

Mapping of the *Heliothis armigera* entomopoxvirus (HaEPV) genome, and analysis of genes encoding the HaEPV spheroidin and nucleoside triphosphate phosphohydrolase I proteins

Alagacone Sriskantha, Rebecca J. Osborne and David J. Dall

CSIRO Division of Entomology, GPO Box 1700, Canberra, ACT 2601, Australia

The genome of *Heliothis armigera* entomopoxvirus (HaEPV) has been mapped with four restriction endonuclease enzymes (*Bam*HI, *Hind*III, *Pst*I and *Xho*I), and its length estimated at 233 kbp. An *Eco*RI-generated HaEPV genomic fragment hybridized to all fragments identified as genomic termini, providing the first experimental evidence for the presence of terminal repeat elements in an EPV genome. The HaEPV spheroidin and nucleoside triphosphate phosphohydrolase I (NPHI) genes have been cloned and sequenced, and their deduced products shown to possess high levels of identity with homologues from other Genus B entomopoxviruses (EPVs). The

genomic locations of these and other HaEPV genes and open reading frames have been determined; comparison of their locations with those of homologues in the *Amsacta moorei* EPV genome largely supports an hypothesis that the Genus B EPVs share a conserved genomic organization which differs from that of chordopoxviruses. It is proposed that genes of EPVs can be assigned to five actual or predicted homology-based groups, a categorization which is useful for directing and interpreting investigations of EPV gene functions and relationships.

Introduction

The family *Poxviridae* contains two subfamilies, the *Chordopoxvirinae* and *Entomopoxvirinae*, whose members infect vertebrates and insects, respectively. Three genera of entomopoxviruses (EPVs) are currently recognized (Genera A, B and C; Murphy *et al.*, 1995); the lepidopteran pathogen *Heliothis armigera* entomopoxvirus (HaEPV; Goodwin *et al.*, 1991; Fernon *et al.*, 1995) is a member of EPV Genus B, together with other caterpillar-infecting viruses such as the *Amsacta moorei*, *Choristoneura biennis* and *C. fumiferana* EPVs (AmEPV, CbEPV and CfEPV, respectively; Murphy *et al.*, 1995).

Synthesis and maturation of EPV virions involve, in their latter stages, incorporation of the developing virus particles into proteinaceous occlusion bodies, or spheroids, whose primary function is believed to be protection of the particles from environmental rigours such as desiccation and exposure

to UV radiation. Horizontal transmission of EPVs involves ingestion of spheroids, their solubilization in the alkaline and reducing insect gut milieu, and subsequent release of large numbers of infectious virions. This process is highly reminiscent of the one by which baculoviruses infect their insect hosts (see, e.g. Adams & McClintock, 1991), in which the virions are also occluded into, and subsequently liberated from, proteinaceous occlusion bodies known as polyhedra. The major matrix protein of spheroids (spheroidin) is a large (100–115 kDa), cysteine-rich protein which is unique to EPVs. Genomic sequences encoding homologues of this protein have been described from AmEPV, CbEPV and CfEPV (Hall & Moyer, 1991, 1993; Li *et al.*, 1997) and from the Genus A *Melolontha melolontha* EPV (MmEPV; Sanz *et al.*, 1994). Comparison of deduced amino acid sequences encoded by these genes has shown a close relationship between the proteins from AmEPV, CbEPV and CfEPV (e.g. 82% identity for AmEPV/CfEPV), and a more distant relationship (approximately 45%) between these and the homologue from MmEPV.

Sequences of other genes and open reading frames (ORFs) adjacent to spheroidin loci have also been reported from AmEPV, CbEPV and CfEPV (Hall & Moyer, 1991, 1993); in

Author for correspondence: David J. Dall.

Fax +61 26246 4173. e-mail davidd@ento.csiro.au

The GenBank accession numbers of sequences in this paper are AF017791 and AF019224.

each of those viruses an ORF encoding a homologue of the chordopoxvirus (ChPV) nucleoside triphosphate phosphohydrolase I protein (NPHI; Johnson *et al.*, 1993) has been found immediately downstream of, and in an opposite orientation to, the spheroidin gene. Comparison of the deduced amino acid sequences of NPHI proteins from AmEPV and CbEPV shows a high degree of conservation (88% identity; Hall & Moyer, 1993), while comparison of those proteins with their homologue from vaccinia virus shows a more distant relationship (37% identity).

Mapping of the locations of genes encoding the NPHI and thymidine kinase (TK) enzymes in the AmEPV genome (Hall & Moyer, 1991) showed that their relative genomic positions differed from those previously observed in ChPVs, and led both to the conclusion that the 'conserved genome core' common to various ChPVs (see, e.g. Gershon *et al.*, 1989; Senkevich *et al.*, 1996) was not present in EPVs, and to the hypothesis that EPVs might instead maintain a conserved genomic organization unique to the subfamily (Hall & Moyer, 1991). Examination of this hypothesis is clearly of considerable interest with respect to evaluation of evolutionary patterns within the *Poxviridae*, but has thus far been hindered by the lack of genome maps for EPVs other than AmEPV (Hall & Hink, 1990), and by the small number of identified genes/ORFs available for use in description of the pattern of EPV genomic organization.

