The Complete Sequence of the First Spodoptera frugiperda Betabaculovirus Genome: A Natural Multiple Recombinant Virus

Paola E. Cuartas 1, Gloria P. Barrera 1,* Mariano N. Belaich 2, Emiliano Barreto 3, Pablo D. Ghiringhelli 2,† and Laura F. Villamizar 1,†

1 Centro de investigación Tibaitatá, Corporación Colombiana de Investigación Agropecuaria CORPOICA, Km 14 Vía Mosquera 250047, Cundinamarca, Colombia; E-Mails: pcuartas@corpoica.org.co (P.E.C.); lvillamizar@corpoica.org.co (L.F.V.)
2 Laboratorio de Ingeniería Genética y Biología Celular y Molecular—Área Virosis de Insectos (LIGBCM—AVI), Dto. de Ciencia y Tecnología, Universidad Nacional de Quilmes, Roque Saenz Peña 352, Bernal, Provincia de Buenos Aires, 1876, Argentina; E-Mails: mbelaich@unq.edu.ar (M.N.B.); pdg@unq.edu.ar (P.D.G.)
3 Centro de Bioinformática, Instituto de Biotecnología, Universidad Nacional de Colombia. Avenida Carrera 30 # 45, Bogotá 11001000, Cundinamarca, Colombia; E-Mail: ebarretoh@unal.edu.co

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: gbarrera@corpoica.org.co; Tel.: +57-1-4227300 (ext. 1328-1476); Fax: +57-1-4227300 (ext. 1000).

Academic Editor: John Burand and Madoka Nakai

Received: 28 November 2014 / Accepted: 26 December 2014 / Published: 20 January 2015

Abstract: Spodoptera frugiperda (Lepidoptera: Noctuidae) is a major pest in maize crops in Colombia, and affects several regions in America. A granulovirus isolated from S. frugiperda (SfGV VG008) has potential as an enhancer of insecticidal activity of previously described nucleopolyhedrovirus from the same insect species (SfMNPV). The SfGV VG008 genome was sequenced and analyzed showing circular double stranded DNA of 140,913 bp encoding 146 putative ORFs that include 37 Baculoviridae core genes, 88 shared with betabaculoviruses, two shared only with betabaculoviruses from Noctuid insects, two shared with alphabaculoviruses, three copies of own genes (paralogs) and the other 14 corresponding to unique genes without representation in the other baculovirus species. Particularly, the genome encodes for important virulence factors such as 4 chitinases and 2 enhancins.
The sequence analysis revealed the existence of eight homologous regions (hrs) and also suggests processes of gene acquisition by horizontal transfer including the SfGV VG008 ORFs 046/047 (paralogs), 059, 089 and 099. The bioinformatics evidence indicates that the genome donors of mentioned genes could be alpha- and/or betabaculovirus species. The previous reported ability of SfGV VG008 to naturally co-infect the same host with other virus show a possible mechanism to capture genes and thus improve its fitness.

Keywords: baculovirus; granulovirus; Spodoptera frugiperda; genome; recombination

1. Introduction

The fall armyworm (FAW), Spodoptera frugiperda (J.E. Smith, 1797) (Lepidoptera: Noctuidae) is a polyphagous insect of wide geographical distribution, considered the most important pest in maize (Zea mays L.) in the Americas [1]. The larvae consume the plant whorl affecting its growth, and complete defoliation could arise when epizooties occur. Control of S. frugiperda in maize crops includes the integration of cultural, physical, biological and chemical methods [2]. However, the control with broad spectrum and highly toxic synthetic chemical insecticides (categories I and II) is the main method for reducing the effects of the pest [3]. To reduce the damage and maintain pest levels below the economic threshold, some strategies, such as biological control, have been developed through the use of entomopathogenic viruses, mainly from Baculoviridae family or baculovirus, principally of the genus Alphabaculoviridae [4–6].

Baculoviridae is a family comprising rod-shaped viruses that infects members of the Phylum Arthropoda. This family includes four genera: Alphabaculovirus [lepidopteran-specific nucleopolyhedroviruses (NPVs)], Betabaculovirus [lepidopteran-specific Granuloviruses (GVs)], Gammabaculovirus (hymenopteran-specific NPVs) and Deltabaculovirus (dipteran specific NPVs) [7–9]. Their genomes vary in size from approximately 81.7 to 178.7 kbp according to available information, are circular covalently closed double stranded DNA (cccdsDNA) and encode ~90 to ~180 open reading frames (ORFs) [10]. The viral cycle presents a biphasic infection process generating progeny with two different phenotypes: budded viruses (BVs), which are produced at the initial stage of the multiplication cycle and are responsible for systemic infection inside the insect host, and occlusion-derived viruses (ODVs), which are produced in the last stage of the cycle within the infected cell and are required for the primary infection that takes place in the midgut epithelium cells of the insect host [9,11,12]. Mature ODVs are finally occluded in a protein matrix (mainly conformed by polyhedrin in alpha-, gamma- and deltabaculoviruses; or mainly conformed by granulin in betabaculoviruses) to form occlusion bodies (OBs), which protect them from the environment [13,14].

The use of mixtures containing granulovirus and nucleopolyhedrovirus has been studied in order to observe a synergic effect in the co-infections and in some cases granuloviruses are able to enhance the infectivity and virulence of NPVs [15,16]. This could be used as strategy for development of a biopesticide to control the FAW in maize crop in America. For this purpose, two baculoviruses (GV and NPV) were isolated from S. frugiperda larva collected in a pasture crop in Colombia (Córdoba) causing a natural co-infection [17]. The Colombian granulovirus (SfGV VG008) was characterized in
Viruses 2015, 7 396

morphologically, biologically and molecularly terms showing its potential use as enhancer of insecticidal activity of a SfMNPV [4,18]. SfGV VG008 was compared with a Brazilian SfGV isolate without finding differences in the insecticidal activity; however, the genomic comparison with restriction profiles showed differences in number and size fragments. To date, no information about the SfGV complete genome exists and there are only 16 full-length Betabaculovirus genomes with respect to 57 reports for Alphabaculovirus species [19]. Betabaculovirus genomes range from between 99,000–180,000 bp encoding between 119 and 183 putative proteins. Particularly, 14 articles reporting GV genome analyses were published since 1999 increasing knowledge of this kind of baculoviruses [20–32].

To add information that collaborates the understanding of the value of co-infections on insecticidal activity against S. frugiperda, the genome sequence and its bioinformatics characterization of SfGV VG008 are presented. In view of that, the genome analysis of SfGV VG008 revealed the presence of several ORFs that encode virulence factors such as enhancins and chitinases. Additionally, this betabaculovirus shows evidence of multiple points of recombination with both alpha- and betabaculoviruses, suggesting one of the main processes involved in baculovirus evolution.

2. Materials and Methods

2.1. Insect’s Source, Rearing and Virus Production

2.1.1. Larvae of S. frugiperda

Larvae of S. frugiperda were obtained from a laboratory colony established in the Biological Control Laboratory of the Colombian Corporation of Agricultural Research (CORPOICA) using larvae collected from maize fields in Villavicencio, Colombia. This insect colony was periodically refreshed with insects collected in field and maintained at 25 °C, 60% RH (relative humidity) and 12:12 h (light:dark) photoperiod on a wheat germ-based semisynthetic diet [33].

