Virus Genes 30:3, 403–417, 2005

Physical and Partial Genetic Map of *Spodoptera frugiperda* Nucleopolyhedrovirus (SfMNPV) Genome

OIHANE SIMÓN,[1] FRANÇOIS CHEVENET,[2] TREVOR WILLIAMS,[1] PRIMITIVO CABALLERO,[1] & MIGUEL LÓPEZ-FERBER[3,*]

[1]*Depto. de Producción Agraria, Universidad Pública de Navarra, 31006 Pamplona, Spain*
[2]*UMR IRD/CNRS 9926-BP 64501, 34394 Montpellier cedex 5, France*
[3]*Laboratoire de Pathologie Comparée, UMR 5087, INRA-CNRS-Université de Montpellier II, 30380 Saint Christol-Lez-Ales, France*

Received October 11, 2004; Revised November 9, 2004; Accepted December 13, 2004

Abstract. A Nicaraguan isolate of *Spodoptera frugiperda* multicapsid nucleopolyhedrovirus (SfMNPV) is undergoing field trials for control of this pest in the Americas. This isolate is composed of multiple genotypes, some of which are deletion mutants. Identification of the genetic changes in deleted genotypes cannot be accomplished without the construction of a detailed physical map. In the present study, combinations of restriction endonuclease analysis and Southern blot analysis was performed. This map was refined by sequencing the termini of cloned restriction fragments. The SfMNPV genome was estimated to be 129.3 kb, 8 kb larger than the previously characterized Sf-2 variant from the United States, due to a deletion between 14.8 and 21.0 m.u. in the physical map described in this study. A total of 27.92 kb were sequenced, which represented 21.5% of the whole genome and included 38 ORFs. Comparison with other sequenced baculoviruses revealed that SfMNPV displayed the highest sequence identity (66%) and gene arrangement (78%) with *Spodoptera exigua* MNPV, sharing 36 putative ORFs. In addition, the genome organization was similar to that of SeMNPV, with minor differences. Phylogenetic analysis confirmed the close relatedness between SeMNPV and SfMNPV, suggesting they evolved from a common ancestor.

Key words: genetic map, nucleopolyhedrovirus, physical map, *Spodoptera frugiperda*

Introduction

Insect-infecting baculoviruses have been reported worldwide from over 600 host species, mainly from the order Lepidoptera [1]. Baculoviruses are enveloped, rod-shaped virions carrying a circular double-stranded DNA genome ranging in size from 90 to 180 kb, depending on the virus species (http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/). The family Baculoviridae comprises two genera, *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV), distinguished by the morphology of the occlusion bodies (OBs) they form in infected cells [1]. The NPVs characteristically form polyhedral OBs, each containing many virions, whereas the GVs typically produce ovoid OBs containing a single virion. NPVs have been divided into two phylogenetic groups (group I and group II), based on genomic characteristics for those NPVs for which the complete genome sequences are currently available [2]. NPVs are also designated as single (S) or multiple (M), depending on the number of nucleocapsids packaged in each virion, although this characteristic does not correlate to genetic relatedness and is not considered to be a phylogenetic trait [3].

* Author for all correspondence:
Present address:
LGEI
Ecole des Mines d'Alès Clavières
6, Av. de Clavières
30319 Alès Cedex, France

Baculoviruses are of interest because of their applications as biological insecticides for control of insect pests [4,5] and for the development of baculovirus-based gene expression systems [6,7]. The genetic improvement of baculoviruses for both these applications and the study of baculovirus diversity and ecology requires detailed knowledge of the physical maps and the molecular biology of these viruses. The complete genome sequences of nine lepidopteran NPVs are currently available, including that of the type species, *Autographa californica* MNPV (AcMNPV) [8]. These data offer a resource for phylogenetic analysis, gene content mapping, as well as for the rapid mapping of other NPV genomes and the identification of homologous genes [2]. Such information is also very helpful for the construction of physical and genetic maps of other NPVs.

An NPV has been reported as a common pathogen in natural populations of the fall armyworm, *Spodoptera frugiperda* [9], which is an important pest of maize and sorghum in much of Latin America [10]. Several multinucleocapsid NPV isolates have been collected from geographically distant populations of *S. frugiperda* (SfMNPV) and have been subjected to DNA restriction endonuclease analysis and SDS-PAGE analysis of structural polypeptides [11,12]. Plaque purification was used to isolate seven distinct naturally occurring genotypic variants in a SfMNPV from the United States [13]. Physical maps were generated using several restriction endonucleases for each of these variants. The prototype variant, SfMNPV-2, was very similar to the isolates examined previously [14]. The physical map of a different SfMNPV isolate from Ohio has also been produced [15]. Genetic studies of baculoviruses isolated from four different *Spodoptera* species (*S. exigua* [SeMNPV], *S. littoralis* [SpliNPV], *S. frugiperda* [SfMNPV] and *S. exempta* [SpexNPV]) revealed that SfMNPV is related, to some extent, to SpliNPV and SpexNPV, but is most closely related to SeMNPV [16]. Relatively few genes have been sequenced from this virus, namely, the *polyhedrin* gene [17], the *gp41* gene [18], and recently, the *egt* region [19] (GenBank accession AY250076).

A SfMNPV isolate from Nicaragua has potential as a bioinsecticide for the control of *S. frugiperda* [12], and it was selected for formulation and field trails in Honduras and Mexico [20]. This isolate was found to be composed of at least nine genotypic variants, some of which are deletion mutants that are not infectious *per os* [21,22]. Identification of the genetic changes in the defective genotypic variants cannot be accomplished without the construction of a physical map. In this report, we present a detailed physical map of the SfMNPV genome. In order to investigate the genomic organization and phylogenetic status of SfMNPV, viral DNA was cloned as restriction fragments into a plasmid library and the inserts were terminally sequenced. Gene organization was deduced by partial genetic mapping. Phylogenetic analysis based on the completely sequenced genes will contribute to elucidating the ecological role of the genotypic variants detected in the Nicaraguan SfMNPV, as well as revealing the phylogenetic relationships of this virus to other NPVs.

Materials and Methods

Insect, Viruses and Viral DNA Isolation

Larvae from a laboratory colony of *S. frugiperda* were maintained on a wheatgerm-based semi-synthetic diet [23] at 26 ± 2°C, 16 h:8 h L:D, 70–80% RH. The wild-type SfMNPV was originally isolated from diseased *S. frugiperda* larvae collected in Nicaragua [12]. The pure genotype SfNIC-B was plaque-purified in *S. frugiperda* (Sf9) cells, as described elsewhere [21]. Both viruses, SfMNPV and SfNIC-B, were amplified by infection of fourth instar *S. frugiperda* using the droplet-feeding method [24]. OB purification and viral DNA isolation were performed as described previously [25,26]. Briefly, infected larvae were homogenized in sterile distilled water, the OBs were filtered through muslin, washed with 0.1% sodium dodecyl sulphate (SDS) and 0.1 M NaCl, pelleted by centrifugation, and resuspended in distilled water. DNA was extracted from OB released virions by incubation with SDS and proteinase K, followed by phenol–chloroform extraction and alcoholic precipitation. DNA concentration was estimated by absorption at 260 nm or by agarose gel electrophoresis.

Restriction Endonuclease (REN) Analysis

Viral DNA was digested with the restriction enzymes *Eco*RI, *Hin*dIII, *Pst*I, *Bam*HI or *Sma*I (Amersham, UK) at 37°C for 4–12 h and the reactions were stopped at 65°C for 15 min. Digested DNA was then electrophoresed in 1% TAE agarose gels at low voltage (20–40 V) for 6 to 20 h. The commercial marker Smart Ladder (Stratagene), containing fragments which ranged from 10 to 0.2 kb in size, was run as a standard for size determination. DNA fragments were stained with ethidium bromide, visualized in a UV transilluminator and photographed using the program Molecular Analyst (Bio-Rad). The resulting fragments from each restriction enzyme were used to estimate the virus genome size.

Constructing the Genomic DNA Libraries

Two genomic libraries of SfMNPV *Eco*RI and *Hin*dIII fragments were constructed in pSP70 plasmid (Promega) using a DNA ligation kit (Rapid Ligation Kit, Roche Diagnostics, Meylan, France). Competent *Escherichia coli* TG1 cells were transformed with the recombinant plasmids and plated on LB agar containing 100 μg/ml ampicillin. Plasmids were purified using standard methods [27] and screened for the presence of inserts by *Eco*RI or *Hin*dIII digestion and electrophoresis in 1% agarose gel. Inserts were authenticated by comparing their migration in agarose gels with the SfMNPV fragments generated by digestion with the same restriction endonucleases. A library of SfNIC-B *Pst*I DNA fragments was also generated in pUC19 (Promega) using the procedure described above, but in this case, the transformed *E. coli* cells were plated on LB agar containing 100 μg/ml ampicillin, 1 μM IPTG and 80 μg/ml X-Gal, for blue/white color selection.

