Available online at www.sciencedirect.com

SCIENCE DIRECT

Virology 347 (2006) 160 – 174

VIROLOGY

www.elsevier.com/locate/yviro

Polydnavirus genomes reflect their dual roles as mutualists and pathogens

Bruce A. Webb [a,*,5], Michael R. Strand [b,*], Stephanie E. Dickey [a], Markus H. Beck [b], Roland S. Hilgarth [a], Walter E. Barney [a], Kristy Kadash [b,1], Jeremy A. Kroemer [a], Karl G. Lindstrom [a], Walaikorn Rattanadechakul [a], Kent S. Shelby [a,2], Honglada Thoetkiattikul [b], Matthew W. Turnbull [a,3], R. Andrews Witherell [b,4]

[a] *Department of Entomology, University of Kentucky, Lexington, KY 40506, USA*
[b] *Department of Entomology, University of Georgia, Athens, GA 30602, USA*

Received 22 September 2005; returned to author for revision 17 October 2005; accepted 8 November 2005
Available online 27 December 2005

Abstract

Symbionts often exhibit significant reductions in genome complexity while pathogens often exhibit increased complexity through acquisition and diversification of virulence determinants. A few organisms have evolved complex life cycles in which they interact as symbionts with one host and pathogens with another. How the predicted and opposing influences of symbiosis and pathogenesis affect genome evolution in such instances, however, is unclear. The Polydnaviridae is a family of double-stranded (ds) DNA viruses associated with parasitoid wasps that parasitize other insects. Polydnaviruses (PDVs) only replicate in wasps but infect and cause severe disease in parasitized hosts. This disease is essential for survival of the parasitoid's offspring. Thus, a true mutualism exists between PDVs and wasps as viral transmission depends on parasitoid survival and parasitoid survival depends on viral infection of the wasp's host. To investigate how life cycle and ancestry affect PDVs, we compared the genomes of *Campoletis sonorensis ichnovirus* (*CsIV*) and *Microplitis demolitor bracovirus* (*MdBV*). CsIV and MdBV have no direct common ancestor, yet their encapsidated genomes share several features including segmentation, diversification of virulence genes into families, and the absence of genes required for replication. In contrast, CsIV and MdBV share few genes expressed in parasitized hosts. We conclude that the similar organizational features of PDV genomes reflect their shared life cycle but that PDVs associated with ichneumonid and braconid wasps have likely evolved different strategies to cause disease in the wasp's host and promote parasitoid survival.

Keywords: Viral evolution; Polydnavirus; Symbiosis; Pathogenesis; Parasitoid

Introduction

Multicellular organisms have evolved diverse strategies to defend themselves against pathogenic microbes and to maintain interactions with symbionts. Reciprocally, symbionts and pathogens experience selection pressures that differentially affect genome properties and the suite of genes they encode. Symbionts typically exhibit reductions in genome complexity and gene loss with dependence on the host (Shigenobu et al., 2000; Akman et al., 2002; Moran, 2003; Brownlie and O'Neill, 2005) while pathogens often exhibit increases in genome complexity through acquisition of new virulence genes or the evolution of novel gene variants (Hacker and Kaper, 2000; Parkhill et al., 2001; Salanoubat et al., 2002; Stewart et al., 2004). Unlike Bacteria, Archaea, and Eukaryotes, viruses are not a monophyletic assemblage and undoubtedly have multiple origins (Villareal, 2005). The evolution of large DNA viruses (genomes 100 kb to greater than 1 mb) such as poxviruses, iridoviruses and phycodnaviruses is of increasing interest because of their ancient origins, potential relatedness to other cellular organisms, and possible roles in the evolution of

* Corresponding authors.
E-mail addresses: bawebb@pop.uky.edu (B.A. Webb), mrstrand@bugs.net.edu (M.R. Strand).
[1] Current address: State Forensic Crime Lab, Albany, NY 12210, USA.
[2] Current address: USDA ARS Beneficial Insects Laboratory, Columbia, MO 65203, USA.
[3] Current address: Department of Entomology, Clemson University, Clemson SC 29634, USA.
[4] Current address: Department of Plant Pathology, University of Wisconsin, Madison 53706, WI, USA.
[5] Fax: +1 859 323 1120.

0042-6822/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2005.11.010

eukaryotic organelles (Blasco, 1995; Iyer et al., 2001; Woese, 2002; Delaroque et al., 2003; Raoult et al., 2004; Villareal, 2005).

As with other viruses, most large DNA viruses have a pathogenic association with their hosts, whereas symbiosis is rare. The exception to this are the large DNA viruses in the family Polydnaviridae that have evolved a complex association with parasitoid wasps and insect hosts (Fig. 1A). PDVs are the only viruses whose circular, double-stranded (ds) DNA

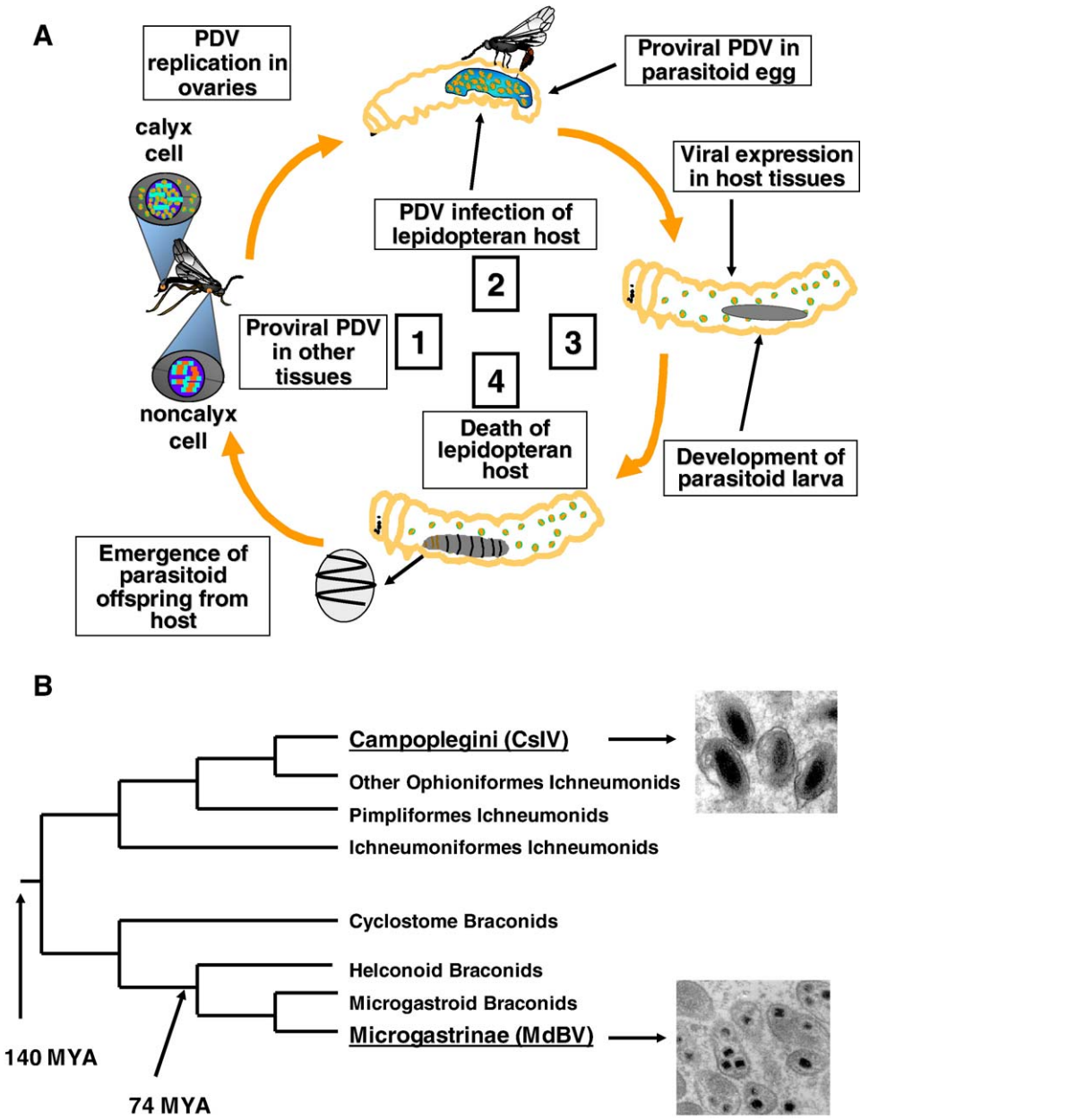


Fig. 1. Ichnoviruses (IVs) and bracoviruses (BVs) share a common life cycle but are associated with different lineages of ichneumonoid wasps. (A) Life cycle. (1) IVs and BVs exist as proviruses in all cells of their associated wasp. Virus replicates specifically in calyx cells of the female ovary and encapsidated virus accumulates in the lumen of the lateral oviducts. (2) At oviposition, the wasp injects one or more eggs into the host along with a quantity of virus and secretions from the venom gland. PDV virions infect hemocytes, fat body and other tissues of the parasitized host. The proviral form of the virus is integrated into the DNA of the wasp's egg. (3) Multiple PDV genes are expressed in host tissues beginning as early as 1 h post-parasitism and continuing as the parasitoid egg and larva develops. (4) The parasitoid offspring completes its juvenile development, pupates and emerges as an adult wasp while the host dies as a consequence of PDV infection and feeding by the parasitoid larva. (B) Phylogenetic organization of the superfamily Ichneumonoidea based on prior analyses (Quicke et al., 2000; Whitfield, 2002). The Ichneumonoidea likely diverged into two families, the Ichneumonidae and Braconidae, approximately 140 million years ago (MYA). Both families subsequently diversified into many subfamilies which collectively contain an estimated 100,000 species today. PDVs are associated with selected derived lineages of ichneumonids and braconids. Most IV-carrying wasps are in the subfamily Campopleginae including *Campoletis sonorensis*. A second subfamily of ichneumonids, Banchinae, also carry a PDV but no molecular data on these viruses is currently available. All BV-carrying wasps belong to a monophyletic microgastroid lineage which is divided into the Microgastrinae, which includes *Microplitis demolitor*, and two other subfamilies (Cardiochilinae, Cheloninae). No PDVs are known from any other ichneumonoid taxa.

genomes are segmented (Webb et al., 2000). PDVs persist as stably integrated proviruses in the genome of associated wasps and replicate in the ovaries of females where virions accumulate to high concentrations. The hosts of PDV-carrying wasps are primarily larval stage moths (Lepidoptera). When a female wasp oviposits into a host, she injects one or more eggs and virions that infect host immune cells and other tissues (Schmidt et al., 2001) (Fig. 1A). PDVs do not replicate in the wasp's host but expression of viral genes prevents the host immune system from killing the wasp's egg and causes other physiological alterations that ultimately cause the host to die (Asgari et al., 1996; Turnbull and Webb, 2002; Tanaka et al., 2000; Beckage and Gelman, 2004; Webb and Strand, 2005). Thus, a true mutualism exists between PDVs and wasps as viral transmission depends on parasitoid survival and parasitoid survival depends on viral infection of the wasp's host. Reciprocally, PDVs are virulent pathogens of the wasp's host.