2.1.2. Occlusion Body Purification

Granulovirus SfGV VG008 was isolated from one S. frugiperda larva collected in a pasture crop in Colombia. For virus isolation, neonate larvae of S. frugiperda were inoculated using the droplet feeding method [34]. For this purpose, starved larvae of S. frugiperda were orally inoculated with an occlusion body (OBs) suspension (10^6 OBs/mL), individually reared at 25 °C and 60% RH, under a natural photoperiod of 12:12 h (light:dark) until death. OBs were extracted from dead diseased larvae by homogenizing cadavers in 0.1% SDS solution (w/v) and purified by filtration and centrifugation on a 30%, 50% and 70% sucrose gradient [35]. To quantify viral suspensions absorbance measurements at 280 nm were carried out [18].

2.2. SfGV VG008 Genome Sequencing

Purified granules were dissolved by alkaline lysis and DNA was extracted according to Caballero et al. [36]. SfGV VG008 was sequenced using the 454 Genome Sequencer (GS) FLX™ Standard (Roche) at the Centro Nacional de Secuenciación Genómica (CNSG; Universidad de Antioquia, Medellín, Colombia), with consensus accuracy of Q20. De novo assembly was generated on
NewBler assembler (GS FLX Data Analysis Software, Branford, Connecticut, USA) obtaining 1 unique contig of 140,917 bp. To verify some loci (the junction of the contig’s ends and other two regions with low quality data), PCR amplification, molecular cloning and subsequent sequencing by the standard dideoxy method of Sanger were performed. Thus, primers used were: V008_gap_122212_Fw (5’-CAT GGTGTTGCCAAAGTCAG-3’) and V008_gap_122916_Rv (5’-GTCCATAGAGGACGCGTTGA-3’); V008_gap_109661_Fw (5’-TTGTGTTTCGCAATCTTCACCTTG-3’) and V008_gap_109667_Rv (5’-GAGTATCAGTGCCGAGATG-3’); V008_gap_138941_Fw (5’-TGCGTGTTGGACACCGT-3’) and V008_gap_139260_Rv (5’-TGACCATAGTGACCAGTCTTG-3’). This strategy allows confirming and correcting previous sequences, including the addition of two nucleotides and the elimination of other six. All the experiments associated with cloning were done using pGEM-T-Easy vector (Promega, Madison, Wisconsin, USA) and standard protocols.

ORFs were identified using ARTEMIS [37] and employing Fickett’s method [38]. ATG initiated ORFs of at least 150 nt (50 amino acids) with minimal overlap were selected for further analysis. All SfGV VG008 putative genes were searched for typical promoter motifs using ad hoc software (P.D. Ghiringhelli, 2012, unpublished) and previous data for the Baculoviridae family [22]. Initially the early CA(G/T)T and late (A/G/T)TAAG initiator (INR) sequences were searched in a sequence space comprising 120 residues upstream the ATG codon. After, a typical TATA box [TATA(A/T)A(A/T)] was searched in genes having an INR motif (25 to 35 bp upstream the INR) and in the upstream sequence of genes in which any INR was detected.

2.3. Phylogenetic Inference for SfGV VG008

2.3.1. ORF Identification

In general, ORFeome and proteome similarity searches were done using BlastN, BlastP, tBlastN, tBlastX and PSI-Blast [39] initially against other betabaculovirus genomes and then against the other baculovirus species. Identities and similarities among homologous genes were obtained by doing global alignments with ClustalX [40,41] using default parameters.

2.3.2. Phylogeny

Phylogeny was inferred using 37 core proteins [42] from 73 baculovirus genomes [43], plus data from SfGV VG008. Each core protein set was independently aligned using ClustalX program [40,41] with the following parameters: Pairwise alignment (Gap Open Penalty = 10, Gap Extension Penalty = 0.1, protein weight matrix: Gonnet 250); Multiple alignment (Gap Open Penalty=10, Gap Extension Penalty = 0.05, protein weight matrix: Gonnet series). Then, a concatemer was generated by addition of the complete individual alignments and phylogeny was inferred using MEGA program [44,45] with the following parameters: Method = Neighbor-Joining; Bootstrap with 1000 replicates; gap/Missing data = pairwise deletion; Model = Amino (Dayhoff Matrix); patterns among sites = Same (Homogeneous); rates among sites = Different (Gamma Distributed); gamma parameter = 0.8764. On the other hand, betabaculovirus core proteins were aligned among them using an all-against-all pairwise strategy, and identity and similarity percentages were obtained.

Baculovirus genomes used are listed (Table 1).
Table 1. Baculovirus genomes used in the phylogenetic analysis.