Physical Mapping of the SfNIC Genome

The construction of the physical map was achieved by ordering the restriction fragments on the viral genome according to the Southern blot hybridization data and multiple digestion of cloned viral REN fragments. All *Eco*RI cloned fragments were digested with *Hin*dIII, *Pst*I, *Bam*HI and/or *Sma*I, all available *Hin*dIII fragments were digested with *Eco*RI, *Pst*I, *Bam*HI and/or *Sma*I and all *Pst*I cloned fragments with *Eco*RI, *Hin*dIII, *Bam*HI and/or *Sma*I. The fragments resulted from multiple digestions were electrophoresed and the fragment sizes were then compared to cloned *Eco*RI, *Hin*dIII or *Pst*I fragments. The order of the restriction fragments was determined by analysis of the overlapping portions of cloned fragments.

Southern blotting was also performed to order the restriction fragments. SfMNPV DNA and cloned fragments were digested with *Eco*RI, *Hin*dIII or *Pst*I and electrophoresed in 1% agarose gels. The separated DNA fragments were blotted onto positively charged membranes (Roche). Plasmids containing cloned virus DNA fragments were labeled with digoxigenin-dUTP by random priming (DIG DNA Labelling kit, Roche) and used as probes in Southern blot hybridizations. Membranes were prehybridized for 2 h at 60°C with (pre)hybridization buffer (Roche). DNA probes were hybridized to membranes overnight under the same conditions. Post-hybridization washings were carried out under low stringency conditions [27] and exposed to film for 1 to 8 h. The cloned fragments were mapped with *Eco*RI, *Hin*dIII, *Pst*I and *Bam*HI and their size estimated from multiple digestion results. The order of the restriction fragments was assessed by analysis of the overlapping portions of the different libraries using the techniques described above.

Mapping of the SfMNPV genome was confirmed by sequencing the termini of all the cloned *Hin*dIII fragments and some of the cloned *Eco*RI and *Pst*I fragments which were used to verify contiguity between clones (Table 1). Nucleotide sequences were determined in an ABI PRISM 377 automated DNA sequencer (MWG-Biotech, Germany, and Genome-Express, France), employing standards SP6 and T7 for fragments cloned into pSP70 and M13 and M13 reverse primers for fragments cloned into pUC19.

Sequence information was analyzed for the presence of open reading frames (ORFs) and for the prediction of domains and sequence analysis using the Clone Manager 5.0 program (Scientific and Educational Software Server). Homology searches were performed both at the nucleotide level and at the amino acid level, using all putative ORFs. DNA and protein comparisons with entries in the updated GenBank/EMBL, SWISS-PROT,

Table 1. Molecular size of *Eco*RI, *Hin*dIII, *Pst*I, and *Bam*HI restriction endonuclease fragments of SfMNPV genomic DNA. The DNA fragments are named alphabetically, starting with A for the largest fragment (see Fig. 1), and their sizes are given in kbp

	Restriction fragments			
Fragment	*Eco*RI	*Hin*dIII	*Pst*I	*Bam*HI
A	12.4	16.8	27.0	52.8
B	11.2	15.4	24.0	28.4
C	10.0*	14.7*	16.7*	15.5
D	10.0	12.5	9.0*	13.9
E	9.6*	10.7	8.2*	9.9
F	9.5	9.7	7.8*	5.1
G	9.0*	8.1*	6.9*	4.2
H	7.5*	7.1*	6.7*	
I	6.1*	6.0*	5.9*	
J	5.1*	6.0*	5.0*	
K	3.8*	5.9*	4.7*	
L	3.6*	5.6*	2.9*	
M	3.2*	3.5*	2.0*	
N	3.1*	2.7*	1.3*	
O	2.6*	2.3*	1.2*	
P	2.4*	1.0*		
Q	2.2*	0.7*		
R	2.0*	0.6*		
S	1.9*			
T	1.8*			
U	1.7*			
V	1.5*			
W	1.4*			
X	1.4*			
Y	1.3*			
Z	1.2*			
a	1.2*			
b	1.0*			
c	0.6*			
d	0.5*			
Total	129.3	129.3	129.3	129.3

*Fragments cloned into pSP70 (those obtained by *Eco*RI and *Hin*dIII genomic DNA digestion) or pUC19 (those obtained by *Pst*I digestion).

and PIR databases were performed using BLASTn, BLASTPn and FASTA programs [28,29]. Multiple sequence alignments were performed with CLUSTALX version 1.7 computer program [30]. Baculovirus sequences used in the comparative analysis were GenBank (accession numbers included): *S. exigua* (Se) MNPV (AF169823), *S. littoralis* (Spli) NPV (X99376), *S. littura* (Splt) NPV (NC0003102), *Mamestra configurata* (Maco) MNPV (NC004117), *Helicoverpa armigera* (Ha) SNPV (NC003094), *Autographa californica* (Ac) MNPV (L22858), *Bombyx mori* (Bm) NPV (L33 180), *Lymantria dispar* (Ld) MNPV (AF081810), *Xestia c-nigrum* (Xc) GV (U70896) and *Cydia pomonella* (Cp) GV (NC002816).

The zero point of this map was the start of the *polyhedrin* (*polh*) gene following generally accepted convention [31] and was determined by sequence information. Previously, this point was determined by heterologous hybridization with the *polh* gene of AcMNPV [13].

Phylogenetic Analysis

The SfMNPV protein sequences deduced from the sequenced fragments and the partially sequenced *polh* gene were aligned with the corresponding available sequences in GenBank database with the CLUSTAL X algorithm [30].

Distance analysis was performed using the PHYLIP software package [32]. Distance matrices were computed using the Prodist program based on the Dayhoff PAM matrix. Unrooted trees were computed using the neighbor joining method (Neighbor program), with the corresponding amino acid sequences from the mosquito baculovirus, *Culex nigripalpus* NPV (CuniNPV) [33] used as an outgroup when available. A graphic analysis was computed using the TreeDyn editor [34].

Results

REN Patterns and Size of SfMNPV Genome

The SfMNPV DNA genome was cleaved into 30, 18, 15 and 7 visible fragments by the restriction enzymes *Eco*RI, *Hin*dIII, *Pst*I and *Bam*HI, respectively (Fig. 1). These fragments were designated alphabetically assigning the letter A to the largest fragment for each endonuclease digest as usual (Fig. 1). The endonuclease *Sma*I did not cleave either uncut DNA or DNA fragments generated by any of the enzymes used in this study (data not shown). Six sets of co-migrating fragments were detected by stain intensity in the *Eco*RI REN pattern (fragments C–D, K–L, R–S, T–U, W–X, and Y–Z), one in the *Hin*dIII REN pattern (fragments I–J), two in the *Pst*I REN pattern (fragments A–B

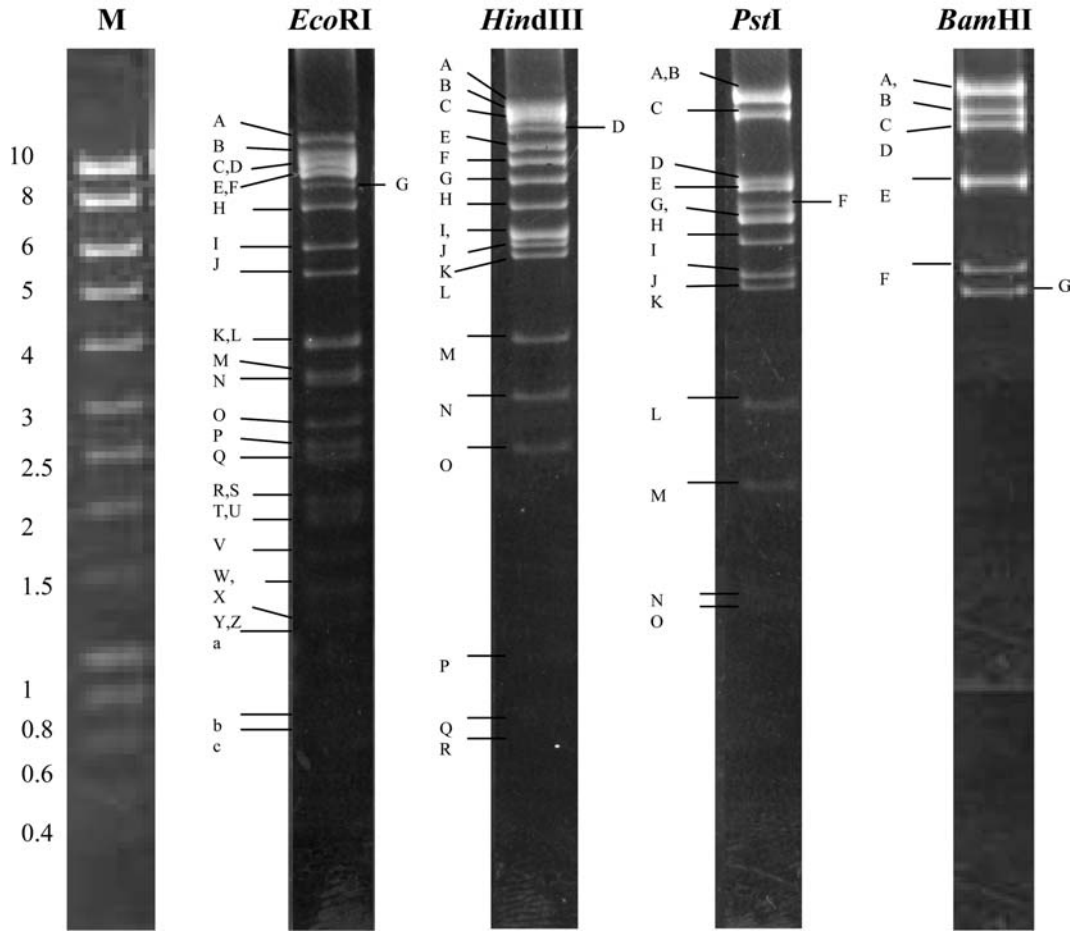

Fig. 1. REN patterns of SfMNPV DNA digested with *Eco*RI, *Hin*dIII, *Pst*I or *Bam*HI as indicated, generated by electrophoresis in 1% agarose gel. The DNA Marker Smart Leader (Stratagene) was used as a molecular size marker (indicated in kbp) (M). All DNA fragments are marked with a letter corresponding to their sizes.

