

Genome of Epinotia aporema granulovirus (EpapGV), a polyorganotropic fast killing betabaculovirus with a novel thymidylate kinase gene.

María Ferrelli, Ricardo Salvador, Marina Biedma, Marcelo Berretta, Santiago Haase, Alicia Sciocco-Cap, Pablo Ghiringhelli, Víctor Romanowski

► **To cite this version:**

María Ferrelli, Ricardo Salvador, Marina Biedma, Marcelo Berretta, Santiago Haase, et al.. Genome of Epinotia aporema granulovirus (EpapGV), a polyorganotropic fast killing betabaculovirus with a novel thymidylate kinase gene.. BMC Genomics, BioMed Central, 2012, 13 (1), pp.548. 10.1186/1471-2164-13-548 . inserm-00751409

HAL Id: inserm-00751409

<https://www.hal.inserm.fr/inserm-00751409>

Submitted on 13 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

Open Access

Genome of *Epinotia aporema* granulovirus (EpapGV), a polyorganotropic fast killing betabaculovirus with a novel thymidylate kinase gene

María Leticia Ferrelli¹, Ricardo Salvador^{1,2}, Marina Elizabeth Biedma^{1,4}, Marcelo Facundo Berretta², Santiago Haase¹, Alicia Sciocco-Cap², Pablo Daniel Ghiringhelli³ and Víctor Romanowski^{1*}

Abstract

Background: *Epinotia aporema* (Lepidoptera: Tortricidae) is an important pest of legume crops in South America. *Epinotia aporema* granulovirus (EpapGV) is a baculovirus that causes a polyorganotropic infection in the host larva. Its high pathogenicity and host specificity make EpapGV an excellent candidate to be used as a biological control agent.

Results: The genome of *Epinotia aporema* granulovirus (EpapGV) was sequenced and analyzed. Its circular double-stranded DNA genome is 119,082 bp in length and codes for 133 putative genes. It contains the 31 baculovirus core genes and a set of 19 genes that are GV exclusive. Seventeen ORFs were unique to EpapGV in comparison with other baculoviruses. Of these, 16 found no homologues in GenBank, and one encoded a thymidylate kinase. Analysis of nucleotide sequence repeats revealed the presence of 16 homologous regions (*hrs*) interspersed throughout the genome. Each *hr* was characterized by the presence of 1 to 3 clustered imperfect palindromes which are similar to previously described palindromes of tortricid-specific GVs. Also, one of the *hrs* (*hr4*) has flanking sequences suggestive of a putative non-*hr* ori. Interestingly, two more complex *hrs* were found in opposite loci, dividing the circular dsDNA genome in two halves. Gene synteny maps showed the great colinearity of sequenced GVs, being EpapGV the most dissimilar as it has a 20 kb-long gene block inversion. Phylogenetic study performed with 31 core genes of 58 baculoviral genomes suggests that EpapGV is the baculovirus isolate closest to the putative common ancestor of tortricid specific betabaculoviruses.

Conclusions: This study, along with previous characterization of EpapGV infection, is useful for the better understanding of the pathology caused by this virus and its potential utilization as a bioinsecticide.

Background

Baculoviruses (family *Baculoviridae*) are rod-shaped, enveloped, insect-specific viruses with double-stranded, circular DNA genomes ranging in size from 80 to 180 kb [1]. The family *Baculoviridae* is subdivided into four genera: *Alphabaculovirus* (lepidopteran-specific nucleopolyhedrovirus, NPVs), *Betabaculovirus* (lepidopteran-specific granulovirus, GVs), *Gammabaculovirus* (hymenopteran-specific NPVs) and *Deltabaculovirus*

(dipteran-specific NPV) [2,3]. GVs have been isolated only from insects belonging to the order Lepidoptera and are classified in three groups according to the pathology caused in their insect hosts. Type 1 pathology is characterized by an infection limited to the host's midgut and fat body resulting in a relatively slow speed of kill. Type 2 pathology is characterized by infection of most of the host's tissues and a rapid speed of kill. There is a third pathology with a single representative, the *Harrisina brillians* granulovirus, that causes an infection constrained to the midgut epithelium that results in the rapid death of the host [4].

* Correspondence: victor@biol.unlp.edu.ar

¹Instituto de Biotecnología y Biología Molecular, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, CONICET, La Plata, Argentina
Full list of author information is available at the end of the article

A highly pathogenic granulovirus was isolated from a larva of the “bean shoot borer” *Epinotia aporema* (Lepidoptera: Tortricidae), one of the major soybean pests in Argentina, and characterized at biological and molecular levels [5]. Further characterization of its pathology demonstrated that this virus belongs to the type 2 GVs meaning that the infection caused by EpapGV in its host is polyorganotropic [6]. All this information has been instrumental to formally propose its use as a microbial control agent with great potential. In order to contribute to a more thorough characterization of EpapGV we set out to determine and analyze its complete genome sequence.

To date, close to 60 baculovirus genomes have been fully sequenced, 12 of them belong to the *Betabaculovirus* genus. Completely sequenced GVs are listed in Table 1 and their pathology types are indicated. In this report, we present the complete sequence and organization of the EpapGV genome and compare them to other baculoviruses using genomic and phylogenetic analyses.

Results and Discussion

General characteristics of the EpapGV genome

The complete EpapGV genome [GenBank: JN408834] was covered 34 times by 454 sequencing. It consists of 119,082 bp in good agreement with the previous estimate of 120.1 kbp based on restriction mapping [19]. Betabaculoviruses have AT-rich genomes ranging between 54.7% (CpGV) and 67.6% (CrleGV). The AT content of EpapGV genome is 58.5%. However, no correlation between these data and biological properties has been found thus far.

Analysis of the EpapGV genome sequence led to the identification of 133 putative protein coding genes. The search was restricted to open reading frames starting with a methionine codon, coding for polypeptides of at least 50 amino acid residues (aa) and minimal overlapping of adjacent ORFs. This information comprises 90.94% of the nucleotide sequence (Additional File 1). The adenine of the *granulin* start codon was designated nucleotide 1 and the sequence was numbered in the direction of *granulin* gene transcription, which defined the clockwise orientation of the circular genome map [20]. The putative ORFs were numbered sequentially in this orientation. Seventy-two ORFs were in the same orientation as the *granulin* ORF, and sixty-one, in the opposite. EpapGV DNA sequence was searched for promoter motifs 150 bp upstream of the starting codon of each ORF. Early promoter motifs including TATA box (TATAWAW, TATAWTW, TATAW) in conjunction with CAKT initiator sequence (INR) [21] were found in the upstream regions of 26 ORFs; 64 ORFs had a late INR motif DTAAG [22] and 11 ORFs had both early and late elements.

Gene content

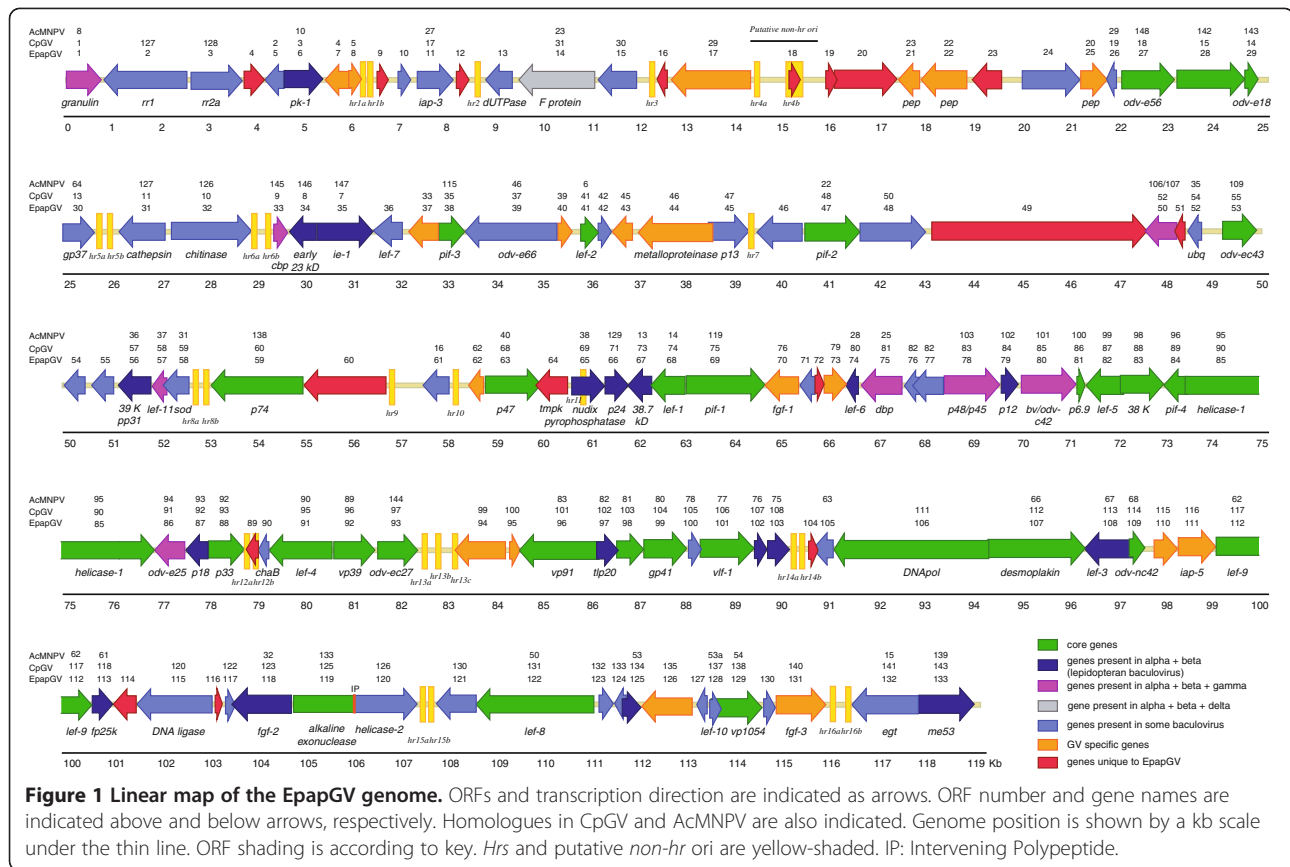
The EpapGV genome contains the 31 core genes present in all baculoviruses. The genes were also classified according to their presence in different genera [23,24] (Figure 1).