In response to this situation, and in order both to better characterize HaEPV, and to assess its degree of relatedness to other members of the subfamily, we have mapped the HaEPV genome with several restriction endonuclease enzymes, cloned and sequenced the HaEPV spheroidin and NPHI genes, and compared the relative locations of these and other genes and ORFs in the genomes of HaEPV, AmEPV and other ChPVs.

Methods

■ **Virus and cells.** Mixed population HaEPV and clonal isolate HaEPV wt#2/011293 (WT2) were propagated *in vitro* in *Helicoverpa zea* or *Spodoptera frugiperda* cells as previously described (Feron *et al.*, 1995), and used as inoculum for *per os* infection of *Helicoverpa armigera* larvae. Infected caterpillars were maintained as previously described (Dall *et al.*, 1993), then collected and stored at 4 °C until use. HaEPV spheroids and spindle bodies were purified from cadavers, and occluded virions were released and purified on sucrose gradients as described (Dall *et al.*, 1993).

■ **Nucleic acid manipulation and characterization.** Methods for preparation of HaEPV genomic DNA have been described (Dall *et al.*, 1993). Genomic DNA from mixed population HaEPV was digested with the restriction endonucleases (RENs) *Bam*HI, *Bgl*II or *Eco*RI, shotgun cloned into appropriately prepared pTZ19R, and propagated and characterized using standard laboratory methods (e.g. Sambrook *et al.*, 1989). Genomic DNA from HaEPV isolate WT2 was digested with *Bam*HI, *Hind*III, *Pst*I or *Xho*I, either alone or in combination, and the resulting fragments were separated by electrophoresis in 0.4 or 1.5% agarose gels, then transferred to nylon membrane (Boehringer Mannheim) according to the manufacturer's alkali transfer methodology. Southern blotting protocols followed standard methods, and used ³²P-labelled random primed probes generated from shotgun-cloned genomic

DNA templates (see above), previously identified HaEPV genes (Dall *et al.*, 1993; Osborne *et al.*, 1996), newly identified HaEPV genes or related PCR products (see below), or isolated restriction fragments of genomic DNA excised and purified from agarose gels. Restriction fragments corresponding to genomic termini of WT2 were identified using a denaturation/'snap-back' protocol (Baroudy *et al.*, 1982).

Oligonucleotides identical to RM58 and RM118 (Hall & Moyer, 1991) were used with HaEPV genomic DNA in PCR amplification procedures (Osborne *et al.*, 1996) to generate a spheroidin-specific probe. After confirmation of the identity of the PCR product by sequencing, the amplicon was used in genomic mapping procedures as described above, and also used to identify fragments suitable for cloning from a panel of restriction digests of HaEPV WT2 genomic DNA. Adjoining 1.9 and 4.3 kb *Cla*I-*Xho*I fragments of HaEPV WT2 genomic DNA were identified and cloned into pBSKS(+), then sequenced in both directions using custom-synthesized oligonucleotides, a PCR-based methodology (using AmpliTaq FS, according to the manufacturer's instructions) and an ABI model 373 automated sequencer. Sequence data were analysed with the Wisconsin GCG package (Devereux *et al.*, 1984) and BLAST-based database searches.

Results

Mapping of the HaEPV genome

Fig. 1 shows the distribution of recognition sites of the RENs *Bam*HI, *Hind*III, *Pst*I and *Xho*I in the genome of HaEPV WT2. The genomic map of HaEPV is presented in an orientation which places the spheroidin locus (see below) in a similar relative position to that previously described for AmEPV (Hall & Moyer, 1991). Estimates of sizes of the genomic fragments identified in Fig. 1 are provided in Table 1; their summation estimates the size of the HaEPV genome as 233 kb. Termini of the HaEPV genome were initially identified by a denaturation/'snapback' renaturation procedure (Baroudy *et al.*, 1982); those results were supported by Southern hybridizations using random-labelled probe from a cloned HaEPV genomic fragment (*Eco*RI clone #12) which apparently represents a portion of an HaEPV terminal repeat element (see below). We identified a single instance of polymorphism in the WT2 isolate of HaEPV during the mapping process, that being the presence of a low-frequency variant containing an additional *Bam*HI site within the 45 kb *Bam*HI B fragment (not shown in Fig. 1 or Table 1), approximately 18 kbp from the leftwards end of the genome. We presume that the low observed frequency of this variant reflects occurrence of a mutation in the WT2 genome subsequent to its isolation.

