *Microplitis demolitor* bracovirus proviral loci and clustered replication genes exhibit distinct DNA amplification patterns during replication. J Virol 89:9511–9523. https://doi.org/10.1128/JVI.01388-15.
31. Savary S, Beckage N, Tan F, Periquet G, Drezen J. 1997. Excision of the polydnavirus chromosomal integrated EP1 sequence of the parasitoid wasp *Cotesia congregata* (Braconidae, Microgastinae) at potential recombinase binding sites. J Gen Virol 78:3125–3134. https://doi.org/10.1099/0022-1317-78-12-3125.
32. Beck MH, Zhang S, Bitra K, Burke GR, Strand MR. 2011. The encapsidated genome of *Microplitis demolitor* bracovirus integrates into the host *Pseudoplusia includens*. J Virol 85:11685–11696. https://doi.org/10.1128/JVI.05726-11.
33. Pasquier-Barre F, Dupuy C, Huguet E, Monteiro F, Moreau A, Poirié M, Drezen J-M. 2002. Polydnavirus replication: the EP1 segment of the parasitoid wasp *Cotesia congregata* is amplified within a larger precursor molecule. J Gen Virol 83:2035–2045. https://doi.org/10.1099/0022-1317-83-8-2035.
34. Albrecht U, Wyler T, Pfister-Wilhelm R, Gruber A, Stettler P, Heiniger P, Kurt E, Schümperli D, Lanzrein B. 1994. Polydnavirus of the parasitic wasp *Chelonus inanitus* (Braconidae): characterization, genome organization and time point of replication. J Gen Virol 75:3353–3363. https://doi.org/10.1099/0022-1317-75-12-3353.
35. Chevignon G, Periquet G, Gyapay G, Vega-Czarny N, Musset K, Drezen J-M, Huguet E. 2018. Cotesia congregata bracovirus circles encoding PTP and *Ankyrin* genes integrate into the DNA of parasitized *Manduca sexta* hemocytes. J Virol 92:e00438-18. https://doi.org/10.1128/JVI.00438-18.
36. Annaheim M, Lanzrein B. 2007. Genome organization of the *Chelonus inanitus* polydnavirus: excision sites, spacers and abundance of proviral and excised segments. J Gen Virol 88:450–457. https://doi.org/10.1099/vir.0.82396-0.
37. Chelliah J, Jones D. 1990. Biochemical and immunological studies of proteins from polydnavirus *Chelonus* sp. near curvimaculatus. J Gen Virol 71:2353–2359. https://doi.org/10.1099/0022-1317-71-10-2353.
38. Allain TW, Stentiford GD, Bass D, Behringer DC, Bojko J. 2020. A novel nudivirus infecting the invasive demon shrimp *Dikerogammarus haemobaphes* (Amphipoda). Sci Rep 10:1–13. https://doi.org/10.1038/s41598-020-71776-3.
39. Kroemer JA, Webb BA. 2005. I$\kappa\beta$-related vankyrin genes in the *Campoletis sonorensis* ichnovirus: temporal and tissue-specific patterns of expression in parasitized *Heliothis virescens* lepidopteran hosts. J Virol 79:7617–7628. https://doi.org/10.1128/JVI.79.12.7617-7628.2005.
40. Bézier A, Louis F, Jancek S, Periquet G, Thézé J, Gyapay G, Musset K, Lesobre J, Lenoble P, Dupuy C, Gundersen-Rindal D, Herniou EA, Drezen J-M. 2013. 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 368:20130047. https://doi.org/10.1098/rstb.2013.0047.
41. Desjardins CA, Gundersen-Rindal DE, Hostetler JB, Tallon LJ, Fuester RW, Schatz MC, Pedroni MJ, Fadrosh DW, Haas BJ, Toms BS, Chen D, Nene V. 2007. Structure and evolution of a proviral locus of *Glyptapanteles indiensis* bracovirus. BMC Microbiol 7:1–17. https://doi.org/10.1186/1471-2180-7-61.
42. Strand M. 1994. *Microplitis demolitor* polydnavirus infects and expresses in specific morphotypes of *Pseudoplusia includens* haemocytes. J Gen Virol 75:3007–3020. https://doi.org/10.1099/0022-1317-75-11-3007.
43. Wyler T, Lanzrein B. 2003. 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 84:1151–1163. https://doi.org/10.1099/vir.0.18830-0.