A distinct feature of the EpapGV genome is that the core gene *alkaline nuclease* (*alk-exo*, *epap119*) is fused in frame with the *helicase-2* ORF (*epap120*). This fusion gene codes for an 886 aa polypeptide with the first 383

Table 1 Completely sequenced Betabaculovirus

Virus	Genome size (bp)	Accession number	Annotated ORFs	Average % id with EpapGV	Host family	Pathology type	Reference
EpapGV	119,082	JN408834	133	-	Tortricidae	II	[6], this work.
AdorGV	99,657	AF547984	119	41.02	Tortricidae	I	[7,8]
AgseGV	131,680	AY522332	132	44.12	Noctuidae	II	Xiulian <i>et al.</i> , 2004, unpublished
ChocGV	104,710	DQ333351	116	44.39	Tortricidae	nr	[9]
CrleGV	110,907	AY229987	128	44.25	Tortricidae	II	[10]
CpGV	123,500	U53466	143	44.16	Tortricidae	II	[11]
HearGV	169,794	EU255577	179	39.71	Noctuidae	I	[12]
PhopGV	119,217	AF499596	130	42.55	Gelechiidae	II	Crozier <i>et al.</i> , 2002, unpublished
PiraGV	108,592	NC_013797	120	44.62	Pieridae	nr	[13]
PlxyGV	100,999	AF270937	120	41.01	Plutellidae	II	[14]
PsunGV	176,677	EU678671	183	40.17	Noctuidae	I	Li <i>et al.</i> , 2008, unpublished
SpliGV	124,121	DQ288858	136	40.96	Noctuidae	nr	[15,16]
XcenGV	178,733	AF162221	181	39.84	Noctuidae	I	[17]
ClanGV*	101,487	NC_015398	123	nd	Notodontidae	nr	[18]

When information is available, GV type pathology is indicated. nr: not reported. (*) The genome of *Clostera anachoreta* GV was published [18] after the present work was completed; it was not included in this analysis.



residues homologous to Alk-Exo and the last 456 to Helicase-2. A 47 aa-long intervening polypeptide of unknown origin and without significant sequence similarity to any protein in GenBank was found between the Alk-Exo and the Helicase-2 regions. The intervening polypeptide may act as a low-structure linker between Alk-Exo and Helicase-2 such that both enzyme domains could fold as if they were independent polypeptides retaining their respective functions (Figure 2 and Additional files 2 and 3). Although this region was confirmed by resequencing, it will be important to study this genomic region in alternative isolates of EpapGV and determine the transcription and translation products in infected larvae. All GVs contain these two genes in the same order, but there are no reports of a fusion. Fusion genes seem to be extremely rare in baculovirus genomes, but there is one report of fused genes encoded by *Spodoptera litura* NPV: the *ubiquitin* ORF is fused in frame with *gp37* and the fusion protein is proteolytically processed [25].

For two EpapGV ORFs (*epap10* and *epap130*) the BlastP search found homologues in only one member of the *Baculoviridae*. *Epap10* is preceded by early and late promoter motifs and codes for a 90 aa protein that shares 34% amino acid identity with a 88 aa protein encoded by *epo28* of *Epiphyas postvittana* NPV. This

ORF was reported to be unique to EppoNPV and has an early promoter motif [26]. *Epap130* codes for a 77 aa protein that matched a 56 aa protein of *Spodoptera litura* GV (*spli32*) with 38% sequence identity according to ClustalW alignment.

ORFs 10, 46, 54, 55 and 105 were found to have homologues in alphabaculovirus genomes but not in other betabaculoviruses. *Epap46* is a 306 aa long protein that seemed to be homologous to *Spodoptera exigua* MNPV ORF 30 by BlastP search ($E = 0.07$) although they have very low amino acid identity (12%). *Epap54* (148 aa) and *epap55* (157 aa) are both similar to ORF 3 of *Adoxophyes honmai* NPV (AdhoNPV) and ORF 3 of *Adoxophyes orana* NPV (AdorNPV). *Epap55* shares 32% identity with the N-terminal portion of AdorNPV ORF 3 and *Epap54* is homologous to the C-terminal region of ORF 3 of AdorNPV (34%) and AdhoNPV (34%). GV homologues of *epap54* and *epap55* were found only in AgseGV (Additional File 1). *Epap105* is similar to *ac63* of *Autographa californica* MNPV (AcMNPV); their predicted proteins are 28% identical. Its homologue in *Bombyx mori* NPV (BmNPV), *bm51*, was reported to be a structural gene associated with the budded virus (BV) envelope [27], but its deletion resulted in a virus with a phenotype similar to the wild type indicating that it might be a nonessential gene [28].

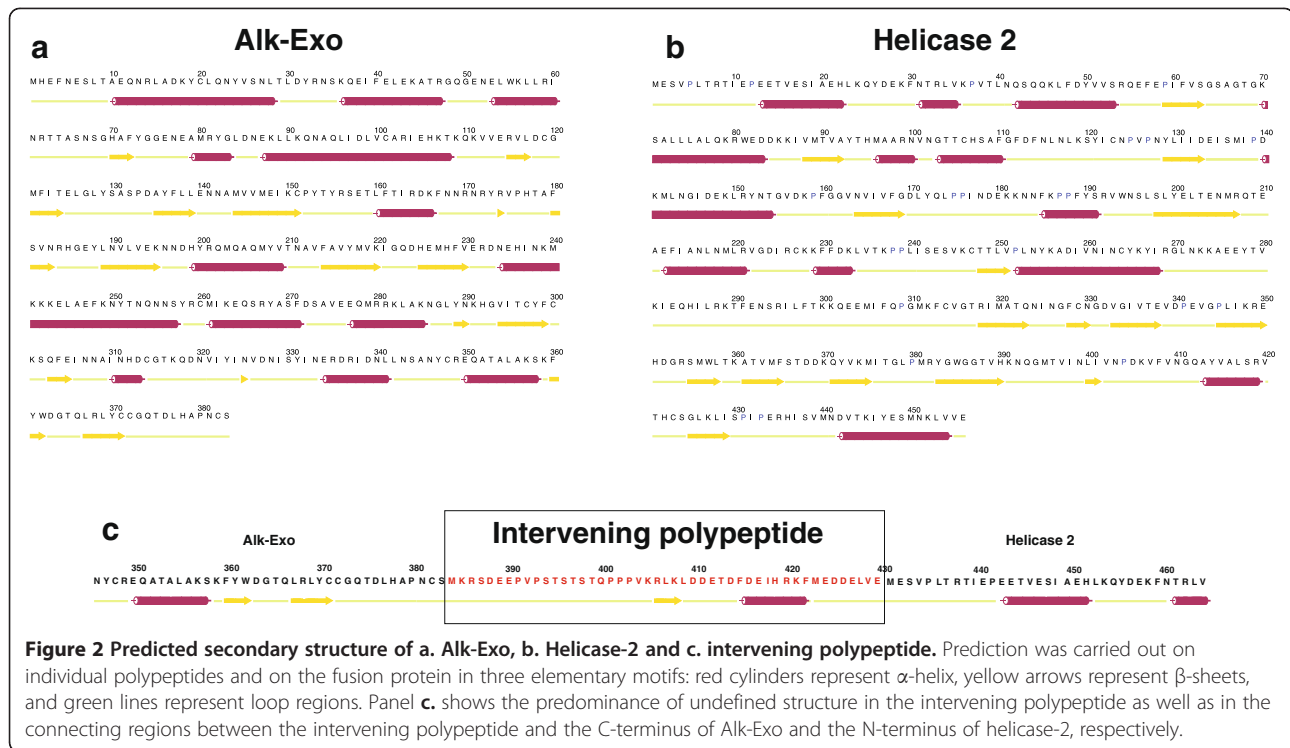


Figure 2 Predicted secondary structure of a. Alk-Exo, b. Helicase-2 and c. intervening polypeptide. Prediction was carried out on individual polypeptides and on the fusion protein in three elementary motifs: red cylinders represent α -helix, yellow arrows represent β -sheets, and green lines represent loop regions. Panel **c.** shows the predominance of undefined structure in the intervening polypeptide as well as in the connecting regions between the intervening polypeptide and the C-terminus of Alk-Exo and the N-terminus of helicase-2, respectively.

Epap24 codes for a 388 aa long protein that is highly similar to ORF 21 of *Cryptophlebia leucotreta* GV (CrleGV) according to BlastP search ($E = 4E-05$). Crle21 is a 308 aa predicted protein reported to be similar to Se43 [10]. These proteins share a protein motif of the DUF1383 superfamily. They have homologues in all alphabaculoviruses [28] and studies conducted with a deletion mutant of the homologue in AcMNPV (*ac18*) indicated that it is not essential for viral replication both *in vitro* and *in vivo*, but it may play a role in efficient virus infection in *Trichoplusia ni* larvae [29]. Homologues of *epap24/crle21* were not found in the rest of the granuloviruses.