The evolutionary success of this association is apparent as PDVs are carried by tens of thousands of parasitoid species that collectively form the most species-rich group of metazoan parasites known (Price, 1980; Godfray, 1994). All PDV-carrying parasitoids reside in the superfamily Ichneumonoidea which diverged into two families, the Ichneumonidae and Braconidae, approximately 140 million years ago (MYA) (Whitfield, 2002; Turnbull and Webb, 2002; Webb and Strand, 2005) (Fig. 1B). PDVs in turn are classified into two genera called the ichno-(IV) and bracoviruses (BV) (Webb et al., 2000). The BV-braconid association arose 74 MYA in a monophyletic lineage (microgastroid braconids) that now consists of more than 17,000 species with a worldwide distribution (Stoltz and Whitfield, 1992; Whitfield, 2002) (Fig. 1B). The evolutionary transitions among subfamilies in the Ichneumonidae are less clear but IVs are largely restricted to the Campopleginae which contains approximately 13,000 species (Quicke et al., 2000).

The complex life cycle of PDVs and restricted association with particular lineages of wasps raises two important questions. First, as obligatory symbionts, we might expect PDV genomes to exhibit reductions in complexity that reflect their reliance on wasps for persistence and replication. In contrast, the severe pathogenic effects PDVs have on parasitized hosts indicate that PDVs are also under strong selection pressure to maintain a suite of virulence genes. How the opposing influences of symbiosis and pathogenesis actually affect PDV genomes, however, is unclear. Second, while the similar life cycle and association of all PDVs with wasps suggests a common origin, distinct differences in virion morphology and packaging (Webb and Strand, 2005) and the absence of PDVs in any basal ichneumonoids that are ancestral to the PDV-carrying lineages suggests IV- and BV-parasitoid associations arose independently (Whitfield, 2002; Webb and Strand, 2005) (Fig. 1B). Delineating species and discerning higher order phylogenetic relationships among viruses is often very difficult (Villareal, 2005). However, PDVs are potentially more tractable because each virus is integrated into the genome of its wasp host and transmission is strictly vertical (Webb et al., 2000). As a result, PDVs persist via mendelian inheritance with each PDV-parasitoid species being a unique and identifiable biological entity.

To further explore the effects of life cycle and ancestry on PDV evolution, we report here the first complete genome sequence of an IV, *Campoletis sonorensis* ichnovirus (CsIV), and second complete sequence of a BV, *Microplitis demolitor* bracovirus (MdBV). Our results indicate that CsIV and MdBV share several genome properties but encode few shared genes. Comparisons to the recently sequenced genome of Cotesia congregata bracovirus (Espagne et al., 2004) and partial sequence data from other PDVs further suggest that gene transfer occurs commonly between PDVs and hosts, and that BVs have evolved different strategies to cause disease in parasitized hosts.

Results and discussion

CsIV and MdBV genomes share several genome properties

As previously noted, PDVs interact with two insect hosts: the wasp in which the virus persists and replicates and the insect

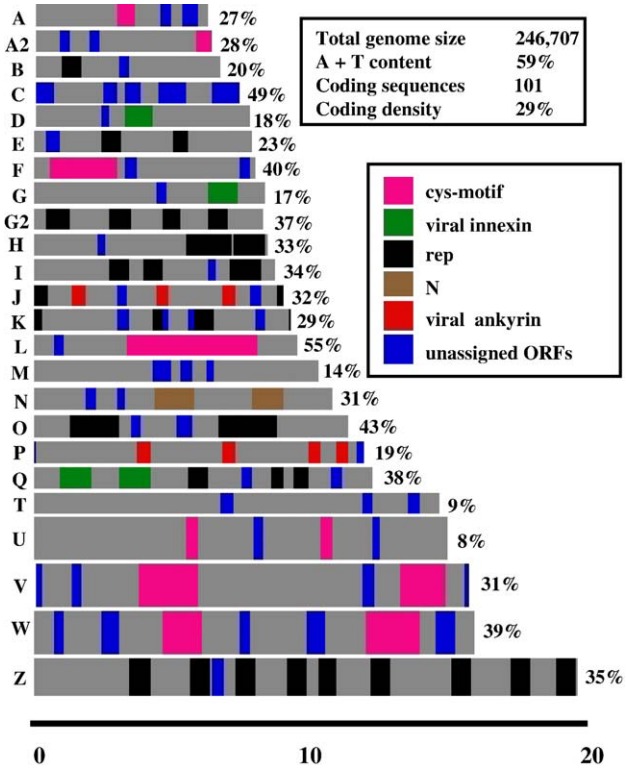


Fig. 2. Diagrammatic representation of sequence and annotation of the *Cs*IV genome. The 24 non-redundant circular *Cs*IV genome segments are represented as linear molecules to visualize segment size relationships within the genome. Individual segments are listed from smallest (*A* = 6.1 kb) to largest (*Z* = 19.6 kb) on left with percentages of coding DNA per segment identified on the right. Wider bars represent high-copy, repetitive type segments. High-copy segments have been identified by densitometry, quantitative Southern blot hybridization and/or quantitative PCR and are a minimum of 5 fold more abundant than low copy segments. ORFs and members of gene families have been annotated by color and location to individual genome segments with color key to *Cs*IV shown. Grey regions represent non-coding DNA. The aggregate *Cs*IV genome is 246,707 bp in size, 71% non-coding, and 29% coding overall. Scale bar is in kilobases.

host(s) that the wasp parasitizes. For the sake of clarity and to avoid confusion in the remainder of the manuscript, we refer to the wasp hosts of CsIV and MdBV as simply "wasp" or "parasitoid" while we refer to insects the wasp infects with virus when ovipositing as "hosts", "host larvae" or "parasitized hosts". *Campoletis sonorensis* is a campoplegine ichneumonid that is endemic to North America while *Microplitis demolitor* is a microgastrine braconid that is native to Australia. Both wasps parasitize the larval stage of related moths and fully depend upon infection of hosts by CsIV and MdBV respectively for survival (Webb and Strand, 2005). The 246,707 bp CsIV genome is divided into 24 ds, circular DNA segments that range in size from 6,138 bp (segment A) to 19,557 bp (Z) (Fig. 2). Eight other DNA segments are also encapsidated in CsIV virions but these 'nested' segments are smaller derivatives of larger segments and encode no novel sequence apart from that generated by recombination (Cui and Webb, 1997). Therefore, these CsIV segments are not described further. The 189,088 bp MdBV genome is divided into 15 ds, circular segments that range from 3,611 bp (A) to 34,334 bp (O) (Fig. 3). These sizes agree with prior estimates generated from agarose gel electrophoresis and restriction mapping (Blissard et al., 1989; Strand et al., 1992).

CsIV and MdBV share other genome properties besides segmentation. First, genomic segments in each virus vary in abundance with some segments being present at higher copy number than others. This was evident from random sequencing of CsIV and MdBV genomic clones and is also consistent with prior electrophoretic and hybridization studies (Blissard et al., 1989; Cui and Webb, 1997; Strand, 1994; Strand et al., 1997; Kroemer and Webb, 2005). The CsIV genome contains three high abundance segments (U, V and W), while the MdBV genome contains one (O). Notably, the high copy segments in each genome are also the largest segments. Since MdBV and other BVs package one viral segment per virion (Albrecht et al., 1994; Strand et al., unpublished), these results suggest that virions containing certain viral segments are more abundant than others in the assemblage of virus that wasps inject when parasitizing hosts. IV virions have capsid dimensions large enough to accommodate packaging of the entire genome (Webb and Strand, 2005) but whether all or only some segments are actually present in any single virion is unclear. Sequence analysis indicates that high abundance segments in the CsIV and MdBV genomes contain large direct repeats that are absent in low abundance segments. These direct repeats are associated with gene duplication which suggests that the high abundance segments in both viral genomes have likely undergone recent recombination events (Cui and Webb, 1997).

The second feature shared by both genomes is a strong A/T bias and gene densities that are the lowest reported for any virus (Tables 1, 2; Fig. 2, 3). We identified 101 open reading frames (ORFs) in the CsIV genome and 61 ORFs in the MdBV genome using gene finder programs and relatively stringent criteria. This yields an average coding density for CsIV of 29% with individual genome segments having coding densities that range from 8% to 40% (Fig. 2). Average coding density for MdBV is 17% with segments ranging from 2% to 43%. The third shared feature is that the majority of predicted genes in each genome are structurally related; with families of related gene variants often residing on the same viral segment. For CsIV, we identified 5 families (rep, cys-motif, vankyrin, vinexin, N-family) containing a total of 53 genes. Another 48 predicted genes are unique and have no detectable homologs in current protein databases (Fig. 2). For MdBV, we also identified 5 families (ptp, vankyrin, tRNA, egf-motif, glc) containing 39 genes plus 21 unique genes predicted to encode for unknown proteins (Fig. 3).

