Rohrmann, G.F., M.A. Erlandson, and D.A. Theilmann, The genome of a baculovirus isolated from Hemileuca sp. encodes a serpin ortholog. Virus Genes, 2013. DOI 10.1007/s11262-013-0951-x

GenBank accession KF158713

The genome of a baculovirus isolated from *Hemileuca* sp. encodes a serpin ortholog

George F. Rohrmann[1], Martin A. Erlandson[2], and David A. Theilmann[3]

[1]Dept. of Microbiology, Oregon State University, Corvallis, OR 97331-3804 USA
[2]Agriculture and Agri-Food Canada, Saskatoon Research Centre, 107 Science Place, Saskatoon, SK, Canada S7N 0X2
[3]Agriculture and Agri-Food Canada, Pacific Agri-Food Research Centre, Summerland, B.C. Canada V0H 1Z0

Phone: 541 737 1793
Email: rohrmanng@orst.edu

Abstract
The genome sequence of a baculovirus from *Hemileuca* sp. was determined. The genome is 140,633 kb, has a G+C content of 38.1%, and encodes 137 putative open reading frames over 50 amino acids. 126 of these ORFs showed similarity to other baculovirus genes in the database including all 37 core genes. Of the remaining 11 predicted genes, one is related to a lepidopteran serpin gene. This is the first report of a baculovirus encoding a member of this family of serine protease inhibitors, and the first report of a viral serpin outside the Poxviridae. The genome also contained 3 homologous repeat sequences. Phylogenetic analysis indicated that the virus is a Group II Alphabaculovirus and belongs to a lineage that includes *Orgyia leucostigma*, *Ectropis obliqua*, *Apocheima cinerarium*, and *Euproctis pseudoconspersa* nucleopolyhedroviruses.

Introduction

In order to determine if a collection of baculovirus samples from the US Forest Service Laboratory in Corvallis, Oregon contained any novel viruses, DNA was isolated from selected samples, and the *lef8* gene was PCR amplified using degenerate primers. 10 viruses were identified that had sequences significantly different from viruses whose genomes had been sequenced (1). Subsequently, a project was undertaken to sequence the genomes of these 10 viruses. Although documentation was limited, one sample originated from a vial of infected insects labeled '*Pseudohazis* polyhedrosis'. This insect genus has been renamed *Hemileuca* and is a member of the Saturniidae. In this report we describe the genome sequence of this virus.

Materials and Methods

Virus

The virus was isolated from a vial of infected insects labeled '*Pseudohazis* polyhedrosis' and was part of the Mauro Martignoni collection from the US Forest Service Laboratory in Corvallis, Oregon. A publication from that lab describes research on a baculovirus from the 'brown day moth, *Pseudohazis eglanterina*' (2). The brown day moth is also called the western sheep moth, (*Hemileuca eglanterina*). The genus *Pseudohazis* is no longer used and has been replaced by *Hemileuca*. Because of the lack of a confirmed species designation we have called this virus *Hemileuca species* NPV- MEM to designate that it originated from the Mauro E. Martignoni collection. Therefore the abbreviation we have used in this study is HespNPV-MEM.

Virus DNA Purification

Virus DNA was purified from occlusion derived virions (ODV) released from approximately $5.0x10^8$ HespNPV occlusion bodies (OB) using previously described methods (3). Briefly, the HespNPV OB suspension was incubated in purification buffer (TE pH7.5, 0.5%SDS, 0.1% Triton-X100) at 37°C for 1h, filtered through cheesecloth and OB pelleted at 1,400xg. OBs were resuspended in sterile ddH_2O at $2.0x10^8$ OB/ml. Approximately, $5.0x10^8$ OB were mixed with an equal volume of OB dissolution buffer (0.1M Na_2CO_3; 0.001M Na_2EDTA; 0.17 NaCl pH 10.8) and incubated for ~ 20 min at RT until OB dissolution was complete. Released ODV were pelleted by centrifugation through a 20% sucrose cushion in SW28 Beckman rotor at 95,000xg for 1 h. The ODV were resuspended in 500ul of TE pH 7.5 and incubated at 60°C to inactivate nucleases. The ODV were then incubated in 1.0% SDS and 5mg/ml proteinase K at 37°C for 2 h. Finally, DNA was extracted twice with an equal volume of phenol:chloroform:isoamylalcohol (50:48:2), once with an equal volume of chloroform:isoamlyalcohol and then the aqueous phase dialyzed against TE pH 7.5.

