Journal of General Virology (2002), **83**, 2393–2402. Printed in Great Britain

Characterization of *Campoletis sonorensis* ichnovirus segment I genes as members of the repeat element gene family

Roland S. Hilgarth[1] and Bruce A. Webb[2]

[1] Department of Biochemistry, University of Kentucky, Lexington, KY 40536, USA
[2] Department of Entomology, University of Kentucky, Lexington, KY 40546, USA

***Campoletis sonorensis* ichnovirus (*Cs*IV) is a symbiotic virus associated with the endoparasitic wasp *C. sonorensis*. The virus is injected into the wasp's host, *Heliothis virescens*, during oviposition. One *Cs*IV gene has been identified as a repeat element (rep) gene and encodes a ubiquitous imperfectly conserved 540 bp sequence. We report the sequencing and mapping of a rep-containing segment, segment I, that hybridizes to a known rep sequence from segment O[1]. Analysis of this 8·6 kbp segment identified three ORFs having high similarity to the 540 bp rep sequence. All three rep sequence ORFs were expressed in parasitized *H. virescens* as well as in *C. sonorensis* tissues. Two of these rep genes, I 0.9 and I 1.1, have single copies of the 540 bp repeat sequence, while the third rep gene, I 1.2, has two imperfect copies, which are more similar to each other than to sequences on the segment I single-motif genes. Like the *Cs*IV BHv 0.9 rep gene, the segment I rep genes lack introns and a signal peptide, suggesting that they are not secreted. Based on their similarity in nucleotide sequence, predicted amino acid sequence and gene structure, the three segment I repeat-containing genes, I 0.9, I 1.1 and I 1.2, are new members of the rep gene family.**

Introduction

Polydnaviruses (PDVs) are obligate endosymbionts of some endoparasitic Hymenoptera in the families Braconidae and Ichneumonidae (Fleming, 1992; Stoltz, 1993). These viruses replicate from integrated proviral DNA in specialized cells in the calyx region of the female reproductive tract (Stoltz, 1993). PDVs are classified into two genera, *Bracovirus* and *Ichnovirus*, based on virus morphology and hymenopteran family association (Webb *et al.*, 2000). Polydnavirus genomes consist of multiple polydisperse, circular, double-stranded DNA molecules. Vertical transmission of viral DNA is exclusively through the integrated proviral genome resident in the wasp germline (Stoltz, 1993). PDV genomes are thought to occur in every cell of affected wasp species, but virus replication is restricted to the reproductive tract of adult females (Fleming & Summers, 1986, 1991; Stoltz *et al.*, 1986) at the late pupal and adult stages (Norton & Vinson, 1983; Webb & Summers, 1992; Albrecht *et al.*, 1994). Virus is introduced into the haemocoel of the lepidopteran host with one or more wasp eggs during oviposition. In parasitized lepidopteran larvae, PDVs induce physiological alterations in host development and immunity that are essential for the survival of the developing parasitoid (Edson *et al.*, 1981; Lavine & Beckage, 1995; Strand & Pech, 1995; Shelby & Webb, 1999). The alteration of lepidopteran host physiology occurs without PDV replication, but viral DNA expression is required, with viral genes being transiently or persistently expressed (Blissard *et al.*, 1987; Theilmann & Summers, 1987; Strand *et al.*, 1992; Asgari *et al.*, 1996; Cui & Webb, 1996). The genome of the *Campoletis sonorensis* ichnovirus (*Cs*IV) is predicted from electrophoretic analysis of viral DNA to consist of at least 28 segments that range in size from 6 to 20 kbp, with some segments being present in higher molar amounts than others (Krell *et al.*, 1982). Viral genome segments are of at least two types, nested and unique (Cui & Webb, 1997). Nested segments produce multiple smaller viral DNA segments by intramolecular recombination. This intramolecular recombination is thought to occur between internal direct repeats following replication of the parent extrachromosomal viral

Author for correspondence: Bruce Webb.
Fax +1 859 323 1120. e-mail bawebb@pop.uky.edu

The GenBank accession number of the sequence reported in this paper is AF362513.

segment. Segment nesting results in amplification of some viral genes, and is linked to an increase in expression of the amplified genes in parasitized insects (Cui & Webb, 1998). By contrast, unique segments are excised from the wasp genome and give rise to single viral segments.

*Cs*IV gene expression has been studied in some detail with RNA blot hybridizations. Viral genes may be expressed in the wasp during virus replication (class I), in the parasitized lepidopteran host (class II), or in both hosts (class III) (Theilmann & Summers, 1988). The published studies have emphasized characterization of genes that are expressed in parasitized lepidopteran hosts (class II) in order to identify those viral genes that may alter their physiology. Two distinct viral gene families have been studied in *Cs*IV: a cysteine-rich gene family with five identified members, which is expressed only in parasitized lepidopteran larvae (Dib-Hajj *et al*., 1993; Cui & Webb, 1996), and a second putative gene family, the repeat element (rep) gene family, which has been proposed based on cross-hybridization of a sequenced rep gene, BHv 0.9, to several mRNAs on Northern blots and on sequence similarity among viral DNA segments (Theilmann & Summers, 1988). BHv 0.9 has a class II gene expression pattern, although some of the cross-hybridizing mRNAs show class III expression (Theilmann & Summers, 1988). The proposed rep gene family is distinguished by an imperfectly conserved 540 bp repeat sequence (Theilmann & Summers, 1987). This rep sequence hybridized to the majority of *Cs*IV segments at reduced stringencies, suggestive of either a widely distributed gene family or a conserved viral DNA structural element. The segment B rep gene and three other expressed rep sequences were identified on viral segments H and O^1 by Theilmann & Summers (1988). The putative H and O^1 rep genes contain multiple 540 bp rep sequences, while BHv 0.9 has a single copy of the repeat sequence (Theilmann & Summers, 1987). Because the 540 bp repeat sequence cross-hybridized with most, possibly all, *Cs*IV segments, Theilmann & Summers (1988) hypothesized that the rep genes may represent a gene family having many members. Whether or not the cross-hybridizing sequences were expressed genes, potentially expressed genes or non-functional pseudogenes was not determined. In the present study, we describe the isolation and sequence of segment I, and identify three new members of the rep gene family. Characterization of these genes suggests that the rep genes have evolved in lineages of single-repeat-containing genes and multiple-repeat-containing genes. Preliminary characterization of the genes indicates that they are functional but expressed at different levels, even though they are found on the same viral segment.