and G–H), and one in the *Bam*HI REN pattern (fragments A–B) (Fig. 1). The size estimates for the REN fragments cloned into pSP70 or pUC19 plasmids and the total genome size for SfMNPV are given in Table 1. The genome of SfMNPV was estimated to be 129.3 kbp and was calculated from the sum of the sizes of cloned restriction fragments, generated either by single or double enzyme digestions. All cloned fragments were mapped with *Eco*RI, *Hin*dIII, *Pst*I or *Bam*HI and their sizes estimated from the multiple digestion results. To minimize the error in estimating the genome size, we used the sum of the relatively smaller fragment sizes that resulted from multiple restriction digests.

Physical Mapping of the SfMNPV Genome

Two incomplete libraries of 27 of the 30 *Eco*RI fragments and 13 of the 18 *Hin*dIII fragments were cloned into pSP70. Another incomplete library of 13 of the 15 *Pst*I fragments was cloned into pUC19 (Table 1). The identity of the cloned virus DNA fragments was confirmed by Southern blot hybridization to immobilized viral DNA. Southern blot hybridization, using recombinant plasmid DNA or, in some cases, viral DNA fragments extracted from agarose gel as probes, and REN-digested viral DNA immobilized on membranes produced the data used to determine the co-linearity of SfMNPV DNA fragments generated by

the different restriction enzymes (Fig. 2). Southern blot hybridization indicated homology among the *Eco*RI-Q, *Hin*dIII-M, *Pst*I-J, and *Bam*HI-A restriction fragments that all contain the *polh* gene and were therefore located at the beginning (left side) of the linear representation of the physical map (Fig. 2). To determine the detailed alignments of the REN DNA fragments in regions where the clones of a particular enzyme were not sufficient, cloned REN fragments from other enzyme digests were used as hybridization probes. For example, the SfMNPV *Pst*I-F (clone p279.51) fragment hybridized to the large *Hin*dIII-C and -D fragments. Whereas, fragment *Pst*I-L (clone p279.19) and *Pst*I-K (clone p279.56) hybridized only to the large *Hin*dIII-C fragment. Thus, it was determined that *Hin*dIII-D was next to and followed by fragment *Hin*dIII-C and that *Pst*I-L and *Pst*I-K were at the right end of *Pst*I-F. Likewise, by hybridization with the *Pst*I-G fragment (p279.146), we determined that *Hin*dIII-E was at the left end of *Hin*dIII-D fragment. These data, as well as those generated by reciprocal double digestion of cloned fragments, or in some cases, of isolated fragments, were used to construct the physical map of the SfMNPV. The map includes 70 restriction sites for the five restriction enzymes (*Eco*RI, *Hin*dIII, *Pst*I, *Bam*HI, and *Sma*I) used in this study (Fig. 2). Several small restriction fragments such as *Eco*RI-a (p264.101), *Eco*RI-b (p264.103), *Eco*RI-c (p264.245), *Eco*RI-d (p264.243), *Hin*dIII-R (p258.72) or *Hin*dIII-Q (p258.75) were ordered by hybridization to the complete genome that had been restricted by the enzymes described previously. Since *Eco*RI-a hybridized to the largest *Hin*dIII and *Pst*I fragments, *Eco*RI-b hybridized to *Hin*dIII-D, *Pst*I-F and *Bam*HI-F and *Eco*RI-G hybridized to *Hin*dIII-C and -D, it was determined that *Eco*RI-b was located between *Eco*RI-C and -G. *Eco*RI-c hybridized completely to *Hin*dIII-L, partially to *Pst*I-E and to a lesser degree to *Pst*I-A, showing that *Pst*I-A was next to *Pst*I-E. *Eco*RI-d hybridized to *Hin*dIII-H and to *Pst*I-I and to a larger *Bam*HI fragment. However, the alignments of some other restriction enzyme fragments could not be determined by hybridization or double

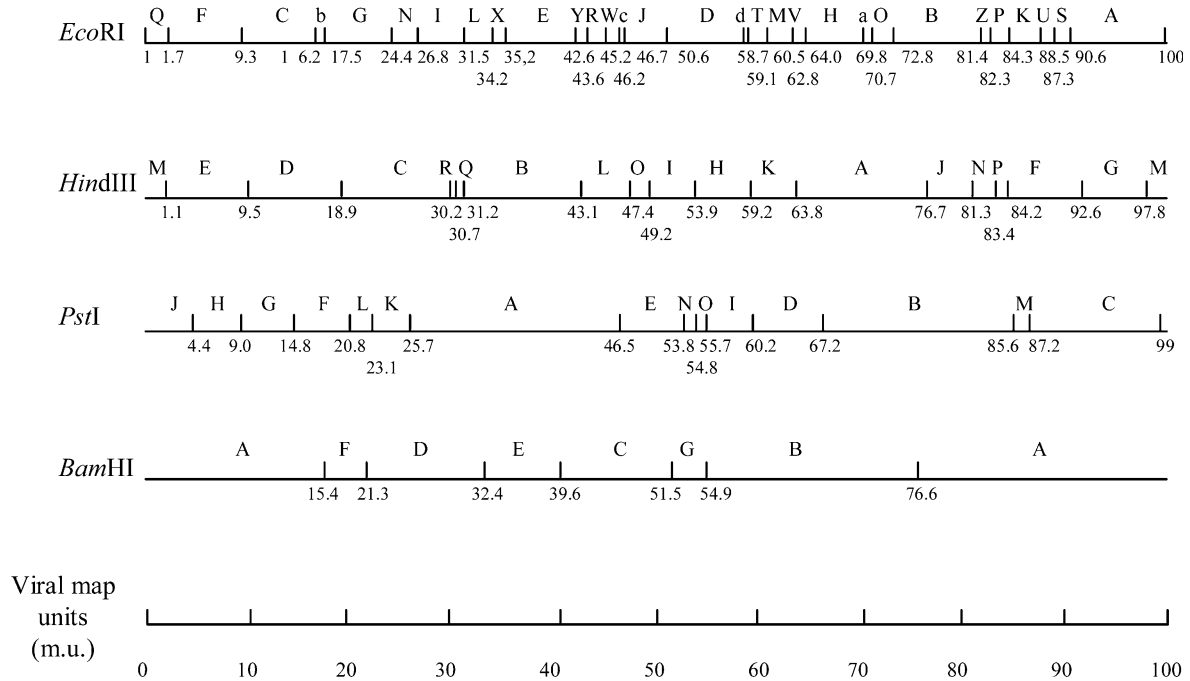

Fig. 2. Physical maps of the SfMNPV–NIC genome. Restriction maps for *Eco*RI, *Hin*dIII, *Pst*I and *Bam*HI are shown. The endonuclease *Sma*I did not cleave SfMNPV DNA (data not shown). The first nucleotide of the map is the first nucleotide of the *Eco*RI-Q fragment that carries the *polyhedrin* gene. The circular SfMNPV-NIC DNA is represented in linear form, and the site between *Eco*RI-A and *Eco*RI-Q was designated site 1. Numbers of map units (m.u.) representing restriction sites are indicating below the maps. The genome size was estimated at 129.3 kbp.

digestion. Fragments *Hin*dIII-R and *Hin*dIII-Q hybridized to *Eco*RI-I and *Pst*I-A suggesting that both fragments were adjacent to each other, but their relative position in the SfMNPV *Hin*dIII physical map was unclear. Both *Hin*dIII-L and *Hin*dIII-K fragments, which have similar sizes, hybridized to multiple *Eco*RI fragments of similar sizes, thus hindering the construction of the SfMNPV *Hin*dIII physical map.