<table>
<thead>
<tr>
<th>Baculovirus</th>
<th>Acc. Number</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antheraea pernyi MNPV Isolate L2</td>
<td>EF207986</td>
<td>AnpeMNPV</td>
</tr>
<tr>
<td>Antheraea pernyi NPV Isolate Z</td>
<td>NC_008035</td>
<td>AnpeNPV</td>
</tr>
<tr>
<td>Anticarsia gemmatalis MNPV</td>
<td>NC_008520</td>
<td>AgMNPV</td>
</tr>
<tr>
<td>Autographa californica MNPV Clone C6</td>
<td>NC_001623</td>
<td>AcMNPV</td>
</tr>
<tr>
<td>Bombyx mandarina NPV S1</td>
<td>NC_012672</td>
<td>BomaNPV S1</td>
</tr>
<tr>
<td>Bombyx mandarina NPV S2</td>
<td>JQ071499</td>
<td>BomaNPV S2</td>
</tr>
<tr>
<td>Bombyx mori NPV Isolate T3</td>
<td>NC_001962</td>
<td>BmNPV</td>
</tr>
<tr>
<td>Choristoneura fumiferana MNPV</td>
<td>NC_004778</td>
<td>CfMNPV</td>
</tr>
<tr>
<td>Choristoneura fumiferana Defective MNPV</td>
<td>NC_005137</td>
<td>CfDEFMNPV</td>
</tr>
<tr>
<td>Choristoneura marina MNPV Strain Darmstadt</td>
<td>NC_023177</td>
<td>ChmuNPV</td>
</tr>
<tr>
<td>Choristoneura occidentalis NPV Isolate BC1</td>
<td>NC_021925</td>
<td>ChocNPV</td>
</tr>
<tr>
<td>Choristoneura rosaceana NPV Isolate NB1</td>
<td>NC_021924</td>
<td>ChroNPV</td>
</tr>
<tr>
<td>Epiphyas postvittana NPV</td>
<td>NC_003083</td>
<td>EppoNPV</td>
</tr>
<tr>
<td>Hyphantria cunea NPV</td>
<td>NC_007767</td>
<td>HycuNPV</td>
</tr>
<tr>
<td>Maruca vitrata NPV</td>
<td>NC_008725</td>
<td>MaviNPV</td>
</tr>
<tr>
<td>Orgyia pseudotsugata MNPV</td>
<td>NC_001875</td>
<td>OpMNPV</td>
</tr>
<tr>
<td>Philosamia cynthia ricini NPV</td>
<td>JX404026</td>
<td>PhcyNPV</td>
</tr>
<tr>
<td>Plutella xylostella MNPV Isolate CL3</td>
<td>NC_008349</td>
<td>PxyMNPV</td>
</tr>
<tr>
<td>Rachiplusia ou MNPV</td>
<td>NC_004323</td>
<td>RoMNPV</td>
</tr>
<tr>
<td>Thysanoplusia orichalcea NPV P2</td>
<td>NC_019945</td>
<td>ThorNPV P2</td>
</tr>
<tr>
<td>Adoxophyes honmai NPV</td>
<td>NC_004690</td>
<td>AdhoNPV</td>
</tr>
<tr>
<td>Adoxophyes orana NPV</td>
<td>NC_011423</td>
<td>AdorNPV</td>
</tr>
<tr>
<td>Agrotis ipsilon MNPV</td>
<td>NC_011345</td>
<td>AgipMNPV</td>
</tr>
<tr>
<td>Agrotis segetum NPV</td>
<td>NC_007921</td>
<td>AgseNPV</td>
</tr>
<tr>
<td>Apocheima cinerarium NPV</td>
<td>NC_018504</td>
<td>ApciNPV</td>
</tr>
<tr>
<td>Buzura suppressaria NPV Isolate Hubei</td>
<td>NC_023442</td>
<td>BusuNPV</td>
</tr>
<tr>
<td>Chrysoideix chalcites NPV</td>
<td>NC_007151</td>
<td>ChchNPV</td>
</tr>
<tr>
<td>Clanis bilineata NPV Isolate DZ1</td>
<td>NC_008293</td>
<td>ClbiNPV</td>
</tr>
<tr>
<td>Ectropis obliqua NPV Strain A1</td>
<td>NC_008586</td>
<td>EcobNPV</td>
</tr>
<tr>
<td>Euproctis pseudoconspersa NPV</td>
<td>NC_012639</td>
<td>EupsNPV</td>
</tr>
<tr>
<td>Helicoverpa armigera MNPV</td>
<td>NC_011615</td>
<td>HaMNPV</td>
</tr>
<tr>
<td>Helicoverpa armigera NPV Isolate Australia</td>
<td>JN584482</td>
<td>HaNPV Aus</td>
</tr>
<tr>
<td>Helicoverpa armigera NPV Strain C1</td>
<td>NC_003094</td>
<td>HaNPV C1</td>
</tr>
<tr>
<td>Helicoverpa armigera NPV Strain G4</td>
<td>NC_002654</td>
<td>HaNPV G4</td>
</tr>
<tr>
<td>Helicoverpa armigera SNPV Strain NNg1</td>
<td>NC_011354</td>
<td>HaSNPV</td>
</tr>
<tr>
<td>Helicoverpa zea NPV</td>
<td>NC_003349</td>
<td>HezeNPV</td>
</tr>
<tr>
<td>Hemileuca sp. NPV</td>
<td>NC_012123</td>
<td>HespNPV</td>
</tr>
<tr>
<td>Leucania separata NPV Strain AH1</td>
<td>NC_008348</td>
<td>LeseNPV</td>
</tr>
<tr>
<td>Lymantria dispar MNPV</td>
<td>NC_001973</td>
<td>LdMNPV</td>
</tr>
<tr>
<td>Lymantria xyline MNPV</td>
<td>NC_013953</td>
<td>LxyxMNPV</td>
</tr>
<tr>
<td>Mamestra brassicae MNPV Isolate Chb1</td>
<td>JX138237</td>
<td>MabrMNPV Chb1</td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Baculovirus</th>
<th>Acc. Number</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mamestra brassicae</em> MNPV Isolate K1</td>
<td>NC_023681</td>
<td>MabrMNPV K1</td>
</tr>
<tr>
<td><em>Mamestra configurata</em> NPV Strain 90-2</td>
<td>NC_003529</td>
<td>MacoNPV 90 2</td>
</tr>
<tr>
<td><em>Mamestra configurata</em> NPV Strain A90-4</td>
<td>AF539999</td>
<td>MacoNPV A90 4</td>
</tr>
<tr>
<td><em>Mamestra configurata</em> NPV Strain B</td>
<td>NC_004117</td>
<td>MacoNPV B</td>
</tr>
<tr>
<td><em>Orgyia leucostigma</em> NPV IsolateCFS77</td>
<td>NC_010276</td>
<td>OrleNPV</td>
</tr>
<tr>
<td><em>Spodoptera exigua</em> MNPV</td>
<td>NC_002169</td>
<td>SeMNPV</td>
</tr>
<tr>
<td><em>Spodoptera frugiperda</em> MNPV Isolate 3AP2</td>
<td>NC_009011</td>
<td>SfMNPV 3AP2</td>
</tr>
<tr>
<td><em>Spodoptera frugiperda</em> MNPV Isolate Nicaraguan</td>
<td>HM595733</td>
<td>SfMNPV Nic</td>
</tr>
<tr>
<td><em>Spodoptera frugiperda</em> MNPV Isolate Nicaraguan DefG</td>
<td>JF899325</td>
<td>SfMNPV NicG</td>
</tr>
<tr>
<td><em>Spodoptera frugiperda</em> MNPV Strain 19</td>
<td>EU258200</td>
<td>SfMNPV 19</td>
</tr>
<tr>
<td><em>Spodoptera littoralis</em> NPV Isolate AN1956</td>
<td>JX454574</td>
<td>SpltNPV AN1956</td>
</tr>
<tr>
<td><em>Spodoptera litura</em> II MNPV</td>
<td>NC_011616</td>
<td>SpliMNPV II</td>
</tr>
<tr>
<td><em>Spodoptera litura</em> MNPV Strain G2</td>
<td>NC_003102</td>
<td>SpliMNPV G2</td>
</tr>
<tr>
<td><em>Trichoplusia ni</em> SNPV</td>
<td>NC_007383</td>
<td>TnSNPV</td>
</tr>
<tr>
<td><em>Adoxophyes orana</em> GV</td>
<td>NC_005038</td>
<td>AdorGV</td>
</tr>
<tr>
<td><em>Agrotis segetum</em> GV</td>
<td>NC_005839</td>
<td>AgseGV</td>
</tr>
<tr>
<td><em>Choristoneura occidentalis</em> GV</td>
<td>NC_008168</td>
<td>ChocGV</td>
</tr>
<tr>
<td><em>Clostera anastomosis</em> GV Strain Henan</td>
<td>NC_022646</td>
<td>CaGV</td>
</tr>
<tr>
<td><em>Cryptophlebia leucotreta</em> GV</td>
<td>NC_005068</td>
<td>CrleGV</td>
</tr>
<tr>
<td><em>Cydia pomonella</em> GV</td>
<td>NC_002816</td>
<td>CpGV</td>
</tr>
<tr>
<td><em>Epinotia aponema</em> GV</td>
<td>NC_018875</td>
<td>EpapGV</td>
</tr>
<tr>
<td><em>Helicoverpa armigera</em> GV</td>
<td>NC_010240</td>
<td>HearGV</td>
</tr>
<tr>
<td><em>Phthorimaea operculella</em> GV</td>
<td>NC_004062</td>
<td>PhopGV</td>
</tr>
<tr>
<td><em>Pseudaletia unipuncta</em> GV Strain Hawaii</td>
<td>NC_013772</td>
<td>PsunGV</td>
</tr>
<tr>
<td><em>Pieris rapae</em> GV</td>
<td>NC_013797</td>
<td>PiraGV</td>
</tr>
<tr>
<td><em>Plutella xylostella</em> GV</td>
<td>NC_002593</td>
<td>PlxyGV</td>
</tr>
<tr>
<td><em>Spodoptera litura</em> GV Strain K1</td>
<td>NC_009503</td>
<td>SpliGV</td>
</tr>
<tr>
<td><em>Xestia c-nigrum</em> GV</td>
<td>NC_002331</td>
<td>XecnGV</td>
</tr>
<tr>
<td><em>Neodiprion abietis</em> NPV</td>
<td>DQ317692</td>
<td>NeabNPV</td>
</tr>
<tr>
<td><em>Neodiprion lecontei</em> NPV</td>
<td>NC_005906</td>
<td>NeleNPV</td>
</tr>
<tr>
<td><em>Neodiprion sertifer</em> NPV</td>
<td>NC_005905</td>
<td>NeseNPV</td>
</tr>
<tr>
<td><em>Culex nigripalpus</em> NPV</td>
<td>NC_003084</td>
<td>CuniNPV</td>
</tr>
</tbody>
</table>