Methods

■ **Insect rearing and nucleic acid isolation.** Insect rearing and isolation of viral DNA were essentially as described by Krell *et al*. (1982). *C. sonorensis* pupae were staged according to Norton & Vinson (1983), as modified by Webb & Summers (1992). RNA isolation from parasitized *H. virescens* and *C. sonorensis* tissues used Tri-Pure Reagent and protocols (Molecular Products). Third-instar *H. virescens* larvae were parasitized by mated female wasps for 1 h with time post-parasitization calculated from the initial exposure of larvae to wasps. RNA extractions were from at least three larvae (~ 100 mg total) per time-point. Tissue-specific RNA extraction was from parasitized and control fourth-instar *H. virescens* larvae. Tissues were washed four times for 30 s each at 4 °C in 2 ml PBS, pH 6·8, containing 100 mM NaCl, 3 mM KCl, 10 mM Na_2HPO_4 and 2 mM KH_2PO. One μg of total RNA was used for RT–PCR experiments. RNA was also extracted from four pairs of *C. sonorensis* ovaries from staged pupae, intact adult females, ovariectomized females and adult males.

■ **Isolation of the segment I clone.** Probes were labelled by the random primer method (Feinberg & Vogelstein, 1983) with [^{32}P]dATP using the Prime-a-Gene labelling system (Promega). Thirteen cDNA clones that hybridized to the rep sequence, HC1185 (Theilmann & Summers, 1987, 1988), were isolated by screening a λgt11 parasitized *H. virescens* cDNA library under low-stringency conditions (30% formamide, 6 × SSC, 5 × Denhardt's solution, 0·5% SDS at 42 °C, with two washes of 20 min in 2 × SSC, 0·5% SDS at 42 °C), followed by two washes of 0·1 × SSC, 0·5% SDS at 65 °C for 20 min with autoradiography at −80 °C overnight (Sambrook *et al*., 1989). Hybridizing phage plaques were amplified by PCR, and then subcloned into pBluescript II KS(−) (Stratagene), as previously described (Cui & Webb, 1996). Five clones representing different insert sizes were sequenced. To isolate genomic clones encoding the expressed rep transcripts, the largest cDNA clone of 1·6 kbp was used to screen *Cs*IV genomic libraries prepared by digesting *Cs*IV genomic DNA with *Xho*I, SphI, *Bam*HI and *Sac*I, and ligating inserts into compatibly digested pZero-1.0 vector (Invitrogen) (L. Cui, unpublished data). Hybridization of cDNA clones to viral DNA was carried out under high stringency conditions (50% formamide, otherwise as above, followed by two washes in 0·1 × SSC, 0·5% SDS at 65 °C for 20 min) with autoradiography at −80 °C overnight.

■ **Sequencing of segment I.** For sequencing, segment I subclones were constructed by digesting SphI35, a positive segment I clone identified by hybridization, with *Xho*I, *Bam*HI, *Sac*I and *Eco*RI individually and cloning fragments into pBluescript II KS (Stratagene). Nested unidirectional deletion subclones were also constructed using the Erase-a-Base System (Promega) after linearizing 10 μg of SphI35 with *Spe*I and *Kpn*I. Some site-specific primers were designed and used to complete the segment I sequence.

■ **Identification of segment I rep genes by rapid amplification of cDNA ends (RACE).** Rep sequences were identified by homology to BHv 0.9, O^1 and H rep sequences from segments B, O^1 and H (Theilmann & Summers, 1987, 1988) with gene expression confirmed by amplification and sequencing of the 3′ end of the segment I rep transcripts. Gene-specific primers were designed to the three putative segment I rep genes: I 1.1 (5′ ATG GAT TCC CCG GTC AAA G 3′, I1.1a-SphI35, nt 3896–3914), I 1.2 (5′ ATG GGA TCA TTG ACT AGT G 3′, I1.2a-SphI35, nt 7062–7080) and I 0.9 (5′ ATG GAG TCC CTG GGC GAA GG 3′, I 0.9a-SphI35, nt 2695–2714). First-strand cDNA synthesis used an oligo(dT) primer [5′ CCA GTG AGC AGG TGA CGA GGA CTC GAG CTC AAG $C(T)_{17}$] (Frohman, 1994), and 4 μg total RNA isolated from parasitized *H. virescens* larvae, non-parasitized *H. virescens* larvae and *C. sonorensis*. Reverse transcription reactions of 1 × cDNA synthesis buffer [250 mM Tris–HCl, pH 8·3, 375 mM KCl, 15 mM $MgCl_2$, 200 μM of each dNTP, 10 mM DDT, 50 pmol oligo(dT) and 200 units of Superscript II (Gibco)] in 20 μl were incubated at 42 °C for 50 min. The amplification of the 3′ end of segment I rep genes followed a procedure

modified from Frohman (1994). One µl of the reverse transcription mixture was added to a 50 µl PCR reaction containing 1 × PCR buffer [10 mM Tris–HCl, pH 9·0, 50 mM KCl], 1·5 mM $MgCl_2$, 50 µM dNTP, 25 pmol of oligo(dT) and 25 pmol of a segment I rep gene-specific primer, and amplified through five amplification cycles of 94 °C for 1 min, 61 °C for 1 min and 72 °C for 1·5 min, followed by 25 amplification cycles of 94 °C for 1 min, 56 °C for 1 min and 72 °C for 1·5 min, in a Perkin-Elmer/Cetus 480 DNA Thermal Cycler. Segment I rep gene PCR reactions were analysed by electrophoresis of 8 µl of each reaction on a 1% agarose gel in 1 × TAE. To detect segment I transcripts in *C. sonorensis* and I 0.9 transcripts in parasitized *H. virescens*, a second round of amplification was required. The initial amplification was diluted 1:100 in water, and 1 µl of this dilution used in a second amplification reaction with a nested gene-specific primer: I 1.1 (5′ TTA GAA GAG GAC GTA TGC CCA 3′, I1.1b-SphI35, nt 4373–4393), I 1.2 (5′ ATG CGA GAT ATT GGC TAC GTC GT 3′, I1.2b-SphI35, nt 7582–7604) and I 0.9 (5′ GCG AAG AAG ACG ACA CTT TC 3′, I0.9b-SphI35, nt 2850–2869) with oligo(dT) as the reverse primer and amplification reactions and conditions as described above. The 3′ RACE amplimers corresponding to the predicted sizes of the transcripts were cloned into pGEM-T Easy (Promega) and sequenced to confirm amplification of segment I rep gene transcripts. Control amplifications used a segment I genomic DNA template with negative controls lacking template (water only) in primer combinations as described above.