The SfMNPV physical map obtained by restriction analysis and Southern blotting was confirmed and completed by using the sequence similarity of the termini of some of the cloned fragments. The termini of all *Hin*dIII cloned fragments as well as two *Eco*RI cloned fragments (*Eco*RI-N and *Eco*RI-X) and four *Pst*I cloned fragments (*Pst*I-G, *Pst*I-F, *Pst*I-L and *Pst*I-K) were sequenced. Terminal sequence information obtained was compared to the GenBank database using the FastBlast software. The most homologous ORFs identified are shown in Table 2. Most of the *Hin*dIII, *Eco*RI, and *Pst*I restriction sites analyzed fell within ORFs that showed similarity to 36 genes from SeMNPV, 37 genes from MacoMNPV, and 32 genes from SpliNPV. This allowed adjacent restriction fragments to be mapped together. Terminal sequence information provided further confirmation for the position of most of these fragments. For example, the *Hin*dIII-R fragment was located in the right part of the *Hin*dIII-C fragment because both fragments presented a contiguous region which comprised an ORF homologue to Se43 (Table 2). Similarly, an ORF homologue to Se44 located at the right terminal of *Hin*dIII-R and at the left terminal of *Hin*dIII-Q fragments and the *gp41* gene located at the right terminal of *Hin*dIII-H and at the left terminal of *Hin*dIII-K determined the relative position and orientation of these fragments in the physical map. We also determined that the *Hin*dIII-O fragment was contiguous and situated at the right part of the *Hin*dIII-L fragment since the ORF homologue to Se68 was located at the termini of both fragments.

Zero Point of the Physical Map

By convention, the zero point of NPV physical maps is determined by the *polh* gene [31,35]. For better comparisons with the SfMNPV physical map [13], we designated the *Eco*RI REN fragment containing the *polh* gene as the zero point of the linearized SfMNPV physical map. Sequence analysis revealed that the *polh* gene was contained completely in the *Hin*dIII-M fragment. This fragment hybridized to *Eco*RI-Q, *Pst*I-J and *Bam*HI-A. Thus, *Eco*RI-Q was confirmed as the zero point of the SfMNPV map. Based on the direction of transcription of this gene the *Eco*RI-F fragment was at the right end of *Eco*RI-Q.

Gene Content

From the terminal sequence information obtained, a total of 38 putative ORFs were identified in the SfMNPV genome (Table 2). Of these, 37 have homologous sequences in other baculoviruses and one (ORF15) appears to be a gene unique to SfMNPV (Table 2). Most of the ORFs were assigned on the basis of their similarity with previously described ORFs in the highly homologous SeMNPV. However, when the gene had been previously described, it was assigned the name previously reported for that gene. The 38 putative baculovirus genes were grouped into five categories: genes coding for structural proteins (*polh*, vp39, *vp80*, two copies of *odv-e66*, *pif*, *pif*-2, *p74*, *gp41* and the ORF homologous to ORF-8 of SeMNPV); DNA replication genes (*helicase*, *lef*-3, *lef*-4 or *lef*-8); genes for regulatory proteins (*p47*, ORF1629 or *ie*-0); genes with auxiliary functions (*fgf*, *egt* and *cathepsin*); genes involved in nucleotide metabolism (*rr*-1); repeated ORFs (*bro*); and ORFs with unknown function (ORF homologues to SeMNPV ORF28, Se28, Se29, Se30, Se37, Se43, Se44, Se67, Se68, Se69, Se79, Se81, Se89, Se90, Se109, Se113 and the ORF unique to SfMNPV). The direction of transcription of all these ORFs relative to that of the *polh* gene, as well as the early and late potential promoter motifs and polyadenylation signals that were localized, are summarized in Table 2.

The percentage of amino acid sequence identity and amino acid sequence similarity of the 38 ORFs identified were compared to the ORFs of other sequenced baculoviruses (Table 3). The most homologous baculovirus to SfMNPV was SeMNPV with 66% identity and 78% similarity. Among baculoviruses, polyhedrin (POLH) was the most homologous protein.

Table 2. The position and orientation of the 38 putative ORFs in the SfMNPV genome

ORF No.	Gene family	Genomic fragment	Most homologous ORF	Size/Position	Dir.*	Promoter Φ	Poly (A) Ψ
1	*polyhedrin*	*Hin*dIII-M	Se1	96/CT	>	ND	−2 +21
2	ORF 1629	*Hin*dIII-M/E	Se2	131/CT	<	ND	+7
3	Se8	*Pst*I-G	Se8	110/Central	>	ND	ND
4	*cath*	*Pst*I-G *Pst*I-F	Se16	77/Central 76/NT	<	TTAAG (−22)	ND ND
5	*egt*	*Hin*dIII-D/C	Se27	111/CT	>	ND	+32
6	Se28	*Hin*dIII-C *Pst*IF-L	Se28	82/NT 106/CT	>	CAGT (−33) TATA (−64)	+234
7	Se29	*Pst*-L	Se29	145/NT	>	ATAAG (−183)	ND
8	Se30	*Pst*I-L/K	Se30	276/CT	<	ND	ND
9	*pif*-2	*Eco*RI-N *Pst*I-K	Se35	286/NT 101/CT	>	CTAAG (−14)	ND
10	*pif*	*Eco*RI-N *Pst*I-K/A	Se36	529/Total	>	CTAAG (−14)	ND
11	Se37	*Eco*RI-N	Se37	85/Total	>	TTAAG (−234)	ND
12	*fgf*	*Eco*RI-N/I	Se38	300/CT	<	ND	ND
13	Se43	*Hin*dIII-C/R	Se43	194/NT	>	TTAAG (−52)	ND
14	Se44	*Hin*dIII-R/Q	Se44	125/NT	<	CTAAG (−89) CAGT (−38) TATA (−73)	ND
15	unique	*Hin*dIII-Q/B	Not homologous	247/CT	<	ND	ND
16	*odv-e66*	*Hin*dIII-B *Eco*RI-X	Lese *odv-e66*	249/Central	>	ND	ND
17	*vp80 capsid*	*Hin*dIII-B/L	Se61	392/Central	<	ND	ND
18	Se67	*Hin*dIII-L	Se67	73/CT	>	ND	+98
19	Se68	*Hin*dIII-L/O	Se68	151/Total	>	GTAAG (−14)	−18
20	*bro-e*	*Hin*dIII-O	Maco90	155/CT	<	ND	−18 +386
21	Se69 p19	*Hin*dIII-O	Se69	168/Total	<	TATA (−4) GATA (−28)	−2
22	*helicase*	*Hin*dIII-O/I	Se70	304/NT	>	ATAAG (−48)	ND
23	*lef-4*	*Hin*dIII-I/H	Se74	459/Total	<	CAGT (−7/−26) GATA (−149) CGT	ND
24	*vp39 capsid*	*Hin*dIII-H	Se75	66/NT	>	ATAAG (−24)	ND
25	Se79	*Hin*dIII-H	Se79	148/CT	>	ND	+89
26	*gp41*	*Hin*dIII-H/K	Se80	336/Total	>	ATAAG (−42)	ND
27	Se81	*Hin*dIII-K	Se81	28/NT	>	TTAAG (−58) CAGT (−33)	ND
28	Se89	*Hin*dIII-K	Se89	56/NT	<	CAGT (−45) TATA (−101)	ND
29	Se90	*Hin*dIII-K	Se90	124/Total	<	CAGT (−37)	ND
30	*lef-3*	*Hin*dIII-K/A	Se91	73/NT	>	CAGT (−245) TATA (−82)	ND
31	Se109	*Hin*dIII-A/J	Se109	165/Central	>	ND	ND
32	*lef-8*	*Hin*dIII-J/N	Se112	178/CT	>	ND	+217
33	Se113	*Hin*dIII-N	Se113	59/Total	<	ATAAG (−5)	ND
34	*odv-e66*	*Hin*dIII-N	Se114	171/CT	<	ND	+6
35	*p47*	*Hin*dIII-N/P	Se115	399/Total	>	ND	ND
36	*p74*	*Hin*dIII-F/G	Se131	251/Central	<	ND	ND
37	*ie*-0	*Hin*dIII-G	Se138	178/NT	<	ATAAG (−51)	ND
38	*rr*	*Hin*dIII-G/M	Se139	258/Central	<	ND	ND

◀

*Direction of transcription.
> Transcription is in the same sense as the *polyhedrin* gene.
< Transcription is the contrary sense to *polyhedrin* gene.
Φ Late (TAAG) and early (CAGT, TATA, GATA, CGT) promoters.
ND not possible to determine due to a lack of sequence information corresponding to the promoter region, or because no classical motifs were identified in the upstream sequence.
Ψ Polyadenylation signal comprising an AATAAA motif.