DNA Sequencing and Assembly

DNA sequencing was done at the National Research Council, Plant Biotechnology Institute (Saskatoon, Saskatchewan, Canada) using Roche 454 FLX-titanium pyrosequencing technology which generated 60,907 reads with an average read length of 606 nucleotides. The sequences were assembled using CLC-Genomics Workbench 6.0.2 into a single contig of 140,633 bp that contained 60,822 reads and represented an average sequence coverage of 258X. The final sequence was analyzed and annotated using Gene Quest (DNASTAR- Lasergene 10) and MacVector software programs. Repeated sequences were identified on Pustel DNA matrices and dot plots. The GenBank accession number is KF158713.

Results and Discussion

The HespNPV genome

A map showing the organization of the open reading frames (ORFs) over 150 nt in the 141kb HespNPV genome is presented in Fig. 1 and details are listed in Table 1. The virus is a member of Group II Alphabaculoviruses as it lacks an ortholog of the GP64 envelope fusion protein, and encodes an F (fusion) protein (HESP130), the hallmark of Group II viruses. The genome has a G+C content of 38.1% and encodes 137 predicted orfs, of which 126 are related to genes reported for other baculoviruses including 101 present in the *Autographa californica* multiple MNPV genome (Table 1). All 37 baculovirus core genes are present, reviewed in (4) (5) (6). There are also three homologous repeat regions present and although they showed similarity to each other, they did not appear to contain palindromic sequences (Fig. 1, Table 1). All the HespNPV orfs were compared to those of a closely related virus from *Orgyia leucostigma* NPV (OrleNPV) and the archetype Group I baculovirus, AcMNPV (Table 1). Polyhedrin was the most conserved orf, 96 and 87% amino acid sequence identity, respectively. Overall, the HespNPV amino acid sequences were on average 53 and 38% identical to OrleNPV and AcMNPV, respectively, indicating that the genome is significantly different from both viruses.

Novel putative genes in the HespNPV genome
A serpin ortholog

Of the 137 predicted orfs, only 11 did not show significant similarity to baculovirus genes previously described. However, three of these showed significant levels of similarity to genes in the Genbank database that have not been previously described for baculoviruses. HESP018 shows about 34% amino acid sequence identity to serpins from *Manduca sexta* (serpin 4a and 4b) and *Bombyx mori* suggesting that it was captured from a host insect. It was predicted to have a serpin domain from aa35-407 by SMART (simple modular architecture research tool) and contained a serpin signature sequence (FTVNPPFIFFI) as predicted by PROSITE. Serpins, (serine protease inhibitors), were named because of their ability to inhibit chymotrypsin-like serine proteases. No other baculovirus has been reported to encode an ortholog of this gene. In addition, the only other virus family reported to encode serpin orthologs is the Poxiviridae. In cowpox virus, a serpin, crmA, has been implicated in mitigating both the inflammatory and apoptotic responses. Although CrmA is a member of the serpin group, in contrast to other serpins, it is capable of inhibiting caspases, a family of cysteine proteases, thereby inhibiting apoptosis, reviewed in (7). Similar to serpins, crmA binds to active caspases as a pseudosubstrate and has been shown to irreversibly bind to caspase-1 and is also capable of inhibiting caspases 4, 5, 9 and 10, reviewed in (8). Serp 1, from Myxoma virus, another member of the *Poxviridae*, functions as a major virulence factor by its ability to block the inflammatory response. When it is defective, the viral infection is self-

limiting rather than lethal in European rabbits, reviewed in (9). A major antiviral response in insects that involves a serine protease cascade is melanization in which prophenoloxidase is cleaved and activated allowing it to catalyze the oxidation of phenols that polymerize and form melanin (10). Melanin can form capsules around pathogens thereby preventing their spread. This process can also result in the production of reactive oxygen species that can be toxic to pathogenic microbes. In *Manduca sexta,* a serpin has been characterized that is capable of inhibiting melanization (11). Therefore, possible roles for this baculovirus serpin in subverting these two pathways of insect defense could be major areas of future investigation.

A diguanylate cyclase domain protein

Orthologs of HESP005 have not been reported for other viruses. It shows 26% amino acid sequence identity over 200 amino acids with a significance (E-value of 8e-04) to a diguanylate cyclase domain protein from *Clostridium difficile*. However, it did not contain the conserved GGEEF domain found in the *C. difficile* proteins. It also did not show high levels of relatedness to anything in the structural database when examined by Hhpred (12). Consequently, the significance of this relationship is not clear.