44. Burke GR, Strand MR. 2012. Deep sequencing identifies viral and wasp genes with potential roles in replication of *Microplitis demolitor* bracovirus. J Virol 86:3293–3306. https://doi.org/10.1128/JVI.06434-11.
45. Burke GR, Thomas SA, Eum JH, Strand MR. 2013. Mutualistic polydnaviruses share essential replication gene functions with pathogenic ancestors. PLoS Pathog 9:e1003348. https://doi.org/10.1371/journal.ppat.1003348.
46. Desjardins CA, Gundersen-Rindal DE, Hostetler JB, Tallon LJ, Fadrosh DW, Fuester RW, Pedroni MJ, Haas BJ, Schatz MC, Jones KM, Crabtree J, Forberger H, Nene V. 2008. Comparative genomics of mutualistic viruses of *Glyptapanteles* parasitic wasps. Genome Biol 9:R183–17. https://doi.org/10.1186/gb-2008-9-12-r183.
47. Burke GR, Strand MR. 2014. Systematic analysis of a wasp parasitism arsenal. Mol Ecol 23:890–901. https://doi.org/10.1111/mec.12648.
48. Serbielle C, Dupas S, Perdereau E, Héricourt F, Dupuy C, Huguet E, Drezen J-M. 2012. Evolutionary mechanisms driving the evolution of a large polydnavirus gene family coding for protein tyrosine phosphatases. BMC Evol Biol 12:253. https://doi.org/10.1186/1471-2148-12-253.
49. Pennacchio F, Strand MR. 2006. Evolution of developmental strategies in parasitic Hymenoptera. Annu Rev Entomol 51:233–258. https://doi.org/10.1146/annurev.ento.51.110104.151029.
50. Strand MR. 2010. Polydnaviruses. Insect virology Caister Academic Press, Norwich, United Kingdom:171–197.
51. Nijhout HF. 1981. Physiological control of molting in insects. Am Zool 21:631–640. https://doi.org/10.1093/icb/21.3.631.
52. Riddiford LM. 1993. Hormones and *Drosophila* development: p 899–939. *In* Bate M, Arias AM (ed), The Development of *Drosophila melanogaster*, vol 2. Cold Spring Harbor Press, Cold Spring Harbor, NY.
53. Riddiford LM, Hiruma K, Zhou X, Nelson CA. 2003. Insights into the molecular basis of the hormonal control of molting and metamorphosis from *Manduca sexta* and *Drosophila melanogaster*. Insect Biochem Mol Biol 33:1327–1338. https://doi.org/10.1016/j.ibmb.2003.06.001.
54. Jones G. 1985. The role of juvenile hormone esterase in terminating larval feeding and initiating metamorphic development in *Trichoplusia ni*. Entomol Exp Appl 39:171–176. https://doi.org/10.1111/j.1570-7458.1985.tb03559.x.
55. Steiner B, Pfister-Wilhelm R, Grossniklaus-Bürgin C, Rembold H, Treiblmayr K, Lanzrein B. 1999. Titres of juvenile hormone I, II and III in *Spodoptera littoralis* (Noctuidae) from the egg to the pupal moult and their modification by the egg–larval parasitoid *Chelonus inanitus* (Braconidae). J Insect Physiol 45:401–413. https://doi.org/10.1016/s0022-1910(98)00139-5.
56. Pfister-Wilhelm R, Lanzrein B. 2009. Stage dependent influences of polydnaviruses and the parasitoid larva on host ecdysteroids. J Insect Physiol 55:707–715. https://doi.org/10.1016/j.jinsphys.2009.04.018.
57. Soller M, Lanzrein B. 1996. Polydnavirus and venom of the egg-larval parasitoid *Chelonus inanitus* (Braconidae) induce developmental arrest in the prepupa of its host *Spodoptera littoralis* (Noctuidae). J Insect Physiol 42:471–481. https://doi.org/10.1016/0022-1910(95)00132-8.
58. Grossniklaus-Bürgin C, Pfister-Wilhelm R, Meyer V, Treiblmayr K, Lanzrein B. 1998. Physiological and endocrine changes associated with polydnavirus/venom in the parasitoid–host system *Chelonus inanitus-Spodoptera littoralis*. J Insect Physiol 44:305–321. https://doi.org/10.1016/s0022-1910(97)00144-3.