The 5′ ends of the segment I rep cDNAs were isolated by 5′ RACE (5′ RACE system, Gibco BRL). Individual segment I rep gene primers were used in the reverse transcription reactions with 3 µg of total *H. virescens* RNA [I 1.1 (5′ CAT GGT CAA TGT GTT TAC AAA TTC 3′; I1.1c-SphI35, nt 3069–3048), I 1.2 (5′ GTC CTC TCG TTC GTA GG 3′, I1.2c-SphI35, nt 7207–7190) and I 0.9 (5′ GTT TTT CTT TGA AAG TGT CG 3′, I0.9c-SphI35, nt 2879–2860)]. PCR amplification of each segment I rep gene reverse transcription product was carried out with specific primers for each rep gene [I 1.1 (5′ GTA GCT CGA CTT CCC CAA GGA TTG 3′, I1.1d-SphI35, nt 4232–4209), I 1.2 (5′ GCG TAC AGA CCA TGA GAA CAT 3′ I1.2d-SphI35, nt 7168–7148) and I 0.9 (5′ TAG TCC TCG ATG TTC CTT CGC 3′, I0.9d-SphI35, nt 2728–2708)] in conjunction with the anchor primer (5′ GGC CAC GCG TCG ACT AGT ACG GGI IGG GII GGG IIG 3′) (Gibco BRL), following the manufacturer's instructions. PCR conditions were 30 cycles of 94 °C for 1 min, 56 °C for 1 min and 72 °C for 1 min. The 5′ RACE-amplified products were analysed on a 1% agaraose gel in 1 × TAE, and then cloned into pGEM-T Easy (Promega) and sequenced.

■ **DNA sequence determination and analysis.** DNA sequencing reactions routinely used the ABI Prism dRhodamine or Big Dye Terminator Cycle Sequencing Ready Reaction kits with the sequence visualized on ABI 377 or ABI 310 DNA sequencers (PE Applied Biosystems). Sequence data were assembled with the Lasergene sequence analysis software (DNASTAR). ORF analysis employed Gene Construction Kit 2 (Textco) and the Baylor College of Medicine Gene Finder Find genes H program set on Drosophila (http://dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html). The segment I rep cDNA sequences were analysed for protein motifs and cellular localization signals using the PSORT prediction program (http://psort.nibb.ac.jp/). Nucleotide and peptide sequence similarity searches utilized BLAST search programs (Altschul *et al.*, 1997) with DNA alignments utilizing the University of Wisconsin Genetics Computer Group DNA analysis software for the VAX computer (release 7.2) and protein alignments used the Clustal X alignment program (Thompson *et al.*, 1997).

Results

A 540 bp repeated sequence was identified by Theilmann & Summers (1987) as hybridizing to the majority of *Cs*IV segments. This sequence was hypothesized to be a gene family or a repeated sequence found on a large majority, possibly all, *Cs*IV segments (Theilmann & Summers, 1987, 1988). To further characterize rep sequences in the *Cs*IV genome, a partial cDNA clone encoding two copies of the segment O[1] repeated sequence, HC1185 (Theilmann & Summers, 1987, 1988), was used to isolate a segment containing cross-hybridizing sequences. An 8·6 kbp clone (SphI35) isolated from *Cs*IV genomic plasmid libraries was mapped with restriction enzymes, subcloned and sequenced. SphI35 contains a complete *Cs*IV segment, as indicated by amplifications across the *Sph*I cloning site using *Cs*IV genomic DNA as a template (data not shown). *Cs*IV Southern hybridizations with SphI35 probes hybridized to a single *Cs*IV segment identified as segment I based on the size of the hybridizing segment (Fig. 1). This is

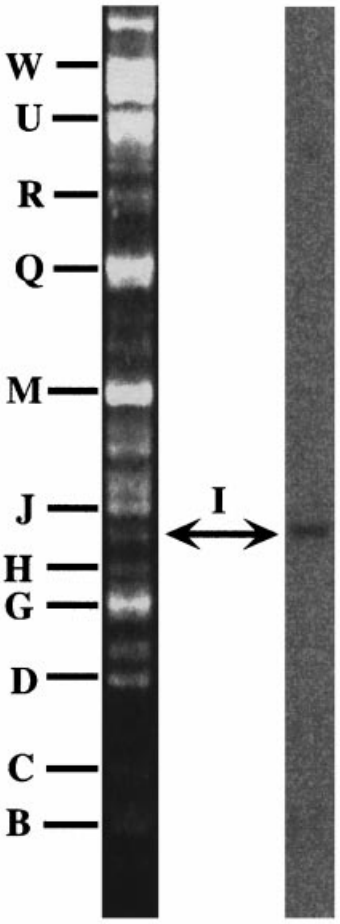

Fig. 1. Hybridization of SphI35 to *Cs*IV genomic DNA. The left panel shows 3 µg of superhelical *Cs*IV DNA separated on a 0·7% agarose gel with DNA and visualized by ethidium bromide staining. Major *Cs*IV segments are labelled on the left. The right panel shows hybridization of the SphI35 probe to the blot of the gel shown on left. The SphI35 probe hybridized to a single segment, identified as segment I.