Transcription genes

EpapGV DNA codes for the RNA polymerase subunits *lef-4* (*epap91*), *lef-9* (*epap112*), *lef-8* (*epap122*) and *p47* (*epap63*), *lef-5* (*epap82*) and *vlf-1* (*epap101*), present in all baculoviruses. Additional genes related to the transcription process found in all lepidopteran baculovirus were also detected in the EpapGV genome: 39 *k/pp31* (*epap56*), *lef-6* (*epap74*), *lef-11* (*epap57*) (present in gammabaculoviruses) and *pk-1* (*epap6*). *Lef-10*, involved in late transcription and present in most alpha- and betabaculoviruses, was also found in EpapGV genome (*epap128*). Of the baculoviral early transcription genes *ie-0*, *ie-1*, *ie-2* and *pe38*, only *ie-1* (*epap35*) is present in all GVs and *pe38* was found in CpGV, CrleGV, PhopGV and PrGV.

Replication genes

Genes involved in DNA replication that belong to the core group were found in EpapGV genome: *dnapol* (*epap106*), *lef-1* (*epap68*), *lef-2* (*epap41*) and *helicase-1* (*epap85*).

In addition, other genes that belong to this category and were found in EpapGV and in other lepidopteran baculoviruses are *dbp* (*epap75*) (also present in gammabaculoviruses), *lef-3* (*epap108*), *ie-1* (*epap35*), *me53* (*epap133*) and *ac38* (*epap65*). A *lef-7* homologue was found in a BlastP search restricted to baculoviruses: the protein encoded by *epap36* has a match with PsunGV LEF-7 ($E = 0.54$). This protein was demonstrated to be a baculoviral replication enhancer in AcMNPV [30] and BmNPV [31]. Homologues of this gene are present in group I NPVs, 3 group II NPVs and 3 GVs (XcGV, HearGV [28] and PsunGV).

EpapGV encodes a *DNA ligase* (*epap115*) as do other members of the *Betabaculovirus* genus and three NPVs of group II (LdNPV, LyxyNPV and OrleNPV). This gene seems to be linked to the presence of a second helicase, *helicase-2* (*epap120*) [32], which is fused with *alk-exo* in EpapGV, but not in the rest of the baculovirus genomes sequenced to date.

Structural genes

EpapGV genome contains all the structural genes corresponding to the core group as well as the lepidopteran baculovirus genes. The structural core group genes are:

p6.9 (epap81), *vp39 (epap92)*, *vp1054 (epap129)*, *vp91 (epap96)*, *gp41 (epap99)*, *odv-ec43 (epap43)*, *odv-e18 (epap29)*, *p74 (epap59)*, *pif-1 (epap69)*, *pif-2 (epap47)*, *pif-3 (epap38)*; *pif-4 (epap84)*; *pif-5/odv-e56 (epap27)* and the recently discovered *pif-6* [33] (*epap109*). Lepidopteran-specific baculovirus structural genes include *granulin (epap1)*; *25 k-fp (epap113)*; *odv-e25 (epap86)*; *bv/odv-c42 (epap80)*, the last two are also present in gammabaculoviruses. *F-protein (epap14)* is the only gene shared by alpha-, beta- and deltabaculoviruses. EpapGV contains 42 of the 47 proteins found in the occlusion derived virus (ODV) of PiraGV [34]. Five of these ORFs were only found in betabaculoviruses: *epap48*, *epap94*, *epap95*, *epap123* and *epap126*.

Auxiliary genes

In addition to core gene *alk-exo (Epap119)*, some other auxiliary genes were found in EpapGV genome. Viral *ubiquitin (epap52)* is present in GVs and all group I alphabaculoviruses. *Cathepsin (epap31)* and *chitinase (epap32)* were found in some GVs and in most alphabaculoviruses. These genes are responsible for the liquefaction of the host in the final stage of infection [35,36]. Their activity is readily apparent in *E. aporema* larvae infected with EpapGV. There is also a *gp37 (epap30)* homologue, which is present in some GVs and most NPVs. GP37 is homologous to the entomopoxvirus (EPV) fusolin which was shown to form spindle-like structures. These spindles enhance the peroral EPV infection by contributing to disruption of the peritrophic membrane [37]. EpapGV *gp37* gene has been characterized and demonstrated to be glycosylated [38]. The EpapGV genome includes three fibroblast growth factor homologues: *fgf-1*, *-2* and *-3 (epap70, epap118* and *epap131*, respectively).

The three *fgf* genes are present in all sequenced GVs but *fgf-2* is also present in all alphabaculoviruses. It is thought to be implicated in the virus dissemination in the insect host [39,40]. *Epap58* encodes a superoxide dismutase homologue (*sod*) which is widely distributed in baculovirus. Its potential role is still unknown and controversial [41].

EpapGV also possesses two *iap* genes (inhibitors of apoptosis), *iap-3 (epap11)* and *iap-5 (epap111)*. *Iap-5* is only present in betabaculovirus whereas *iap-3* is also present in some NPVs. No *p35* homologue was found.

Granulovirus-specific genes

The number of genes considered to be GV-specific has changed in the literature and will be more accurate when more complete genome sequences become available. These genes could be the basis to the differences between granuloviruses and nucleopolyhedroviruses. Taking into account the analyses presented by Lange

et al. [10], Wormleaton et al. [7], Escasa et al. [9], Van Oers & Vlak [42], Miele et al. [23] and the present report, a set of 19 genes has been identified in betabaculovirus genomes which were never found in alpha-, gamma- or deltabaculoviruses (Figure 1, Additional file 1). These are EpapGV ORFs 7, 8, 17, 21, 22, 25, 37, 40, 43, 44 (*metalloproteinase*), 62, 70 (*fgf-1*), 73, 94, 95, 110, 111 (*iap-5*), 126 and 131 (*fgf-3*).

Other genes formerly considered as part of the GV-specific set, were dismissed from the list in the present report: CpGV ORFs 30, 32, 45, 50, 56, 77, 82, 119, 121, 122 and 136. All, except cp27, 56, 77, 121 and 136, have homologues in EpapGV but they are absent in some other GV (see Additional file 1).

Unique genes

Seventeen ORFs appear to be unique to EpapGV compared to the rest of the members of *Baculoviridae* (ORFs 4, 9, 12, 16, 18, 19, 20, 23, 49, 51, 60, 64, 72, 89, 104, 114 and 116). *Epap4* codes for a 144 aa long protein with a conserved motif (COG5152) in its N-terminal region. This motif is an uncharacterized conserved domain that contains RING and CCCH-type Zn-fingers [43]. An early promoter motif was found 150 nt upstream *epap4* ORF. *Epap9* encodes an 81 aa long polypeptide and has no significant BlastP matches. The upstream region contains a GATA motif (TGATAG) and two TATAWAW early promoter elements, but no CAKT INR. *Epap12* codes for a 90 aa protein which shares 23% identity and 43% similarity with a small portion of a 2123 aa protein of *Drosophila ananassae* (XP_001953497); however, no speculation on function can be made. *Epap16* gives no significant BlastP hit, and has early promoter elements upstream of the first ATG (TATAW+3 CATT elements). Something similar happens with *Epap18* which codes for a hypothetical 76 aa protein and a TATAW element upstream. *Epap19* (94 aa) has no significant BlastP hits and has elements of an early promoter. *Epap20* (422 aa) has no significant BlastP hits and shows elements of a late promoter. *Epap23* codes for a hypothetical protein of 197 aa with no significant similarity with any protein of the GenBank under the control of a putative late promoter and a GATA motif (TGATAG).

Epap49 codes for the longest hypothetical protein of EpapGV genome (1465 aa). As it lacks characteristic promoter elements and exhibits no similarity with other baculovirus genes it is difficult to predict if it is actually transcribed. *Epap49* is located between the conserved genes *pif-2 (epap47)*, core gene) and *epap50* (homologue to *cp52*). It is worth mentioning that at least in two GVs (HearGV and ChocGV) a similar situation emerged in the same locus. Although a 1144 aa ORF with 27 leucine zippers was initially found in ChocGV, it was not considered a coding sequence but a non-*hr* ori-like region

instead; the speculation was based upon its very high AT content (81%), the lack of homology with baculovirus ORFs, and the possibility of sequencing errors (for further details see Escasa *et al.*, [9]). In contrast, the 1279 aa ORF in HearGV (*hear44*) was considered a coding sequence resulting from a fusion of the homologues *xc47* and *xc48* of *Xestia c-nigrum* GV [12].

Epap51 codes for a 69 aa peptide under the control of an early promoter and showed no significant matches in BlastP search. Two GATA motifs (TGATAT and AGATAG) were also found in the region upstream its ATG. *Epap60* (582 aa) shows no significant hits with any protein in the GenBank. TATATAA and TATAA motifs were found upstream the ATG, but without the initiator sequence CAKT, characteristic of early promoters. *Epap64* codes for a thymidylate kinase (described below). *Epap72* codes for a 61 aa peptide with no significant matches in the GenBank under the control of a putative late promoter which is overlapped with a GATA motif (AGATAAG). *Epap89* predicted protein (86 aa) did not have significant BlastP hits either and lacks known promoter motifs except for a TATAAAA sequence 86 nt upstream its ATG overlapped with a GATA motif. Similarly, ORFs 104 (63 aa), 114 (162 aa) and 116 (51 aa) contain TATA box-like motifs upstream the initial ATG and show no significant BlastP hits. *Epap114* also presents a ATAAG sequence, and *Epap116* a GATA motif (AGATAA).