59. Jones D, Wache S. 1998. Preultimate 4th/5th instar *Trichoplusia ni* naturally-injected with venom/calyx fluid from *Chelonus curvimaculatus* precociously metamorphose, rather than obey the metamorphic size threshold that would normally compel molting to a 5th/6th instar. J Insect Physiol 44:755–765. https://doi.org/10.1016/s0022-1910(98)00008-0.
60. Johner A, Lanzrein B. 2002. Characterization of two genes of the polydnavirus of *Chelonus inanitus* and their stage-specific expression in the host *Spodoptera littoralis*. J Gen Virol 83:1075–1085. https://doi.org/10.1099/0022-1317-83-5-1075.
61. Bonvin M, Kojic D, Blank F, Annaheim M, Wehrle I, Wyder S, Kaeslin M, Lanzrein B. 2004. Stage-dependent expression of *Chelonus inanitus* polydnavirus genes in the host and the parasitoid. J Insect Physiol 50:1015–1026. https://doi.org/10.1016/j.jinsphys.2004.09.002.
62. Strand MR, Burke GR. 2012. Polydnaviruses as symbionts and gene delivery systems. PLoS Pathog 8:e1002757. https://doi.org/10.1371/journal.ppat.1002757.
63. Bitra K, Zhang S, Strand MR. 2011. Transcriptomic profiling of *Microplitis demolitor* bracovirus reveals host, tissue and stage-specific patterns of activity. J Gen Virol 92:2060–2071. https://doi.org/10.1099/vir.0.032680-0.
64. Chevignon G, Thézé J, Cambier S, Poulain J, Da Silva C, Bézier A, Musset K, Moreau SJ, Drezen J-M, Huguet E. 2014. Functional annotation of *Cotesia congregata* bracovirus: identification of viral genes expressed in parasitized host immune tissues. J Virol 88:8795–8812. https://doi.org/10.1128/JVI.00209-14.
65. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
66. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
67. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
68. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
69. Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. F1000Res 6:1287. https://doi.org/10.12688/f1000research.12232.1.

70. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
71. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
72. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. 2017. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics 33:574–576. https://doi.org/10.1093/bioinformatics/btw663.
73. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D. 2010. Tablet—next generation sequence assembly visualization. Bioinformatics 26:401–402. https://doi.org/10.1093/bioinformatics/btp666.
74. Solovyev V, Kosarev P, Seledsov I, Vorobyev D. 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol 7:S10–12. https://doi.org/10.1186/gb-2006-7-s1-s10.
75. Katoh K, Misawa K, Kuma K-i, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436.
76. Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188–1190. https://doi.org/10.1101/gr.849004.
77. Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360. https://doi.org/10.1038/nmeth.3317.
78. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11:1650–1667. https://doi.org/10.1038/nprot.2016.095.
79. Chen Y-f, Gao F, Ye X-q, Wei S-j, Shi M, Zheng H-j, Chen X-X. 2011. Deep sequencing of *Cotesia vestalis* bracovirus reveals the complexity of a polydnavirus genome. Virology 414:42–50. https://doi.org/10.1016/j.virol.2011.03.009.
80. Lapointe R, Tanaka K, Barney WE, Whitfield JB, Banks JC, Béliveau C, Stoltz D, Webb BA, Cusson M. 2007. Genomic and morphological features of a banchine polydnavirus: comparison with bracoviruses and ichnoviruses. J Virol 81:6491–6501. https://doi.org/10.1128/JVI.02702-06.
81. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
82. Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772–772. https://doi.org/10.1038/nmeth.2109.
83. Miller MA, Pfeiffer W, Schwartz T. 2010. Proceedings of the gateway computing environments workshop (GCE), p 1–8, Creating the CIPRES science gateway for inference of large phylogenetic trees. IEEE New Orleans.
84. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
85. Bézier A, Harichaux G, Musset K, Labas V, Herniou EA. 2017. Qualitative proteomic analysis of *Tipula oleracea* nudivirus occlusion bodies. J Gen Virol 98:284–295. https://doi.org/10.1099/jgv.0.000661.
86. Hou D, Zhang L, Deng F, Fang W, Wang R, Liu X, Guo L, Rayner S, Chen X, Wang H, Hu Z. 2013. Comparative proteomics reveal fundamental structural and functional differences between the two progeny phenotypes of a baculovirus. J Virol 87:829–839. https://doi.org/10.1128/JVI.02329-12.