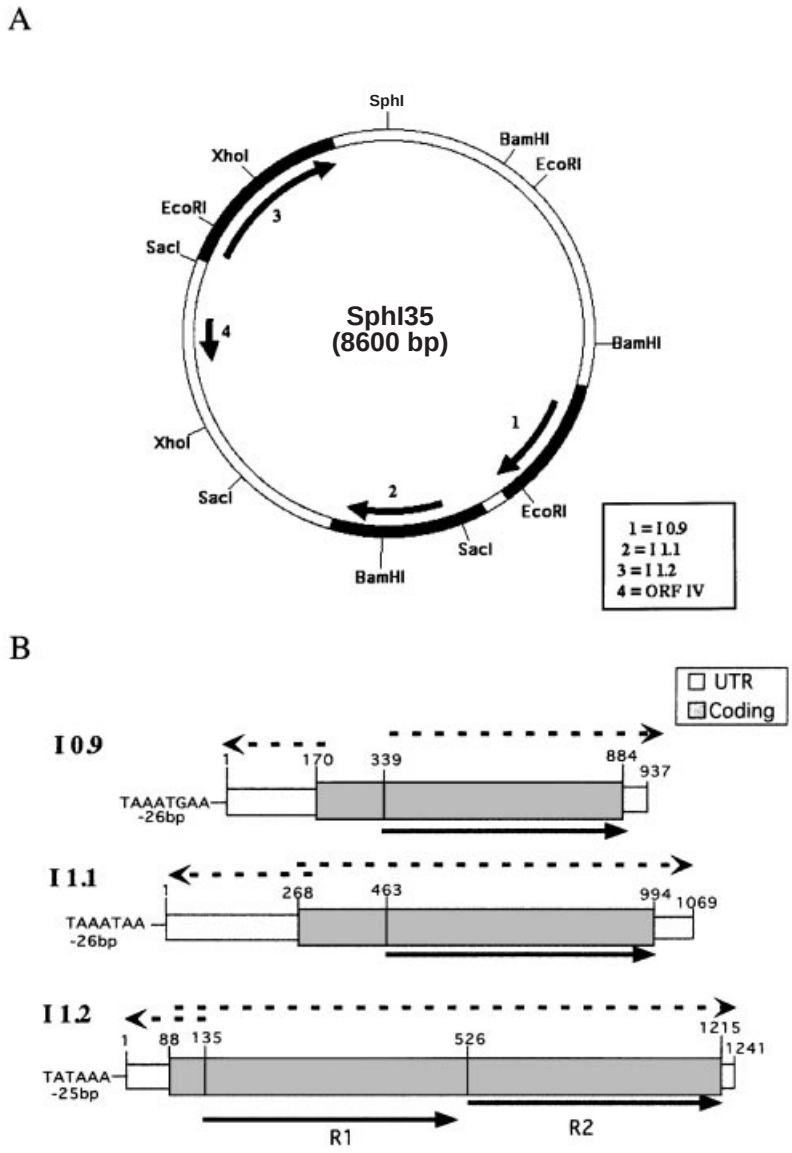

Fig. 2. Segment I restriction and ORF map and schematic diagram of the three segment I rep genes. (a) Restriction map of segment I. Arrows indicate the coding regions of the four predicted ORFs longer than 100 amino acids. The three predicted rep gene transcripts, I 1.2 (nt 6974–8215), I 1.1 (nt 3628–4697) and I 0.9 (nt 2524–3461) are shown as black regions on the segment. (b) Schematic representation of the three rep genes identified on segment I. The I 0.9, I 1.1 and I 1.2 rep genes are shown. The solid arrows indicate the region of the gene having similarity to the 540 bp rep sequence. Dashed arrows indicate the 5′ and 3′ RACE products that were obtained for each gene with the respective I0.9a, I1.1a, I1.2a forward and I0.9b, I1.1b, I1.2b reverse primers.

similar to the other known rep gene contained on segment B (Fleming & Summers, 1986).

Identification of rep genes on segment I

Segment I was mapped with restriction enzymes and sequenced to confirm the presence of the rep sequence (Fig. 2a). ORF analysis of the genomic sequence identified four ORFs over 100 amino acids in length beginning with methionine (Fig. 2a). Eight other ORFs of less than 50 amino acids were present, but had no significant similarity to NCBI database sequences, and were not identified as ORFs by gene finder programs. Therefore, the small ORFs were not examined further. The predicted proteins from segment I ORFs I, II and III were similar to the *Cs*IV rep gene BHv 0.9 in BLAST X searches of the databases. The three ORFs had e-values of $1e^{-42}$, $3e^{-49}$ and $7e^{-45}$ and amino acid identity levels of 38, 47 and 51%, respectively. ORF IV had no similarity to NCBI database sequences, and as RT–PCR did not detect transcription in either parasitized *H. virescens* or *C. sonorensis*, this ORF was not considered further.

Transcripts from segment I were not detected in total RNA Northern blots. Therefore, gene-specific primers were designed and used to analyse expression of the three predicted segment I ORFs. ORF II and ORF III RT–PCR products of 1100 and 650 bp, respectively, were amplified from parasitized *H. virescens* RNA. Nested gene-specific primers were required to detect ORF II and ORF III transcription in *C. sonorensis* and ORF I transcripts in both parasitized *H. virescens* and *C. sonorensis*. The cloned amplimers were sequenced to confirm amplification of segment I rep gene transcripts and to delineate transcript termini. The 5′ amplimers were then cloned and sequenced to

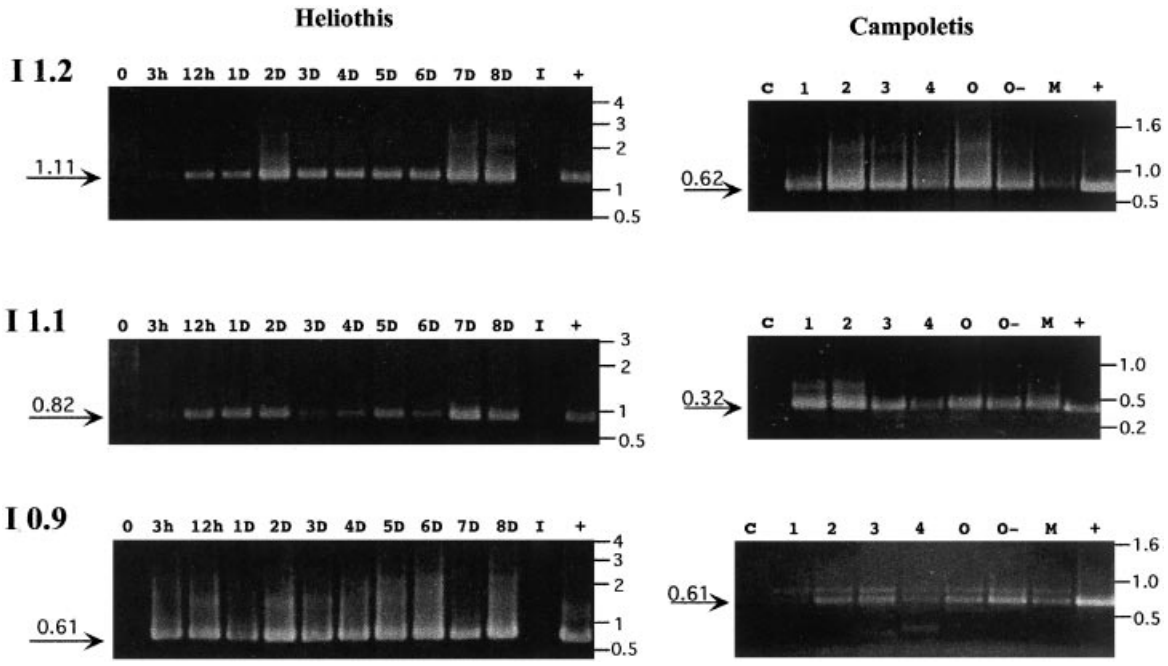

Fig. 3. Temporal expression of segment I rep genes in parasitized *Heliothis virescens* larvae and *Campoletis sonorensis*. The three segment I rep genes were detected by RT–PCR in parasitized *H. virescens* and *C. sonorensis*. The left panel shows transcripts amplified from parasitized *H. virescens* larvae with non-parasitized *H. virescens* larvae controls (0) at various time-points post-parasitization (p.p.) (h, hours p.p; D, days p.p.). As a positive control, gene-specific RT–PCR clones (+) were amplified, while segment I was used as a negative control to show that DNA templates were not amplified (I). The right panel shows amplification of segment I gene transcripts from *C. sonorensis* wasp pupae and adults, with controls (M, male *C. sonorensis*; O, adult *C. sonorensis* ovaries; O−, adult *C. sonorensis* without ovaries; 1–4, *C. sonorensis* ovaries from the pupal stages described by Webb & Summers (1992); +, gene-specific RT–PCR clones; c, water control). Molecular mass markers are indicated in kbp on the right, with amplicon sizes from the rep gene transcripts indicated on the left. The I 1.2 and I 1.1 transcripts were detected after one round of amplification while the I 0.9 rep transcript was detected after two rounds of PCR amplification using a nested primer. All transcripts in *C. sonorensis* are shown after two rounds of amplification.

delineate full-length transcripts of ORF I, ORF II and ORF III. Based on transcript sizes, the segment I genes were designated I 0.9, I 1.1 and I 1.2, respectively, in accordance with existing nomenclature (Theilmann & Summers, 1988).