Nucleotide metabolism genes

Genes coding for enzymes involved in nucleotide metabolism have been reported in baculovirus genomes. Ribonucleotide reductase (RNR) catalyses the reduction of ribose in ribonucleotide diphosphates to yield deoxyribonucleotides, the building blocks of DNA [44]. In most eukaryotes the active RNR is a tetrameric complex made up of homodimers of two subunits coded by genes *rr1* and *rr2*, which have been also found in some NPVs and 4 GVs: CpGV, AgseGV, PhopGV and EpapGV (*epap2* and *epap3*). On the other hand, dUTPase catalyses the dephosphorylation of dUTP to yield dUMP. As dUTP can be mutagenic if incorporated in DNA, this enzyme helps to keep levels of dUTP low and prevents its incorporation in DNA, in lieu of dTTP. This gene is present in some NPVs and in the betabaculoviruses AgseGV, SpltGV and EpapGV (*Epap13*). The presence of *rr1*, *rr2* and *dutpase* appears to be linked in the genomes of OpMNPV, SpltNPV, SeMNPV and LdMNPV [32]. This linkage appears in the betabaculoviruses AgseGV and EpapGV but not in SpltGV. Both enzymes participate in the pathway of *de novo* dTTP biosynthesis (Figure 3a).

EpapGV codes for a novel enzyme in the family *Baculoviridae*, which also takes part in this pathway: *epap64* codes for a predicted 224 aa protein homologous

to thymidylate kinase, also known as thymidine monophosphate kinase (TMPK), that catalyses the phosphorylation of dTMP to produce dTDP. BlastP hits included different eukaryotic organisms and several viruses representing the families *Poxviridae* (Variola Virus), *Iridoviridae* (Invertebrate iridescent virus 6, II6), *Herpesviridae* (Cyprinid Herpes 3, CyHV3), *Nimaviridae* (White Spot Syndrome Virus, WSSV) and *Asfaviridae* (African swine fever virus, ASFV) that were used in the ClustalW alignment with EpapGV TMPK (Figure 3b). EpapGV TMPK showed the highest identity (40%) with TMPK from the insect *Drosophila ananassae* and the least (22%) with TMPK from ASFV. The degree of identity with the other viruses was 35% (II6); 32% (Variola and Vaccinia); and 33% (WSSV). Besides *Baculoviridae*, other viral families that encode nucleotide metabolism genes include *Herpesviridae*, *Poxviridae* and *Asfaviridae*. The alphaherpesvirus pyrimidine deoxynucleoside kinase, popularly known as thymidine kinase (TK) phosphorylates a wide range of nucleoside substrates, as well as TMP (TK + TMPK activity), and is responsible for the rise in the TTP pool characteristic of HSV-infected cells [45]. In poxviruses these TK and TMPK activities are present in separate enzymes as happens in cellular organisms. Vaccinia virus TMPK was found to be nonessential for virus replication in cultured cells and able to complement a *tmpk*-*Saccharomyces cerevisiae* mutant [46].

The White Spot Syndrome Virus (WSSV; *Nimaviridae*) genome contains a mosaic gene that encodes a *tk-tmpk* fusion of both homologues, *i.e.* cellular-type thymidine kinase TK1 and cellular-type TMPK [47]. However, only TK activity, but not TMPK, could be demonstrated for WSSV TK-TMK [48]. TMPK substrate specificity was studied in vaccinia virus and it was found to phosphorylate dTMP, dUMP and, unlike human TMPK, dGMP as well [49]. EpapGV TMPK expression and substrate specificity, as well as its role in infection, remain to be elucidated.

Repeated sequences

A common feature in baculovirus genomes is the presence of nucleotide sequence repeats known as homologous regions (*hrs*). These regions function as enhancers of early gene transcription and are thought to play a role as origins of replication. They are characterized by tandem copies of sequence motifs that include an imperfect palindromic core. Although they present significant sequence similarity within a genome they are highly variable when compared between any two different species (Reviewed in [50]).

In a first screening of the EpapGV genome for repeated sequences with Blast2seq we found two palindromic regions, of 128 bp (58352–58479) and 122 bp (116114–116235), respectively. Both sequences are very likely to exist in equilibrium between double stranded DNA and opposite hairpin-loops constituted by

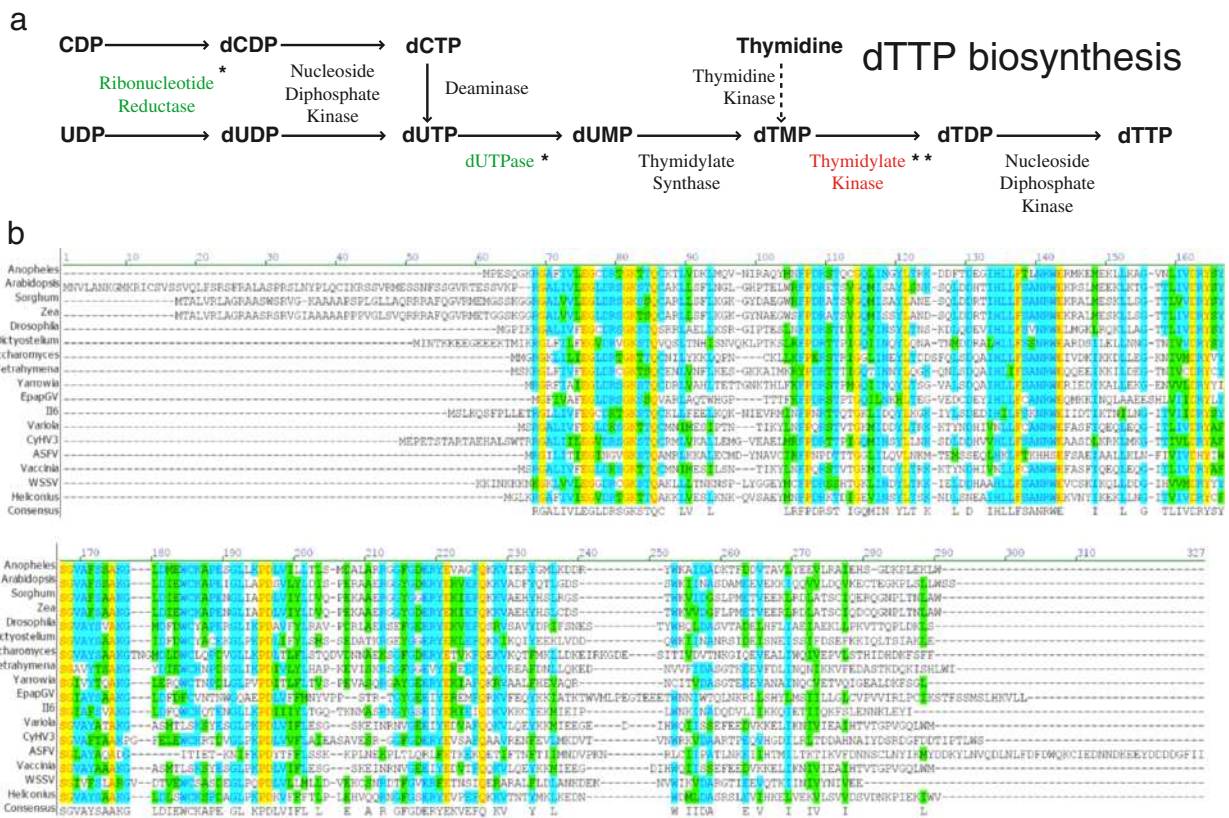


Figure 3 dTTP biosynthesis and TMPK. a. Cellular *de novo* pathway of dTTP biosynthesis. Enzymes present in EpapGV and other baculoviruses are marked with an asterisk (*). TMPK gene present in EpapGV and absent in other baculoviruses is highlighted with two asterisks (**). Solid arrows correspond to the *de novo* pathway of dTTP biosynthesis and the dashed arrow to the salvage pathway **b.** Alignment of EpapGV TMPK with other TMPKs from several organisms (only the genus is indicated) and viruses: *Anopheles gambiae* (XP_314179.3); *Arabidopsis thaliana* (NP_001078772.1); *Sorghum bicolor* (XP_002461104.1); *Zea mays* (NP_001150303.1); *Drosophila ananassae* (XP_001960115.1); *Dictyostelium discoideum* (XP_635930.1); *Saccharomyces cerevisiae* (NP_012591.1); *Tetrahymena thermophila* (XP_001009062.1); *Yarrowia lipolytica* (XP_501790.1); Il6 (NP_149714.1); Variola Major Virus (NP_042204.1); CyHV3 (YP_001096175.1); ASFV (NP_042729.1); Vaccinia Virus (AAW23610.1); WSSV (AAG40728.1); *Heliconius melpomene* (CBH09285.1). Identical residues conserved in all the sequences are shaded in yellow; those conserved in most sequences are shaded in light blue, and conservative changes are shaded in green.

each complementary strand (*hr10*, -75.30 kcal/mol and *hr16a*, -86.50 kcal/mol, respectively) forming a cruciform-like structure (Figure 4a,b). Whether this feature is biologically relevant remains to be elucidated. Using these two sequences as profiles we searched the rest of the genome for similar sequences. Twenty-four AT-rich sequences of similar size were detected dispersed throughout the genome. The alignment of these sequences revealed the presence of conserved palindromes of about 58 bp that correspond with the central part of the two initially identified largest palindromes, with a mean of 70% AT content. The alignment of these shorter palindromes (Figure 4d) shows that they have an AT rich core flanked by 15 bp conserved inverted repeats (Figure 4d,e). This structure is similar to that of *hrs* found in all GV's (those sequenced to date) that infect other insects of the Tortricidae family (CpGV, CrleGV, AdorGV, ChocGV): palindromic sequences of about 63–76 bp characterized

by conserved blocks of 13 bp located at both ends. These ends were found to be similar in sequence not only among the palindromes of a single genome but among the different genomes, as well [51]. EpapGV palindromic ends are similar to the consensus 13 bp sequence for these GV's, including PhopGV (Figure 4f). PhopGV, that infects a member of the Gelechiidae family but appears in the same clade as these GV's, was also found to have this kind of palindromic repeats, although their length was different (142–320 bp) [51]. Palindromes sharing these features were analyzed in infection-dependent replication assays in cells susceptible to CpGV infection. All 14 CpGV palindromes were found to act as origins of replication of plasmids in infection-dependent assays in *Cydia pomonella* cells [51,52], whereas the *hrs* from other GV's tested in the same way did not show this ability, with the only exception of 2 palindromes of the most closely related CpGV virus, CrleGV [51].