Table 3. Percentage of amino acid sequence identity and similarity (in parenthesis) to homologous ORFs of other baculoviruses

	ORF number/% Identity (% Similarity)								
Protein family	SeMNPV	MacoNPV	HaSNPV	SpltNPV	AcMNPV	LdMNPV	BmNPV	XcGV	CpGV
POLH	**1**/86 (87)	**1**/82 (86)	**1**/78 (83)	**1**/85 (86)	**8**/75 (85)	**1**/70 (75)	**1**/77 (83)	**1**	**1**
ORF 1629	**2**/57 (64)	**2**/39(53)	**2**	**2**	**9**	**2**	**2**	**2**	–
Se8	**8**/54 (71)	**9**/43 (61)	**133**/39 (53)	**136**/41 (64)	**23**	**130**/42 (59)	**14**	**27**/32 (48)	**31**/45 (58)
CATH	**16**/71 (71)	**33**/62 (62)	**56**/41 (67)	**54**/36 (67)	**127**/47 (67)	**78**/52 (65)	**104**/45 (61)	**58**/38 (67)	**11**
EGT	**27**/74 (91)	**39**/65 (83)	**126**/45 (60)	**121**/38 (60)	**15**/42 (61)	**135**/41 (63)	**7**/40 (62)	–	**141**
Se28	**28**/56 (69)	**40**/39 (66)	–	–	–	**127**/37 (57)	–	–	–
Se29	**29**/60 (74)	**41**/45 (65)	**128**/27 (50)	**119**/39 (62)	**17**/31 (55)	**128**/34 (54)	**9**/32 (59)	–	–
Se30	**30**/54 (77)	**42**/49 (70)	**129**/29 (51)	**118**/30 (51)	–	**129**/30 (53)	–	–	–
PIF-2	**35**/75 (86)	48/65 (76)	**132**/46 (69)	**135**/40 (69)	**22**/48 (70)	**119**/41 (64)	**13**/52 (71)	**45**/41 (67)	**48**/43 (64)
PIF	**36**/64 (77)	**49**/54 (68)	**111**/42 (58)	**124**/48 (64)	**119**/45 (61)	**155**/43 (59)	**97**/44 (61)	**84**/34 (51)	**75**/36 (53)
Se37	**37**/63 (79)	**50**/45 (65)	**112**	**123**	**120**	**155a**	**98**	–	–
FGF	**38**/44 (57)	**51**/40 (55)	**113**/32 (54)	**122**/31 (60)	**32**/26 (54)	**156**/27 (49)	**24**/28 (51)	**144**	**123**
Se43	**43**/64 (80)	**56**/55 (73)	–	**113**/34 (56)	**18**/24 (41)	**158**/23 (400	**10**/23 (41)	–	–
Se44	**44**/56 (82)	**57**/47 (70)	–	–	–	–	–	–	–
unique	–	–	–	–	–	–	–	–	–
ODV-E66	**57**/46 (54) **114**/39 (52)	**78**	**96**/63 (72)	**98**/43 (64)	**46**/49 (63)	**131**/63 (75)	**37**/49 (63)	**149**/60 (74)	**37**/50 (64)
VP80	**61**/59 (60)	**82**/38 (56)	**92**/24 (41)	**94**/32 (56)	**104**/24 (38)	**105**/30 (57)	**88**/24 (38)	–	–
Se67	**67**/78 (92)	**88**/72 (88)	**86**/60 (80)	**88**/64 (81)	**98**/55 (75)	**99**/68 (81)	**82**/51 (73)	**96**/48 (75)	**88**/45 (70)
Se68	**68**/45 (58)	**43**/33 (50)	**83**	–	–	–	–	–	–
BRO	– **122**/37 (56)	**90**/44 (61)	**105**/24 (45)	**128**/28 (53)	**2**/43 (64)	**153**/37 (58)	**131**/33 (55)	**114**/22 (43)	–
Se69 p19	**69**/77 (87)	**92**/70 (81)	**85**/57 (74)	**87**/52 (80)	**96**/56 (76)	**98**/59 (75)	**79**/57 (76)	**97**/36 (61)	**89**/41 (54)
HELICASE	**70**/55 (67)	**93**/53 (66)	**84**/37 (50)	**86**/28 (45)	**95**/32 (51)	**97**/37 (55)	**78**/33 (52)	**98**/40 (58)	**90**/26 (42)
LEF-4	**74**/74 (84)	**98**/62 (75)	**79**/50 (66)	**82**/45 (65)	**90**/45 (61)	**93**/49 (64)	**73**/45 (62)	**110**/34 (50)	**95**/31 (45)
VP39	**75**/81 (89)	**99**/60 (75)	**78**/50 (66)	**81**/40 (59)	**89**/50 (59)	**92**/53 (71)	**72**/52 (60)	**111**/45 (53)	96/38 (52)
Se79	**79**/73 (83)	**103**/56 (76)	**74**/59 (75)	**77**/55 (75)	**81**/53 (69)	**89**/52 (69)	**67**/56 (70)	**120**/50 (64)	103/46 (66)
GP41	**80**/94 (94)	**104/**79 (90)	**73**/57 (75)	**76**/86 (88)	**80**/55 (74)	**88**/56 (74)	**66**/55 (73)	**121**/37 (59)	104/31 (53)
Se81	**81**/100(100)	**105**/67 (74)	**72**/70 (88)	**75**/73 (88)	**78**/74 (81)	**87**/72 (80)	**64**/74 (81)	**122**	
Se89	**89**/52 (74)	**111**/40 (62)	**63**	–	**69**	–	**57**	–	
Se90	**90**/81 (92)	**112**/77 (87)	**66**/60 (77)	**64**/54 (75)	**68**/46 (65)	**80**/49 (67)	**56**/47 (65)	**135**/40 (60)	114/30 (62)
LEF-3	**91**/70 (78)	**113**/56 (77)	**65**	**67**	**67**	**81**	**55**	–	
Se109	**109**/46 (62)	**138**/46 (60)	**42**/28 (43)	**44**/29 (14)	**52**	**53**/28 (45)	**41**/20 (40)	–	
LEF-8	**112**/96 (98)	**141**/80 (87)	**38**/83 (90)	**38**/80 (90)	**50**/78 (86)	**51**/74 (87)	**39**/80 (87)	**148**/62 (76)	131/61 (76)
Se113	**113**/66 (79)	**143**/49 (81)	**37**/29 (58)	–	**43**/38 (63)	–	**34**/34 (59)	**147**	
ODV-E66	**114**/55 (70) **57**/30 (51)	**144**/35 (56)	**96**/30 (56)	**98**/34 (55)	**46**	**131**	**37**	**149**/33 (55)	
P47	**115**/82 (91)	**145**/75 (87)	**35**/59 (76)	**36**/56 (71)	**40**/52 (68)	**48**/63 (78)	**31**/52 (68)	**78**/47 (66)	68/45 (61)
P74	**131**/60 (72)	**160**/52 (66)	**20**/48 (61)	**21**/42 (62)	**138**/47 (64)	**27**/46 (64)	**115**/46 (66)	**77**/30 (46)	60/44 (58)
IE-0	**138**/72 (89)	**168**/59 (78)	**8**/31 (53)	**8**/34 (59)	**141**/24 (52)	**21**/31 (62)	**117**/24 (51)	–	
RR	**139**/34 (56)	**169**	–	**23**	–	**148**	–	–	
Total %Id (Sim)	**66 (78)**	**55 (71)**	**46 (64)**	**46 (65)**	**46 (64)**	**47 (64)**	**45 (63)**	**40 (60)**	**41 (60)**

The homologous ORF numbers are represented in bold.

Gene Order

To investigate whether the genome organization in SfMNPV was co-linear with the genome of other baculoviruses, a comparison between the SfMNPV and seven previously sequenced nucleopolyhedroviruses was performed (Fig. 3). Although a limited number of SfMNPV ORFs have been mapped to date, these ORFs cover practically all the genome and it is readily apparent that the gene order of the SfMNPV genome shares greater similarity with those of SeMNPV and MacoNPV than with any other NPVs included in this study (Fig. 3). The gene organization is most conserved in the central region (30–70 m.u.) of the linearized baculovirus genomes. The left part of the SfMNPV genome (1–30 m.u.) displays considerable differences to those of AcMNPV, BmNPV, OpMNPV, but not to SeMNPV. The right part (70–100 m.u.) of the SfMNPV genome also shows a high degree of homology with SeMNPV. The orientation of the partial or completely sequenced SfMNPV genes appears to be the same as those of SeMNPV. The gene order of SfMNPV and SeMNPV is practically the same; the main observable difference between both NPVs is that the genes appear to be displaced to the right in the SeMNPV genome.