Segment I rep gene structure

The I 0.9 transcript is 937 bp excluding the poly(A) tract (Fig. 2b). This gene encodes one copy of the 546 bp sequence characteristic of the rep gene family. The I 0.9 ORF encodes a predicted protein size of 28·6 kDa having a 56 amino acid sequence upstream of the conserved repeat. This transcript has a 170 bp 5′ UTR with a consensus TATA box 27 bp upstream of its 5′ RACE terminus. The 3′ UTR is 51 bp with a polyadenylation signal (AATAAA) 15 bp upstream of the poly(A) tract.

ORF II encodes the I 1.1 transcript of 1069 bp excluding the poly(A) tract (Fig. 2b). This gene contains a 531 bp sequence with repeat sequence homology and a 195 bp coding region upstream of the repeat sequence. The ORF in this transcript encodes a predicted protein of 28·9 kDa. The 5′ UTR of this transcript is at least 268 bp, with a putative TATA box 26 bp upstream of its terminus and a 75 bp 3′ UTR having a consensus polyadenylation signal (AATAAA) 13 bp upstream of the 3′ poly(A) tract.

The third segment I rep gene, I 1.2, encodes a transcript 1241 bp in length excluding the poly(A) tract (Fig. 2b). This gene contains two copies of the repeat sequence in a 525 bp tandem repeat. A 48 bp coding region is present upstream of the repeat sequences resulting in a predicted protein of 43·6 kDa. The 525 bp tandem repeats have 89% nucleotide sequence similarity. The 5′ UTR of the I 1.2 transcript is at least 88 bp, with a putative TATA box present 25 bp upstream of the transcript's 5′ terminus. A 26 bp 3′ UTR contains a potential polyadenylation signal (AATAAA) 15 bp upstream of the 3′ poly(A) tract.

Consensus translation initiation sequences (PuNNATGPu) (Kozak, 1983) flank the first methionine codon of each segment I rep gene ORF described above. The N termini of these predicted proteins do not encode a consensus signal peptide (von Heijne, 1986), suggesting that the proteins encoded by these ORFs are not secreted.

Expression of the segment I gene

To evaluate segment I rep gene expression, RT–PCR was performed at selected time-points after parasitization with

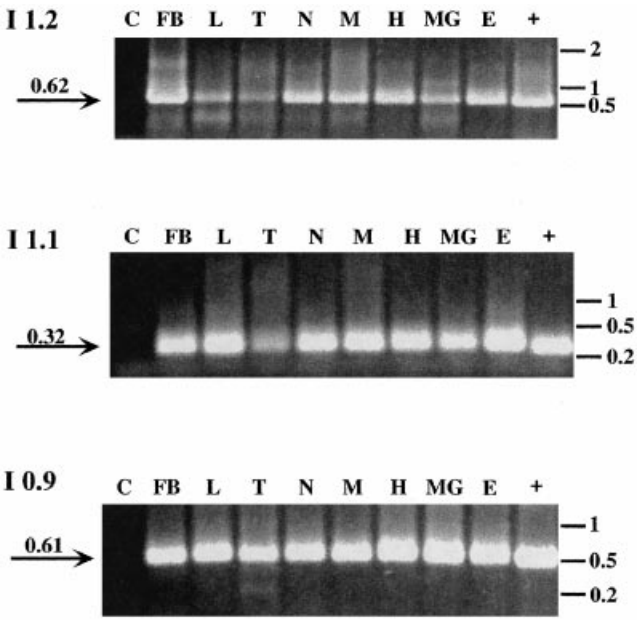

Fig. 4. Amplification of segment I rep transcripts from parasitized *H. virescens* tissues. The transcript amplified and size in kbp are indicated on the left (I 1.2, I 1.1 and I 0.9). Negative (C, water) and positive (+, 10 ng of a cDNA) controls are indicated. RNAs from non-parasitized insects did not produce amplicons (Fig. 3). Parasitized insect RNA (1 µg; 48 h post-parasitization) was used in RT–PCR reactions. Tissues assayed were fat body (FB), labial glands (L), testes (T), nerve tissue (N), Malpighian tubules (M), haemocytes (H), midgut (MG) and epidermis (E). Molecular mass markers are indicated in kbp on the right.

primers specific for each of the segment I rep genes. Consistent with the expression of other *Cs*IV genes, transcripts from I 0.9, I 1.1 and I 1.2 were detected 3 h after oviposition and throughout 8 days of parasitization (Fig. 3). The I 1.1 and I 1.2 rep genes were readily amplified from parasitized *H. virescens* mRNA, but two rounds of amplification were required to detect I 0.9 rep gene transcripts. To determine if segment I rep genes are transcribed in *C. sonorensis*, RNAs were extracted from adult, stage I, stage II, stage III and stage IV *C. sonorensis* ovaries for RT–PCR analyses (Webb & Summers, 1992). RNA was also isolated from ovariectomized females and male *C. sonorensis*. Transcripts were detected in all stages of ovarian development, in males, and in ovariectomized females (Fig. 3). Negative controls using the segment I clone (SphI35) showed that a DNA template would not produce an amplicon in these reactions.

To study tissue-specific expression of the segment I rep genes in parasitized *H. virescens*, the segment I rep gene transcripts were amplified from RNAs isolated from selected host tissues. The three segment I rep gene transcripts were detected from all parasitized *H. virescens* tissues assayed (Fig. 4). Although tissues were washed extensively, the possibility that some haemocytes adhered to tissues and contributed to these uniformly positive results was not eliminated. However, the fact that washed testes and nerve tissue were positive for transcription and are unlikely to harbour haemocytes makes this possibility less likely, at least for these tissues.