All members of the cys-motif gene family of CsIV and the glc and egf-motif gene families of MdBV contain introns while members of the other gene families do not. Strikingly, nearly all of the intron-containing gene families are encoded on high

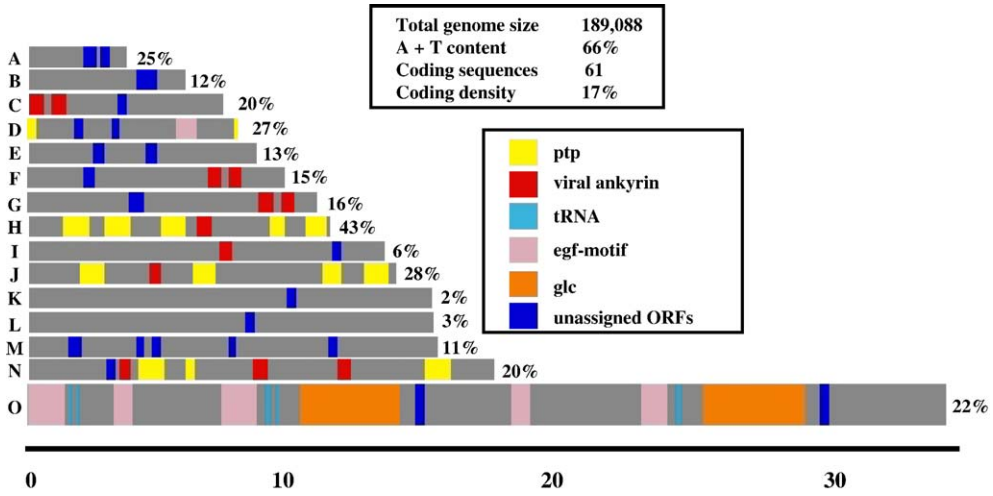


Fig. 3. Diagrammatic representation of sequence and annotation of the *Md*BV genome. The 15 non-redundant circular *Md*BV genome segments are represented as linear molecules to visualize segment size relationships within the genome. Individual segments are listed from smallest (*A* = 3.6 kb) to largest (0 = 34.3 kb) on left with percentages of coding DNA per segment identified on the right. Wider bars represent high-copy segments as defined in Fig. 1. ORFs and members of gene families have been annotated by color and location to individual genome segments with color key to *Cs*IV shown. Grey regions represent non-coding DNA. The aggregate *Md*BV genome is 189,088 bp in size, 83% non-coding, and 17% coding overall. Scale bar is in kilobases.

abundance segments while intron-less family members are on low abundance segments. Overall, sequence analysis during the current study combined with prior analyses of CsIV and MdBV transcripts (Cui and Webb, 1997; Hilgarth and Webb, 2002; Kroemer and Webb, 2005; Turnbull et al., 2005; Strand et al., 1997; Trudeau et al., 2000; Beck and Strand, 2003; Beck and Strand, 2005; Thoetkiattikul et al., 2005) indicate that 10% of CsIV genes and 14% of MdBV genes are spliced. In sharp contrast, Espagne et al. (2004) reported that the CcBV genome contains 156 predicted coding sequences of which 69% contain introns. This large difference in the percentage of spliced genes prompted us to reexamine the CcBV genome using the same gene identification criteria as applied to CsIV and MdBV. Our analysis identified 148 predicted coding sequences in the CcBV genome of which only 6.8% contain introns. From this, we conclude that while some PDV genes are spliced most likely are not.

CsIV and MdBV encode multiple virulence genes but lack essential genes for replication

The primary role of PDVs as symbionts of wasps is to suppress the immune system of the parasitized host so that it cannot kill the parasitoid's offspring (Schmidt et al., 2001; Webb and Strand, 2005). As previously noted, PDVs also affect the growth and development of parasitized hosts which may also facilitate parasitoid survival. The five gene families identified in the CsIV genome all encode predicted virulence factors that are expressed in parasitized host insects (Table 1, Fig. 2). The largest gene family is the rep genes whose members contain a 540 bp sequence present in 1 to 5 copies (Theilmann and Summers, 1987; Hilgarth and Webb, 2002; Webb and Strand, 2005). A total of 30 rep genes are located on 10 low copy segments (Fig. 2, Tables 1, 3). Ten of these genes are expressed in CsIV-infected host caterpillars but the function of Rep proteins is unknown (Hilgarth and Webb, 2002). The second gene family consists of ten spliced genes distinguished by a conserved cysteine-knot motif (Blissard et al., 1989; Cui and Webb, 1997). Four of these genes (WHv1.0, WHv1.6, VHv1.1, VHv1.4) located on high copy segments V and W were previously identified in screens of cDNA libraries made from virus-infected host caterpillars (Cui and Webb, 1997). Genome sequencing during the current study identified six others on high copy segment U and low copy segments A, A2, F, and L (Fig. 2, Table 3) and their expression and splicing have also been confirmed. All cys-motif family members possess signal peptides and share structural features with ω-conotoxins which are produced by molluscs and have diverse pharmacological properties (Dib-Hajj et al., 1993; Turnbull and Webb, 2002; Webb and Strand, 2005). Functional analysis of the protein encoded by VHv1.4 indicates that it is secreted into the hemocoel of infected host caterpillars where it interferes with capsule formation by host immune cells and disrupts proteins synthesis (Li and Webb, 1994; Fath-Goodin et al., in preparation). The third gene family contains seven vankyrin genes on two segments (I2, P). Vankyrin genes encode for ankyrin-repeat containing proteins that resemble the inhibitory domains of inhibitor κβ (Iκβ) proteins (Kroemer and Webb, 2005). Iκβs are critical regulators of the insect and mammalian immune system that bind NF-κβ transcription factors (Ghosh et al., 1998; Hoffmann and Reichhart, 2002). The CsIV vankyrin genes possess a truncated ankyrin repeat domain (ARD) that is required for NF-κβ binding but lack the N- and C-terminal regulatory regions required for dissociation from NF-κβ targets and degradation. All of the vankyrin genes are expressed in parasitized insects with some family members being preferentially expressed in the host fat body and others in hemocytes (Kroemer and Webb, 2005). The fourth gene family encodes four homologs of insect innexins on segments D, G, and Q. Innexins are proteins that form invertebrate gap junctions (Turnbull and Webb, 2002). Two of these viral innexins, or vinnexin, genes are specifically expressed in hemocytes of CsIV-infected host caterpillars and form functional gap junctions in *Xenopus* oocytes (Turnbull et al., 2005). This suggests that CsIV vinnexins may interfere with the cellular immune system of infected hosts by disrupting hemocyte communication required for defense responses like encapsulation. The final CsIV gene family identified during the current study consists of two structurally related genes on segment N that encode predicted proteins that have no detectable homologs in current protein databases. Whether these N genes or the other 48 unique predicted coding sequences present in the CsIV genome are expressed in parasitized host insects, in *C. sonorensis*, or in both is unknown.

The largest gene family in the MdBV genome is the 13-member ptp family on segments D, H, J, and N that encode predicted proteins related to protein tyrosine phosphatases (PTPs) (Tables 2, 3, Fig. 3). PTPs regulate tyrosyl phosphorylation in numerous signaling pathways and have also been usurped by selected prokaryotic pathogens to disrupt phagocytosis by mammalian immune cells (Neel and Tonks, 1997; Cornelis, 2002). All of the MdBV ptp genes are expressed in parasitized insects with some members preferentially expressed in virus-infected hemocytes (A. Pruijssers and M. R. S., unpublished). PDV-encoded PTPs have also been suggested to play a role in disrupting ecdysteroid biosynthesis in parasitized hosts (Provost et al., 2004). The second gene family contains 12 *vankyrin* genes that are located on 7 segments (C, F, G, H, I, J and N) (Fig. 3, Table 1). Similar to CsIV, the MdBV *vankyrins* also lack control domains in their N- and C-termini and possess truncated ARDs that align to ankyrin repeats 3–6 of known Iκβs from insects and mammals (Thoetkiattikul et al., 2005). Two of the MdBV vankyrin proteins, Iκβ–H4 and Iκβ–N3, are functional Iκβs that bind insect NF-κβs and block inducible expression of NF-κβ-regulated immune genes (Thoetkiattikul et al., 2005). Whether the other MdBV *vankyrin* genes encode functional Iκβs or proteins with different activities is unknown. The third MdBV gene family consists of 7 near identical tRNA genes on segment O that include cognates for only serine (Fig. 3). These 7 tRNA genes are expressed in parasitized host caterpillars but their functional significance is unclear given that MdBV genes do not exhibit any serine bias (S. D., M. H. B., B. W. W., M. R. S., unpublished). The fourth and fifth MdBV gene

Table 1
Segments, open reading frames and identified gene families in the *CsIV* genome

Segment	ORF	Strand	Position	Length (bp)	Name	Gene family
A1	A1-1	+	2912–3622	711	*AHv1.0*	Cys-motif
	A1-2	–	4488–4835	348		
	A1-3	–	5286–5585	300		
	A1-4	+	5551–5859	309		
A2	A2-1	–	865–1203	339		
	A2-2	+	1921–2232	312		
	A2-3	+	5786–6283	498	AHv0.8	Cys-motif
B	B-1	+	953–1657	705	*Rep-B1*	Rep-motif
	B-2	+	2983–3372	390		
	C-1	–	1–616			
	C-2		6357–7276	1536		
	C-3	–	2413–2892	480		
	C-4	+	3207–3815	609		
	C-5	–	4472–5410	939		
D	D-1	+	2314–2613	300		
	D-2	+	3170–4258	1089	*Vnx-D1*	Vinnexin
E	E-1	+	432–944	512		
	E-2	+	2398–3075	678	*Rep-E1*	Rep-motif
	E-3	+	4990–5559	572	*Rep-E2*	Rep-motif
F	F-1	+	575–2977	2403	*FHV2.8*	Cys-motif
	F-2	–	3272–3640	369		
	F-3	+	7395–7751	357		
G	G-1	–	4440–4802	363		
	G-2	+	6327–7355	1029	*Vnx-G1*	Vinnexin
G2	G2-1	+	413–1258	846	*Rep-G2-1*	Rep-motif
	G2-2	+	2688–3428	741	*Rep-G2-1*	Rep-motif
	G2-3	+	4613–5314	702	*Rep-G2-1*	Rep-motif
	G2-4	+	6243–6941	699	*Rep-G2-1*	Rep-motif
H	H-1	–	2221–2523	303		
	H-2	+	5500–8313	2814	*Rep-H-1*	Rep-motif
I	I-1	+	2695–3408	714	*Rep-I1*	Rep-motif
	I-2	+	3896–4621	726	*Rep-I2*	Rep-motif
	I-3	–	6245–6565	321		
	I-4	+	7062–8189	1128		
	I2-1	+	8755–8866	630	*Rep-I2-1*	Rep-motif
	I2-2	+	1359–1823	465	*Ank-I2-1*	Vankyrin
	I2-3	+	2955–3278	324		
	I2-4	+	4330–4833	504	*Ank-I2-2*	Vankyrin
	I2-5	+	6748–7260	513	*Ank-I2-1*	Vankyrin
	I2-6	–	7763–8185	423		
J	J-1	+	7–426	420	*Rep-J1*	Rep-motif
	J-2	+	3039–3404	366		
	J-3	+	4223–4672	450	*Rep-J2*	Rep-motif
	J-4	+	4569–4892	324		
	J-5	+	5590–5922	333		
	J-6	–	5782–6483	702	*Rep-J1*	Rep-motif
	J-7	+	7939–8298	360		
L	L-1	–	695–1090	396		
	L-2	+	3285–8021	4737	*LHv2.8*	Cys-motif
M	M-1	+	4317–4952	636		
	M-2	+	5276–5710	435		
	M-3	–	6172–6489	318		
N	N-1	–	1857–2192	336		
	N-2	+	3016–3345	330		
	N-3	+	4395–5828	1434	*NHv1.4*	N gene
	N-4	+	7826–9025	1200	*NHv1.2*	N gene
O	O-1	+	1300–3072	1773	*RepO1*	Rep-motif
	O-2	+	3508–3852	345		
	O-3	+	5127–5426	300		
	O-4	+	5434–5748	315		
	O-5	+	6605–8761	2157	*Rep-O2*	Rep-motif