An ascovirus ortholog

The third gene not previously reported for baculoviruses is *hesp112*.
The predicted protein exhibits 32% identity (7e-09) to ORF006 of *Trichoplusia ni* ascovirus 2c (accession number YP_803229) (13). The function of the gene is unknown.

Other novel genes

There are 8 predicted genes that have no orthologs in the database indicating that they are either novel genes unique to HespNPV or are not functional.

Antiapoptotic genes

HespNPV encodes three *iap* orthologs and an ortholog of *apsup*. HESP039 showed the closest similarity to *Cydia pomonella* granulovirus (CypoGV) IAP-3 with 30% identity over 287 amino acids. In contrast, HESP062 was most similar to iap-2 from *Apocheima cinerarium* NPV (ApciNPV) APCI040 (41%) and OrleNPV ORLE062 (37%). HESP101 was most similar to *Euproctis pseudoconspersa* NPV (EupsNPV) IAP-3 EUPS099 (41%), but also showed significant similarity to an entomopox IAP (*Amsacta moorei*, 33% identical). Another putative apoptosis inhibitor is HESP104 that displayed 40% amino acid sequence identity with the *Lymantria dispar* MNPV LD109, also known as APSUP (apoptosis suppressor) (14).

Additional repeated genes

There are 2 DNA binding protein proteins DBP-1and DBP-2 (HESP019) and 030), 2 baculovirus repeated orfs (Bro) (HESP 049,107), 2 P26 (Ac136) (HESP

022, 060) and 3 CHA-B (HESP 047, 048, 127) orthologs. Although DBP has been characterized in other baculoviruses (15-17), little is known about the role of these genes in baculovirus infections.

The ubiquitin gene is longer than normal

Although viral ubiquitin proteins have been identified in many alpha- and betabaculovirus, the one encoded by the HespNPV (HESP025) is unusual in that it is over twice as large (167 amino acids) as normal ubiquitin due to a C terminal fusion of 90 amino acids. The 77-amino acid N-terminal region is related to ubiquitin, whereas the 90 amino acid C-terminal region shows 48% identity over the complete 73-amino acid sequence of orf 097 from *Apocheima cinerarium* NPV. This suggests that the ubiquitin gene of HespNPV fused to an orf related to APCI097 in the viral genome. The fusion of ubiquitin genes with ribosomal proteins has been reported and, although subsequently the product is cleaved and released, this association facilitates their incorporation into ribosomes (18).

HespNPV phylogeny

In order to investigate the phylogenetic relationships of HespNPV, the LEF8 and PIF2 proteins of selected baculoviruses were used. As previously suggested by the analysis of the partial LEF8 sequence HespNPV (1), the virus is a member of the group II Alphabaculoviruses and its closest relatives are in a group consisting of *Orgyia leucostigma* NPV*, Ecotropis obliqua* NPV (EcobNNPV)*, Euproctis pseudoconspersa* NPV (EupsNPV), and *A. cinerarium* NPV (ApciNPV). Several other viruses in the Martignoni collection also are members of this clade, suggesting that it is a major lineage of Group II viruses. As seen in Fig. 2, a similar phylogenic relationship was observed for the concatenated LEF8 and PIF2 orfs. The bootstrap values indicate that this lineage is distinct from both Group I, and other members of the Group II lineage. In addition, the depth and similarity of the distance of the divisions, indicated that this lineage of viruses underwent a major radiation early in its development.

Acknowledgements

The research was supported by a Tartar grant to GR and a Canadian Crop Genomics Initiative grant to ME and DT. The technical assistance of Doug Baldwin in the preparation of HespNPV DNA for sequencing and the assembly of the initial virus sequence is gratefully acknowledged.

Refences

1. Rohrmann G.F., J Invertebr Pathol, in press, 2013.
2. Hughes K.M., J Invertebr Pathol 31, 217-224, 1978.
3. Erlandson M., J Invertebr Pathol 56, 47-56, 1990.
4. Miele S.A., Garavaglia M.J., Belaich M.N., and Ghiringhelli P.D., Int J Evol Biol 2011, 379424, 2011.
5. Garavaglia M.J., Miele S.A., Iserte J.A., Belaich M.N., and Ghiringhelli P.D., J Virol 86, 12069-12079, 2012.