2.4. Protein Synteny

BlastP analysis of SfGV VG008 proteome against the closest betabaculovirus proteomes (HearGV, PsunGV, SpliGV and XecnGV) were carried out using 0.001 as expected value (e-value) threshold. Once the corresponding orthologous proteins were detected, each SfGV VG008 protein was aligned against its ortholog in a pairwise fashion manner using ClustalW2 program and Gonnet 250 matrix of conservative changes [46]. Identity and similarity percentages were calculated for each alignment. The protein synteny graphs were generated using personal routines (P.D. Ghiringhelli, 2014, unpublished) with a color scale and similarity cut offs indicated in the corresponding figure. The length of the upper and lower prism sides is proportional to the length of the respective polypeptides.
2.5. Non-Coding Region Analyses

2.5.1. Homologous Regions (hrs)

Canonical nucleotide decamers previously described in hrs of HearGV, PsunGV and XecnGV were used as computational probes to search the corresponding hrs in the SfGV VG008 genome. All non-coding regions containing repeats similar to probes were recovered and manually inspected. For each selected region, the secondary DNA structure prediction of the main sequence was obtained using the Mfold server of Michael Zuker website [47] and using RNADraw program [48]. In order to construct sequences logos, multiple alignments of palindromes were performed using ClustalX algorithm [41,49] with default parameters, and then sequences logos were obtained using the WebLogo server [50].

2.5.2. A + T-Rich Regions

A + T-content was profiled using a partially overlapped sliding window strategy (window = 500 nucleotides, displacement = 50 nucleotides) (P.D. Ghiringhelli, 2012, unpublished). Relationships between each point and genomic average A + T-content were obtained and peaks of 1.12 or above were considered as A + T-rich regions.

2.6. Analyses of Genes Putatively Derived from Horizontal Transfer

2.6.1. BlastP Relationships

First, a general baculoviral proteome database (GBPD) was constructed and BlastP searches using the SfGV VG008 proteome against GBPD were carried out (e-value threshold = 0.001). Candidate proteins acquired by horizontal transfer were selected. Then, individual specific protein databases (ISPDs) containing similar proteins of all related species were constructed to determine the relationships among selected species. Then, relaxed BlastP (e-value = 0.01) using individual proteins against the corresponding ISPD as a query was performed. Finally, the minimum e-values that showed reciprocity between pairs of viruses were selected and illustrated, involving all related viral species in network graphs.

2.6.2. Recombination Analyses

In order to detect potential recombination events partial genome sequence of SfGV VG008 comprising the recombinant candidate genes (SfGV VG008 ORFs 059/099) and flanking regions was compared with the corresponding sequence in other baculoviruses (HearGV, PsunGV, XecnGV and SpltNPV II or SfMNPV) by running two alternative methods. In the first one (P.D. Ghiringhelli, 2008, unpublished), alignments were carried out with the ClustalX program (default parameters) [40,41,49] between sequence pairs, always involving the putative recombinant candidate from SfGV VG008 and one of the other sequences. The relative similarities were calculated using the ClustalX consensus symbol (* and blank space) as the input sequence, in an overlapping windows-based strategy. Arbitrary values of +1 for identical (*) and −0.5 for non-identical (blank spaces) residues were used. The sum of assigned values for each residue in each window (35 nucleotides) was divided by the window width and allotted to the central position to generate the plots. Profiles were drawn and analyzed with the aim of
detecting crosspoints between them. In order to find a good relation between graph complexity and crosspoint detection sensitivity, different windows lengths were scanned. The second method was the bootscan analysis available in the Simplot program (version 3.5.1) [51,52] using the following parameters: (Window: 500 residues; Step: 50 residues; Gaps strip: on; Replicates: 100; Model: Kimura 2-parameters; Transition and transversion ratio: 2.0; Phylogenetic method: Neighbor-Joining). The breakpoints were estimated.

2.7. Characterization of SfGV VG008 ORFs 047/059/089/099

To determine the nature of SfGV VG008 ORF047, ORF059, ORF089 and ORF099 theoretical proteins, different bioinformatics tools were used. Hydrophobicity profiles were constructed using a sliding windows strategy (window = 17 amino acids; sliding 1 residue each time) and ad hoc program (P.D. Ghiringhelli, 2004, unpublished). Several hydrophobicity scales were assayed [53–55]. Signal peptide presence or absence was assessed by using SignalP [56]. Putative functions were evaluated using the HHpred server [57]. Secondary and tertiary structures were predicted using the LOcal MEta-Threading-Server [58], and the I-TASSER server [59] or the QUARK server [60]. Finally, the assessment of closest neighbors was carried out through phylogenetic inference of related sequence collections using MEGA program [44,45] with the following parameters: Method = Neighbor-Joining; Bootstrap with 500 replicates; gap/ Missing data = pairwise deletion; Model = Amino (Dayhoff Matrix); patterns among sites = Same (Homogeneous); rates among sites = Uniform Rates.

3. Results and Discussion

3.1. Genome of SfGV VG008 and Gene Content

The genome of SfGV VG008 (GenBank: KM371112) was covered 20 times and consists of 140,913 bp showing 53.8% of A + T content, a value very close to the lowest one estimated for betabaculovirus members which range between 53.2% in Clostera anastomosis GV and 67.5% in Cryptophlebia leucotreta GV [25]. However, no correlation was found between these data and virus biological properties impeding make predictions about features such as host range, pathogenicity or virulence [10].

The SfGV VG0008 genome contains 146 putative ORFs, all encoding theoretical polypeptides with at least 50 amino acid lengths and considering a minimal sequence overlapping among flanking regions. In view of the above, the ORFeome would cover 95.6% of the whole nucleotide sequence. ORFs were consecutively numbered from the first nucleotide of the granulin start codon resulting in 82 encoding regions in the granulin polarity and other 64 in the opposite one. The identity of genes was established by Blast (Figure 1).

To extend the previous study, typical promoter motifs located up to 120 bp upstream to the initial ATG and similarity comparison among orthologous genes from Noctuidae isolates (HearGV, PsunGV, SpliGV and XecnGV) were analyzed (Table S1). Therefore, that, early CAKT initiator sequence (INR) [61] was found in 47 ORFs, including or not TATA-box. Late INR motif [62] was detected in 21 ORFs; other 59 showed both early and late elements and 2 had only a TATA-box. The remaining 17 ORFs do not have any of the mentioned motifs, but could be transcribed from other regulatory elements [63].
Figure 1. Circular map of the SfGV VG008 genome. The illustration shows all predicted SfGV ORFs (1–145 including 143a and 143b) and their transcription direction (with respect to granulin) indicated as arrows above a line that represent the genome (140,913 bp). The arrow colors represent different features such as presence of homologous sequences in other species of Baculoviridae and/or highlighting activities that include virulence factors (enhancins and chitinases). ORF’s numbers are indicated above or below the arrows. The names of the genes (assigned by significantly similarity with sequences form other baculoviruses) can be seen in Table S1. The regions without arrows represent non-coding regions.
Similarity analysis with the reported proteomes of each other baculovirus revealed that 125 proteins are shared with the all other betabaculoviruses meanwhile 2 are only shared with HearGV, PsunGV, XecnGV and SpltNPV II (ORF047 and ORF089). Other results showed that 2 proteins are shared with alphabaculoviruses (ORF059 and ORF099), 14 are unique (ORF024, ORF029, ORF031, ORF040, ORF049, ORF054, ORF055, ORF098, ORF118, ORF125, ORF129, ORF133, ORF134 and ORF141) and 3 seems to be product of divergent copies of some SfGV VG008 genes (sets of paralogous: ORF046 and ORF047; ORF057 -named Bro c- with ORF052 –Bro a-, ORF053 –Bro b-, ORF060 –Bro d-, ORF090 –Bro e-, ORF094 –Bro f- and ORF095 –Bro g-; ORF143a and ORF143b). One unique protein (ORF134) has similarity with a non-baculoviral *chitinase*-2c. In particular, the genome analysis revealed that SfGV VG008 would encode virulence factors associated to the enhancement of insecticidal activity [15]. Thus, this betabaculovirus contains four *chitinases* (ORFs 010, 071, 072, 134) and two *enhancins* (ORFs 127 and 132). Reviewing the 14 betabaculovirus genome analyses (papers cited in the introduction), only ChocGV, AgseGV, PsunGV, HearGV and XecnGV have genes encoding to enhancins (one for first two, three for the next and four ORFs for the others). Moreover, chitinase genes are present in two copies into ChocGV and EpapGV, and only one copy into CaLGV, ClanGV, CpGV, HearGV, PlxyGV and XecnGV genomes (CrleGV has a truncated chitinase gene). In this sense, SfGV VG008 is the betabaculovirus that would express more proteins associated to virulence.