(*continued on next page*)

Table 1 (*continued*)

Segment	ORF	Strand	Position	Length (bp)	Name	Gene family
P	P-1	–	1–85			
	P-2		11,623–11,841			
	P-3	+	3740–4252	513	*Ank-P1*	Vankyrin
	P-4	+	6787–7266	480	*Ank-P2*	Vankyrin
	P-5	+	9908–10,390	483	*Ank-P3*	Vankyrin
	P-6	+	10,895–11,374	480	*Ank-P4*	Vankyrin
Q	Q-1	–	932–2026	1095	*Vnx-Q2*	Vinnexin
	Q-2	+	3107–4213	1107	*Vnx-Q1*	Vinnexin
	Q-3	+	5592–6299	708	*Rep-Q1*	Rep-motif
	Q-4	+	7456–7797	342		
	Q-5	+	8557–8943	387	*Rep-Q2*	Rep-motif
	Q-6	+	9328–9909	582	*Rep-Q3*	Rep-motif
	Q-7	+	10,740–11,141	402		
T	T-1	–	6750–7223	474		
	T-2	–	11,861–12,211	351		
	T-3	–	13,499–13,879	381		
U	U-1	+	5478–5936	459	*UHv0.8a*	Cys-motif
	U-2	+	7926–8228	303		
	U-3	+	10,323–10,697	375	*UHV0.8b*	Cys-motif
	U-4	+	12,190–12,492	303		
V	V-1	+	1–225			
			15,424–15,510	312		
	V-2	–	1293–1634	342		
	V-3	+	3666–5835	2170	*VHv1.4*	Cys-motif
	V-4	+	11,748–12,176	429		
	V-5	+	13,132–14,746	1615	*VHv1.1*	Cys-motif
W	W-1	+	726–1028	303		
	W-3	+	4660–6101	1442	*WHv1.0*	Cys-motif
	W-4	+	7417–7800	384		
	W-5	–	9786–10,205	420		
	W-6	–	10,075–10,467	393		
	W-7	+	11,922–13,876	1955	*WHv1.6*	Cys-motif
	W-8	–	14,474–15,184	711		
Z	Z-1	+	3465–4259	795	*Rep-Z1*	Rep-motif
	Z-2	+	5671–6357	687	*Rep-Z2*	Rep-motif
	Z-3	+	6416–6829	414		
	Z-4	–	7260–8012	753	*Rep-Z3*	Rep-motif
	Z-5	+	9078–9809	732	*Rep-Z4*	Rep-motif
	Z-6	+	10,293–10,970	678	*Rep-Z5*	Rep-motif
	Z-7	+	2162–12,845	684	*Rep-Z6*	Rep-motif
	Z-8	+	5022–15,738	717	*Rep-Z7*	Rep-motif
	Z19	+	7232–17,966	735	*Rep-Z2*	Rep-motif

families consist of spliced genes that with one exception are located on segment O. The *egf-motif* gene family consists of a total of 6 members that share an identical cysteine-rich, epidermal growth factor (EGF)-like motif at their deduced N-termini. Three of these genes (*egf1.5, 1.0, 0.4*) were previously identified in screens of cDNA libraries made from MdBV-infected host larvae (Strand et al., 1997; Trudeau et al., 2000). During the current study, we found that two identical copies of *egf0.4* and *egf1.5* exist on segment O, and that a third near identical copy of *egf0.4* exists on segment D. Proteins encoded by the family members *egf1.5, 1.0*, and *0.4* are all secreted into the hemolymph of parasitized host larvae. The EGF-like domain of these proteins is also 43% similar to the cysteine-rich domain of a unique serine protease inhibitor, AcAPc2, from the hookworm *Ancylostoma caninum* (Trudeau et al., 2000). The *glc* gene family consists of two identical genes (*glc1.8*) on segment O that encode for a mucin. In parasitized hosts, the Glc1.8 protein localizes to the cell surface of MdBV-infected hemocytes and blocks both encapsulation and phagocytosis (Beck and Strand, 2003, 2005). None of the 21 putatively unique genes in the MdBV genome share any homology with any known genes including any factors from CsIV, other PDVs, or other viruses. Whether any of these unique genes are expressed in parasitized hosts or *M. demolitor* is unknown.

In contrast to the large number of genes encoding virulence factors, database searches failed to identify any genes in the CsIV or MdBV genomes with homology to genes required for DNA replication (polymerases or helicases) or that encode structural proteins. The exception to this is that CsIV segment C encodes the structural protein p12. By contrast, another CsIV structural protein, p44, is encoded in the *C. sonorensis* genome (Deng et al., 2000). Thus, while it is possible that some of the orphan genes in the CsIV and MdBV genomes could encode structural proteins, our results overall suggest that the encapsidated genomes of both viruses lack genes required for replication and virus assembly. Given

Table 2
Segments, open reading frames and identified gene families in the *Md*BV genome

Segment	ORF	Strand	Position	Length (bp)	Name	Gene family
A	A-1	–	2037–2555	519		
	A-2	–	2701–3078	378		
B	B-1	–	1497–2243	747		
C	C-1	+	31–555	525	*IκB-C1*	Vankyrin
	C-2	+	867–1412	546	*IκB-C2*	Vankyrin
	C-3	+	3328–3717	390		
D	D-1	–	7778–7823			
			1–377	423	*ptp-D1*	PTP
	D-2	+	1781–2146	366		
	D-3	–	3148–3471	324		
	D-4	+	5585–5639			
			5755–5885			
			6177–6302	312	*egf0.4*	Egf-motif
E	E-1	–	2379–2732	354		
	E-2	+	2494–2820	327		
	E-3	–	4370–4816	447		
F	F-1	+	2127–2534	408		
	F-2	+	6771–7265	495	*IκB-F1*	Vankyrin
	F-3	+	7544–8047	504	*IκB-F2*	Vankyrin
G	G-1	+	3805–4158	354		
	G-2	–	3975–4349	375		
	G-3	+	8692–9222	531	*IκB-G1*	Vankyrin
	G-4	+	9531–10,037	507	*IκB-G2*	Vankyrin
H	H-1	+	1250–2260	1011	*ptp-H1*	PTP
	H-2	–	2846–3823	978	*ptp-H2*	PTP
	H-3	+	4941–5903	963	*ptp-H3*	PTP
	H-4	+	6313–6891	579	*IκB-H1*	Vankyrin
	H-5	+	9043–9591	549	*ptp-H4*	PTP
	H-6	+	10,409–11,194	786	*ptp-H5*	PTP
I	I-1	+	7101–7604	504	*IκB-I1*	Vankyrin
	I-2	+	11,371–11,700	330		
J	J-1	+	1931–2833	903	*ptp-J1*	PTP
	J-2	–	4462–4930	469	*IκB-J1*	Vankyrin
	J-3	–	6126–7019	894	*ptp-J2*	PTP
	J-4	–	10,999–11,754	756	*ptp-J3*	PTP
	J-5	–	12,570–13,469	900	*ptp-J4*	PTP
K	K-1	+	9717–10,025	309		
L	L-1	+	8107–8484	378		
M	M-1	+	1506–1841	336		
	M-2	–	4003–4320	318		
	M-3	–	4568–4900	333		
	M-4	+	7448–7783	336		
	M-5	+	11,214–11,525	312		
N	N-1	–	2914–3222	309		
	N-2	–	3382–3801	420	*IκB-N1*	Vankyrin
	N-3	–	4114–5070	957	*ptp-N1*	PTP
	N-4	–	5825–6223	399	*ptp-N2*	PTP
	N-5	–	8394–8945	552	*IκB-N2*	Vankyrin
	N-6	–	11,571–12,062	492	*IκB-N3*	Vankyrin
	N-7	–	14,837–15,796	960	*ptp-N3*	PTP
O	O-1	–	71–865			
			1062–1225			
			1325–1397	1032	*egf1.5*	Egf-motif
	O-2	+	1653–1724	72		tRNA-Ser
	O-3	+	1875–1946	72		tRNA-Ser
	O-4	+	3238–3292			
			3408–3538			
			3831–3956	312	*egf0.4*	Egf-motif
	O-5	–	7290–8093			
			8290–8453			
			8553–8625	1041	*egf1.5*	Egf-motif
	O-6	+	8881–8952	72		tRNA-Ser
	O-7	+	9104–9175	72		tRNA-Ser

(*continued on next page*)

Table 2 (*continued*)