6. Rohrmann G.F., Baculovirus Molecular Biology. National Library of Medicine (US), NCBI.http://www.ncbi.nlm.nih.gov/books/NBK49500/, Bethesda (MD), 2011.
7. Ekert P.G., Silke J., and Vaux D.L., Cell Death Differ 6, 1081-1086, 1999.
8. Turner S., Kenshole B., and Ruby J., Immunol Cell Biol 77, 236-241, 1999.
9. Zheng D., Chen H., Bartee M.Y., Williams J., Davids J.A., Huang E., Moreb J., and Lucas A., Trends Mol Med 18, 304-310, 2012.
10. Lemaitre B., and Hoffmann J., Annu Rev Immunol 25, 697-743, 2007.
11. Zhu Y., Wang Y., Gorman M.J., Jiang H., and Kanost M.R., J Biol Chem 278, 46556-46564, 2003.
12. Soding J., Bioinformatics 21, 951-960, 2005.
13. Wang L., Xue J., Seaborn C.P., Arif B.M., and Cheng X.W., Virology 354, 167-177, 2006.
14. Yamada H., Shibuya M., Kobayashi M., and Ikeda M., J Virol 85, 5237-5242, 2011.
15. Mikhailov V.S., Okano K., and Rohrmann G.F., Virology 346, 469-478, 2006.
16. Mikhailov V.S., Vanarsdall A.L., and Rohrmann G.F., Virology 370, 415-429, 2008.
17. Vanarsdall A.L., Mikhailov V.S., and Rohrmann G.F., Virology 364, 475-485, 2007.
18. Finley D., Bartel B., and Varshavsky A., Nature 338, 394-401, 1989.