### 3.2. Phylogenetic Inference for SfGV VG008

As previously mentioned, ORFs encoding the 37 described core proteins for the *Baculoviridae* family [42] were found in the genome of SfGV VG008, covering the essential functions of: replication; transcription; cell cycle arrest and/or interaction with host proteins; viral packaging, assembly and release; and oral infectivity [64]. The phylogenetic analysis based on the 37 concatenated core proteins of 73 baculovirus genomes plus SfGV VG008 was performed (Figure 2).

The obtained cladogram reproduced the grouping of four genera recognized in the current classification of the family *Baculoviridae* [7]. As expected, SfGV VG008 isolated from Noctuidae insect, was a novel member of the *Betabaculovirus* genus grouping with HearGV, PsunGV, XecnGV and SpltGV. In previous reports it was observed that Noctuidae specific betabaculoviruses tend to be located in a separated group with respect to the members that infect the Tortricidae [8,65].

The theoretical proteins of core genes present in *Betabaculovirus* were compared in order to obtain pairwise identity and similarity values (Figure 3). According to this, the SfGV VG008 showed identity values ranges between 9% (SfGV ORF104 vs. EpapGV ORF043) and 83.3% (SfGV ORF103 vs. XecnGV ORF121) with a median of 48.4% (SfGV ORF078 vs. ClanGV ORF072). The equivalent study focused on similarity showed values ranges between 34.3% (SfGV ORF104 vs. CaLGV ORF037) and 98.2% (SfGV ORF085 vs. HearGV ORF098) with a median of 77.6% (SfGV ORF037 vs. CpGV ORF047 and SfGV ORF069 vs. EpapGV ORF069).

The accepted function of each core protein and the ORF number according to genome annotations are detailed (Supplementary material Table S2). This study revealed that 37 core proteins are a set of factors strongly conserved into *Betabaculovirus* genus because they probably play the essentials roles needed to complete the virus cycle. So that, this set of ancestral sequences remains the best option for phylogenetic inference in *Baculoviridae*.
**Figure 2.** Phylogenetic inference for SfGV VG008. Cladogram based on a concatemer built with the 37 core proteins obtained from 74 baculoviral genomes. The phylogenetic tree was inferred using MEGA 6 program. The four *Baculoviridae* genera are indicated and in order to preserve space, *Alpha-* (Groups I and II) and *Gammabaculovirus* clades were collapsed.

**Figure 3.** Identity and similarity analyses of SfGV VG008 core proteins. The box plot shows the amino acid identity (light orange) and similarity (light blue) percentages of the 37 core proteins present in SfGV VG008 respect to the orthologs located in the other betabaculoviruses. Core proteins names are the currently accepted (*Sim.* abbreviates “similar to”). The boundary of boxes closest to zero indicates the 25th percentile, a line within the box marks the median, and the boundary of the box farthest from zero indicates the 75th percentile. Error bars above and below boxes indicate the 90th and 10th percentiles, respectively. The filled circles indicate outlying points.
3.3. Genome Collinearity Analysis

To characterize the genome organization of SfGV VG008, a gene collinearity study based on proteins with respect to the closely related betabaculoviruses (those isolated from Noctuides and which clustered in described phylogeny) was performed using synteny graphs (Figure 4, Supplementary Material Table S3). Thus, a great gene order correlation among the SfGV VG008, HearGV, PsunGV, SpliGV and XecnGV was observed, with some inversions and drifts. In general, the synteny maps are conserved among betabaculovirus species differing from alphabaculoviruses [21,25,28,30].

Figure 4. Protein synteny graphs. The illustration shows the comparison of gene collinearity based on genome physical positions and protein similarities between SfGV VG008 and each one of the most related betabaculoviruses (HearGV, PsunGV, SpliGV and XecnGV). The grey lines represent genomes and their lengths are proportional among them (bp scale). Colored lines between genomes (grey lines) relate homologous sequences indicating the percentage of similarity according to the key.
3.4. Homologous Regions (hrs) and A + T-Rich Regions

Baculovirus genomes have nucleotide sequence repeats known as homologous regions (hrs) that could act as replication starting points and/or as enhancers of transcription. Moreover, it is considered that these segments increase the genome plasticity and may mediate the intra and inter molecular recombination [63]. In addition to this, intergenic A + T-rich regions have been identified as putative non-hr origins of replication [66,67]. Considering the importance of this kind of sequences, the genome of SfGV VG008 was analyzed. Particularly, it contains 8 hrs [hr-1 (1 repeat), hr-2 (7 repeats), hr-3 (8 repeats), hr-4 (5 repeats), hr-5 (2 repeats), hr-6 (4 repeats), hr-7 (6 repeats) and hr-8 (1 repeat)]. Reviewing closest beta baculovirus genomes, 9 hrs has been described in HearGV [hr-1 (5 repeats), hr-2 (4 repeats), hr-3 (4 repeats), hr-4 (7 repeats), hr-5 (2 repeats), hr-5a (1 repeat), hr-6 (4 repeats), hr-7 (3 repeats) and hr-8 (4 repeats)], 9 in PsunGV [hr-1 (5 repeats), hr-2 (3 repeats), hr-3 (4 repeats), hr-4 (5 repeats), hr-5 (2 repeats), hr-5a (1 repeat), hr-6 (3 repeats), hr-7 (4 repeats) and hr-8 (3 repeats)], and 9 in XcenGV genomes [hr-1 (5 repeats), hr-2 (5 repeats), hr-3 (4 repeats), hr-4 (6 repeats), hr-5 (3 repeats), hr-5a (1 repeat), hr-6 (4 repeats), hr-7 (5 repeats) and hr-8 (4 repeats)].

A deep analysis showed that repeats are stretches of 40–48 nucleotides length, in which the 10 bp at each end are perfect direct or inverted sequences, and only in few cases one of the flank is an imperfect repetition. All hrs were found within A + T-rich regions (Figure 5). It is important to note that SpliGV contains one A + T-rich region but sequences like hrs were not found [68]. This observation suggests that hrs may not be essentials in baculovirus cycle but surely their presence positively contribute in the other important processes previously mentioned, such as gene acquisition, genome replication or transcription enhancement.