Analysis of the HaEPV spheroidin/NPHI locus

We cloned and sequenced adjoining *Cla*I-*Xho*I and *Xho*I-*Cla*I fragments from the genome of HaEPV WT2, from a region at the rightward end of the *Hind*III D fragment, and extending into the F fragment (Fig. 1). When conjoined those sequences make up a 6210 nucleotide (nt) genomic fragment containing the gene encoding the HaEPV spheroidin protein, and (following the convention of using the *Hind*III map as the basis of poxvirus genome fragment nomenclature) gene F1L, which encodes an NPHI protein, and ORFs F2L and F3R, and

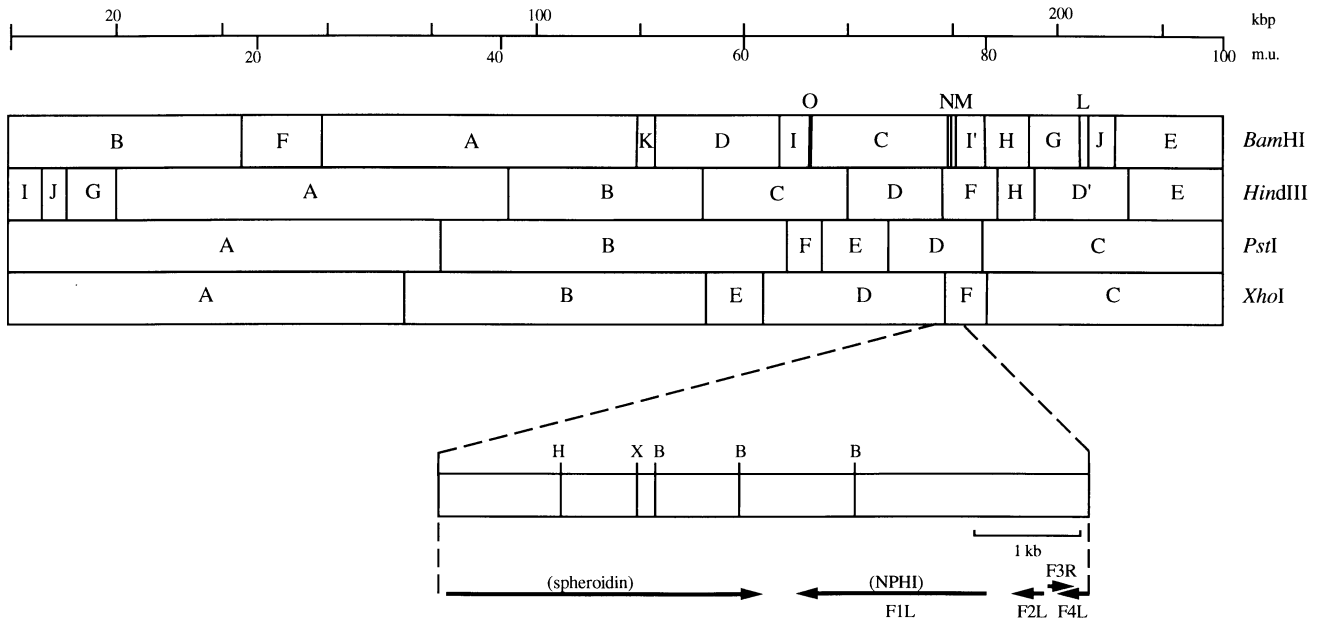


Fig. 1. Recognition sites of enzymes *Bam*HI, *Hind*III, *Pst*I and *Xho*I in the genome of HaEPV WT2, and fine mapping of a 6.2 kb segment from *Hind*III fragments D and F. Recognition sites of *Bam*HI (B), *Hind*III (H) and *Xho*I (X) in the latter are shown in relation to positions of the HaEPV spheroidin and NPHI (F1L) genes. Upper scale shows genome size (kbp) and genomic map units (m.u.).

Table 1. Designations and estimated sizes (kbp) of restricted fragments of the genome of HaEPV WT2.

Fragment	<i>Bam</i> HI	<i>Hind</i> III	<i>Pst</i> I	<i>Xho</i> I
A	61	76	82	75
B	45	37	67	58
C	26	28	44	43
D	24	18 (× 2)	18	36
E	20	17.5	12.5	11.5
F	15	11.0	6.5	7.0
G	9.8	9.5		
H	8.1	7.2		
I	5.8 (× 2)	7.0		
J	5.2	4.7		
K	3.4			
L	1.7			
M	1.0			
N	0.8			
O	0.3			
Total	233	234	230	231

a candidate ORF F4L. This sequence has been deposited in GenBank as accession AF019224. Several additional *Cla*I sites were observed to be present within the cloned fragments, suggesting that the digestion reaction from which they were isolated had not gone to completion.

The gene encoding the HaEPV spheroidin protein is 3024 nt in length (including the translation termination

Table 2. Relationships (% sequence identity; GCG Gap algorithm) of spheroidin and NPHI proteins of HaEPV with those of related viruses

(a) HaEPV and other EPV spheroidin proteins (Hall & Moyer, 1991; Sanz *et al.*, 1994; Li *et al.*, 1997); (b) NPHI proteins of HaEPV and other EPVs (Hall & Moyer, 1991, 1993; Yuen *et al.*, 1991) compared with African swine fever virus (ASFV; Freije *et al.*, 1993) and vaccinia, fowlpox and molluscum contagiosum chordopoxviruses (VV, Goebel *et al.*, 1990; FPV, Binns *et al.*, 1990; MCV, Senkevich *et al.*, 1996; respectively).