Comparison of segment I rep genes with other repeat element sequences

The segment I rep genes were significantly related only to the *Cs*IV rep sequences in database searches. The segment I rep genes had an average nucleotide identity of ~ 50% over the repeat region compared with published sequences from segments B, O[1] and H (Theilmann & Summers, 1988). Segments B and I rep element sequences were also similar in an area 5′ of the repeat region of the rep gene (nt 1–192). The overall nucleic acid similarity within this region between BHv 0.9, I 1.1, I 1.2 and I 0.9 was 62%. The BHv 0.9 N terminus had the highest nucleotide similarity to I 1.1 (83%) and I 0.9 (70%). Interestingly, the I 1.1 and I 0.9 N termini were less similar, having only a 64% nucleotide identity. Within the rep sequence family, the I 1.2 repeats had an identity of 89%, while I 0.9 and I 1.1 repeats were also quite similar at the nucleotide level (74%). Interestingly, repeat regions of the tandem I 1.2 repeats were more similar at a nucleotide level to the repeat on segment B, BHv 0.9 (~ 70%), than they were to the single-repeat-containing genes that are also present on segment I. By contrast, the two single-repeat-containing rep genes on segment I, I 0.9 and I 1.1, have repeat sequences that are slightly more similar at a nucleotide level to one of the segment H repeats (~ 55%) than to the single repeat sequence containing the BHv 0.9 rep gene (~ 52%).

The BHv 0.9 predicted protein (Theilmann & Summers, 1988) and the three predicted segment I rep proteins were compared at the amino acid level. I 0.9, I 1.1 and BHv 0.9 had 56 amino acids located at the N terminus with an identity ranging from 58–53%. The I 1.2 gene lacks 40 amino acids in the non-repeat portion of the N terminus relative to the BHv 0.9, I 0.9 and I 1.1 genes, with the rest of the predicted peptides encoded by the rep sequence. Within the repeat sequences, the four predicted proteins had an average amino acid identity of 42%. The two I 1.2 repeats were most similar, having amino acid identities of 79%. Alignments of the proteins encoded by BHv 0.9, I 0.9, I 1.1 and I 1.2 revealed two amino acid sequence motifs conserved within the repeat coding portion of these genes (Fig. 5). A 16 amino acid conserved sequence of F-N---IEVEY-Y-RE was encoded by the N-terminal portion of the repeat, while the 19 amino acid consensus sequence D-CP--HFHH--P-H---W was present near the C terminus. Two additional motifs were observed in comparisons of the predicted proteins of the non-repeat coding portion of the single-repeat rep genes, BHv 0.9, I 0.9 and I 1.1 (Fig. 5). A 13 amino acid consensus sequence, TP-YFSTRQ--LP, and a 10 amino acid consensus sequence, FR-F-RAMWP, appeared in the BHv 0.9, I 1·0 and I 0.9 genes near the N terminus. The second motif was at the boundary of the 540 bp repeat sequence. In the I 1.2 gene, the two sequences present outside the repeat region were absent (Fig. 5).

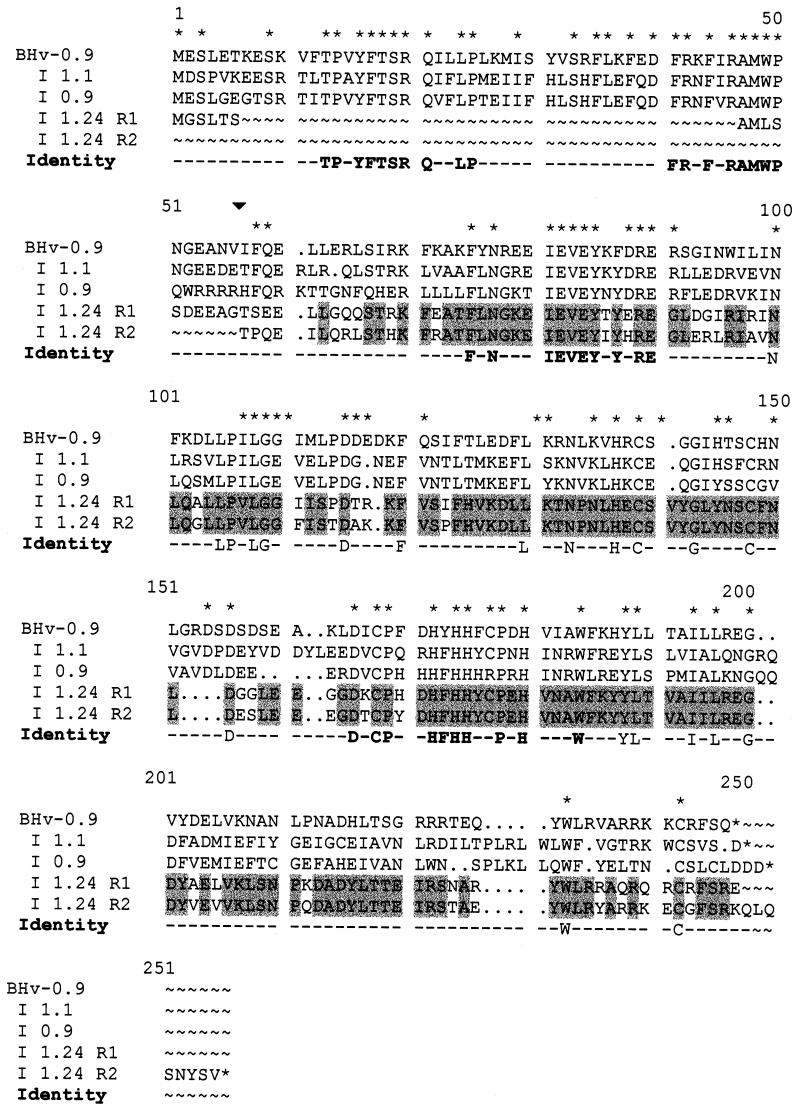

Fig. 5. Alignment of the three proteins encoded by segment I rep genes with the BHv 0.9 protein. The I 1.2 predicted protein encodes two copies of the repeat sequence, I 1.2 R1 and I 1.2 R2, which are encoded by repeats 1 and 2. The arrow indicates the start of the 540 bp rep coding sequence from Theilmann & Summers (1988). Asterisks above the protein sequence indicate amino acid identity between the BHv 0.9, I 0.9 and I 1.1 proteins. The shaded amino acids between the I 1.2 repeats show amino acid identity between the two tandem repeats in the I 1.2 gene. The bold amino acid sequences indicate conserved amino acid motifs in the rep genes.

Discussion

Theilmann & Summers (1988) proposed that the rep genes make up a second *Cs*IV gene family. Here we report identification of three members of the rep gene family located on segment I: I 0.9, I 1.1 and I 1.2. Prior to this study, only one rep gene member had been completely sequenced, BHv 0.9, while three other rep genes were suggested based on Northern blot hybridizations and partial sequences (Theilmann & Summers, 1987, 1988). Identification of three new rep genes from segment I demonstrates that the multiple rep sequences are expressed and constitute a *Cs*IV gene family. The segment I genes are members of the rep gene family based on amino acid sequence similarity to BHv 0.9. The four isolated rep genes, BHv 0.9, I 0.9, I 1.1 and I 1.2, lack introns and have a 540 bp rep sequence that defines this gene family. These characteristics may facilitate identification of other rep gene family members in the *Cs*IV genome, as hybridization studies suggest that other members exist (Theilmann & Summers, 1987).