HearGV, PsunGV and XcenGV have hrs of variable length (50–58 nt) where the first and last 10 nucleotides are perfect or imperfect copies (direct or inverted) of a core oligonucleotide [TTAAT(G/A)TGCA] which flank variable regions (30–38 nt) rich in AT content. Besides, it is possible to detect, by clustering analyses, 3 hr variants for HearGV and XcenGV, and 4 for PsunGV. In contrast, SfGV VG008 has other organization because each hr contains one sequence unit (TTAATGTGC) located into A + T rich regions of about 50 nucleotides (Table S4).

It is known that these regions may differ in location within genomes, number of copies and nucleotide sequences between different baculovirus species; however, their generalized distribution suggest that functions are conserved [66,69]. In general, the non-coding regions of baculovirus genomes represent less than 10% of whole sequence and this fraction usually contains the hrs [67]. In this sense, six hrs of SfGV VG008 were found into non-coding regions [hr-2 (39,744–40,073 bp), hr-3 (46,924–47,327 bp), hr-4 (53,816–54,064 bp), hr-5 (87,869–87,960 bp), hr-6 (122,550–122,734 bp) and hr-8 (133,555–133,603 bp)] and the other two hrs [hr-1 (19,328–19,376 bp) and hr-7 (128,872–129,176 bp)] were located in zones that show a minimal overlapping with encoding regions (SfGV VG008 ORF 024 for hr-1 and ORFs 130/131 for hr-7).
Figure 5. A + T-rich profiles and characterization of hrs. The plot graphs represent the profiles of A + T-content indicated as relative percentages along some betabaculovirus genomes. Blue lines show the cut off used to define the A + T-rich regions. The hrs sequences are indicated in the corresponding positions of each genome (for SpliGV, hrs were not reported). (A) A + T profile of SfGV VG008 genome; (B) A + T profile of HearGV genome; (C) A + T profile of PsunGV genome; (D) A + T profile of SpliGV genome; (E) A + T profile of XecnGV genome; (F) Sequence logo of the motif of hrs from SfGV VG008; (G). Sequence logo of the three main motifs of hrs from HearGV; (H) Sequence logo of the four main motifs of hrs from PsunGV; (I) Sequence logo of the three main motifs of hrs from XecnGV.
3.5. Analyses of Genes Putatively Derived from Horizontal Transfer

Core genes in *Baculoviridae* are vertically transferred from the last virus common ancestor [42]. In contrast, other genes were later incorporated in particular species by horizontal transmission by processes that include recombination and transposition events. In SfGV VG008 three genes (ORF046, ORF047 and ORF089) are only shared with some betabaculoviruses and some alphabaculoviruses of Group II. Similarly, 2 other genes (ORF059 and ORF099) are shared only with alphabaculoviruses suggesting origins by horizontal transfer.

In order to add information that clarify this supposition, different studies were performed. First, all species were related in a network where the minimum BlastP e-values obtained for pairs of viruses were selected (Figure 6). It is important to note that the reciprocal e-values were not always coincident because some proteins have different sizes.

**Figure 6.** Protein relationships for SfGV VG008 ORFs 046/047/059/089/099. The relationships among some ORFs of SfGV VG008 and their orthologous genes contained in other viruses were calculated by BlastP. The illustration shows the related baculovirus species indicating a name abbreviation for each one (three letters for species and ORF number) into filled circles (yellow for betabaculoviruses and green for Group II alphabaculoviruses). The BlastP e-value between pairs of species is indicated above each arrow. (A) Protein relationships for SfGV VG008 ORFs 046/047; (B) Protein relationships for SfGV VG008 ORF059; (C) Protein relationships for SfGV VG008 ORF089; (D) Protein relationships for SfGV VG008 ORF099. AIN: AgipNPV; HAG: HearGV; HAN: HearMNPV; MC2: MacoNPV 90-2; MC4: MacoNPV A90-4; MBC: MabrNPV CHb1; MBK: MabrNPV K1; MCB: MacoNPV B; PUG: PsunGV; SEN: SeMNPV; SF2: SiMNPV 3AP2; SF9: SiMNPV 19; SFA: SiMNPV Nic DefG; SFG: SfGV VG 008 (in red letters); SFN: SiMNPV Nic; SL2: SpliNPV G2; SLN: SpliNPV II; XCG: XecnGV.
3.5.1. SfGV VG008 ORFs 046/047/089

Detailed analysis of SfGV VG008 ORF047 with respect to the most similar proteins present in HearGV, PsunGV, XecnGV and SpltNPV II shows identity/similarity values of 27.5%/61.2%, 29.1%/61.1%, 27.4%/63.3% and 37.6%/60.9%, respectively. Although the highest identity value is between SfGV VG008 and SpltNPV II, the origin of the gene that encodes this protein is unclear because the significance of differences in the similarity values between SfGV VG008 ORF047 and the other proteins are doubtful. However, the BlastP searches for similar proteins in the corresponding ISPD shows a stronger relationship between SfGV VG008 ORF047 and SpltNPV II ORF089 than can be found between SfGV VG008 ORF047 and the similar proteins of the most related betabaculoviruses. This relationship could be derived from an ancient acquisition and parallel evolution.

The analysis of SfGV VG008 ORF089 with the similar proteins from HearGV, PsunGV, XecnGV, SpltNPV II, MacoNPV 90-2 and MacoNPV A90-4 shows identity/similarity values of 31.4%/67.9%, 31.7%/67.7%, 33.3%/69.8%, 48.5%/79.1%, 44.7%/66.5% and 44.7%/66.5%, respectively. The highest values are those obtained when the comparison was with SpltNPV II, although the differences in identity between SfGV VG008 ORF089 and the corresponding proteins of SpltNPV II, MacoNPV 90-2 and MacoNPV A90-4 are minimal. However, the BlastP searches for similar proteins in the corresponding ISPD shows a higher relationship between SfGV VG008 ORF089 and SpltNPV II ORF020 than can be found between SfGV VG008 ORF089 and the similar proteins of the most related betabaculoviruses, MacoNPV 90-2 or MacoNPV A90-4. According to this, at least two different evolutive scenarios could have occurred. In the first one, SpltNPV II or MacoNPV 90-2 or MacoNPV A90-4 has acted as the source of the gene in independent events of horizontal transfer to betabaculoviruses; differences in the ancestriality of this process could explain the present identity/similarity values before showed. In the second one, SpltNPV II or MacoNPV 90-2 or MacoNPV A90-4 has acted as the donor species in a horizontal transfer event, which has occurred with a hypothetical common ancestor of HearGV, PsunGV, SfGV and XecnGV. The differences in the speciation and divergency times could explain the present identity/similarity values.

3.5.2. SfGV VG008 ORFs 059/099

The situation is most clear in the case of proteins SfGV VG008 ORF059 and SfGV VG008 ORF099 because they are only shared with alphabaculoviruses of Group II. SfGV VG008 ORF059 is only similar to the protein encoded in the ORF137 of SpltNPV II, whereas SfGV VG008 ORF099 is similar to the proteins encoded in: HearMNPV ORF024, MacoNPV 90-2 ORF030, MacoNPV A90-4 ORF030, MacoNPV B ORF025, MabrMNPV Chb1 ORF025, MabrMNPV K1 ORF024, SfMNPV 3AP2 ORF023, SiMNPV 19 ORF022, SiMNPV Nic ORF023, SiMNPV Isolate Nic DefG ORF023 and SpltNPV G2 ORF106. The highest identity/similarity values is between SfGV VG008 ORF099 and ORF023 of the SiMNPV isolates (ORF022 in SiMNPV 19), although the similarity between SfGV VG008 ORF099 and the corresponding proteins of SiMNPV or MacoNPV isolates converged to nearest values. However, the BlastP searches for similar proteins in the corresponding ISPD shows a higher relationship among SfGV VG008 ORF099 and similar proteins of SiMNPV Nic, SiMNPV 3AP2 and SiMNPV 19 isolates than can be found among SfGV VG008 ORF099 and the similar proteins of the other related
alphabaculoviruses of Group II. In both cases, the obtained results suggest the occurrence of independent horizontal transfer events during the evolution.