(a)

	spheroidin		
	MmEPV	CfEPV	AmEPV
HaEPV	44	79	82
AmEPV	45	77	
CfEPV	44		

(b)

	NPHI					
	ASFV	VV	FPV	MCV	CbEPV	AmEPV
HaEPV	23	37	39	35	82	84
AmEPV	24	37	37	34	89	
CbEPV	22	37	39	35		
MCV	20	62	61			
FPV	22	62				
VV	26					

codon); conceptual translation of the sequence describes a 1007 amino acid (aa) polypeptide with a deduced relative molecular mass (M_r) of 115 438. The protein contains 35 cysteine residues (3.5 mol%), but no unusual sequence motifs. Comparison of the deduced aa sequence with those in databases revealed its relationship to four other proteins,

namely, the spheroidin proteins of AmEPV, CbEPV, CfEPV and MmEPV (Hall & Moyer, 1991, 1993; Sanz *et al.*, 1994; Li *et al.*, 1997). Pairwise comparisons of complete spheroidin sequences using the GCG Gap algorithm revealed close relationships (of approximately 80% identity) between the protein from HaEPV and the other Genus B viruses [AmEPV and CfEPV; Table 2(a)], but more distant relationships between those proteins and the homologue from the Genus A MmEPV (about 45% identity). Alignment of the four homologues (GCG Gap algorithm, default settings; Fig. 2) further emphasized the degree of similarity between the Genus B-derived polypeptides; whereas long continuous regions of homology were obvious in the polypeptides encoded by those viruses (overlined), their comparison with the MmEPV homologue revealed the existence of several points of relative sequence insertion and deletion, and other regions (dashed underline) where there was little or no conservation between sequences from Genus A and B viruses. Comparison of nt sequences (PileUp algorithm, default settings) immediately 5' of the HaEPV, AmEPV and CfEPV spheroidin genes also revealed a high degree of conservation in this 'upstream' region, with 51 of the 56 nt immediately preceding the translation initiation codons being identical in all three sequences. The HaEPV sequences thus included the tandem consensus early gene transcription termination signals (T_5 NT) and multiple translation termination codons described from AmEPV by Hall & Moyer (1991), and the consensus late promoter motif TAA-ATG at (and including) the HaEPV spheroidin gene translation initiation codon.

Gene F1L, which is predicted by sequence homology to encode an HaEPV NPHI protein, was found to be separated from the spheroidin gene by a 321 bp intergenic region, and to code in an opposite orientation. Nucleotide sequence 5' to this gene (a 117 bp A-T rich intergenic region separating F1L and F2L; Fig. 1) also contained the TAAATG motif at (and including) the predicted translation initiation codon, immediately preceded by the T_5 NT motif described above. The F1L gene is 1941 nt in length (including the translation termination codon) and encodes a 646 aa polypeptide with a deduced M_r of 75 763. Comparison of the deduced HaEPV polypeptide sequence with those in databases revealed relationships with NPHI enzymes of AmEPV, CbEPV, and vaccinia, variola, fowlpox, molluscum contagiosum and African swine fever viruses [for which complete sequences were available; see Table 2(b)], and malignant rabbit fibroma and swinepox viruses (for which only incomplete sequences were available; Strayer *et al.*, 1991; Massung *et al.*, 1993). Pairwise comparisons of the HaEPV and other complete sequences (excluding that of variola virus, which is 98.9% identical to that of vaccinia virus; Massung *et al.*, 1994) revealed similar and close levels of relationship between those from the Ha-, Am- and CbEPVs (82–89% identity), and more distant relationships between those and homologues from the other viruses [Table 2(b)]. Alignment of these sequences with the PileUp algorithm (not

shown) produced an outcome essentially identical to that previously demonstrated with the AmEPV, CbEPV and vaccinia virus homologues (see Fig. 8 in Arif, 1995); the HaEPV aa sequence proved identical to that of AmEPV through the first of two domains previously noted as being conserved in ATP-binding proteins (Walker *et al.*, 1982; Broyles & Moss, 1987; Arif, 1995), and differed from the Am-/CbEPV consensus at only three of 29 residues in the second domain. In addition, the HaEPV sequence differed from the Am-/CbEPV consensus at only two of 36 residues in a third C-terminal region previously noted as being highly conserved (Arif, 1995).

The HaEPV F2L ORF was observed to potentially encode a 101 aa product with a deduced M_r of 11 567, and possess a late poxvirus consensus promoter sequence at its translation initiation codon. The deduced F2L polypeptide was related only to the deduced product of a previously undescribed and incomplete ORF located in an homologous position adjacent to the CbEPV NPHI gene (57% identity over their 56 C-terminal residues; Yuen *et al.*, 1991). Sequence from the same location of the AmEPV genome (Hall & Moyer, 1993) was unrelated to that of the deduced HaEPV and CbEPV sequences. Two other ORFs, F3R and F4L, potentially encoding products of 61 and > 85 aa respectively, were also recognized (Fig. 1); neither putative product showed homology to any known protein.