Segment	ORF	Strand	Position	Length (bp)	Name	Gene family
O						
	O-8	+	9326–9397	72		tRNA-Ser
	O-9	–	10,191–10,691			
			11,099–11,332			
			11,740–11,973			
			12,381–12,614			
			13,022–13,255			
			13,663–13,724			
			13,911–13,959	1548	*glc1.8*	Glc
	O-10	+	14,533–14,859	327		
	O-11	+	18,136–18,190			
			18,306–18,436			
			18,727–18,852	312	*egf0.4*	Egf-motif
	O-12	–	23,001–23,465			
			23,662–23,825			
			23,925–23,997	702	*egf1.0*	Egf-motif
	O-13	+	24,254–24,325	72		tRNA-Ser
	O-14	+	24,477–24,548	72		tRNA-Ser
	O-15	–	25,344–25,844			
			26,252–26,485			
			26,893–27,126			
			27,534–27,767			
			28,175–28,408			
			28,816–28,877			
			29,064–29,112	1548	*glc1.8*	Glc
	O-16	+	29,685–30,011	327		

that the CcBV genome also lacks replication and structural genes, (Espagne et al., 2004), these results well explain why IVs and BVs do not replicate in parasitized hosts (Kroemer and Webb, 2004; Webb and Strand, 2005). These results in combination with the types of genes that are encoded in the encapsidated genomes of PDVs also suggest that significant lateral gene transfer events have occurred between PDVs and wasps. Virulence genes in the CsIV and MdBV genomes of likely insect origin include the *vankyrin*s (CsIV and MdBV), *vinnexins* (CsIV), cys-motif (CsIV) and *ptp* genes (*Md*BV). Some of the virulence genes in the MdBV or CsIV genomes are also flanked by inverted repeats suggesting horizontal acquisition. In contrast, the absence of replication and assembly genes in the encapsidated genomes of CsIV and MdBV strongly suggests that these factors reside in the genomes of *C. sonorensis* and *M. demolitor.*

After infecting hosts, most viruses increase gene copy number by replicating. Since PDVs cannot replicate in parasitized host caterpillars, we suggest the organizational features CsIV and MdBV share are compensatory adaptations for increasing copy number and/or regulating gene expression (Table 3). The first of these is the production of high and low abundance genomic segments during replication in the wasp. How segment abundance is regulated during replication is unknown but its net effect increases copy number of CsIV genes on segments U, V and W (*Cys-motif* genes) and MdBV genes on segment O (*egf-motif* and *glc* gene family members) relative to genes on low abundance segments. CsIV high abundance segments also undergo intramolecular recombination during replication to produce smaller segments that encode most of the same genes present on the parental high abundance segments. This process is called segment nesting and it further increases

Table 3
Distribution of CsIV and MdBV genes among high copy and low copy segments, protein targeting and representation in other PDV genomes

CsIV genes	N	High copy	Low copy	Localization
rep	30	–	B, I2,.E, H, 0; I, J, Q, G2, Z	Intracellular (Hilgarth and Webb, 2002)
cys-motif	10	V U W	A, A2, F, L	Secreted and membrane (Cui and Webb, 1997; Li and Webb, 1994)
vankyrin	7	–	I2, P	Nuclear and cytoplasmic (Kroemer and Webb, 2005)
vinnexin	4	–	D, G, Q	Intracellular and membrane bound (Turnbull et al., 2005)
unassigned	48			
MdBV genes	N	High copy	Low copy	Localization representation
ptp	13	–	D, H, J, I, N	Intracellular (Pruijssers, A., and M. R. S., unpublished)
egf-motif	6	O	–	Secreted (Trudeau et al., 2000)
glc	2	O	–	Membrane bound (Beck and Strand, 2003; 2005)
vankyrin	12	–	C, F, G, H, J	Intracellular (Thoetkiattikul et al., 2005)
tRNA(ser)	7	O	–	Intracellular (S. D., M. H. B., M. R. S., unpublished)
unassigned	21			

Table 4
Comparison of CsIV and MdBV gene families to selected other IVs and BVs[a]

Family	Rep	Cys-motif	Vinnexin	Vankyrin	PTP	Egf-motif	Glc
CsIV	30	10	4	7	0	0	0
HfIV	36	5	8	7	0	0	0
HdIV	3	0	1	1	0	0	0
TrIV	10	5	2	1	0	0	0
MdBV	0	0	0	12	13	6	2
CcBV	0	0	0	5	23	0	0
TnBV	0	0	0	4	8	0	0

[a] Several unique genes are known from each IV and BV besides those listed in the table. See text for examples.

copy number of the encapsidated *cys-motif* genes to levels 67% higher than would exist in the absence of nesting (Cui et al., 2000). Segment nesting does not occur in the MdBV genome but every gene on segment O exists in multiple copies due to intrasegmental duplication events. Second, despite sharing no homology, all CsIV and MdBV genes located on high abundance segments are all spliced and encode secreted or cell surface proteins that are the most abundant viral gene products in parasitized hosts (Cui et al., 2000; Strand et al., 1997; Trudeau et al., 2000; Webb and Strand, 2005). Gene families on low abundance segments (*vankyrins, vinnexins,* and *rep* genes (CsIV), *vankyrins* and *ptp* genes (MdBV)) in contrast lack both signal sequences and introns (Table 3) and appear to encode protein products that localize to the cytoplasm or nucleus of infected host cells (Kroemer and Webb, 2005; Thoetkiattikul et al., 2005; Webb and Strand, 2005; Le Hir et al., 2003). These features strongly suggest that virulence genes on high abundance segments are targeted for secretion and high-level expression in parasitized hosts while genes on low abundance segments interact with host intracellular signaling pathways.

IVs and BVs encode few related genes with one another

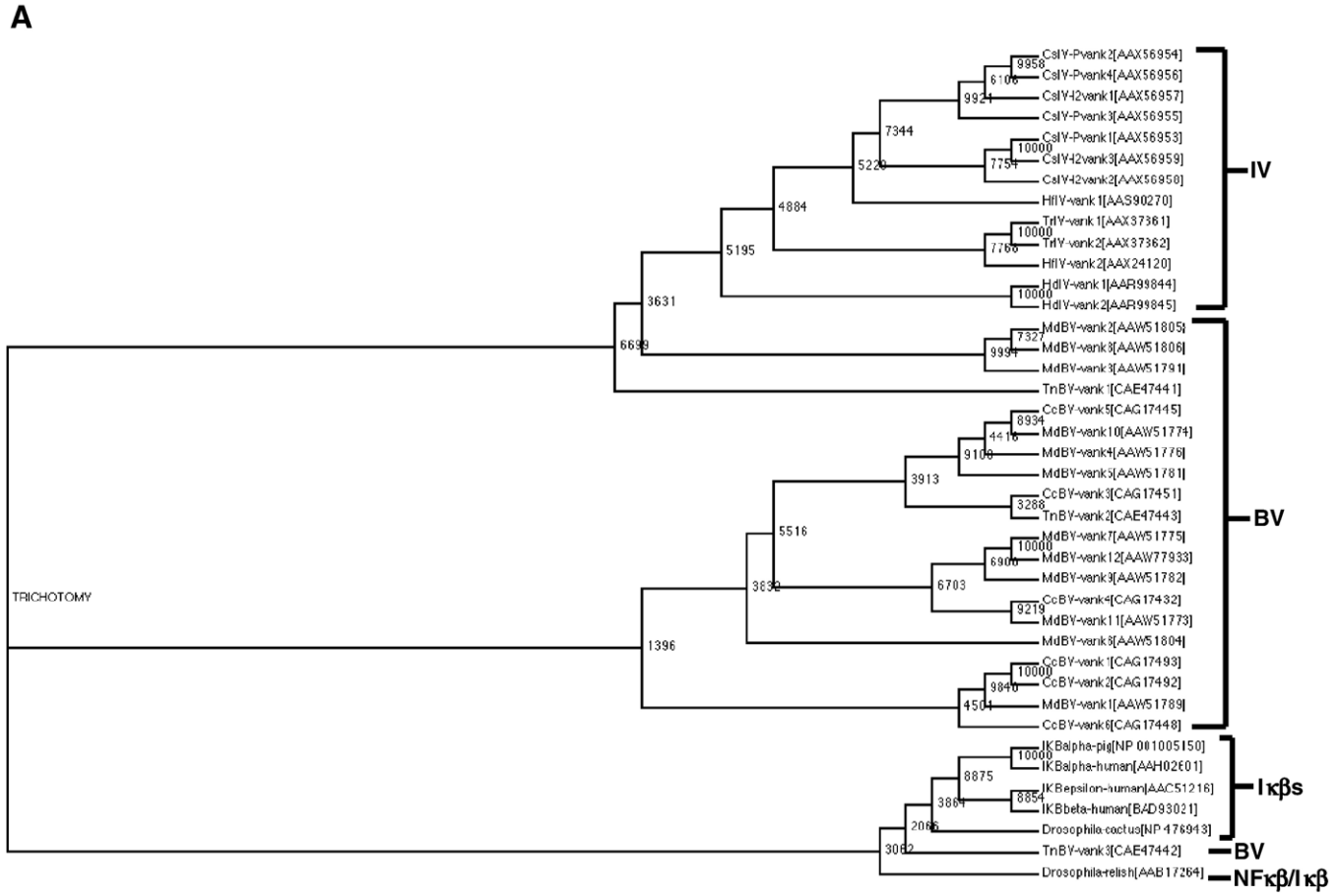
Comparison of the CsIV and MdBV genomes suggests that they share only one gene family (*vankyrins*) in common (Figs. 2, 3). As previously noted, the CcBV genome is fully sequenced while partial genomic sequence data is available for IVs from the campoplegenes *Hyposoter didymater* (HdIV) (N. Volkhoff and B. A. W., unpublished) *H. fugitivus* (HfIV) (D. Stoltz and B. A. W., unpublished), and *Tranosema rostrale* (TrIV) (M. Cusson and B. A. W., unpublished) and a BV from the microgastroid *Toxoneuron nigriceps* (TnBV) (Provost et al., 2004; GenBank direct submissions). Comparison of these IVs to CsIV reveals that each encodes gene families that are orthologous to the *rep, cys-motif, vankyrin*, and *vinnexin* genes (Table 4). In contrast, CsIV, HdIV, HfIV and TrIV share no genes with MdBV except for *vankyrins*. Comparison of CcBV and TnBV to MdBV indicates that all encode multi-member *ptp* and *vankyrin* gene families but, excluding *vankyrins,* do not appear to share any genes with IVs (Table 4, also see below). CcBV and TnBV lack genes related to the MdBV *glc* family and reciprocally encode genes absent from MdBV (Table 4). For example, CcBV encodes *ep1*, *ccv1*, and *cysteine rich* genes for which homologs are known from other *Cotesia* BVs but not BVs from other microgastroids (Whitfield and Asgari, 2003; Le et al., 2003). TnBV, like other PDVs we examined, also encodes two unique genes *TnBV1* and -2 (Falabella et al., 2003; Lapointe et al., 2005) that are absent from MdBV, CcBV and the other PDVs included in this comparison.