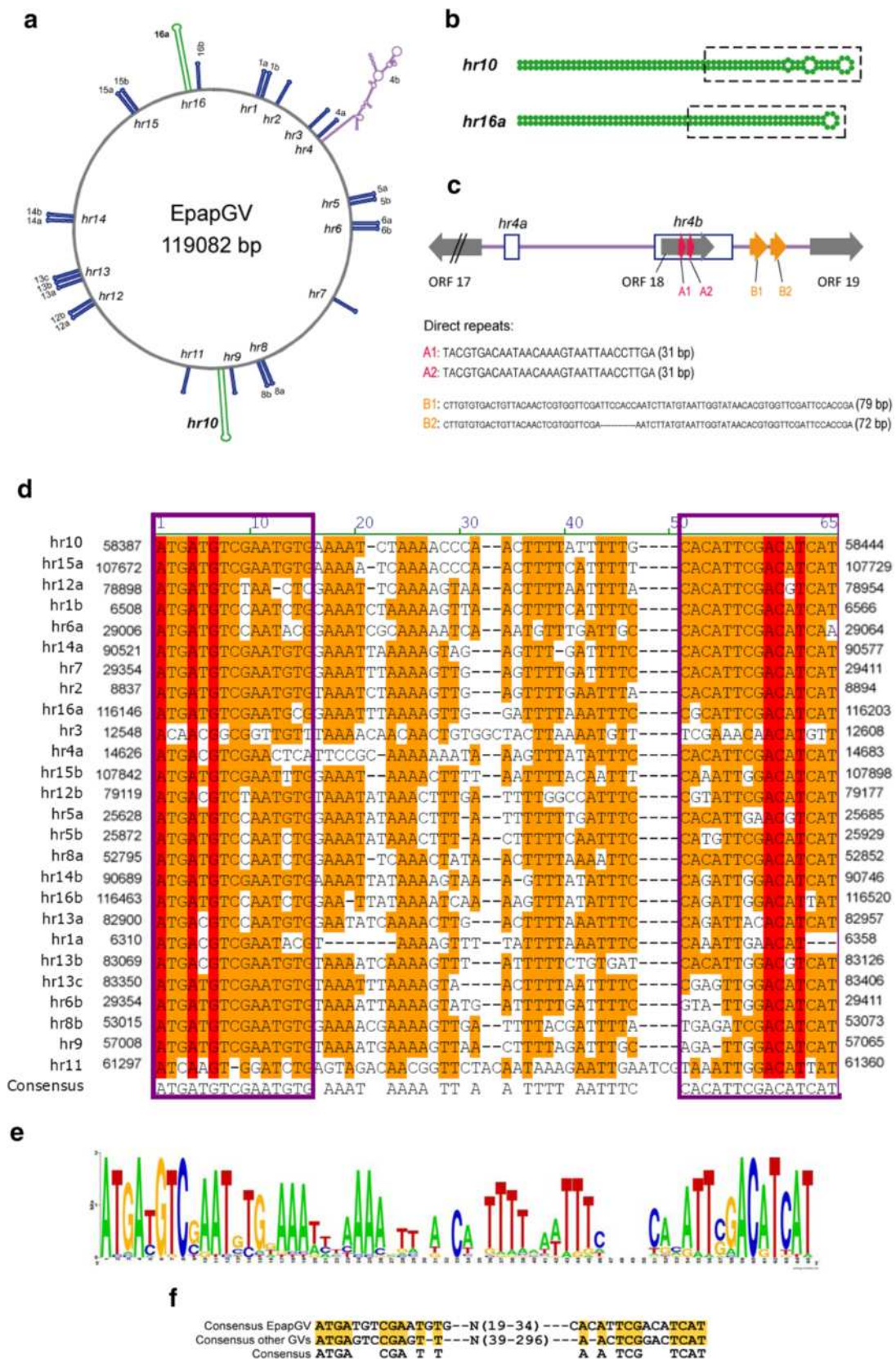


Figure 4 (See legend on next page.)

(See figure on previous page.)

Figure 4 Repetitive sequence analysis. **a.** Distribution of the putative *hrs* in the EpapGV genome with their respective palindromes. **b.** Predicted secondary structure of the two largest palindromes. The central palindrome used in the multiple alignment is boxed with a dashed line. **c.** Structure of the putative *non-hr* ori. **d.** Alignment of all the palindromes excluding *hr4b*. Genome positions are indicated at both ends of the aligned sequences. Conserved 15 bp ends are boxed. **e.** Sequence logo performed with the multiple alignment of the palindromes. **f.** Comparison between consensus EpapGV palindromic ends with the consensus 13 bp ends reported by Hilton & Winstanley (2008) [51] (see Figure 3 of this reference).

These palindromic sequences are found with a much greater frequency in EpapGV DNA compared to the other tortricid-specific GVs. EpapGV contains 26 palindromes (within 16 *hrs*), whereas the others have up to 17 palindromes as is the case of CrleGV [51]. In contrast, only four *hrs* (each one containing only 2–3 direct repeats) were reported in the most recently published betabaculovirus genome sequence (ClanGV; [18]).

There seem to be some conserved locations for the *hrs* in GV genomes. For example, the region between *sod* and *p74* and downstream of the CpGV ORF 5 [51].

It has been reported for AcMNPV that VLF-1 (a protein present in all the baculoviruses sequenced to date) binds with high affinity to cruciform DNA structures and it was suggested that this may play an important role in the replication/packaging process [53]. These cruciform structures, formed by the two largest palindromes or by the smaller ones interspersed in the EpapGV genome, may as well interact with VLF-1 and play a role in the replication or packaging.

In addition to the 26 palindromes mentioned above, there is a large structure consisting of 327 bp flanked by the 15 bp conserved ends predicted to form the secondary structures shown in Figure 4(a). This structure is located in the *hr4* region (including *hr4a* and *hr4b*), an AT-rich sequence between ORFs 17 and 19. The sequence organization is depicted in Figure 4(c) showing ORF 18 within *hr4b*, which also contains two 31 bp direct repeats (A1, A2), and an intergenic region with a second pair of imperfect direct repeats of 79 and 72 bp (B1, B2), respectively. Interestingly, this region is located in the same relative position where a putative *non-hr* ori was described in CpGV spanning ORFs 24, 25 and 26 (which are absent in EpapGV) [11] and in CrleGV [54].

Relationships with other baculoviruses

Strong colinearity is observed in granulovirus genomes sequenced to date [7,10,11,51]. Baculovirus gene colinearity has been analysed mainly by Gene Parity Plot [55]. In this study we used the Artemis Comparison Tool to analyse the gene synteny of EpapGV compared to all other sequenced GVs and the type species of the *Alphabaculovirus* genus, AcMNPV. This tool enables to construct synteny maps through a tBlastX comparison between genomes, where inversions are easily detected as well as the different percentages of identities that correlate with different colour intensity. Figure 5 shows the

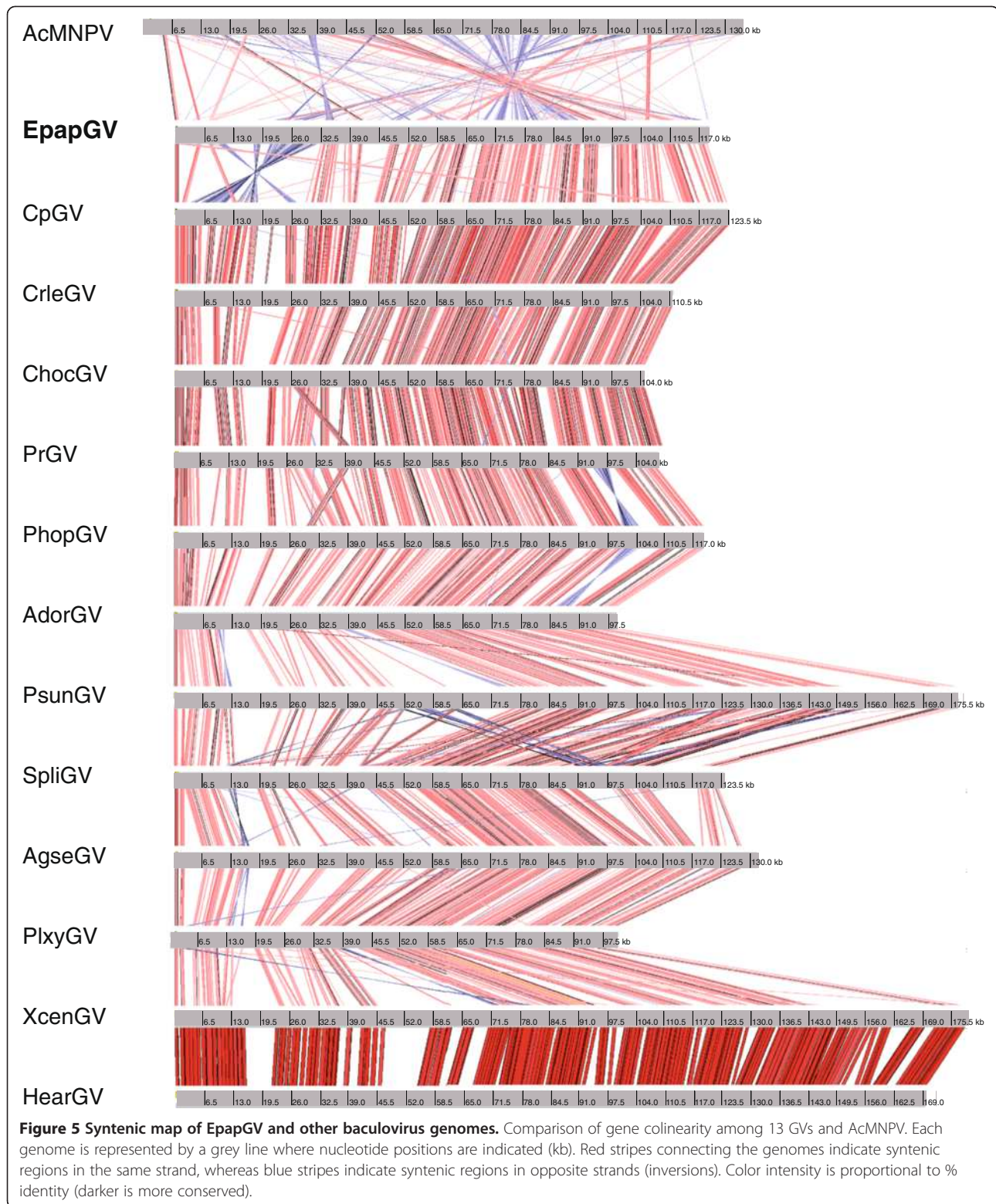
conserved gene colinearity of all 13 sequenced GV genomes and the poorly conserved synteny between GVs and AcMNPV. Notably EpapGV differs from the rest of the GVs by a *ca.* 20 kb gene block inversion, as we noted previously in a physical map [19].