Analysis of a repeated near-terminal HaEPV genomic element

We sequenced a 972 bp *Eco*RI-generated fragment isolated from a preparation of mixed clone HaEPV genomic DNA. This fragment hybridized to all fragments of the HaEPV WT2 genome identified as genomic termini (see above), and is thus considered to contain part of a terminal repeat (TR) element of the virus. Sequence data from this fragment (GenBank accession AF017791) showed that 521 bp (at the 5' end of the sequence, as submitted) comprised a non-coding or intergenic region, and that most of the remaining sequence comprised a 447 bp ORF potentially encoding a 148 aa polypeptide with a deduced M_r of 17 522. Southern hybridizations using random-labelled probe from a PCR-generated template corresponding to the ORF sequence alone showed that sequence corresponding to this '17K' ORF was present at both genomic termini. Comparison of the deduced aa sequence of the 17K ORF with those in databases revealed homologies with several ChPV genes/ORFs from ITR and near-terminal genomic locations of vaccinia (strains WR and IHD-W), fowlpox, Shope fibroma and ectromelia viruses (Kotwal & Moss, 1988; Tomley *et al.*, 1988; Upton *et al.*, 1994; Senkevich *et al.*, 1995). Pairwise comparisons (GAP algorithm, default settings) of the HaEPV 17K sequence with homologues from fowlpox virus (FPV ORF 6; Tomley *et al.*, 1988) and vaccinia virus strain WR (21.7K gene; Kotwal & Moss, 1988), showed identities of 31 and

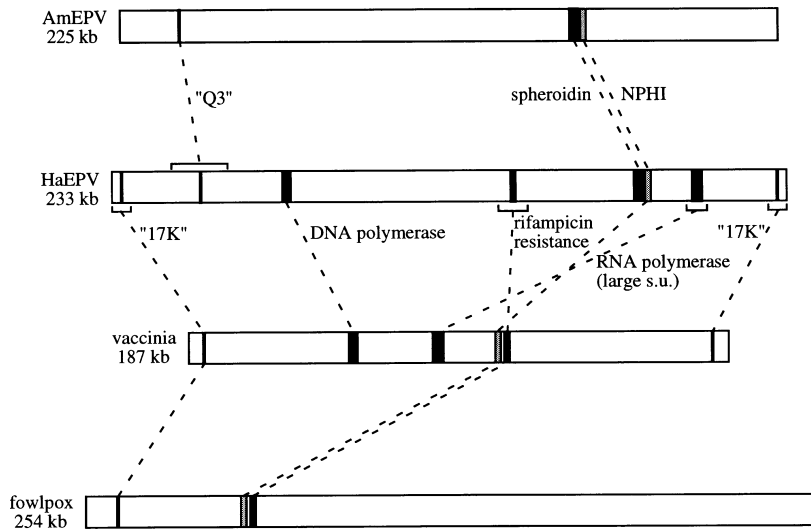


Fig. 3. Comparison of genomic positions of homologous loci in HaEPV, AmEPV and vaccinia and fowlpox viruses. Homologous loci are named and linked by vertical lines; horizontal bars denote limits of mapping accuracy currently possible for HaEPV. See text for further discussion.

24%, respectively. Alignment of the deduced aa sequence of HaEPV 17K with those homologues (PileUp algorithm) showed that a relatively small number of residues was conserved in all three sequences; of interest was the observation that six of those 14 residues were amino acids with bulky aromatic side-chains (i.e. tyrosine and tryptophan). Nevertheless, the significance of this observation, if any, remains to be determined. Other homologues (see above) of the HaEPV 17K ORF, FPV ORF6 and the VV WR 21·7K gene have additional C-terminal extension sequences which give both a greater M_r (approximately 26–33 000), and, in some instances, contain a zinc-finger motif. The deduced aa sequence of the HaEPV 17K protein does not, however, contain this or any other unusual motif.

Random-labelled probes prepared from cloned HaEPV WT2 *Hind*III fragment J (Fig. 1, Table 1) did not hybridize to genomic fragments from the rightwards end of the genome, allowing us to conclude that the leftwards HaEPV TR is confined to the *Hind*III fragment I, and that it is thus no larger than 7 kb.

Comparative alignment of HaEPV and other poxvirus genomic loci

As shown in Fig. 3, we have determined the genomic positions of HaEPV genes encoding the 'rifampicin resistance factor' protein (Osborne *et al.*, 1996), the spheroidin and NPHI genes, and the 17K ORF (see above) within accuracy limits permitted by the genome map presented here (Fig. 1). The locations of genes encoding the large subunit of HaEPV RNA polymerase, and an ORF homologous to AmEPV Q3 (Gruidl *et al.*, 1992), have also been determined (Fig. 3; A. Sriskantha, R. J. Osborne & D. J. Dall, unpublished data). Fig. 3 also compares, where possible, the positions of these loci with those of homologues in AmEPV, vaccinia and fowlpox viruses (Hall & Moyer, 1991; Kotwal & Moss, 1988 *a*; Johnson *et al.*,

1993; Mockett *et al.*, 1992). We have observed that where comparisons are possible there is a relatively good correspondence between the locations of homologous genes in the two EPV genomes, but that this syntony apparently does not extend to either of the two ChPVs examined here.

Discussion

Data presented here contribute to an emerging definition of the molecular characteristics of Genus B EPVs. Thus, we estimate the size of the HaEPV genome as 233 kb, a figure close to the estimated size of the genome of AmEPV (225 kb; Hall & Hink, 1990), which is both the type member of the genus, and the only other member of the subfamily for which a genomic map has previously been available.