These results clearly indicate that CsIV shares more genes with other IVs than any BV, while MdBV usually shares genes with other bracoviruses. These results are fully consistent with the phylogeny of associated wasps which predicts a priori a closer relationship among IVs and BVs, respectively, under conditions of mendelian inheritance as proviruses. Functionally, the very different suite of genes IVs and BVs encode suggests each lineage has evolved broadly different strategies to immunosuppress and/or alter the development of parasitized hosts. The large number of unique genes present in CsIV, MdBV, and other PDVs, also indicates that each PDV-parasitoid species has also undergone many novel gene acquisition and/or loss events that likely further refine the types of physiological interactions that occur with hosts.

Our results provide little support for campoplegine IVs and microgastroid BVs being related given that their encapsidated genomes share only one gene family. Alignment of all PDV vankyrin proteins using mammalian and *Drosophila* IκβS as outgroups produced incompletely resolved trees that provide no clear evidence of shared ancestry (Fig. 4A). Given that the IV and BV vankyrins cluster separately, however, indicates that the IV *vankyrins* are more related to one another than they are to corresponding genes in BVs. Espagne et al. (2004) suggested that the *cysteine-rich* gene family identified in the CcBV genome is related to the *cys-motif* genes of CsIV and a parasitoid produced virulence protein, TSP 14, from the braconid *Microplitis croceipes*. As previously noted, the *egf-motif* genes of MdBV likewise contain a cysteine rich domain. Alignment of these genes, however, also produced incompletely resolved trees (Fig. 4B). Distinctly separate clustering of the proteins encoded by *cys-motif* genes from CsIV and the *cysteine rich* genes of CcBV suggest no significant relationship. The CcBV genes also appear to be unrelated to TSP 14 but do cluster with cysteine rich homologs from other *Cotesia* BVs. In contrast, the MdBV EGF proteins 1.5 and 1.0 do cluster with high bootstrap values to TSP-14 (Fig. 4B). This intriguingly points to a potentially more recent lateral gene transfer event between wasp and virus since *M. demolitor* and *M. croceipes* are congenerics.

CsIV and MdBV repetitive DNA

One of the most striking features of the CsIV and MdBV genomes is the large amount of repetitive DNA. In total, 86 and 74 repeats of greater than 50 bp in length were identified in the CsIV and MdBV genomes respectively using the REPuter program. This comprises approximately 15% of the genomic DNA in each virus. In CsIV, the CENSOR server detected a total of 23 sequences, 40 bp in length or longer, that showed at least 65% similarity to transposable elements (TEs). Twenty-one of these elements were in non-coding domains that were distributed on both the low copy segments A, D, F, G2, G, I2, L, M, N, P, G, T, and Z and on high copy segment W. Most of these elements


A
CsIV-Pvank2[AAX56954]
CsIV-Pvank4[AAX56956]
CsIV-f2vank1[AAX56957]
CsIV-Pvank3[AAX56955]
CsIV-Pvank1[AAX56953]
CsIV-f2vank3[AAX56959]
CsIV-f2vank2[AAX56950]
HfIV-vank1[AAS90270]
TrIV-vank1[AAX37361]
TrIV-vank2[AAX37362]
HfIV-vank2[AAX24120]
HdIV-vank1[AAR99844]
HdIV-vank2[AAR99845]
MdBV-vank2[AAW51805]
MdBV-vank3[AAW51806]
MdBV-vank3[AAW51791]
TnBV-vank1[CAE47441]
CcBV-vank5[CAG17445]
MdBV-vank10[AAW51774]
MdBV-vank4[AAW51776]
MdBV-vank5[AAW51781]
CcBV-vank3[CAG17451]
TnBV-vank2[CAE47443]
MdBV-vank7[AAW51775]
MdBV-vank12[AAW77933]
MdBV-vank9[AAW51782]
CcBV-vank4[CAG17452]
MdBV-vank11[AAW51773]
MdBV-vank8[AAW51804]
CcBV-vank1[CAG17493]
CcBV-vank2[CAG17492]
MdBV-vank1[AAW51789]
CcBV-vank6[CAG17448]
IKBalpha-pig[NP 001005150]
IKBalpha-human[AAH02601]
IKBepsilon-human[AAC51216]
IKBbeta-human[BAD93021]
Drosophila-cactus[NP 476943]
TnBV-vank3[CAE47442]
Drosophila-relish[AAB17264]
IV
BV
IκβS
BV
NFκβ/Iκβ
TRICHOTOMY
B
MdBV-Egf1.0[AAY24530]
MdBV-Egf1.5[AAY24525]
Microplitis.sp-TSP14[AAN17943]
CsIV-Fhv1.4[AAO43444]
CsIV-Lhv2.8[AAO43445]
CsIV-Vhv1.4[AAB65441]
CcIV-P1.0[BAC55881]
CsIV-Vhv1.1[AAA79240]
CsIV-Uhv0.8[AY197494]
CsIV-Uhv0.8a[AAO43446]
CsIV-Ahv0.8[AAO43447]
CsIV-Ahv1.0[AAO43442]
CsIV-Whv1.0[AAC58527]
CsIV-Whv1.6[AAC58528]
CcBV-CRP1[YP 184840]
CgBV-P3[AAR37024]
CcBV-CRP2[YP 184841]
CcBV-CRP3[CAG18434]
BV
TSP 14
IV
BV
TRICHOTOMY

are present as single copies but two 131-bp sequences that are 71% similar to an internal domain ACCORD1 retrotransposon from *Drosophila* flank members of the *rep* gene family on segments G2 and Z. Other TEs appear to be randomly distributed. We identified 16 TE sequences in non-coding domains of the MdBV genome on low copy segments D, E, F, G, N, M, J, K and high copy segment O but none directly flank any genes. Almost no TE sequences are shared between the CsIV and MdBV genomes suggesting invasion occurred independently. Nonetheless, the large number of repeats combined with the presence of some TE sequences and active processes like segment nesting (see above) suggest that PDV genomes have likely undergone extensive shuffling and that viral genomes from even very closely related wasps may show little colinearity.

While most intracellular bacteria and viruses have little or no non-coding repetitive DNA, recent studies identify exceptions to this trend such as *Wolbachia* that infect a variety of arthropods and nematodes (Wu et al., 2004; Foster et al., 2005). The high percentage of repeats and non-coding sequence in *Wolbachia* have been suggested to either reflect very weak selection for its elimination or increased exposure to mobile elements as a consequence of a parasitic lifestyle (Wu et al., 2004; Moran, 2003). Alternatively, repetitive DNA may actually provide a selective benefit by contributing to genome plasticity and long-term phenotypic evolution (Meagher and Costich, 2004; Foster et al., 2005). It is also possible that some non-coding sequence in PDV genomes contains non-translated elements like microRNAs that have functional but currently unknown roles in parasitism.

Conclusions

Genome sequence analyses have been most used and valued as catalogs for the entire gene set of an organism. Our results document the number and type of genes encoded by an IV and BV, while also providing new insights on the organization and possible origins of these unusual viruses. Our results reveal viral genomes that are comparatively large but that have low coding densities dominated by a small number of genes that have diverged into multiple variants which are expressed in parasitized host insects. Furthermore, highly expressed gene families that encode secreted or cell surface proteins are all spliced and encoded on high copy genomic segments while low abundance intracellular gene products lack introns and are associated with low copy segments. This suggests that in the absence of replication both viruses have evolved similar compensatory strategies for regulating the abundance and stability of different viral transcripts. The diversification of virulence genes into families is also likely a key adaptation that reflects the role of these viruses in parasitism. Variants within gene families of insects and mammals often exhibit cell, tissue or developmental stage specific functions. Similarly, the *vankyrin* genes from CsIV and MdBV, and *ptp* genes from CcBV and TnBV exhibit tissue specific patterns of expression (Provost et al., 2004; Kroemer and Webb, 2005; Thoetkiattikul et al., 2005) while genes from other PDVs are differentially expressed as a function of host stage (Johner and Lanzrein, 2002). The evolution of gene families in PDVs also likely involves functional divergence as demonstrated by the *vankyrin H4* and *N3* genes of MdBV which differentially bind different types of host NF-κBs (Thoetkiattikul et al., 2005).

As previously noted, prior evidence supporting the independent origin of IVs and BVs include: (1) viruses in each genus exhibit different replication patterns and virion morphologies and (2) no PDV-carrying wasp ancestor to campoplegines and microgastroids is known to exist (Whitfield, 2002; Webb and Strand, 2005). Our results overall also support that these viruses evolved independently. Comparison of the CsIV and MdBV genomes to other PDVs indicates that IVs share several genes with one another as do BVs. In contrast, few genes are shared between IVs and BVs and those that are (*vankyrins*, *cysteine-rich*) align poorly suggesting IVs and BVs have independently recruited a small number of related wasp genes in response to similar functional requirements in parasitism. By extension, we conclude that segmentation, non-equimolar segments, and other genomic properties shared by IVs and BVs also reflect convergence. Lastly, if IVs and BVs are unrelated, they should not be classified into a single family (Polydnaviridae) as exists currently (Webb et al., 2000).

Much less clear is the origin of IVs or BVs themselves. Viruses with large, dsDNA genomes are generally thought to evolve by capturing multiple genes from their cellular hosts (Villareal, 2005). Based on a limited number of shared replication and structural genes, large DNA viruses like poxviruses, ascoviruses, iridoviruses, asfarviruses and phycodnaviruses may share a common, ancient viral ancestor (Iyer et al., 2001; Delaroque et al., 2003; Federici and Bigot, 2003; Raoult et al., 2004), whereas the origins of other large DNA viruses like baculoviruses is unknown (Herniou et al., 2003). Two models have been suggested for the origin of PDVs. The first (PDVs from host) proposes that PDVs evolved from a wasp ancestor(s) that developed the ability to produce and package circular DNA intermediates containing wasp genes encoding proteins that already had functional roles in parasitism (Stoltz and Whitfield, 1992; Whitfield and Asgari,

Fig. 4. Phylogenetic alignments of PDV *vankyrin* (*vank*) and *cys-motif* proteins suggest independent acquisition and evolutionary convergence of gene families between IV and BV genera. (A) Phylogenetic tree (bootstrapped from 10,000 independent runs) constructed from alignment of *Cs*IV and *Md*BV vankyrin proteins, selected vankyrins from other PDV genomes (*Cotesia congregata* BV (CcBV), *Toxoneuron nigriceps* BV (TnBV), *Tranosema rostrale* IV (TrIV), *Hyposoter fugitivus* IV (HfIV), *Campoletis chloridae* ichnovirus (CcIV) and *Hyposoter didymator* IV (HdIV), selected mammalian Iκβ proteins, and the Iκβ proteins Relish and Cactus from *Drosophila*. Each vankyrin and Iκβ along its accession number is listed to the right of the dendrograms. All IV vankyrins, the majority of BV vankyrins, and eukaryotic Iκβ cluster into independent groups while only a few BV proteins show significant identities with the IV cluster. (B) Phylogenetic tree (bootstrapped from 10,000 independent runs) constructed from alignment of CsIV, HfIV, MdBV and CcBV cys-motif protein family variants plus TSP14 from *Microplitis croceipes*. Cys motif family proteins exhibit similar relationships to those observed within the vankyrin family with the majority of IV and BV variants clustering independently in separate clades.