Phylogenetic analysis based on 31 concatenated core genes of 58 baculovirus genomes was performed (Figure 6). The obtained cladogram reproduced the grouping of four genera recognized in the current classification of the family [1]. Division in two main groups of the *Alphabaculovirus* genus agrees with the group I and II. Two clades (Ia and Ib) previously described in group I by using concatenated amino acid sequences of the partial *polh/gran*, *lef-8* and *lef-9* genes [56] were also confirmed in our analysis.

As expected, EpapGV grouped in the *Betabaculovirus* genus. In previous reports it was observed that tottriciidae and noctuidae specific GVs tend to be in separated groups [56,57]. The cladogram obtained in this work confirms previous observations, and the additional complete genomes considered here allowed the division of betabaculoviruses in two well separated monophyletic clades as reported previously [23]. Clade “a” includes six species: PxGV, AgseGV, SpliGV, PsunGV and XcGV, which were isolated mainly from *Noctuidae* hosts. PxGV is the exception; its host belongs to the *Plutellidae* family. Clade “b” includes seven species: EpapGV, AdorGV, PhopGV, CpGV, CrleGV, PiraGV and ChocGV; five of them were isolated from *Tortricidae*, whereas PiraGV was isolated from *Pieridae* and PhopGV, from *Gelechiidae*. EpapGV seems to be the GV isolate closest to the common ancestor of Clade “b”. Both clades includes slow killing (type 1 GVs) and fast killing (type 2 GVs), reinforcing the concept of that this biological feature is not phylogenetically informative [57].

Conclusions

In this study the complete genome of EpapGV was characterized. It includes genes that are common to all baculovirus, and others that have been found only in some of the isolates; in addition, it contains 17 genes that are not shared with the rest of the family: 16 with unknown functions, and one encoding a TMPK homologue which may have been captured from of a host genome or a different coinfecting pathogen. Also a set of 19 betabaculovirus-specific genes, was determined. The information collected and analyzed in this study provides ground for further investigations to



improve the understanding of the molecular steps involved in EpapGV infection.

The analyses of gene order and identity suggest that evolution of baculoviruses occurred via acquisition of

both individual genes (or gene fragments) and larger blocks of host DNA sequences followed by events of inversions, deletions, and re-acquisitions of previously lost sequences. Interestingly, the phylogenetic analysis

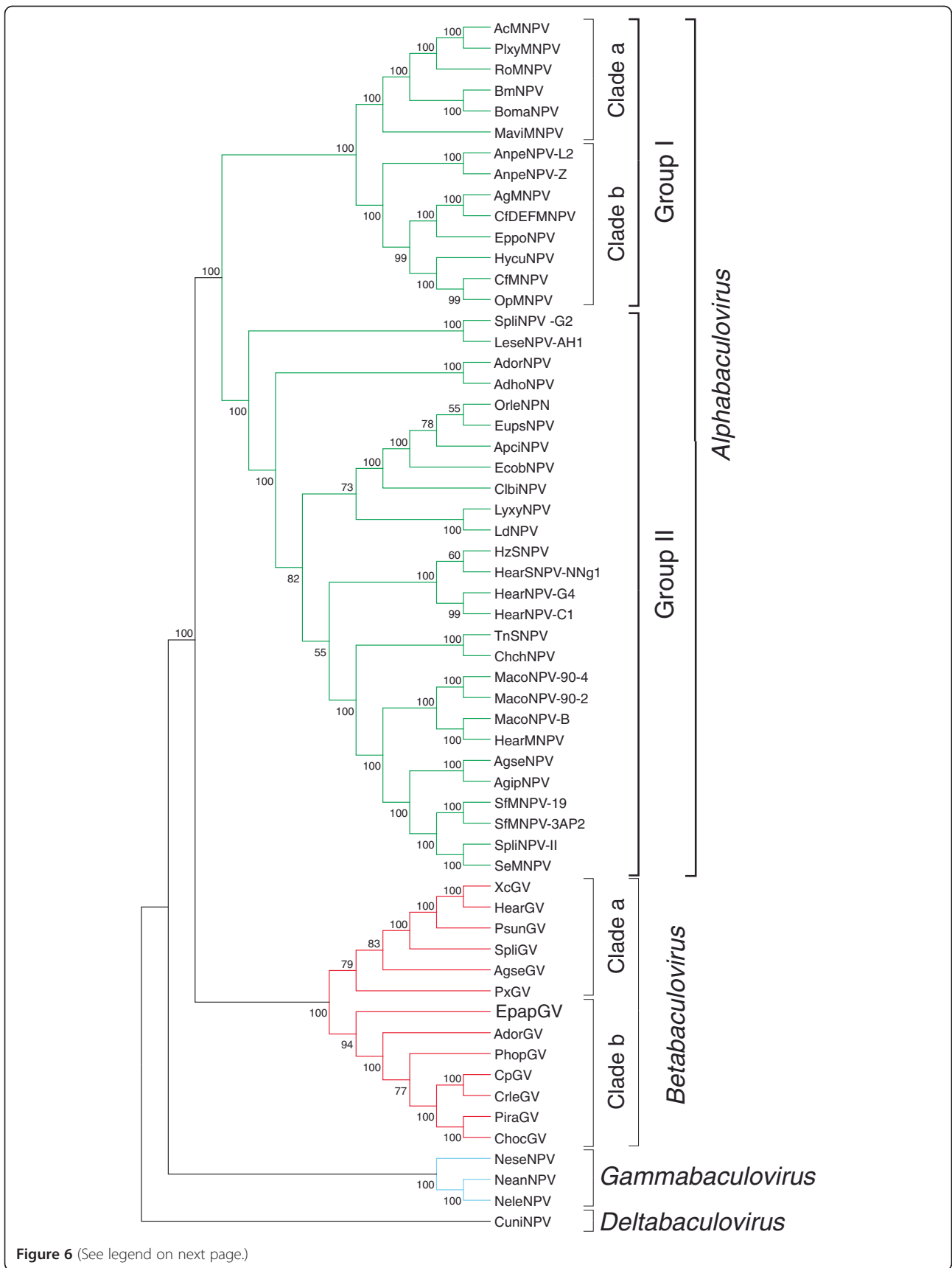


Figure 6 (See legend on next page.)

(See figure on previous page.)

Figure 6 Phylogenetic tree. Cladogram based on amino acid sequences of 31 core genes obtained from 58 available complete baculoviral genomes were individually aligned, concatenated and a phylogenetic tree was inferred with MEGA 5 program. The four Baculovirus genera and the different subgroups are indicated.

suggested that EpapGV is situated closest to the common ancestor of clade b *Betabaculovirus*.

Methods

Insects, virus and viral DNA

EpapGV was originally isolated from a larva of the bean shoot borer *Epinotia aporema* collected in Oliveros (Santa Fe, Argentina) [5]. It was amplified allowing fourth instars to feed on artificial diet superficially contaminated with EpapGV occlusion bodies (OBs). Moribund larvae were collected and processed according to Parola *et al.* [19]: viral DNA was isolated from sucrose gradient purified OBs. Its integrity and identity was checked by restriction digestion and agarose gel electrophoresis.

Nucleotide sequence determination and analysis

EpapGV genomic DNA was sequenced with the 454 Genome Sequencer (GS) FLX™ Standard (Roche) at the Interdisciplinary Center for Biotechnology Research (ICBR), University of Florida (Gainesville, US). *De novo* assembly was generated on newBler assembler (GS FLX Data Analysis Software).

Open reading frames (ORFs) were identified using VectorNTI software (Invitrogen) and ORF Finder <http://www.ncbi.nlm.nih.gov/gorf/gorf.html> [58]. ATG initiated ORFs of at least 150 nt (50 aa) with minimal overlap were selected for further analysis. Homology searches were done using Blast [59]. Percentage identities between homologous genes were obtained by global alignments with ClustalW [60] using default parameters. Early (E) and late (L) Promoter motifs within 150 bp upstream of the putative ORFs were screened. E indicates the presence of a TATA-box (TATAW, TATAWAW, TATAWTW) with a CAKT mRNA start site 20–40 nucleotides downstream; whereas L denotes a DTAAG sequence [7,10,61]. Also GATA motifs WGATAR [62] and WGATAY [63] were searched for the unique genes.