Availability of a genomic map of HaEPV has enabled us to demonstrate that the HaEPV spheroidin/NPHI locus is in a broadly similar genomic location to that previously identified for the locus in AmEPV [about 77 and 68 map units (m.u.), respectively; Hall & Hink, 1990; Hall & Moyer, 1991], as is an HaEPV homologue of the AmEPV 'Q3' ORF (located between 9 and 17 m.u. in HaEPV, and at about 9 m.u. in AmEPV; Gruidl *et al.*, 1992; Sriskantha & Dall, unpublished data). Cloning and sequencing of the former locus has also shown that the relative order and orientations of the spheroidin and NPHI genes of HaEPV are conserved with respect to those in AmEPV, CfEPV, and CbEPV (Hall & Moyer, 1991, 1993; Li *et al.*, 1997). In addition, we have observed that HaEPV and CbEPV possess another homologous ORF (HaEPV F2L) at the 5' (rightwards) end of the NPHI gene (Yuen *et al.*, 1991). While this ORF is not present in the same location in the AmEPV genome (Hall & Moyer, 1993), we have also found that the ORF immediately upstream of the HaEPV spheroidin locus is closely related to the G4R ORF present in the analogous location in the AmEPV and CfEPV genomes (Hall & Moyer, 1991; Li *et al.*, 1997; A. Sriskantha & D. J. Dall, unpublished data). Collectively, these

data support the hypothesis (Hall & Moyer, 1993) that Genus B EPVs (at least) share a largely conserved genomic organization.

Hall & Moyer (1991) also noted that the arrangement of genes in the AmEPV genome appeared distinct from the essentially collinear gene arrangement common to chordopoxviruses (ChPVs) such as the ortho- and capripoxviruses, and molluscum contagiosum virus (Gershon *et al.*, 1989; Senkevich *et al.*, 1996), and which is also largely conserved in the rearranged gene blocks of avipoxvirus genomes (Mockett *et al.*, 1992). We have shown that the NPHI and 'rifampicin resistance factor' loci, which constitute vaccinia virus genes D11L and D13L, respectively, and which are present in an identical arrangement in fowlpoxvirus (Binns *et al.*, 1990), are widely separated in the HaEPV genome (see Fig. 3), adding support to the hypothesis that the pattern of gene organization within the genome of Genus B EPVs is distinct from that in ChPVs.

Comparison of the deduced sequences of proteins encoded by the HaEPV NPHI, spheroidin and 17K loci with those of other poxvirus homologues clearly differentiated the Genus B viruses both from a Genus A EPV (MmEPV), and various ChPVs. Thus, while the deduced amino acid sequences of NPHI proteins from the Genus B EPVs showed 82–89% identity to each other [Table 2(b)], comparison of those sequences with homologues from members of three ChPV genera showed identities of only 34–39%. Similarly, deduced amino acid sequences of spheroidin proteins from Genus B EPVs showed 77–79% identity with each other, but shared only some 45% identity with that of the Genus A MmEPV [Table 2(a)].

Our observation that common sequences are present in both HaEPV genomic termini (as judged by Southern hybridization using the 17K ORF) is consistent with the presence of inverted terminal repeat elements (ITRs) in the HaEPV genome. Although ITRs are likely to be present in all poxvirus genomes, this work provides the first direct experimental support for their presence in a member of the EPV subfamily. Furthermore, the location of the 17K ORF in the putative HaEPV ITRs is consistent with the observed location of vaccinia virus homologues at the distal ends of the ITRs of strain WR (Kotwal & Moss, 1988a). ITR elements of ChPV genomes vary widely in size, ranging from less than 1 kb in variola virus (Massung *et al.*, 1994) to more than 12 kb in the closely related vaccinia virus (Goebel *et al.*, 1990); our estimate of a maximum size of 7 kb for the ITRs of HaEPV is thus consistent with other observations from the family. The function(s) of the putative products of poxvirus 17K ORF homologues have not been determined either in this instance, or for any other poxvirus. The fact that the gene is absent from the genome of the Copenhagen strain of vaccinia virus (Goebel *et al.*, 1990) shows that its function is not essential to poxvirus replication; there is, however, evidence that although an ectromelia virus homologue (which possesses a zinc-finger motif lacking from

the putative HaEPV product) is non-essential for growth, expression of its product is a determinant of both the tissue tropism of the virus, and its *in vivo* pathogenicity (Senkevich *et al.*, 1994, 1995).