In order to explore the recombination hypothesis for SfGV VG008 ORF059 and ORF099 two approaches were used: relative similarity and bootscanning analyses (Figure 7). Thus, the results obtained with mentioned studies for both genes supported the hypothesis, which showed the occurrence of putative recombination events. Particularly, studies of relative similarity showed that any betabaculoviruses from Noctuide insects (HearGV, PsunGV and XcenGV) present homologous sequences to SfGV VG008 ORF059. In contrast, when the comparison was with SpltNPV II the relative similarity was significantly high with SpltNPV II ORF137. In fact, this observation was confirmed with high percentage of permuted trees by bootscanning plot against previously mentioned betabaculoviruses and SpltNPV II suggesting a recombination event involving two different virus species.

Figure 7. Cont.
Figure 7. Origin by horizontal transfer of SfGV VG008 ORFs 059 and 099. Similarity plots and bootscanning analysis for possible recombination processes between the ancestors of SfGV VG008 and other baculoviruses are studied. The genome regions considered are those containing SfGV VG008 ORFs 059 and 099 (about 3900 bp and 5600 bp, respectively). In all cases, colored arrows according to the key indicated in Figure 1 represent ORFs of SfGV VG008. In contrast, ORFs of the other baculoviruses are represented as black arrows excepting homologous genes to SfGV VG008 ORFs 059 and 099 (yellow arrows). All the genome positions in the studies are indicated at the beginning and the end of considered regions (bp scale). The similarity plots are indicated in black, and in the bootscanning analyses different colors are used being referenced into the graphs. (A) Possible origin of SfGV GV008 ORF059. 1. Similarity plot between SfGV VG008 and HearGV; 2. Similarity plot between SfGV VG008 and PsunGV; 3. Similarity plot between SfGV VG008 and XecnGV; 4. Similarity plot between SfGV VG008 and SpltNPV II; 5. Bootscanning using SfGV VG008, HearGV, PsunGV, XecnGV and SpltNPV II; 6. Partial genomic maps of SpltNPV II and SfGV VG008, and putative recombination event; (B) Possible origin of SfGV GV008 ORF099. 1. Similarity plot between SfGV VG008 and HearGV; 2. Similarity plot between SfGV VG008 and PsunGV; 3. Similarity plot between SfGV VG008 and XecnGV; 4. Similarity plot between SfGV VG008 and SfMNPV-3AP2; 5. Bootscanning using SfGV VG008, HearGV, PsunGV, XecnGV and SmNPV-3AP2; 6. Partial genomic maps of SfMNPV-3AP2 and SfGV VG008, and putative recombination event.
In the case of SfGV VG008 ORF099 similar results were observed, although the relative similarities and the bootscanning plots showed that the sequence donor probably be the SfMNPV genome or some relative ancestors. Besides, it is important to note that upstream to recombination site there is a SfGV GV008 unique ORF (ORF098). Both recombination events seem to have occurred in non-coding regions by insertion or by replacement of a stretch of non-coding sequence. Importantly, very close to SfGV VG008 ORF59 (53,537–53,124 bp) is the hr-4 (53,816–54,064 bp) that could be associated with the recombination event as previously reported [70].

Recombination may serve as a mechanism for baculoviruses to rapidly obtain the genetic variation required for survival in a non-static environment without the potential loss of viability that may occur with a high mutation rate [70]. These events in the Baculoviridae family have been observed between different viral species (heterologous recombination) involving genomic DNA exchange in natural co-infections and in cell cultures exposed to different viruses, and between very close related viruses (homologous recombination) [70, 71].

It has been shown that recombination events allow some baculovirus expand its host range, as is the case between AcMNPV and BmNPV [72]. This feature has been studied as an alternative to extend the usefulness of baculovirus as biotechnological tool or as biopesticides [73].

It is important to note that the SfGV VG008 isolate was found in S. frugiperda larvae collected in the field, which presented a natural co-infection with a Colombian isolate of SfMNPV (SfCOL). In this sense, events of intra- and inter-specific recombination during co-infection in the same host have been described for some isolates of CrleGV and CpGV [74]. These processes have also been observed in alphabaculoviruses, where heterologous recombination between isolates of AcMNPV and RoMNPV in natural populations of Galleria mellonella (Linnaeus, 1758) (Lepidoptera: Pyralidae) was reported [75].

Similar to this study, ORFs derived from possible independent recombination events with SeMNPV and/or SpltMNPV has been reported for Colombian SfMNPV isolate [4]. These sequences did not present homologies with the other SfMNPV isolates. Specifically, the ORFs 4 and 5 share high similarities with ORFs splt20 and splt21 of the SpltMNPV genome respectively, and with ORFs se21 and se22 + se23 of SeMNPV genome. In contrast, Colombian SfMNPV lacked the sf23 ORF (unknown function) which is found in the other reported SfMNPV genomes [4].

### 3.6. Characterization of Proteins Encoded by SfGV VG008 ORFs 047/059/089/099

Considering the previous results, the genes obtained by horizontal transfer were particularly analyzed in terms of protein structure and phylogenetic relationships with other baculovirus species (Figure 8).

SfGV VG008 ORF047 protein consists of 263 residues, with 33 negatively charged (Asp + Glu) and 38 positively charged (31 Arg + Lys, and 7 His) amino acids. Based on sequence, the molecular weight is 31,401.9 kDa and its net charge is +1.5. The hydrophobicity profile suggests that is a soluble protein with an average hydrophobicity of −0.58. Secondary structure prediction using the LOMETS server shows that 216 residues (82.1%) are distributed in at least 12 alpha helices (74.0%) and 3 beta sheets (6.1%), while the other ones are part of loops or turns. According to previously mentioned material, the same software predicted that SfGV VG008 ORF047 is a globular protein. On the other hand, the polypeptide of SfGV and the SpltNPV II ORF089 protein share 64.6% of the three-state secondary structure motifs. To assess the hypotheses that SpltNPV II ORF089 protein is the closest sequence as
was previously determined by the BlastP relationship, an evolutionary history of all related baculoviral proteins was inferred using the Neighbor-Joining method. According to this, the phylogenetic tree showed a strong correlation with the above adding evidence for this assumption. Moreover, HHpred software did not identify any specific domain but showed some similarity with proteins that participate in the transcription process, such as the Ribbon-helix-helix motif acting in regulation of transcription DNA-templated, late transcription factor VLTF-2 (also called Poxvirus trans-activator protein A1), and transcription initiation factor IIA (gamma chain in Homo sapiens and small chain in Saccharomyces cerevisiae). Anyway, the biological role of SfGV VG008 ORF047 in the transcription process requires experimental confirmation.