Prediction of secondary structure of Alk-Exo_Helicase-2 fused protein was performed with the Jpred3 server [64]; <http://www.compbio.dundee.ac.uk/www-jpred/>) using default parameters and single sequence submit. Actually, the prediction accuracy of Jnet (main Jpred3 algorithm) raised 81.5% in blind tests with soluble proteins. C-terminal end of Alk-Exo and N-terminus of Helicase-2 were selected on the basis of multiple alignments of the respective GV proteins.

Repeated sequences were searched first aligning EpapGV genome to itself through Blast2seq program

from NCBI [65]. The first hit which corresponds to the 100% match of the complete genome was ignored and the following hits were used for further analysis. The consensus alignment obtained from two palindromes was used to find similar sequences along the genome with the VectorNTI program (Invitrogen). The secondary DNA structure prediction of these sequences were performed in the Mfold server of The Vienna RNA website [66]. The alignment of all the palindromes found was performed with ClustalW algorithm with default parameters. The sequence logo of this alignment was carried out at the WebLogo server (<http://weblogo.berkeley.edu/>) [67].

EpapGV genome was compared with other baculovirus genomes by constructing syntenic maps with the Artemis Comparison Tool (ACT) [68] (The Sanger Institute; <http://www.sanger.ac.uk/resources/software/act/>), using tBlastX program.

Phylogenetic analysis was performed using 31 core genes from 58 baculovirus genomes (Additional File 4) which were independently aligned using ClustalX program [69], with the following parameters: Pairwise alignment (Gap Open Penalty = 10, Gap Extension Penalty = 0.1, protein weight matrix: Blosum 30); Multiple alignment (Gap Open Penalty = 10, Gap Extension Penalty = 0.05, protein weight matrix: Blosum series). Then a concatemer was generated by addition of the complete individual alignments and phylogeny was inferred using MEGA 5 program [70] with the following parameters: UPGMA; Bootstrap with 1000 replicates; gap/Missing data = complete deletion; Model = Amino (Dayhoff Matrix); patterns among sites = Same (Homogeneous); rates among sites = Different (Gamma Distributed); gamma parameter = 2.25. The obtained data was deposited in TreeBASE (<http://purl.org/phylo/treebase/phylovs/study/TB2:S12862>).

Additional files

Additional file 1: Predicted ORFs in the genome of EpapGV. This file lists the ORFs predicted in the genome of EpapGV and their homologues in other completely sequenced betabaculoviruses.

Additional file 2: Multiple alignment of betabaculovirus alkaline exonuclease. This file shows the alignment of betabaculovirus alkaline exonuclease amino acid sequences.

Additional file 3: Multiple alignment of betabaculovirus helicase 2. This file shows the alignment of betabaculovirus helicase 2 amino acid sequences.

Additional file 4: Characteristics of baculovirus genomes. This file lists the characteristics of baculovirus genomes, including virus name, genome size, number of ORFs and GeneBank accession number.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MLF obtained and purified the viral DNA, analyzed the sequence and wrote the manuscript. RS and MEB participated in viral amplification and purification from insect larvae and revised the manuscript. RS also assisted with the phylogenetic analysis. SH participated in the sequence analysis and revised the manuscript. MFB participated in the hrs analysis and discussion and revised the manuscript. ASC is one of the senior scientists in the project, she provided the viral isolate and equipment for virus and DNA purification, advised on the study and revised the manuscript. PDG participated in the sequence analysis, the phylogenetic tree and supplementary figures, advised on the study and revised the manuscript. VR conceived the study and revised the manuscript. All authors approved the final manuscript.

Authors' information

MLF and HS hold fellowships from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET); RS and MEB held fellowships from ANPCyT and CONICET, respectively, when they participated in this study. VR, PDG and MFB hold research career awards from CONICET; ASC and MFB are staff researchers at IMYZA (CICVyA, INTA).

Acknowledgements

The authors thank Dr. Lorena Passarelli for her helpful comments and suggestions on the manuscript and Mrs. M. Cornacchio and Mrs. D.B. Moreyra for the maintenance of the insect colony at INTA. This work was supported by grants from the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) and UNLP to VR and ASC and the Instituto Nacional de Tecnología Agropecuaria (INTA) to ASC.

Author details

¹Instituto de Biotecnología y Biología Molecular, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, CONICET, La Plata, Argentina.
²Instituto de Microbiología y Zoología Agrícola, INTA, Castelar, Argentina.
³Laboratorio de Ingeniería Genética y Biología Celular y Molecular - Área Virosis de Insectos, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Bernal, Argentina. ⁴Present address: INSERM U748, Institut de Virologie, Faculté de Médecine, Université de Strasbourg, 3 rue Koeberlé, 67000, Strasbourg, France.

Received: 23 November 2011 Accepted: 22 September 2012

Published: 11 October 2012

References

- Herniou EA, Arif BM, Becnel JJ, Blissard GW, Bonning B, Harrison R, Jehle JA, Theilmann DA, Vlak JM: **Baculoviridae**. In *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. Edited by King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. San Diego: Elsevier Academic Press; 2011:163–173.
- Jehle JA, Blissard GW, Bonning BC, Cory JS, Herniou EA, Rohrmann GF, Theilmann DA, Thiem SM, Vlak JM: **On the classification and nomenclature of baculoviruses: a proposal for revision**. *Arch Virol* 2006, **151**(7):1257–1266.
- Carstens EB, Ball LA: **Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2008)**. *Arch Virol* 2009, **154**(7):1181–1188.
- Federici BA: **Baculovirus Pathogenesis**. In *The Baculoviruses*. Edited by Miller LK. New York and London: Plenum Press; 1997:33–59.
- Sciocco-Cap A, Parola AD, Goldberg AV, Ghiringhelli PD, Romanowski V: **Characterization of a granulovirus isolated from *Epinotia aporema* Wals. (Lepidoptera: Tortricidae) larvae**. *Appl Environ Microbiol* 2001, **67**(8):3702–3706.
- Goldberg AV, Romanowski V, Federici BA, Sciocco de Cap A: **Effects of the Epap granulovirus on its host, *Epinotia aporema* (Lepidoptera: Tortricidae)**. *J Invertebr Pathol* 2002, **80**(3):148–159.
- Wormleaton S, Kuzio J, Winstanley D: **The complete sequence of the *Adoxophyes orana* granulovirus genome**. *Virology* 2003, **311**(2):350–365.
- Winstanley D, O'Reilly DR: **Granuloviruses**. In *The Encyclopedia of Virology, vol 1*. second edition. Edited by Webster R, Granoff A. London: Academic Press; 1999:140–146.
- Escasa SR, Lauzon HA, Mathur AC, Krell PJ, Arif BM: **Sequence analysis of the *Choristoneura occidentalis* granulovirus genome**. *J Gen Virol* 2006, **87**(Pt 7):1917–1933.
- Lange M, Jehle JA: **The genome of the *Cryptophlebia leucotreta* granulovirus**. *Virology* 2003, **317**(2):220–236.
- Luque T, Finch R, Crook N, O'Reilly DR, Winstanley D: **The complete sequence of the *Cydia pomonella* granulovirus genome**. *J Gen Virol* 2001, **82**(Pt 10):2531–2547.
- Harrison RL, Popham HJ: **Genomic sequence analysis of a granulovirus isolated from the Old World bollworm, *Helicoverpa armigera***. *Virus Genes* 2008, **36**(3):565–581.
- Zhang B-Q, Cheng R-L, Wang X-F, Zhang C-X: **The Genome of *Pieris rapae* Granulovirus**. *J Virol* 2012, **86**(17):9544.
- Hashimoto Y, Hayakawa T, Ueno Y, Fujita T, Sano Y, Matsumoto T: **Sequence analysis of the *Plutella xylostella* granulovirus genome**. *Virology* 2000, **275**(2):358–372.
- Wang Y, Choi JY, Roh JY, Woo SD, Jin BR, Je YH: **Molecular and phylogenetic characterization of *Spodoptera litura* granulovirus**. *J Microbiol* 2008, **46**(6):704–708.
- Wang Y, Choi JY, Roh JY, Liu Q, Tao XY, Park JB, Kim JS, Je YH: **Genomic sequence analysis of granulovirus isolated from the tobacco cutworm, *Spodoptera litura***. *PLoS One* 2011, **6**(11):e28163.
- Hayakawa T, Ko R, Okano K, Seong SI, Goto C, Maeda S: **Sequence analysis of the *Xestia c-nigrum* granulovirus genome**. *Virology* 1999, **262**(2):277–297.
- Liang Z, Zhang X, Yin X, Cao S, Xu F: **Genomic sequencing and analysis of *Clostera anachoreta* granulovirus**. *Arch Virol* 2011, **156**:1185–1198.
- Parola AD, Manzan MA, Lozano ME, Ghiringhelli PD, Sciocco-Cap A, Romanowski V: **Physical and genetic map of *Epinotia aporema* granulovirus genome**. *Virus Genes* 2002, **25**(3):329–341.
- Vlak JM, Smith GE: **Orientation of the genome of *Autographa californica* nuclear polyhedrosis virus: a proposal**. *J Virol* 1982, **41**(3):1118–1121.
- Friesen PD: **Regulation of baculovirus early gene expression**. In *The Baculoviruses*. Edited by Miller LK. New York and London: Plenum Press; 1997:141–170.
- Lu A, Miller LK: **Regulation of baculovirus late and very late gene expression**. In *The Baculoviruses*. Edited by Miller LK. New York: Plenum Press; 1997:193–216.
- Miele SAB, Garavaglia MJ, Belaich MN, Ghiringhelli PD: **Baculovirus: molecular insights on their diversity and conservation**. *Int J Evol Biol* 2011, **2011**:15.
- Rohrmann G: **Introduction to the baculoviruses, their taxonomy, and evolution**. In *Baculovirus Molecular Biology*. 2nd edition. Bethesda (MD): National Library of Medicine (US), NCBI; 2011.
- Li Z, Gong Y, Yin C, Wang L, Li C, Pang Y: **Characterization of a novel ubiquitin-fusion gene *Uba256* from *Spodoptera litura* nucleopolyhedrovirus**. *Gene* 2003, **303**:111–119.
- Hyink O, Dellow RA, Olsen MJ, Caradoc-Davies KM, Drake K, Herniou EA, Cory JS, O'Reilly DR, Ward VK: **Whole genome analysis of the *Epiphyas postvittana* nucleopolyhedrovirus**. *J Gen Virol* 2002, **83**(Pt 4):957–971.
- Tian CH, Tang XD, Xu HJ, Ge JQ, Miao YG, Zhang CX: ***Bombyx mori* nucleopolyhedrovirus ORF51 encodes a budded virus envelope associated protein**. *Virus Genes* 2009, **38**(1):171–177.
- Rohrmann G: **The AcMNPV genome: gene content, conservation, and function**. In *Baculovirus Molecular Biology*. 2nd edition. Bethesda (MD): National Library of Medicine (US), NCBI; 2011.
- Wang Y, Wu W, Li Z, Yuan M, Feng G, Yu Q, Yang K, Pang Y: **ac18 is not essential for the propagation of *Autographa californica* multiple nucleopolyhedrovirus**. *Virology* 2007, **367**(1):71–81.
- Lu A, Miller LK, Krell P, Vlak JM, Rohrmann G: **Baculovirus DNA Replication**. In *The Baculoviruses*. Edited by Miller LK. New York and London: Plenum Press; 1997.
- Gomi S, Zhou CE, Yih W, Majima K, Maeda S: **Deletion analysis of four of eighteen late gene expression factor gene homologues of the baculovirus, BmNPV**. *Virology* 1997, **230**(1):35–47.
- Herniou EA, Olszewski JA, Cory JS, O'Reilly DR: **The genome sequence and evolution of baculoviruses**. *Annu Rev Entomol* 2003, **48**:211–234.
- Nie Y, Fang M, Erlandson MA, Theilmann DA: **Analysis of the *autographa californica* multiple nucleopolyhedrovirus overlapping gene pair *lef3* and *ac68* reveals that *AC68* is a per os infectivity factor and that *LEF3* is critical, but not essential, for virus replication**. *J Virol* 2012, **86**(7):3985–3994.