Consideration of EPV genes and ORFs described here, or available elsewhere, has led us to conclude that the EPV genome can usefully be regarded as being composed of a small number of 'classes' of genes, whose delineations rest on relationships with genes of other viruses. Thus we consider that known EPV genes/ORFs (as described in Arif, 1995, unless otherwise stipulated) form four such classes, namely, those that are

- (1) unique to EPVs, including those that encode spheroidin and FALPE proteins (Alaoui-Ismaïli & Richardson, 1996), and many other putative products of unknown function;
- (2) unique to poxviruses, but common to EPVs and ChPVs, including those that encode proteins associated with virion structure/formation, such as the HaEPV 'rifampicin resistance factor' and major virion core proteins (Osborne *et al.*, 1996; R. Crnov & D. J. Dall, unpublished data), and other proteins such as HaEPV poly(A) polymerase (R. Crnov & D. J. Dall, unpublished data), whose roles reflect localization of the poxvirus replicative cycle to the host-cell cytoplasm;
- (3) common to EPVs and other unrelated viruses that also infect insects, including those that encode the fusolin/p34.8 proteins common to EPVs and nuclear polyhedrosis baculoviruses (NPVs), and the HaEPV 11K and 57K proteins which have homologues in NPV and granulosis virus genomes, respectively (R. J. Osborne, R. Crnov & D. J. Dall, unpublished data);
- (4) common to unrelated viruses from a wide variety of hosts, including those that encode 'housekeeping' and salvage pathway enzymes such as nucleoside kinases (e.g. NPHI), and DNA polymerases.

Given that ChPVs have apparently and repeatedly demonstrated a capacity to acquire gene sequences from their hosts (see, e.g. Kotwal & Moss, 1988b; Hu *et al.*, 1994; Lyttle *et al.*, 1994; Senkevich *et al.*, 1996), we assume that a fifth class of EPV genes – containing viral homologues of insect host genes – will ultimately also be recognized; given the relatively small data-set currently available for both the viruses and their insect hosts, it is not surprising that no such EPV gene has yet been identified.

Description of the EPV genome in this manner recognizes the common evolutionary origin of EPVs and other poxviruses, as well as the existence of identical or functionally equivalent features in both EPVs and the unrelated baculovirus group. Detailed comparisons of EPV genes from the second of the classes described above (i.e. the 'poxvirus genes') with their ChPV homologues is likely to enhance understanding of evolutionary patterns and comparative phylogenetic distances within the *Poxviridae*. In contrast, it is clear that EPVs and baculoviruses are related only through utilization of common host taxa, and, as a corollary, that genes common to the two

groups (i.e. those in the third class described above) likely encode proteins of adaptive significance which have been evolved or acquired in response to common evolutionary pressures. Further investigation of the functions of such proteins might then be expected to assist in the elucidation of factors that are crucial to the effective utilization of insect hosts by viral pathogens.

We thank Dr Basil Arif for making sequence data available to us prior to publication.