The three segment I rep genes are expressed suggesting that most, possibly all, rep sequences are part of expressed genes. Previous studies have suggested that the rep genes have two expression patterns. The BHv 0.9 rep gene is expressed only in parasitized larvae making it a class II *Cs*IV gene (Theilmann & Summers, 1988). Two other rep sequences from segments H and O^1 hybridize to mRNAs isolated from parasitized lepidopteran larvae and *C. sonorensis* adult female ovaries, suggesting that the putative segment H and O^1 rep genes are class III *Cs*IV genes. Detection of segment I rep gene transcripts in both parasitized *H. virescens* larvae and *C. sonorensis* confirm that rep gene family members belong to both class II and class III *Cs*IV genes. This result is in contrast with the *Cs*IV cysteine-motif gene family, which is only expressed in parasitized *H. virescens* (Blissard *et al.*, 1987; Cui & Webb, 1996). The temporal expression in parasitized Lepidop-

tera of the segment I genes is identical to that of the BHv 0.9 rep gene, with transcripts detected as early as 2 h post-parasitization (Theilmann & Summers, 1988). The segment I rep genes may be expressed at different levels in parasitized insects, since two rounds of PCR amplification were required to detect the I 0.9 transcripts while one was sufficient for detection of the I 1.1 and I 1.2 rep gene transcripts. Quantitative studies are required to confirm this difference in transcript level, but different levels of expression may suggest that the virus has functional genes that are transcriptionally inactive but available for activation by genetic rearrangement. Alternatively, these genes may be differentially expressed in a tissue- or developmental stage-dependent manner, although there is no experimental evidence to support this hypothesis.

Expression of the BHv 0.9 rep gene in parasitized tissues has not been investigated, as prior studies utilized homogenates of whole larvae and evaluated expression in Northern blots. In this study, segment I rep gene transcripts were detected by RT–PCR in all tissues assayed, suggesting that rep genes are expressed ubiquitously in parasitized tissues, albeit at low levels for some genes and some tissues. Segment I rep gene expression was also detected in *C. sonorensis* in tissues that do and do not support virus replication. Expression of PDV genes in tissues that do not support virus replication has been previously reported by Johner *et al.* (1999) who detected PDV gene expression in *Chelonus inanitus* males by RT–PCR, although virus replication was not detected.

The known rep sequences have an overall nucleotide identity of 50%, with some areas more highly conserved (up to 90% identical). Segment I has two rep genes with single repeat sequences and one with two copies of the rep sequence in tandem. The two single-repeat segment I rep genes, I 0.9 and I 1.1, have a 56 amino acid N terminus upstream of the repeat sequence. This organization is similar to that of the BHv 0.9 rep gene. Interestingly, the single-repeat segment I genes are more similar to the BHv 0.9 repeat than to the I 1.2 rep gene, which has two copies of the rep sequence. These two genes, I 0.9 and I 1.1, have a 74% nucleotide identity, suggesting that gene duplication has occurred to produce these two genes, whereas the I 1.2 gene, which contains two rep sequences in tandem, has an origin more divergent than the repeat gene on segment B.

*Cs*IV segment O^1 may have multiple rep genes, as two regions of the segment hybridize to the 540 bp rep sequence and to transcripts on Northern blots (Theilmann & Summers, 1988). Theilmann & Summers (1988) also showed that many *Cs*IV segments, possibly all, hybridize to repeat sequence probes in genomic Southern blots. They also identified multiple repeat sequences in tandem arrays on segment H (five tandem repeats) and O^1 (three tandem repeats). Interestingly, the I 1.2 tandem repeats are more similar to the sequences from segments H and O^1 that exist in tandem arrays than to the rep sequences from genes encoding a single repeat (BHv 0.9, I 1.1 and I 0.9). This suggests that the I 1.2 rep gene may have arisen from other multiple-repeat-sequence-containing rep genes rather than from duplication of a single repeat sequence on segment I. Thus, two types of rep genes seem to exist and may represent two major lineages within this gene family. One lineage is comprised of rep genes encoding a single repeat sequence. The second lineage is comprised of those genes that have repeat sequences in tandem array. Segment I may be unusual in that it encodes both single-repeat-containing and multiple-repeat-containing rep genes.

Northern blot analysis of mRNAs from *H. virescens* larvae parasitized by *C. sonorensis* detects at least 10 viral mRNAs that hybridize to *Cs*IV DNA (Blissard *et al.*, 1986). The number of transcripts expressed in parasitized *H. virescens* larvae is probably greater because of co-migrating transcripts that would not be differentiated by this method. Four of the expressed genes in parasitized *H. virescens* belong to the cysteine-motif gene family (Blissard *et al.*, 1987; Dib-Hajj *et al.*, 1993; Cui & Webb, 1996). These genes are highly expressed, with gene amplification occurring through an intramolecular recombination process known as segment nesting (Cui & Webb, 1998). With the three rep genes reported here, there are four rep genes now known to be expressed in parasitized larvae, giving a total of seven identified *Cs*IV genes. This number is likely to rise as rep sequences on segments H and O^1 also hybridize to transcripts in parasitized lepidopteran larvae.

The number of *Cs*IV genes (class I and class II) expressed in *C. sonorensis* has not been studied in detail. One *Cs*IV-encoded nucleocapsid gene, p12, has been isolated (Deng & Webb, 1999). Presently, at least six rep genes appear to be expressed in *C. sonorensis*: the three identified segment I rep genes, two O^1 repeat-hybridizing transcripts and one H repeat-hybridizing transcript. Segment B is also predicted to contain a class I gene that does not belong to the rep gene family, but this gene has not yet been fully isolated (Theilmann & Summers, 1988). The p12 gene is expressed only in tissues that support *Cs*IV replication, with expression coincident with virus replication (Deng & Webb, 1999). Segment I rep gene expression appears to differ from that of the p12 gene, as expression is detected prior to virus replication and in non-replicative tissues. However, the RT–PCR technique is more sensitive than Northern analysis previously used to assay for *Cs*IV gene expression, so it is conceivable that our understanding of *Cs*IV transcriptional regulation may be revised as this technique is applied to analyses of additional viral genes.

Although rep gene expression may differ in a host- and tissue-specific manner, similar portions of the rep proteins are conserved among all known rep genes. Two conserved amino acid motifs are evident within the 180 amino acid repeat. One motif is located near the N terminus of the repeat (F-N---IEVEY-Y-RE). The second motif is encoded near the repeat's C terminus (D-CP--HFHH--P-H---W). The presence of these conserved amino acid motifs in the segment B and segment I rep genes suggests that these regions are important for function of these proteins.