Phylogenetic Analysis

To determine the relationship between SfMNPV and the other NPVs, two different approaches were employed. First, a phylogenetic study was performed using fifteen protein sequences (POLH, PTP-2, LEF-4, P47, EGT, GP41, PIF and the ORFs homologues to Se028, Se029,

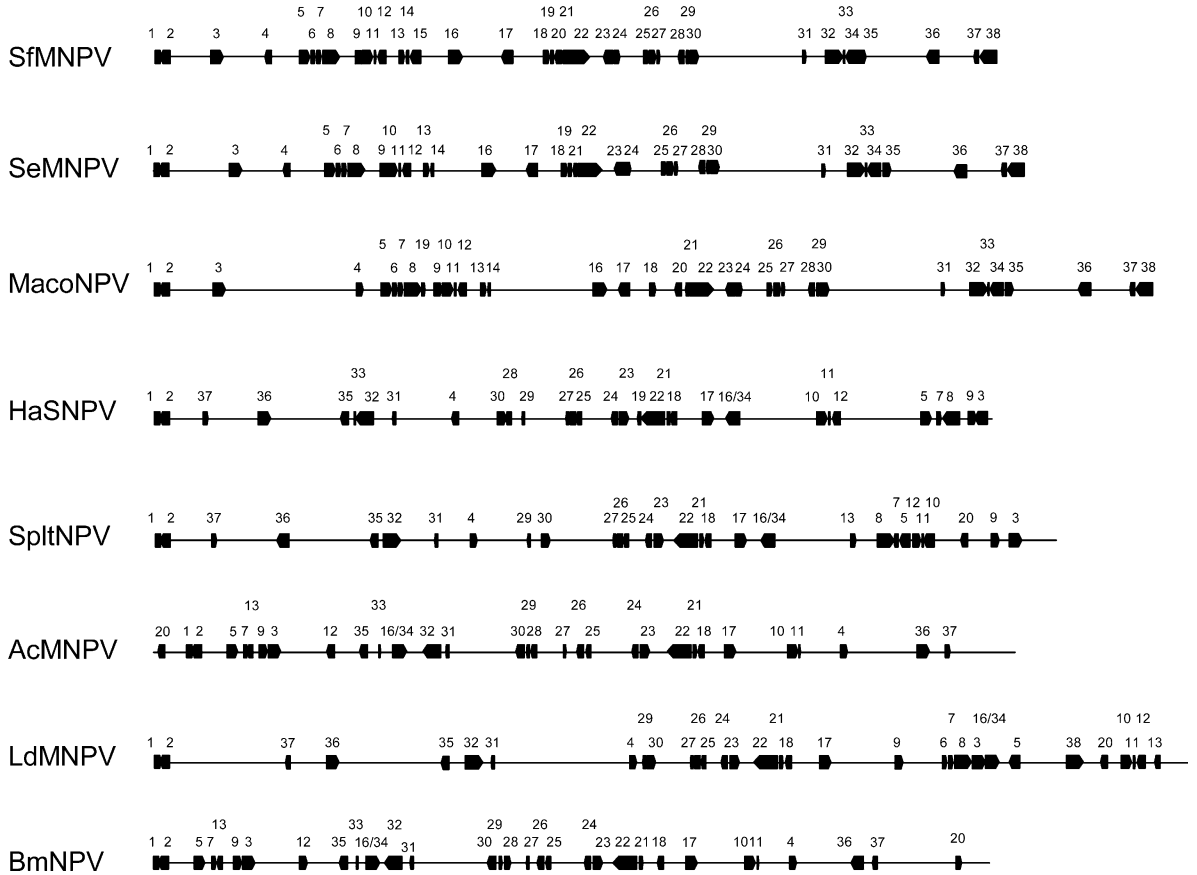

Fig. 3. A comparison of SfMNPV determined ORFs with their homologues in SeMNPV, MacoNPV, HaSNPV, SpltNPV, AcMNPV, LdMNPV and BmNPV. Black boxes with arrowheads indicate location and direction of transcription of the SfMNPV and other NPVs ORFs compared. The numbers above each ORF correspond to ORFs described in Table 2. The sizes of SfMNPV ORFs that were not completely sequenced were assumed to be equal to the homologous SeMNPV ORF.

Se030, Se037, Se068, Se069, Se090, Se113), derived from the genomes of 22 NPVs and 5 GVs (Fig. 4). Second, the gene content and gene order of the SfMNPV genome were analyzed. The protein-based phylogeny was examined to determine if it was congruent with the results on the genomic organization and other biological characteristics of these baculoviruses.

All phylogenetic trees (Fig. 4 and http://viradium.mpl.ird.fr/tredyn/Simon2005) agree with the separation of NPVs and GVs. In all trees but one (*lef4*), GVs originate from a single branch. The analysis of multiple trees also confirms the separation of group I and group II NPVs. In all trees except the *polh* tree, group I NPVs originate from a single branch. Sequence analysis of the PTP-2 protein revealed that the two *ptp-2* genes in CpGV are not closely related. For all genes examined, SfMNPV clusters close to SeMNPV. The average distances between those two viruses are slightly greater that between the two *Mamestra configurata* NPVs.

The use of a parallel analysis for 15 genes revealed apparent differences in the evolutionary rate, both between genes and between virus groups. POLH is a very conserved protein (except in CuniNPV), but LEF4 or SF68 are far less conserved. In general, the distances between viruses are greater between GVs than between NPVs.

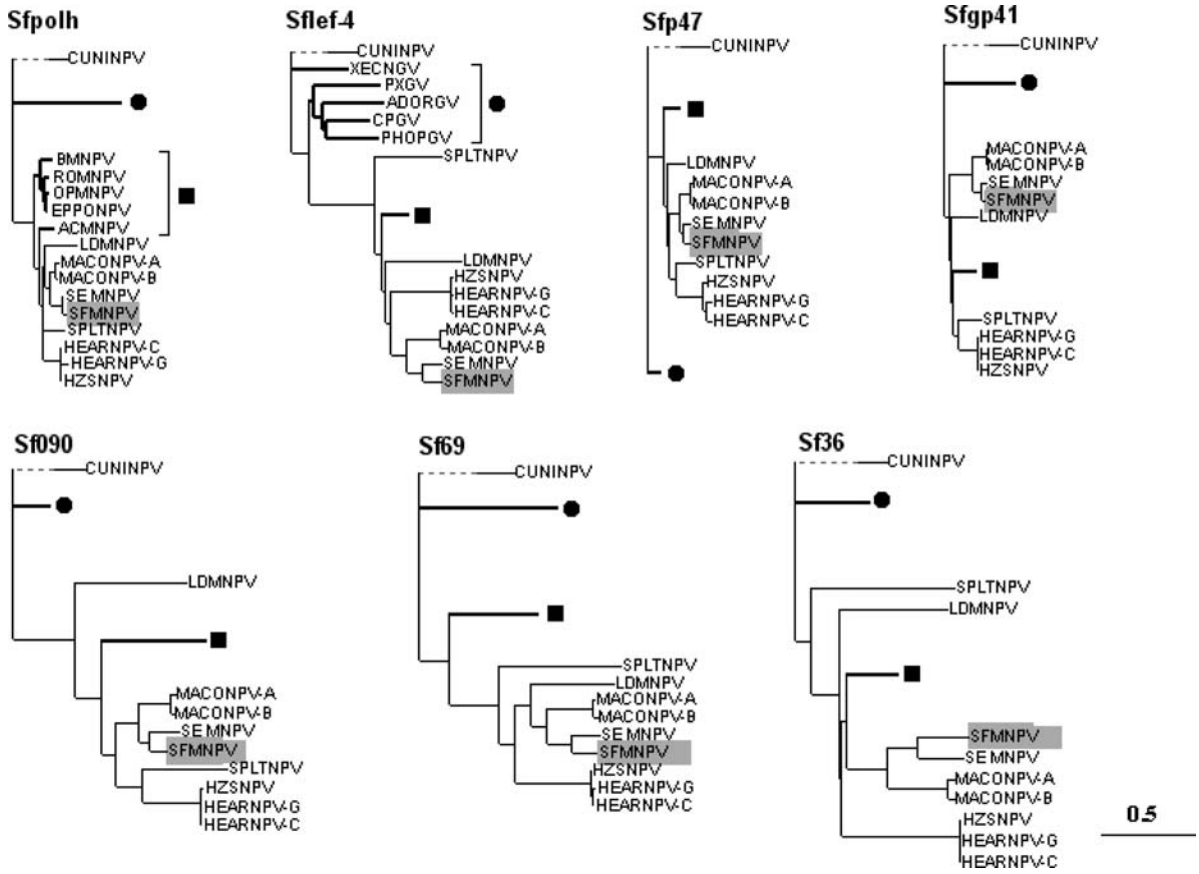

Fig. 4. Phylogenetic analysis of baculovirus sequences. Fifteen amino acid sequence families were studied using the ClustalX software for alignment and the PHYLIP package for distance analysis (Prodist program with Dayhoff PAM matrix and Neighbor with CuniNPV sequence used as an outgroup when available). Graphical analyses were performed using the TreeDyn editor. ORF sequence names are shown as their respective abbreviated virus species names. Common branchings were collapsed to the first intra-group branch: a circle represents all the GVs, whereas a square represents all group I NPVs. The complete trees can be found at http://viradium.mpl.ird/treedyn/Simon2005/. All trees are drawn to the same scale. The outgroup branch length (indicated by a dashed line) is not proportional. A radial conformation was used for trees without an outgroup. SfMNPV and SeMNPV were located together in all trees.