Figure 8. Cont.
Figure 8. Characterization of proteins derived from SfGV VG008 ORFs 047/059/089/099. The theoretical proteins encoded by SfGV VG008 ORFs 047 (panel A), 059 (panel B), 089 (panel C) and 099 (panel D) were analyzed. Pairwise sequence alignment between previously mentioned proteins from SfGV and their closest homologs (highlighting the predicted secondary structures) are shown (A1; B1; C1; D1). Alpha helices are represented as red cylinders and beta sheets as green arrows. Particular amino acids are also indicated (cysteins in red and histidines in blue) for SfGV VG008 ORFs 059 and 099 indicating conserved residues putatively involved in a ring-finger motif. Besides, the evolutionary histories were inferred using the protein sequence collection derived from previous BlastP relationship analyses (Figure 6). Cladograms contain on the branches the percentage of replicated trees in which the associated taxa clustered in the bootstrap test, only indicating those values greater than 60% (A2; C2; D2). In all cases, the most related sequences are highlighted in bold letter. Moreover, 3D structures were predicted using LOMETES (A3; C3) or QUARQS (B3) or I-TASSER (C3) servers. SF2: SfMNPV 3AP2; SFG: SfGV VG 008; SLN: SpltNPV II.
SfGV VG008 ORF089 protein consists of 156 residues, with 24 negatively charged (Asp + Glu) and 31 positively charged (28 Arg + Lys, and 3 His) amino acids. Based on the sequence, the molecular weight is 18,648.8 kDa and its net charge is +5.5. The hydrophobicity profile suggests that it is a soluble protein with an average hydrophobicity of −1.06. Secondary structure prediction using the QUARK server shows that 91 residues (58.3%) are distributed in at least 4 alpha helices (37.8%) and 6 beta sheets (20.5%), while the other ones are part of loops or turns. Besides, this software predicted a globular topology. On the other hand, SfGV VG008 ORF089 and SpltNPV II ORF020 proteins share 81.0% of the three-state secondary structure motifs, result that was supported by the phylogenetic inference. In SpltNPV II ORF020 protein HHpred identified a phosphatase domain comprised between the amino acids 35 and 96, similar to Polynucleotide kinase 3 phosphatase of Schizosaccharomyces pombe (PNK1) [76], which play a role in the repair of single breaks in DNA induced by several DNA-damaging agents. HHpred finding on SfGV VG008 ORF089 protein and local sequence similarity studies performed for both proteins supported the previous function assignment and detect the phosphatase domain between residues 33 to 94. Nevertheless, if SfGV VG008 ORF089 acts in DNA repairing processes requires further confirmation.

The SfGV VG008 ORF059 protein consists of 137 residues, with 12 negatively charged (Asp + Glu) and 22 positively charged (17 Arg + Lys, and 5 His) amino acids. Based on the sequence, the molecular weight is 16,518.4 kDa and its net charge is +5.5. The hydrophobicity profile suggests that is a soluble protein with an average hydrophobicity of −0.71. Secondary structure prediction using the QUARK server shows that 91 residues (56.9%) are distributed in at least 4 alpha helices (48.2%) and 2 beta sheets (8.8%), while the other ones are part of loops or turns. Besides, that software predicted a globular topology. On the other hand, SfGV VG008 ORF059 and SpltNPV II ORF137 proteins share 66.9% of the three-state secondary structure motifs and due to both polypeptides are similar and unique within Baculoviridae, the assessment of which sequence is the phylogenetically closest was not necessary. HHpred identified the typical presence and distribution of Cys (C) and His (H) residues that occur in the Ring-finger family domains. Particularly, SfGV VG008 ORF059 protein contains 9 cysteines and 5 histidines while SpltNPV II ORF137 protein has 8 cysteines and 4 histidines. Besides, a high conservation in polypeptide position was detected by pairwise alignment (7 C and 3 H) constituting the following motif: x(39-40)-C-x(22)-C-H-x-C-x(9)-C-x(2)-H-x(18-19)-C-x(2)-C-x(3)-C-x(22)-H-x(9). It is possible that some of these residues could coordinate Zn$^{+2}$ (e.g., Zinc-fingers) or other divalent cations. In most cases, Zinc-fingers are typical motifs distributed in DNA/RNA regulatory proteins whereas the coordination of heavy metals is often a characteristic of different metallothioneins. Any of these assumptions must be experimentally corroborated.

Finally, the SfGV VG008 ORF099 protein consists of 198 residues, with 37 negatively charged (Asp + Glu) and 28 positively charged (23 Arg + Lys, and 5 His) amino acids. Based on the sequence, the molecular weight is 23,055.3 kDa and its net charge is −11.5. The hydrophobicity profile suggests that is a soluble protein with an average hydrophobicity of −0.72. Secondary structure prediction using the I-TASSER server shows that 134 residues (67.7%) are distributed in at least 6 alpha helices (59.1%) and 3 beta sheets (8.6%), while the other ones are part of loops or turns. Besides, that software predicted a globular topology. On the other hand, SfGV VG008 ORF099 and SfMNPV 3AP2 ORF023 proteins share 70.9% of the three-state secondary structure motifs, a result that was supported by the phylogenetic inference. As occurred with the SfGV VG008 ORF099 protein, HHpred identified C and H residues...
distributed like Ring-finger domains. In particular, this protein contains 17 C and 5 H whereas SfMN
PV 3AP2 ORF023 protein has 16 C and 5 H. Besides, a high conservation in polypeptide position was
detected by pairwise alignment (15 C and 2 H) constituting the following motif: x(16-30)-C-x(14-16)-
C-C-x(2)-[CH]-x(19-20)-C-x(7)-H-x(4)-C-x(30)-C-x(2)-C-x(3)-C-x(22)-C-x-C-H-x-C-x(15-16)-C-x-C-
x(10)-[CH]-x(3)-C-x(2)-C-x(12-29). Similar to previous results, speculations can be made about the
biological role of this polypeptide any of which will require experimental assays.

4. Concluding Remarks

This work presents the first genome analysis of a betabaculovirus from a S. frugiperda, named SfGV
VG008, revealing that Noctuide insects are usually infected by phylogenetically closely related GVs.
However, this virus contains 14 unique genes, including one encoding a putative chitinase protein.
Additionally, the gene content shows that SfGV VG008 possesses encoding sequences for other
virulence factors to insecticidal activity, such as 3 chitinases and 2 enhancins homologs to other
baculoviruses. This observation offers the possibility to postulate it as enhancer factor in NPVs
bio-insecticide formulations for S. frugiperda. Moreover, the SfGV VG008 genome analysis suggests
that its ancestors acquired some genes by horizontal transfer from alpha- and betabaculoviruses. In view
of the high number of unique genes, the evidence here observed about recombination events and the
ability to co-infect a host with other baculovirus species, SfGV VG008 is an example of how the
Baculoviridae family members maintain the diversity, improve their fitness, spread host range and assure
their perpetuation in nature.

Acknowledgments

This work was supported by research funds from COLCIENCIAS (Colombia) and MinCyT (Ministry
of Science and Technology; Argentina), Cooperation Program. Gloria P. Barrera, Paola E. Cuartas
and Laura F. Villamizar are researchers from CORPOICA (Corporación Colombiana de Investigación
Agropecuaria; Colombia). Emiliano Barreto is professor from Universidad Nacional de Colombia
(Colombia). Pablo D. Ghiringhelli and Mariano N. Belaich are members of the Research Career of
CONICET (Consejo Nacional de Ciencia y Tecnología, Argentina) and professors of UNQ (Universidad
Nacional de Quilmes, Argentina).

Author Contributions

Gloria P. Barrera carried out all the experiments needed to acquire the genome sequence.
Pablo D. Ghiringhelli coordinated and conducted all the bioinformatics studies. Paola E. Cuartas
and Mariano N. Belaich collaborated in the bioinformatics studies. Emiliano Barreto collaborated in the
discussion of results. Laura F. Villamizar managed the work and together with Pablo Ghiringhelli
and Gloria P. Barrera conceived the work. Pablo D. Ghiringhelli, Gloria P. Barrera, Paola E. Cuartas
and Mariano N. Belaich drafted the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.
References


3. Flores, F. *Manejo de Plagas en el Cultivo de Maíz*; Estación Experimental Agropecuaria: INTA, Russia, 2010; p. 7.