References

- Adams, J. R. & McClintock, J. T. (1991). Nuclear polyhedrosis viruses of insects. In *Atlas of Invertebrate Viruses*, pp. 87–204. Edited by J. R. Adams & J. R. Bonami. Boca Raton: CRC Press.
- Alaoui-Ismali, M. H. & Richardson, C. D. (1996). Identification and characterisation of a filament-associated protein encoded by *Amsacta moorei* entomopoxvirus. *Journal of Virology* **70**, 2697–2705.
- Arif, B. M. (1995). Recent advances in the molecular biology of entomopoxviruses. *Journal of General Virology* **76**, 1–13.
- Baroudy, B. M., Venkatessan, S. & Moss, B. (1982). Incomplete base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. *Cell* **28**, 315–324.
- Binns, M. M., Britton, B. S., Mason, C. & Bournsnel, M. E. G. (1990). Analysis of the fowlpox genome region corresponding to the vaccinia virus D6 to A1 region: location of, and variation in non-essential genes in poxviruses. *Journal of General Virology* **71**, 2873–2881.
- Broyles, S. S. & Moss, B. (1987). Identification of the vaccinia virus gene encoding nucleoside triphosphate phosphohydrolase I, a DNA-dependent ATPase. *Journal of Virology* **61**, 1738–1742.
- Dall, D. J., Sriskantha, A., Vera, A. P., Lai-Fook, J. & Symonds, T. M. (1993). A gene encoding a highly expressed spindle body protein of *Heliothis armigera* entomopoxvirus. *Journal of General Virology* **74**, 1811–1818.
- Devereux, J., Haeblerli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the Vax. *Nucleic Acids Research* **12**, 387–395.
- Fernon, C. A., Vera, A. P., Crnov, R., Lai-Fook, J., Osborne, R. J. & Dall, D. J. (1995). Replication of *Heliothis armigera* entomopoxvirus *in vitro*. *Journal of Invertebrate Pathology* **66**, 216–223.
- Gershon, P. D., Ansell, D. M. & Black, D. N. (1989). A comparison of the genome organisation of capripoxvirus with that of the orthopoxviruses. *Journal of Virology* **63**, 4703–4708.
- Goebel, S. J., Johnson, G. P., Perkus, M. E., Davis, S. W., Winslow, J. P. & Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. *Virology* **179**, 247–266.
- Goodwin, R. H., Milner, R. J. & Beaton, C. D. (1991). Entomopoxvirinae. In *Atlas of Invertebrate Viruses*, pp. 393–434. Edited by J. R. Adams & J. R. Bonami. Boca Raton: CRC Press.
- Gruidl, M. E., Hall, R. L. & Moyer, R. W. (1992). Mapping and molecular characterization of a functional thymidine kinase from *Amsacta moorei* entomopoxvirus. *Virology* **186**, 507–516.
- Hall, R. L. & Hink, W. F. (1990). Physical mapping and field inversion gel electrophoresis of *Amsacta moorei* entomopoxvirus DNA. *Archives of Virology* **110**, 77–90.
- Hall, R. L. & Moyer, R. W. (1991). Identification, cloning, and sequencing of a fragment of *Amsacta moorei* entomopoxvirus DNA containing the spheroidin gene and three vaccinia virus-related open reading frames. *Journal of Virology* **65**, 6516–6527.
- Hall, R. L. & Moyer, R. W. (1993). Identification of an *Amsacta* spheroidin-like protein within the occlusion bodies of *Choristoneura* entomopoxviruses. *Virology* **192**, 179–187.
- Hu, F., Smith, C. A. & Pickup, D. J. (1994). Cowpox virus contains two copies of an early gene encoding a soluble secreted form of the Type II TNF receptor. *Virology* **204**, 343–356.
- Johnson, G. P., Goebel, S. J. & Paoletti, E. (1993). An update on the vaccinia virus genome. *Virology* **196**, 381–401.
- Kotwal, G. J. & Moss, B. (1988a). Analysis of a large cluster of non-essential genes deleted from a vaccinia virus terminal transposition mutant. *Virology* **167**, 524–537.
- Kotwal, G. J. & Moss, B. (1988b). Vaccinia virus encodes a secretory polypeptide structurally related to complement control proteins. *Nature* **335**, 176–178.
- Li, X., Barrett, J. W., Yuen, L. & Arif, B. M. (1997). Cloning, sequencing and transcriptional analysis of the *Choristoneura fumiferana* entomopoxvirus spheroidin gene. *Virus Research* **47**, 143–154.
- Lyttle, D. J., Fraser, K. M., Fleming, S. B., Mercer, A. A. & Robinson, A. J. (1994). Homologs of vascular endothelial growth factor are encoded by the poxvirus orf virus. *Journal of Virology* **68**, 84–92.
- Massung, R. F., Jayarama, V. & Moyer, R. W. (1993). DNA sequence analysis of conserved and unique regions of swinepox virus: identification of genetic elements supporting phenotypic observations including a novel G protein receptor-coupled homologue. *Virology* **197**, 511–528.
- Massung, R. F., Liu, L. I., Qi, J., Knight, J. C., Yuran, T. E., Kerlavage, A. R., Parsons, J. M., Venter, J. C. & Esposito, J. J. (1994). Analysis of the complete genome of smallpox variola major virus strain Bangladesh-1975. *Virology* **201**, 215–240.
- Mockett, B., Binns, M. M., Bournsnel, M. E. G. & Skinner, M. A. (1992). Comparisons of the locations of homologous fowlpox and vaccinia virus genes reveals major genome reorganization. *Journal of General Virology* **73**, 2661–2668.
- Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A. & Summers, M. D. (editors) (1995). *Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses*. Vienna & New York: Springer-Verlag.
- Osborne, R. J., Symonds, T. M., Sriskantha, A., Lai-Fook, J., Fernon, C. A. & Dall, D. J. (1996). An entomopoxvirus homologue of the vaccinia virus D13L-encoded 'rifampicin resistance' protein. *Journal of General Virology* **77**, 839–846.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sanz, P., Veyrunes, J. C., Cousserans, F. & Bergoin, M. (1994). Cloning and sequencing of the spherulin gene, the occlusion body major polypeptide of the *Melolontha melolontha* entomopoxvirus (MmEPV). *Virology* **202**, 449–457.
- Senkevich, T. G., Koonin, E. V. & Buller, R. M. L. (1994). A poxvirus protein with a RING zinc finger motif is of crucial importance for virulence. *Virology* **198**, 118–128.
- Senkevich, T. G., Wolffe, E. J. & Buller, R. M. L. (1995). Ectromelia virus RING finger protein is localized in virus factories and is required for virus replication in macrophages. *Journal of Virology* **69**, 4103–4111.
- Senkevich, T. G., Bugert, J. J., Sisler, J. R., Koonin, E. V., Darai, G. & Moss, B. (1996). Genome sequence of a human tumorigenic poxvirus: prediction of specific host response-evasion genes. *Science* **273**, 813–816.

Strayer, D. S., Jerng, H. H. & O'Connor, K. (1991). Sequence and analysis of a portion of the genomes of Shope fibroma virus and malignant rabbit fibroma virus that is important for viral replication in lymphocytes. *Virology* **185**, 585–595.

Tomley, F., Binns, M., Campbell, J. & Bournnell, M. (1988). Sequence analysis of an 11.2 kilobase, near-terminal, *Bam*HI fragment of fowlpox virus. *Journal of General Virology* **69**, 1025–1040.

Upton, C., Schiff, L., Rice, S. A., Dowdeswell, T., Yang, X. & McFadden, G. (1994). A poxvirus protein with a RING finger motif binds zinc and localizes in virus factories. *Journal of Virology* **68**, 4186–4195.

Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). Distantly related sequences in the α - and β -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO Journal* **1**, 945–951.

Yuen, L., Noiseux, M. & Gomes, M. (1991). DNA sequence of the nucleoside triphosphate phosphohydrolase I (NPH I) of the *Choristoneura biennis* entomopoxvirus. *Virology* **182**, 403–406.

Received 21 March 1997; Accepted 21 August 1997