Discussion

The size of the SfMNPV genome estimated in this study was 129.3 kb, in very close agreement with our previous estimates [12]. This size estimate is larger than the estimate of 121.76 kb for the Sf-2 genomic variant plaque-purified from a SfMNPV wild-type isolate from the United States [13]. This difference was due to a deletion of approximately 8 kb in the Sf-2 genome, located between 14.8 and 21.0 m.u. in the physical map described in this study. The size of the SfMNPV genome estimated in this study is similar to some of the better studied NPVs, including AcMNPV (133 kb) [8] or SeMNPV (135 kb) [36]. The occurrence of geographical variants has been demonstrated for several other NPVs [12,25,37]. Such studies have generally indicated that geographical variants are related strains of the same virus that show limited differences in the presence and distribution of restriction cleavage sites.

By convention, the orientation of the physical map is set by the location and the orientation of the *polh* gene [31]. Terminal sequence information revealed the presence of the *polh* gene in the restriction fragment *Hin*dIII-M that hybridized with *Eco*RI-Q. This partial sequence allowed the orientation of the physical map. In Sf-2, *polh* is located in *Eco*RI-P, due to the deletion observed in this variant. However, the physical map orientation of Sf-2 [13] was opposite to that of our map. These authors determined the location of the *polh* gene in Sf-2 by hybridization, using the AcMNPV *polh* gene as a probe [13], so they were unable to determine the orientation of the gene. Consequently, although the alignment of the restriction fragments was practically identical, their orientation was opposite.

A total of 27.92 kb were sequenced, which represent 21.5% of the whole SfMNPV genome. This sequence information was compared with those of SeMNPV [36], MacoNPV 90/2 (MacoNPV-A) [38], MacoNPV-96B (MacoNPV-B) [39], HzSNPV [40] HaNPV-C1 (Zhang, C.X. and Jin, W.R. 2000. GenBank accession NC_003094), HaNPV-G4 [41], SpltNPV [42], AcMNPV [8], LdMNPV [43], BmNPV [44], OpMNPV [45], EppoNPV [46], RoNPV (Bonning, B.C. and Harrison, R.L. 2002. GenBank accession NC_004323) and those of the granuloviruses of *Xecn*GV [47], PxGV [48], AdorGV [49], PhopGV (Croizier, L., Taha, A., Croizier, G. and López-Ferber, M. 2002. GenBank accession NC_004062), CuniNPV [33] and *Cp*GV [50]. SfMNPV, SeMNPV and MacoNPV displayed the highest sequence identities (66 and 55%, respectively) and gene arrangements (78 and 71%, respectively), with these three NPVs sharing 36 common putative ORFs. There are a number of other common features of these three viruses. One of the most interesting is the presence of two homologues of *odv-e66*. Terminal sequencing confirmed that the first *odv-e66* copy was located in the *Hin*dIII-N fragment and showed high homology to Se114. By probing the genomic *Hin*dIII fragments with the *Hin*dIII-N fragment, that only included the *odv-e66* gene, a second copy of *odv-e66* was located in the *Hin*dIII-B fragment (also present in the *Eco*RI-X fragment). The hybridization signal from the *Hin*dIII-N fragment was significantly stronger than that observed in the *Hin*dIII-B fragment. Terminal sequencing of the *Eco*RI-X fragment indicated that the second *odv-e66* copy had a higher homology to Se57, and to *odv-e66* of LeseNPV, than to Se114. The *odv-e66* gene located in the *Hin*dIII-N fragment presented 55/60% identity/similarity with Se114, compared to 30/51% identity/similarity with Se114. In contrast, the *odv-e66* located in *Hin*dIII-B fragment was most homologous to LeseNPV *odv-e66* (AB009613) with 64/76% identity/similarity, and presented lower homology with the two *odv-e66* genes of SeMNPV, with 46/54% identity/similarity to Se57 and 39/52% identity/similarity to Se114. Despite the lower homology to the SeMNPV *odv-e66* ORF, the positions of the two copies of SfMNPV *odv-e66* genes in relation to nearby ORFs were the same as described in SeMNPV and MacoNPV (Fig. 3). It is probable that the two copies of the *odv-e66* gene were acquired independently and the SfMNPV ORF homologue to Se57 (Maco78) originated from a source that was more closely related to LeseNPV and LdMNPV than to the SfMNPV ORF homologue to Se114 (Maco144). Similar observations have been reported for the two copies of *odv-e66* in SeMNPV [36] and MacoNPV [38].

For all genes analyzed, SfMNPV, SeMNPV, MacoNPV-A and -B clustered together. In addition, an ORF homologue to Se044 was present in SfMNPV. Such ORFs have only been described in

these four viruses. This suggests that the NPV species SeMNPV, SfMNPV and MacoMNPV may form a group distinguishable from other baculoviruses.

SeMNPV and MacoNPV lack a homologue of the budded virus (BV) surface glycoprotein gene *gp64*. This is characteristic of group II NPVs including LdMNPV [43]. An ORF homologue to Se8 was identified in the sequenced fragments of the SfMNPV genome which also showed homology to Ld130 [43]. This suggests that SfMNPV probably lacks the *gp64* gene.

One unique SfMNPV ORF was located between *Hin*dIII-B and *Hin*dIII-Q fragments and was negatively orientated compared to the *polh* gene. It presents homology with hypothetical proteins of *Arabidopsis thaliana* and *Oryza sativa* (26% amino acid identity) and also with transcriptional regulators of the LysR family from *Sinorhizobium meliloti* (25% amino acid identity). Characterization of the genes that are found in only one or a few members of the family define the individuality of each virus and will influence individual phenotypic traits such as host or tissue tropism, and virulence. The genomic organization, i.e. the order of genes, is similar in OpMNPV, BmNPV and AcMNPV except for a small number of rearrangements [8,44,45], whereas SeMNPV presents a highly characteristic and distinct genome organization compared with these other NPVs [36]. The gene order of SfMNPV is practically the same as that of SeMNPV, but in the SeMNPV genome the genes appear to be slightly displaced to the right. This probably will be due to the presence of certain SeMNPV genes that are absent in SfMNPV, resulting in a larger SeMNPV genome.

Comparison of the relative gene order between SfMNPV and the other baculoviruses revealed the presence of certain clusters that are conserved in all the baculovirus genomes [36,38]. Using the information obtained by partial sequencing of SfMNPV, we conclude that clusters 1, 5, 9 and 16 [36] are also conserved in SfMNPV, whereas cluster 8 is interrupted in SfMNPV by the insertion of the *bro* gene that is not present between the ORF homologues to Se68 and Se69 in the other baculoviruses for which genome sequences are available. Additional clusters were identified when comparing SfMNPV with SeMNPV. One cluster includes the ORF homologues to Se35, Se36, Se37 and Se38, whereas in the genomes of other baculoviruses, these genes are divided in two different clusters. Another cluster comprises the ORFs homologous to Se112, Se113, Se114 and Se115. Furthermore, clusters 6 and 16 are extended to include genes homologous to Se27 and Se44, respectively. The additional and the enlarged clusters of SfMNPV and SeMNPV suggest that the genomic organization of SfMNPV is more closely related to that of SeMNPV than to AcMNPV, LdMNPV, SpltNPV or even MacoNPV. This agrees with the phylogenetic analysis.

Different phylogenetic trees have been constructed for the baculovirus based on single proteins, usually polyhedrin. Until recently, this was the only gene characterized for SfMNPV. Single gene phylogenies led to inconsistencies in the definition of the NPV group II. Disagreements between different single-gene analyses may reflect inaccurate phylogenetic inferences due to unequal rates of evolution or due to lack of a robust phylogenetic signal. Alternatively, they could indicate genuine differences in the phylogeny of individual genes as a result of recombination events, duplications and gene losses [51].

An alternative is the analysis of linked multiple trees. Such multiple visualization facilitates the detection of branching inconsistencies and assists in the comparison of putative evolutionary rates. Using the TreeDyn [34] multiple tree editor for the *polh* gene, we detected the recently described hybrid character of the AcMNPV *polh* gene [52]. As a result, a single branch is not observed for Group I NPVs for that gene. A similar discrepancy appears in the *lef-4* tree. Figure 4 shows the rooted trees in a schematic form. The complete rooted trees and the six unrooted trees can be found at http://viradium.mpl.ird/treedyn/Simon2005/. An additional approach is to consider the conservation of gene order [2]. To circumvent the problem of conflicting single-gene trees, entire genomes have been used to reconstruct baculovirus phylogenies [2,53]. The rationale for using such characters for phylogenetic inference is that two genomes sharing similar organization are more likely to have inherited that organization from a common ancestor than by evolutionary change. In the case of SfMNPV and SeMNPV, the overall homology detected from our work and from

previously published sequences pointed to a close relationship between the two viruses. Studying gene order reveals their true relatedness as distinct virus species with a probable recent common ancestor.

The geographic distributions of the natural hosts for SfMNPV and SeMNPV overlap significantly in the tropical and subtropical regions of the Americas [10,54]. This suggests that SfMNPV and SeMNPV might have co-infected hosts and exchanged gene sequences. In a recent study, we observed that SeMNPV was able to enter and to replicate in *S. frugiperda* larvae [55], but did not cause lethal polyhedrosis disease. This contrasts with the generally accepted notion that SeMNPV exhibits an extremely restricted host range, limited to its own host species, *S. exigua* [56]. In that study, a positive replication signal from the *polh* gene indicated that all genes required for viral DNA replication in heterologous hosts of the genus *Spodoptera* [57], were present in the SeMNPV genome. It is likely that SeMNPV replication genes are homologous with those of SfMNPV, as SfMNPV and SeMNPV are closely related species. This means that both NPVs could replicate in the same host, thus facilitating mutual gene exchange.