2003; Espagne et al., 2004). Cooption of proteins, of wasp or possibly viral origin, for capsid formation allowed for packaging of these DNA intermediates and infection of parasitized host cells. Gene families also could have evolved either prior to or after the evolution of packaging. Evidence supporting this model include the near absence of replication and structural genes in encapsidated IV and BV genomes, the large amounts of non-coding DNA, and presence of introns in several PDV genes which is common for insect genes but rare for viruses. In addition, almost no PDV genes (see below) have been identified that share any similarity with any genes from other viruses.

The second model (PDVs from viruses) proposes that functional viruses associated with the ancestor of campoplegine and microgastroid wasps respectively evolved a beneficial association that resulted in integration of the viral genome into the wasp genome, followed by the loss of viral replication and structural genes from the encapsidated genome and acquisition of virulence genes from the wasp (Federici and Bigot, 2003; Webb and Strand, 2005). This model is supported by the existence of very distinctive particle morphologies, differences in DNA packaging and replication between IVs and BVs (Federici and Bigot, 2003). Federici and Bigot (2003) note that IV virion morphology resembles that of ascoviruses and that the major capsid protein of the ascovirus SfAV-1a shares features with the CsIV p44 structural protein. BVs in contrast have been noted to resemble nudivirus and baculovirus virions (Federici and Bigot, 2003) but no BV structural genes with homology to any baculovirus have been identified. Given the absence of replication and structural genes from the encapsidated genomes of IVs and BVs, distinguishing between these models and firmly establishing whether IVs and BVs are unrelated will require the identification and analysis of these genes from the genome of associated wasps. Studies with BVs suggest all genomic segments may be integrated in tandem at a single location in the wasp genome (Belle et al., 2002; Wyder et al., 2002). If so, genes required for replication and packaging may be in close proximity to encapsidated viral segments or part of a replication intermediate (Drezen et al., 2003).

In summary, the genome sequences of CsIV and MdBV present evidence of genome expansion by lateral transfer of virulence genes involved in parasitism of insect hosts and genome reduction in relation to genes regulating viral replication in wasps. Given that CsIV and MdBV have no detectable sequence homology and exist in phylogenetically distinct lineages of wasps, their shared features most likely reflect convergence driven by their similar roles in parasitism. The long-term relationship between PDVs and parasitoids raises the possibility that neither IVs nor BVs derive from any extant virus families but instead reflect an ancient association of wasps with a virus that no longer exists. This obligate association may have enabled bilateral transfer of virus replication and virion assembly machinery to the wasp genome and concurrent acquisition and packaging of wasp virulence genes in PDV genomes. The result is a versatile, replication-defective expression system for delivery of virulence genes to parasitized host insects.

Materials and methods

Genomic sequencing of CsIV and MdBV

C. sonorensis and *M. demolitor* were maintained separately as previously described (Cui and Webb, 1997; Strand et al., 1997). Viral DNAs were isolated from each wasp species by established methods for genome sequencing (Cui and Webb, 1997). We found conventional approaches to genome sequencing that entail production of shotgun plasmid libraries poorly suited to the nonequimolar, segmented, genomes of PDVs that also undergo intramolecular recombination to produce partial duplicates of nested segments. Therefore, for sequence analyses, we used some *Md*BV and *Cs*IV genomic libraries constructed by standard procedures into *E. coli* plasmid vectors (Cui and Webb, 1997; Beck and Strand, 2003). However, our primary approach exploited libraries that were constructed utilizing TN5-mediated rescue cloning (Dennis and Zylstra, 1998) to introduce transposons containing an origin of replication that relies on *pir* gene function and Kan^R as a selectable marker into circular PDV segments (Dennis and Zylstra, 1998). Clones were screened by colony and Southern blot hybridization with DNA from individual segments and sequenced using vector and transposon primer sites with some primers also designed to specific PDV sequences. Larger clones were sequenced using in vitro transposition to insert a second, TN7-based transposon (Genome Priming System, New England BioLabs).

Analysis of sequence data

Automated DNA sequencing utilized both random sequence approaches and analyses of clones selected by segment-specific hybridization to produce fivefold coverage for each genome segment for the assembly. Reactions were analyzed on Applied Biosystems 310 and Beckman Coulter CEQ DNA sequencers. Sequence-specific primers were designed to verify regions of ambiguity and demonstrate circularity of individual segments. Sequence electropherograms were imported into Lasergene DNA analyses programs (DNASTAR, Madison, WI) for editing and assembly into contiguous regions. Sequence variants were detected and if variation was below 0.5% of an assembled sequence were treated as polymorphisms of the same segment. Regions considered to be potential frame shifts or sequencing errors after the first round of annotation were resequenced from direct genomic PCR products.

For the identification of protein coding genes, genome sequences were conceptually translated in six frames to generate potential protein products of ORFs longer than 100 codons. These potential protein sequences were compared to the database of proteins in the GenBank/EMBL, SWISSPROT and PIR databases using FASTA, BLASTX, and BESTFIT (www.ncbi.nlm.nih.gov). tRNA genes were identified using the tRNA-SCAN program (Lowe and Eddy, 1997). Segment by

segment comparisons of each viral genome to other PDVs and large DNA viruses at reduced stringency to assess these relationships. After manual verification of ORF assignments, validated members were referred to as protein-coding genes. Signal peptides in proteins were predicted using the SignalP program (www.cbs.dut.dk/services/SignalP), transmembrane domains were identified using TMHMM (http://www.lcbs.dtu.dk/services: TMHMM), and regulatory motifs were identified using the PFAM database (www.sanger.ac.uk/Software/Pfam). In addition, searches for repeats in each genome were performed using the REPuter program (http://bibiserv.techfak.uni-bielefeld.de/reputer/) and for invertebrate transposon and retrotransposon associated elements using CENSOR (Jurka et al., 1996). Sequences were deposited in Genbank under accession numbers: U41655, AF004378, AF361487, AF361488, AF361869, AF362507–AF362517, AF411011, AF411012, AY029394, AY029400, AF361869 for CsIV and AY887894, AY875680–AY875690, AY848690, AY842013, DQ000240 for MdBV.

Alignments of PDV vankyrin and cys-motif proteins were performed using Clustal X software (40; http://www-igbmc.u-strasbg.fr/BioInfo/; Thompson et al., 1997). Alignments were bootstrapped from 10,000 independent trials using a random number generator of 200 for vankyrin proteins and 700 for cys-motif proteins. Phylogenetic trees were constructed from alignments using TreeView software (R.D. Page, 2001; http://taxonomy.zoology.gla.ac.uk/rod/treeview.html) with nodal placement of bootstrap values and a ladderise left orientation of branches.

Acknowledgments

We thank Torrence Gill, Yongyuung Kim, Lingkang Huang, Michelle Graham, D. Trudeau, and D. Donnell for input during the study. This work was supported in part by the United States Department of Agriculture, National Science Foundation, and Universities of Kentucky and Georgia Experiment Stations.

References

Albrecht, U., Wyler, T., Pfister-Wilhelm, R., Gruber, A., Stettler, P., Heiniger, P., Schumperli, D., Lanzrein, B., 1994. PDV of the parasitic wasp *Chelonus inanitus* (Braconidae): characterization, genome organization and time point of replication. J. Gen. Virol. 75, 3353–3363.

Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M., Aksoy, S., 2002. Genome sequence of the endocellular symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat. Genet. 32, 402–407.

Asgari, S., Hellers, M., Schmidt, O., 1996. Host haemocyte inactivation by an insect parasitoid: transient expression of a polydnavirus gene. J. Gen. Virol. 77, 2653–2662.

Beck, M., Strand, M.R., 2003. RNA interference silences *Microplitis demolitor* bracovirus genes and implicates glc1.8 in disruption of adhesion in infected host cells. Virology 314, 521–535.

Beck, M., Strand, M.R., 2005. Glc1.8 from *Microplitis demolitor* bracovirus induces a loss of adhesion and phagocytosis by Insect High Five and S2 cells. J. Virol. 79, 1861–1870.

Beckage, N.E., Gelman, D.B., 2004. Wasp parasitoid disruption of host development: implications for new biologically based strategies of pest control. Annu. Rev. Entomol. 49, 299–330.

Belle, E., Beckage, N.E., Rousselet, J., Poirie, M., Lemeunier, F., Drezen, J.-M., 2002. Visualization of polydnavirus sequences in a parasitoid wasp chromosome. J. Virol. 76, 5793–5796.

Blasco, R., 1995. Evolution of poxviruses and African swine fever virus. In: Gibbs, A.J., Calisher, C.H., Garcia-Arehanl, F. (Eds.), Molecular Basis of Viral Evolution. Cambridge Univ. Press, Cambridge, UK, pp. 255–268.

Blissard, G.W., Theilmann, D.A., Summers, M.D., 1989. Segment W of *Campoletis sonorensis* virus: expression, gene products, and organization. Virology 169, 78–89.

Brownlie, J.C., O'Neill, S.L., 2005. *Wolbachia* genomes: insights into an intracellular lifestyle. Curr. Biol. 15, R507–R509.

Cornelis, G.R., 2002. *Yersinia* type III secretion: send in the effectors. J. Cell Biol. 158, 401–408.

Cui, L., Webb, B.A., 1997. Homologous sequences in the *Campoletis sonorensis* polydnavirus genome are implicated in replication and nesting of the W segment family. J. Virol. 71, 8504–8513.

Cui, L., Soldevila, A.I., Webb, B.A., 2000. Relationships between PDV gene expression and host range of the parasitoid wasp *Campoletis sonorensis*. J. Insect Physiol. 46, 1397–1407.