Figure Legends

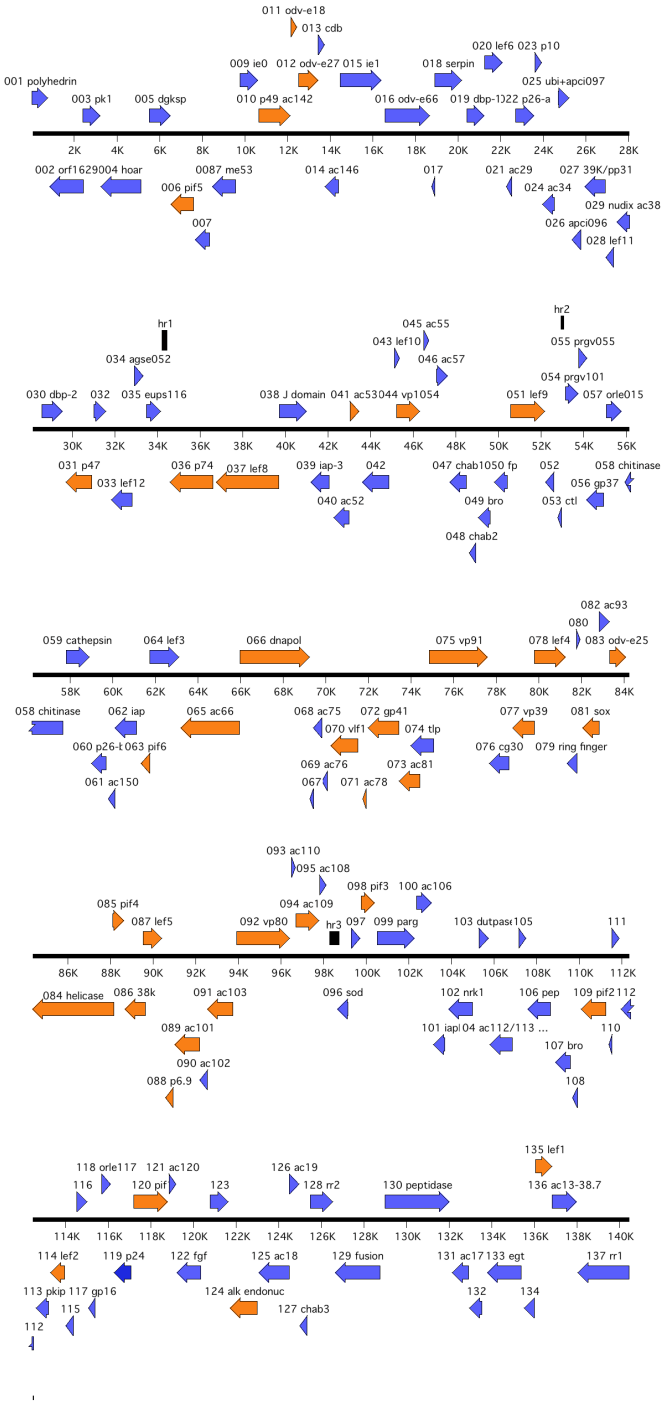

Figure 1 Linear map of the HespNPV genome. Core genes are indicated in red. The three homologous repeat sequences (hrs) are shown as black rectangles. The orf numbers and common names are indicated. If a common name does not exist, the orfs were designated by their AcMNPV orthologous orfs if present.

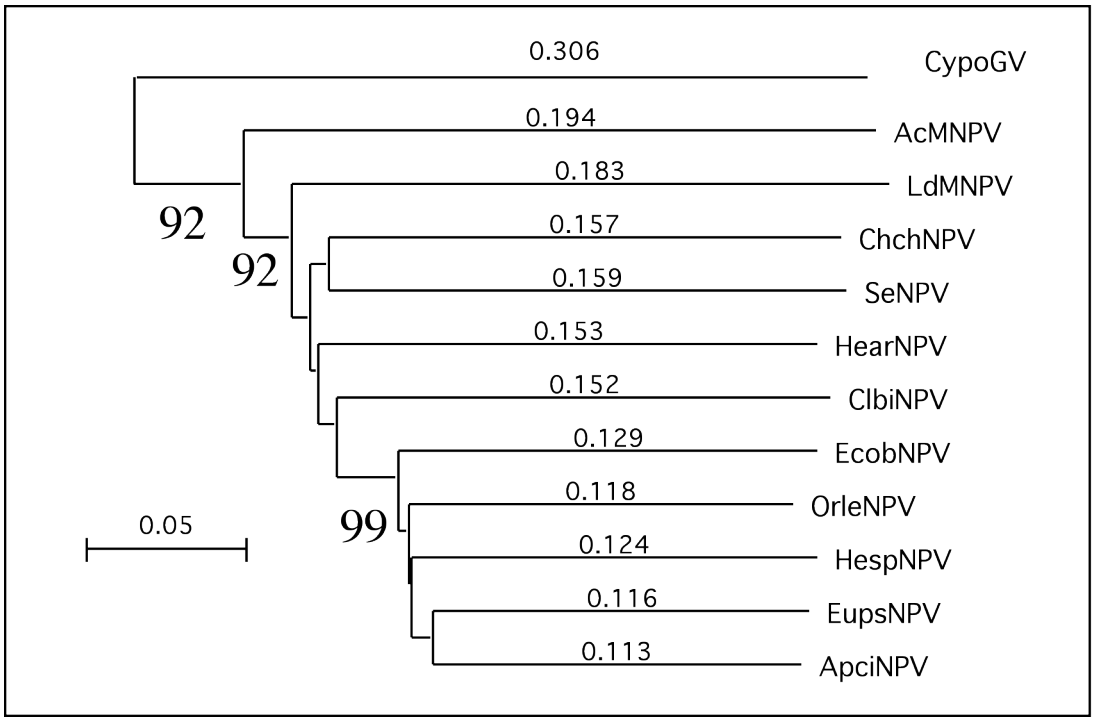

Figure 2. Phylogenetic relatedness of concatenated LEF-8 and PIF2 orfs from selected baculoviruses. *Cydia pomonella* GV (CypoGV) and AcMNPV (Betabaculovirus and Alphabaculovirus Group 1, respectively) were included as outliers. The other sequences are from Group 2 Alphabaculoviruses and include *Lymantria dispar* (LdMNPV), *Chrysodeixis chalcites* NPV (ChchNPV), *Spodoptera exigua* MNPV (SeMNPV), *Helicoverpa armigera* NPV (HearNPV), *Clanis bilineata* (ClbiNPV) *Ectropis oblique* (EcobNPV), *O. leucostigma* NPV (OrleNPV), *E. pseudoconspersa* NPV *(*EupsNPV), and *A. cinerarium* NPV (ApciNPV). The tree was developed using Neighbor joining; best tree; tie breaking = systematic distance: Uncorrected ("p"). Gaps distributed proportionally. Bootstrap values over 90 are shown. The CypoGV leg is not proportional.