In summary, phylogenetic analysis of 15 viral genes, partial gene content and gene order, revealed that the genome of SfMNPV is distinct from those of other baculoviruses, both in gene content and arrangement. Two copies of *odv-e66* are present in the genome, as in SeMNPV, but these NPVs differ in that SfMNPV presents at least one *bro* gene absent in SeMNPV. Putative protein sequences present high values of identity/similarity with those of SeMNPV, both in amino acid and nucleotide sequences. Finally, the gene order is practically the same in both viruses. These observations suggest that SfMNPV and SeMNPV may have a recent common ancestor.

Acknowledgments

We thank J. Jolivet, N. Grard (INRA) and I. Maquirriain (UPNA) for technical assistance and B. Limier (INRA) and N. Gorria (UPNA) for insect rearing. O. S. received a Spanish Ministry of Education and Culture studentship.

References

1. van Regenmortel M.H.V., Fauquet C.M., Bishop D.H.L., Carstens E.B., Estes M.K., Lemon S.M., Maniloff J., Mayo M.A., McGeoch D.J., Pringle C.R., and Wickner R.B. (eds.), *Virus Taxonomy: The Classification and Nomenclature of Viruses*. The Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, 2000.
2. Herniou E.A., Olszewski J.A., Cory J.S., and O'Reilly D.R., Annu Rev Entomol *48*, 211–234, 2003.
3. Hayakawa T., Rohrmann G.F., and Hashimoto Y., Virology *278*, 1–12, 2000.
4. Fuxa J.R., Biotechnol Adv *9*, 425–442, 1991.
5. Moscardi F., Ann Rev Entomol *44*, 257–289, 1999.
6. López-Ferber M., Sisk W.P., and Possee R.D., Methods Mol Biol *39*, 25–63, 1995.
7. Fernandez J.M. and Hoeffler J.P., *Gene Expression Systems: Using Nature for the Art of Expression*. Academic Press, San Diego, 1999, pp. 332–359.
8. Ayres M.D., Howard S.C., Kuzio J., López-Ferber M., and Possee R.D., Virology *202*, 586–605, 1994.
9. Gardner W.A. and Fuxa J.R., Fla Entomol *63*, 439–447, 1980.
10. Sparks A.N., Fla Entomol *62*, 82–87, 1979.
11. Shapiro D.I., Fuxa J.R., Braymer H.D., and Pashley D.P., J Invertebr Pathol *58*, 96–105, 1991.
12. Escribano A., Williams T., Goulson D., Cave R.D., Chapman J.W., and Caballero P., J Econ Entomol *92*, 1079–1085, 1999.
13. Maruniak J.E., Brown S.E., and Knudson D.L., Virology *136*, 221–234, 1984.
14. Knell J.D. and Summers M.D., Virology *112*, 190–197, 1981.
15. Loh L.C., Hamm J.J., and Huang E.S., J Virol *38*, 922–931, 1981.
16. Kelly D.C., Virology *76*, 468–471, 1977.
17. Gonzalez M.A., Smith G.E., and Summers M.D., Virology *170*, 160–175, 1989.
18. Liu J.C. and Maruniak J.E.1995J Gen Virol7614431450.
19. Tumilasci V.F., Leal E., Zanotto P.M., Luque T., and Wolff J.L., Virus Genes *27*, 137–144, 2003.
20. Williams T., Goulson D., Caballero P., Cisneros J., Martínez A.M., Chapman J.W., Roman D.X., and Cave R.D., Biol Control *14*, 67–75, 1999.
21. Simon O., Williams T., López-Ferber M., and Caballero P., Appl Environ Microbiol *70*, 5579–5588, 2004.
22. López-Ferber M., Simon O., Williams T., and Caballero P., Proc R Soc Lond B Biol Sci *270*, 2249–2255, 2003.
23. Greene G.L., Leppla N.C., and Dickerson W.A., J Econ Entomol *69*, 487–488, 1976.
24. Hughes P.R. and Wood H.A., J Invertebr Pathol *37*, 154–159, 1981.
25. Caballero P., Zuidema D., Santiago-Alvarez C., and Vlak J.M., Biocontrol Sci Technol *2*, 145–157, 1992.
26. Croizier G. and Ribeiro C.P., Virus Res *26*, 183–196, 1992.
27. Maniatis T., Fritsch E.F., and Sambrook J., *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1982.
28. Pearson W.R., Methods Enzymol *183*, 63–98, 1990.

29. Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J., J Mol Biol *215*, 403–410, 1990.
30. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., and Higgins D.G., Nucleic Acids Res *24*, 4876–4882, 1997.
31. Vlak J.M. and Smith G.E., J Gen Virol *41*, 1118–1121, 1982.
32. Felsenstein J., PHYLIP: Phylogenetics Inference Package, Version 3.5c., Seattle, Department of Genetics, University of Washington, 1993.
33. Afonso C.L., Tulman E.R., Lu Z., Balinsky C.A., Moser B.A., Becnel J.J., Rock D.L., and Kutish G.F., J Virol *75*, 11157–11165, 2001.
34. Chevenet F., TreeDyn: a dynamic graphics tree editor V1.62. http://viradium.mpl.ird.fr/treedyn, 2003.
35. Smith I.R. and Crook N.E., Virology *166*, 240–244, 1988.
36. Ijkel W.F., van Strien E.A., Heldens J.G., Broer R., Zuidema D., Goldbach R.W., and Vlak J.M., J Gen Virol *80*, 3289–3304, 1999.
37. Laitinen A.M., Otvos I.S., and Lewis D.B., J Econ Entomol *89*, 640–647, 1996.
38. Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., and Erlandson M., Virology *294*, 106–121, 2002.
39. Li L., Donly C., Li Q., Willis L.G., Keddie B.A., Erlandson M.A., and Theilmann D.A., Virology *297*, 226–244, 2002.
40. Chen X., Zhang W.J., Wong J., Chun G., Lu A., McCutchen B.F., Presnail J.K., Herrmann R., Dolan M., Tingey S., Hu Z.H., and Vlak J.M., J Gen Virol *83*, 673–684, 2002.
41. Chen X., Ijkel W.F.J., Tarchini R., Sun X., Sandbrink H., Wang H., Peters S., Zuidema D., Lankhorst R.K., Vlak J.M., and Hu Z., J Gen Virol *82*, 241–257, 2001.
42. Pang Y., Yu J., Wang L., Hu X., Bao W., Li G., Chen C., Han H., Hu S., and Yang H., Virology *287*, 391–404, 2001.
43. Kuzio J., Pearson M.N., Harwood S.H., Funk C.J., Evans J.T., Slavicek J.M., and Rohrmann G.F., Virology *253*, 17–34, 1999.
44. Gomi S., Majima K., and Maeda S., J Gen Virol *80*, 1323–1337, 1999.
45. Ahrens C.H., Russell R.L., Funk C.J., Evans J.T., Harwood S.H., and Rohrmann G.F., Virology *229*, 381–399, 1997.
46. Hyink O., Dellow R.A., Olsen M.J., Caradoc-Davies K.M., Drake K., Herniou E.A., Cory J.S., O'Reilly D.R., and Ward V.K., J Gen Virol *83*, 957–971, 2002.
47. Hayakawa T., Ko R., Okano K., Seong S.I., Goto C., and Maeda S., Virology *262*, 277–297, 1999.
48. Hashimoto Y., Hayakawa T., Ueno Y., Fujita T., Sano Y., and Matsumoto T., Virology *275*, 358–372, 2000.
49. Wormleaton S., Kuzio J., and Winstanley D., Virology *311*, 350–365, 2003.
50. Luque T., Finch R., Crook N., O'Reilly D.R., and Winstanley D., J Gen Virol *82*, 2531–2547, 2001.
51. Page R.M.D., Mol Phylogenet Evol *14*, 89–106, 2000.
52. Jehle J.A., Virus Genes *29*, 5–8, 2004.
53. Herniou E.A., Luque T., Chen X., Vlak J.M., Winstanley D., Cory J.S., and O'Reilly D.R., J Virol *75*, 8117–8126, 2001.
54. Balachowsky A.S., (ed.), *Famille des Noctuidae*. Masson et Cie, Paris, 1972, pp. 1255–1520.
55. Simon O., Williams T., López-Ferber M., and Caballero P., J Gen Virol *85*, 2845–2855, 2004.
56. Heldens J.G., Liu Y., Zuidema D., Goldbach R.W., and Vlak J.M., Virus Res *55*, 187–198, 1998.
57. Kool M., Voeten J.T., Goldbach R.W., and Vlak J.M., Virology *198*, 680–689, 1994