34. Wang X-F, Zhang B-Q, Xu H-J, Cui Y-J, Xu Y-P, Zhang M-J, Han YS, Lee YS, Bao Y-Y, Zhang C-X: **ODV-associated proteins of the pieris rapae granulovirus.** *J Proteome Res* 2011, **10**(6):2817–2827.
35. Ohkawa T, Majima K, Maeda S: **A cysteine protease encoded by the baculovirus Bombyx mori nuclear polyhedrosis virus.** *J Virol* 1994, **68**(10):6619–6625.
36. Slack JM, Kuzio J, Faulkner P: **Characterization of v-cath, a cathepsin L-like proteinase expressed by the baculovirus Autographa californica multiple nuclear polyhedrosis virus.** *J Gen Virol* 1995, **76**(Pt 5):1091–1098.
37. Mitsuhashi W, Kawakita H, Murakami R, Takemoto Y, Saiki T, Miyamoto K, Wada S: **Spindles of an entomopoxvirus facilitate its infection of the host insect by disrupting the peritrophic membrane.** *J Virol* 2007, **81**(8):4235–4243.
38. Salvador R, Ferrelli ML, Berretta MF, Mitsuhashi W, Biedma ME, Romanowski V, Sciocco-Cap A: **Analysis of EpapGV gp37 gene reveals a close relationship between granulovirus and entomopoxvirus.** *Virus Genes* 2012, **45**:1–4. doi:10.1007/s11262-012-0800-3.
39. Detvisitsakun C, Cain EL, Passarelli AL: **The Autographa californica M nucleopolyhedrovirus fibroblast growth factor accelerates host mortality.** *Virology* 2007, **365**(1):70–78.
40. Means JC, Passarelli AL: **Viral fibroblast growth factor, matrix metalloproteases, and caspases are associated with enhancing systemic infection by baculoviruses.** *Proc Natl Acad Sci* 2010, **107**(21):9825–9830.
41. Tomalski MD, Eldridge R, Miller LK: **A baculovirus homolog of a Cu/Zn superoxide dismutase gene.** *Virology* 1991, **184**(1):149–161.
42. van Oers MM, Vlask JM: **Baculovirus genomics.** *Curr Drug Targets* 2007, **8**(10):1051–1068.
43. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, et al: **CDD: a Conserved Domain Database for the functional annotation of proteins.** *Nucleic Acids Res* 2011, **39**(Database issue):D225–D229.
44. Nelson DL, Cox MM: *Lehninger Principles of Biochemistry*. 5th edition. New York: W. H. Freeman Publishers; 2008.
45. Hay J, Ruyechan W: **Alphaherpesvirus DNA replication.** In *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*. Edited by Arvin AC-FG, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K. Cambridge: Cambridge University Press; 2007.
46. Hughes SJ, Johnston LH, de Carlos A, Smith GL: **Vaccinia virus encodes an active thymidylate kinase that complements a cdc8 mutant of Saccharomyces cerevisiae.** *J Biol Chem* 1991, **266**(30):20103–20109.
47. Tsai MF, Yu HT, Tzeng HF, Leu JH, Chou CM, Huang CJ, Wang CH, Lin JY, Kou GH, Lo CF: **Identification and characterization of a shrimp white spot syndrome virus (WSSV) gene that encodes a novel chimeric polypeptide of cellular-type thymidine kinase and thymidylate kinase.** *Virology* 2000, **277**(1):100–110.
48. Tzeng H-F, Chang Z-F, Peng S-E, Wang C-H, Lin J-Y, Kou G-H, Lo C-F: **Chimeric polypeptide of thymidine kinase and thymidylate kinase of shrimp white spot syndrome virus: thymidine kinase activity of the recombinant protein expressed in a baculovirus/Insect cell system.** *Virology* 2002, **299**(2):248–255.
49. Topalis D, Collinet B, Gasse C, Dugue L, Balzarini J, Pochet S, Deville-Bonne D: **Substrate specificity of vaccinia virus thymidylate kinase.** *FEBS J* 2005, **272**(24):6254–6265.
50. Berretta M, Romanowski V: **Baculovirus homologous regions (hrs): pleiotropic functional cis elements in viral genomes and insect and mammalian cells.** *Current Topics in Virology* 2008, **7**:47–56.
51. Hilton S, Winstanley D: **The origins of replication of granuloviruses.** *Arch Virol* 2008, **153**(8):1527–1535.
52. Hilton S, Winstanley D: **Identification and functional analysis of the origins of DNA replication in the Cydia pomonella granulovirus genome.** *J Gen Virol* 2007, **88**(Pt 5):1496–1504.
53. Mikhailov VS, Rohrmann GF: **Binding of the baculovirus very late expression factor 1 (VLF-1) to different DNA structures.** *BMC Mol Biol* 2002, **3**:14.
54. Jehle JA: **The expansion of a hypervariable, non-hr ori-like region in the genome of Cryptophlebia leucotreta granulovirus provides in vivo evidence for the utilization of baculovirus non-hr oris during replication.** *J Gen Virol* 2002, **83**(Pt 8):2025–2034.
55. Hu ZH, Arif BM, Jin F, Martens JW, Chen XW, Sun JS, Zuidema D, Goldbach RW, Vlask JM: **Distinct gene arrangement in the Buzura suppressaria single-nucleocapsid nucleopolyhedrovirus genome.** *J Gen Virol* 1998, **79**(Pt 11):2841–2851.
56. Jehle JA, Lange M, Wang H, Hu Z, Wang Y, Hauschild R: **Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera.** *Virology* 2006, **346**(1):180–193.
57. Herniou EA, Jehle JA: **Baculovirus phylogeny and evolution.** *Curr Drug Targets* 2007, **8**(10):1043–1050.
58. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, et al: **Database resources of the National Center for Biotechnology.** *Nucleic Acids Res* 2003, **31**(1):28–33.
59. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403–410.
60. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**(22):4673–4680.
61. Nai YS, Wu CY, Wang TC, Chen YR, Lau WH, Lo CF, Tsai MF, Wang CH: **Genomic sequencing and analyses of Lymantria xyliina multiple nucleopolyhedrovirus.** *BMC Genomics* 2010, **11**:116.
62. Ko LJ, Engel JD: **DNA-binding specificities of the GATA transcription factor family.** *Mol Cell Biol* 1993, **13**(7):4011–4022.
63. Krappa R, Behn-Krappa A, Jahnel F, Doerfler W, Knebel-Mörsdorf D: **Differential factor binding at the promoter of early baculovirus gene PE38 during viral infection: GATA motif is recognized by an insect protein.** *J Virol* 1992, **66**(6):3494–3503.
64. Cole C, Barber JD, Barton GJ: **The Jpred 3 secondary structure prediction server.** *Nucleic Acids Res* 2008, **36**(Web Server issue):W197–W201.
65. Tatusova TA, Madden TL: **BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences.** *FEMS Microbiol Lett* 1999, **174**(2):247–250.
66. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL: **The Vienna RNA websuite.** *Nucleic Acids Res* 2008, **36**(Web Server issue):W70–W74.
67. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14**(6):1188–1190.
68. Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J: **ACT: the Artemis comparison tool.** *Bioinformatics* 2005, **21**(16):3422–3423.
69. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25**(24):4876–4882.
70. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28**(10):2731–2739.

doi:10.1186/1471-2164-13-548

Cite this article as: Ferrelli et al.: Genome of *Epinotia aporema* granulovirus (EpapGV), a polyorganotropic fast killing betabaculovirus with a novel thymidylate kinase gene. *BMC Genomics* 2012 **13**:548.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