Delaroque, N., Boland, W., Muller, D.H., Knippers, R., 2003. Comparisons of two large phaeoviral genomes and evolutionary implications. J. Mol. Evol. 57, 613–622.

Deng, L., Stoltz, D.B., Webb, B.A., 2000. A gene encoding a polydnavirus structural polypeptide is not encapsidated. Virology 269, 440–450.

Dennis, J.J., Zylstra, G.J., 1998. Plasposons: modular self-cloning minitransposon derivatives for rapid genetic analysis of gram-negative bacterial genomes. Appl. Env. Microbiol. 64, 2710–2715.

Dib-Hajj, S.D., Webb, B.A., Summers, M.D., 1993. Structure and evolutionary implications of a "cysteine-rich" *Campoletis sonorensis* polydnavirus gene family. Proc. Natl. Acad. Sci. U.S.A. 90, 3765–3769.

Drezen, J.-M., Provost, B., Espagne, E., Cattolico, L., Dupuy, C., Poirie, M., Periquet, G., Huguet, E., 2003. Polydnavirus genome: integrated vs. free virus. J. Insect Physiol. 49, 407–417.

Espagne, E., Dupuy, C., Huguet, E., Cattolico, L., Provost, B., Martins, N., Poirie, M., Periquet, G., Drezen, J.M., 2004. Genome sequence of a polydnavirus: insights into symbiotic virus evolution. Science 306, 286–289.

Falabella, P., Varricchio, P., Gigliotti, S., Tranfaglia, A., Pennacchio, F., Malva, C., 2003. *Toxoneuron nigriceps* polydnavirus encodes a putative aspartyl protease highly expressed in parasitized host larvae. Insect Mol. Biol. 12, 9–17.

Federici, B.A., Bigot, Y., 2003. Origin and evolution of polydnaviruses by symbiogenesis of insect DNA viruses in endoparasitic wasps. J. Insect Physiol. 49, 419–432.

Foster, J., et al., 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. Plos Biol. 3, 599–614.

Ghosh, S., May, M.J., Kopp, E.B., 1998. NF-κB and rel proteins: evolutionarily conserved mediators of immune responses. Annu. Rev. Immunol. 16, 225–260.

Godfray, H.J.C., 1994. Parasitoids. Princeton Univ. Press, Princeton.

Hacker, J., Kaper, J.B., 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. 54, 641–679.

Herniou, E.A., Olszewski, J.A., Cory, J.S., O'Reilly, D.R., 2003. The genome sequence and evolution of baculoviruses. Annu. Rev. Entomol. 48, 211–234.

Hilgarth, R.S., Webb, B.A., 2002. Characterization of *Campoletis sonorensis* segment I genes that are members of the repeat element family. J. Gen. Virol. 83, 2393–2402.

Hoffmann, J.A., Reichhart, J.-M., 2002. *Drosophila* innate immunity: an evolutionary perspective. Nat. Immunol. 3, 121–126.

Iyer, L.M., Aravind, L., Koonin, E.V., 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. 75, 11720–11734.

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.

Jurka, J., Klonowski, P., Dagman, V., Pelton, P., 1996. CENSOR—A program for identification and elimination of repetitive elements from DNA sequences. Comput. Chem. 20, 119–122.

Kroemer, J.A., Webb, B.A., 2004. Polydnavirus genes and genomes: emerging gene families and new insights into polydnavirus replication. Annu. Rev. Entomol. 49, 431–456.

Kroemer, J.A., Webb, B.A., 2005. Iκβ-related vankyrin genes in the *Campoletis sonorensis* ichnovirus exhibit temporal and tissue-specific expression in parasitized *Heliothis virescens* lepidopteran hosts. J. Virol. 79, 7617–7628.

Lapointe, R., Wison, R., Vilaplana, L., O'Reilly, D.R., Falabella, P., Douris, V., Bernier-Cardou, M., Pennacchio, F., Iatrou, K., Malva, C., Olszewski, J.A., 2005. Expression of a *Toxoneuron nigriceps* polydnavirus-encoded protein causes apoptosis-like programmed cell death in lepidopteran insect cells. J. Gen. Virol. 86, 963–971.

Le, N.T., Asgari, S., Amaya, K., Tan, F.F., Beckage, N.E., 2003. Persistence and expression of *Cotesia congregata* polydnavirus in host larvae of the tobacco hornworm, *Manduca sexta*. J. Insect Physiol. 49, 533–543.

Le Hir, H., Nott, A., Moore, M.J., 2003. How introns influence and enhance eukaryotic gene expression. Trends Biochem. Sci. 28, 215–220.

Li, K., Webb, B.A., 1994. Apparent functional role for a cysteine-rich polydnavirus protein in suppression of the insect cellular immune response. J. Virol. 68, 7482–7489.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Meagher, T.R., Costich, D.E., 2004. 'Junk' DNA and long-term phenotypic evolution in *Silene* section *Elisante* (Caryophyllaceae). Proc. R. Soc. Biol. Lett. 271, S493–S497.

Moran, N.A., 2003. Tracing the evolution of gene loss in obligate bacterial symbionts. Curr. Opin. Microbiol. 6, 512–518.

Neel, B.G., Tonks, N.K., 1997. Protein tyrosine phosphatases in signal transduction. Curr. Biol. 9, 193–204.

Parkhill, J., et al., 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature 413, 523–527.

Provost, B., Varricchio, P., Anara, E., Espagne, E., Falabella, P., Huguet, E., La Scaleia, R., Cattolico, L., Poirie, M., Malva, C., Olszewski, J.A., Penacchio, F., Drezen, J.-M., 2004. Bracoviruses contain a large multigene family coding for protein tyrosine phosphatases. J. Virol. 78, 13090–13103.

Price, P.W., 1980. Evolutionary Biology of Parasites. Princeton Univ. Press, Princeton.

Quicke, D.L.J., Fitton, M.G., Notton, D.G., Broad, G.R., Dolphin, K., 2000. Phylogeny of the subfamilies of Ichneumonidae (Hymenoptera): a simultaneous molecular and morphological analysis. In: Austin, A.D., Dowton, M. (Eds.), Hymenoptera: Evolution, Bodiversity and Biological Control. CSIRO Publishing, Collingwood, Victoria, Australia, pp. 74–83.

Raoult, d., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Claverie, J.M., 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344–1350.

Salanoubat, M., et al., 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature 415, 497–502.

Schmidt, O., Theopold, U., Strand, M.R., 2001. Innate immunity and evasion by insect parasitoids. BioEssays 23, 344–351.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., Ishikawa, H., 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp, APS. Nature 407, 81–86.

Stewart, P.E., Byram, R., Grimm, D., Tilly, K., Rosa, P.A., 2004. The plasmids of *Borrelia burgdorferi*: essential genetic elements of a pathogen. Plasmid 53, 1–13.

Stoltz, D.B., Whitfield, J.B., 1992. Viruses and virus-like entities in the parasitic Hymenoptera. J. Hymen. Res 1, 125–139.

Strand, M.R., 1994. *Microplitis demolitor* PDV infects and expresses in specific morphotypes of *Pseudoplusia includens* haemocytes. J. Gen. Virol 75, 3007–3020.

Strand, M.R., McKenzie, D.I., Grassl, V., Dover, B.A., Aiken, J.M., 1992. Persistence and expression of *Microplitis demolitor* PDV in *Pseudoplusia includens*. J. Gen. Virol. 73, 1627–1635.

Strand, M.R., Witherell, R.A., Trudeau, D., 1997. Two *Microplitis demolitor* PDV mRNAs expressed in hemocytes of *Pseudoplusia includens* contain a common cysteine-rich domain. J. Virol. 71, 2146–2156.

Tanaka, K., Tsuzuki, S., Matsumoto, H., Hayakawa, Y., 2000. Expression of *Cotesia kariyai* polydnavirus genes in lepidopteran hemocytes and Sf9 cells. J. Insect Physiol. 49, 433–440.

Theilmann, D.A., Summers, M.D., 1987. Physical analysis of the *Campoletis sonorensis* virus multipartite genome and identification of a family of tandemly repeated elements. J. Virol. 61, 2589–2598.

Thoetkiattikul, H., Beck, M.H., Strand, M.R., 2005. Inhibitor κb-like proteins from a polydnavirus inhibit NF-κB activation and suppress the insect immune response. Proc. Natl. Acad. Sci. U.S.A. 102, 11426–11431.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows intergace: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Trudeau, D., Witherell, R.A., Strand, M.R., 2000. Characterization of two novel *Microplitis demolitor* PDV mRNAs expressed in *Pseudoplusia includens* haemocytes. J. Gen. Virol. 81, 3049–3058.

Turnbull, M.W., Webb, B.A., 2002. Perspectives on polydnavirus origin and evolution. Adv. Virus Res. 58, 203–254.

Turnbull, M.W., Volkoff, A.-N., Webb, B.A., Phelan, P., 2005. Functional gap-junction genes are encoded by insect viruses. Curr. Biol. 15, R491–R492.

Villareal, L.P., 2005. Viruses and the Evolution of Life. ASM Press, Washington, DC.

Webb, B.A., Strand, M.R., 2005. The biology and genomics of polydnaviruses. In: Gilbert, L.I., Iatrou, K., Gill, S.S. (Eds.), Comprehensive Molecular Insect Science, vol. 5. Elsevier Press, San Diego, pp. 260–323.

Webb, B.A., Beckage, N.E., Hayakawa, Y., Krell, P.J., Lanzrein, B., Stoltz, D.B., Strand, M.R., Summers, M.D., 2000. Polydnaviridae. In: van Regenmortel, M.H.V, et al. (Eds.), Virus Taxonomy: The Classification and Nomenclature of Viruses. San Diego, Academic Press, pp. 253–259.

Whitfield, J.B., 2002. Estimating the age of the polydnavirus/braconid wasp symbiosis. Proc. Natl. Acad. Sci. U.S.A. 99, 7508–7513.

Whitfield, J.B., Asgari, S., 2003. Virus or not? Phylogenetics of polydnaviruses and their wasp carriers. J. Insect Physiol. 49, 397–405.

Woese, C.R., 2002. On the evolution of cells. Proc. Natl. Acad. Sci. U.S.A. 99, 8742–8747.

Wu, M., et al., 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. Plos Biol. 3, 327–341.

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.