All the *Cs*IV rep proteins lack secretion signals suggesting that these proteins act intracellularly. However, more directed study of the function of this gene family is required to provide an understanding of their roles in parasitization and/or virus replication. The identification and analysis of three members of the rep gene family from segment I confirm that the rep sequences do constitute a *Cs*IV gene family. Studies on the levels of expression of these rep genes should be undertaken to determine the possible role and biological significance of these rep genes in this system. The evidence that rep genes are a large gene family suggests they may have an important, albeit presently unknown, function in the *Cs*IV life cycle (Theilmann & Summers, 1987, 1988).

We gratefully acknowledge support from the National Science Foundation (MCB 9726966, 0094403) and the US Department of Agriculture, National Research Initiative (98-35302) and Hatch act. This is paper number 01-08-185 of the Kentucky Agricultural Experiment Station, Lexington, KY, USA.

References

Albrecht, U., Wyler, T., Pfister-Wilhelm, R., Gruber, A., Stettler, P., Heiniger, P., Kurt, E., Schumperli, D. & Lanzrein, B. (1994). Polydnavirus of the parasitic wasp *Chelonus inanitus* (Braconidae): characterization, genome organization and time point of replication. *Journal of General Virology* **75**, 3353–3363.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

Asgari, S., Hellers, M. & Schmidt, O. (1996). Host haemocyte inactivation by an insect parasitoid: transient expression of a polydnavirus gene. *Journal of General Virology* **77**, 2653–2662.

Blissard, G. W., Fleming, J. G. W., Vinson, S. B. & Summers, M. D. (1986). *Campoletis sonorensis* virus: expression in *Heliothis virescens* and identification of expressed sequences. *Journal of Insect Physiology* **32**, 351–359.

Blissard, G. W., Smith, O. P. & Summers, M. D. (1987). Two related viral genes are located on a single superhelical DNA segment of the multipartite *Campoletis sonorensis* virus genome. *Virology* **160**, 120–134.

Cui, L. & Webb, B. A. (1996). Isolation and characterization of a member of the cysteine-rich gene family from *Campoletis sonorensis* polydnavirus. *Journal of General Virology* **77**, 797–809.

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. *Journal of Virology* **71**, 8504–8513.

Cui, L. & Webb, B. A. (1998). Relationships between polydnavirus genomes and viral gene expression. *Journal of Insect Physiology* **44**, 785–793.

Deng, L. & Webb, B. A. (1999). Cloning and expression of a gene encoding a *Campoletis sonorensis* polydnavirus structural protein. *Archives of Insect Biochemistry and Physiology* **40**, 30–40.

Dib-Hajj, S. D., Webb, B. A. & Summers, M. D. (1993). Structure and evolutionary implications of a 'cysteine-rich' *Campoletis sonorensis* polydnavirus gene family. *Proceedings of the National Academy of Sciences, USA* **90**, 3765–3769.

Edson, K. M., Stoltz, D. B. & Vinson, S. B. (1981). Virus in a parasitoid wasp: suppression of the cellular immune response in the parasitoid's host. *Science* **211**, 582–583.

Feinberg, A. P. & Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**, 6–13.

Fleming, J. A. (1992). Polydnaviruses: mutualists and pathogens. *Annual Review of Entomology* **37**, 401–425.

Fleming, J. A. & Summers, M. D. (1986). *Campoletis sonorensis* endoparasitic wasps contain forms of *C. sonorensis* virus DNA suggestive of integrated and extrachromosomal polydnavirus DNAs. *Journal of Virology* **60**, 552–562.

Fleming, J. G. & Summers, M. D. (1991). Polydnavirus DNA is integrated in the DNA of its parasitoid wasp host. *Proceedings of the National Academy of Sciences, USA* **88**, 9770–9774.

Frohman, M. A. (1994). Cloning PCR Products. In *The Polymerase Chain Reaction*, pp. 14–37. Edited by K. B. Mullis, F. Ferre & R. A. Gibbs. Boston: Birkhauser.

Johner, A., Stettler, P., Gruber, A. & Lanzrein, B. (1999). Presence of polydnavirus transcripts in an egg–larval parasitoid and its lepidopterous host. *Journal of General Virology* **80**, 1847–1854.

Kozak, M. (1983). Comparison of initiation of protein synthesis in prokaryotes, eukaryotes, and organelles. *Microbiological Reviews* **47**, 1–45.

Krell, P. J., Summers, M. D. & Vinson, S. D. (1982). A virus with a multipartite superhelical DNA genome from the ichneumonid parasitoid *Campoletis* sonorensis. *Journal of Virology* **43**, 859–870.

Lavine, M. D. & Beckage, N. E. (1995). Polydnaviruses: potent mediators of host insect immune dysfunction. *Parasitology Today* **11**, 368–378.

Norton, W. N. & Vinson, S. B. (1983). Correlating the initiation of virus replication with a specific pupal developmental phase of an ichneumonid parasitoid. *Cell and Tissue Research* **231**, 387–398.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Shelby, K. S. & Webb, B. A. (1999). Polydnavirus-mediated suppression of insect immunity. *Journal of Insect Physiology* **45**, 507–514.

Stoltz, D. B. (1993). The polydnavirus life-cycle. In *Parasites and Pathogens of Insects*, 1st edn, vol. 1, pp. 167–187. Edited by N. E. Beckage, B. A. Federici & S. N. Thompson. San Diego, CA: Academic Press.

Stoltz, D. B., Guzo, D. & Cook, D. (1986). Studies on polydnavirus transmission. *Virology* **21**, 120–131.

Strand, M. R. & Pech, L. L. (1995). Immunological basis for compatibility in parasitoid–host relationships. *Annual Review of Entomology* **40**, 31–56.

Strand, M. R., McKenzie, D. I., Grassl, V., Dover, B. A. & Aiken, J. M. (1992). Persistence and expression of *Microplitis demolitor* polydnavirus in *Pseudoplusia includens*. *Journal of General Virology* **73**, 1627–1635.

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. *Journal of Virology* **61**, 2589–2598.

Theilmann, D. A. & Summers, M. D. (1988). Identification and comparison of *Campoletis sonorensis* virus transcripts expressed from four genomic segments in the insect hosts *Campoletis sonorensis* and *Heliothis virescens*. *Virology* **167**, 329–341.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for

multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**, 4876–4882.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research* **14**, 4683–4690.

Webb, B. A. & Summers, M. D. (1992). Stimulation of polydnavirus replication by 20-hydroxyecdysone. *Experientia* **48**, 1018–1022.

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 *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 253–260. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.

Received 10 January 2002; Accepted 11 June 2002